



Università  
Ca'Foscari  
Venezia

Corso di Laurea magistrale in  
Economia e Finanza

Tesi di Laurea

**Misurazione del Rischio: Analisi e  
confronto  
tra Z-Score e Reti Neurali Artificiali**

**Relatore:** Pizzi Claudio

**Laureando:** Tommaso Collet, Matricola 871064

**Anno Accademico:** 2018 / 2019

*Ai miei genitori*

## Indice

Titolo	N° pagina
<b>Introduzione</b>	1
<b>Capitolo 1: Il modello Z-Score di Altman</b>	4
1.2 Il modello Z-score	4
1.3 I modelli di scoring del credito	5
1.4 Analisi tradizionale degli indici	5
1.5 L'analisi discriminante	6
1.6 Lo sviluppo del modello Z-Score	8
1.6.1 La selezione del campione	8
1.6.2 La selezione delle variabili e i loro pesi	9
1.6.3 Test del modello su altri campioni di imprese in difficoltà	13
1.6.4 Ricalibrazione del modello	14
1.7 Z-Score per le imprese private	14
1.8 Z-Score per le imprese non manifatturiere o operanti nei mercati emergenti.	15
1.9 Z-Score per le PMI	17
1.10 L'equivalenza con i rating obbligazionari (Bond Rating Equivalents)	18
1.11 Stima la Probabilità di Default mediante l'equivalenza con i rating obbligazionari	20
1.12 La comparazione dei modelli	23
1.13 La stima della Loss Given Deafault (Perdite date dal default)	24
<b>Capitolo 2: Approfondimento quantitativo sull'analisi discriminante e la regressione logistica</b>	26
2.1 Elementi di Analisi statistica discriminante	26
2.2 Caso gaussiano univariato	30
2.3 Caso multivariato	31
2.4 Analisi discriminante classica	33
2.5 Valutazione della correttezza del modello dato dall'analisi discriminante	35
<b>Capitolo 3: Le Reti Neurali</b>	41
3.1 Introduzione e caratteristiche generali delle Reti Neurali Artificiali	41
3.2 La letteratura riguardante le reti neurali	48
3.3 Confronto tra il modello a rete neurale e altre metodologie statistiche	52
3.4 Le reti neurali ibride	55

<b>Capitolo 4: Descrizione quantitativa del neurone artificiale</b>	<b>59</b>
4.1 La funzione di attivazione	60
4.2 Apprendimento	63
4.3 Il neurone artificiale: analisi vettoriale	65
4.4 L'analisi di una rete neurale e gli strumenti per la sua valutazione	69
4.5 L'apprendimento della rete neurale: la regola di Hebb	72
4.5.1 La regola postsinaptica, presinaptica e della covarianza	73
4.6 L'apprendimento supervisionato e il perceptrone semplice	75
4.7 Regola di apprendimento con un unico nodo di uscita ad attivazione bipolare	76
4.8 Perceptrone semplice: dimostrazione sulla convergenza	78
4.9 La regola delta	82
4.9.1 La regola delta applicata ad unità lineari	82
4.9.2 La regola delta applicata ad unità continue non-lineari	86
4.10 Back-propagation	87
<b>Capitolo 5: Confronto tra Z-Score e Reti Neurali</b>	<b>92</b>
5.1 Z'-Score	92
5.1.1 Z'-Score con il campione di sole imprese solventi	92
5.1.2. Z'-Score con il campione di sole imprese insolventi	94
5.1.3 Risultati complessivi Z'-Score	95
5.2 Z''-Score	97
5.2.1. Z''-Score con il campione di sole imprese solventi	97
5.2.2. Z''-Score con il campione di sole imprese insolventi	99
5.2.3 Risultati complessivi Z''-Score	100
5.3 Reti neurali	102
5.3.1 Modello di rete neurale con 2 nodi nascosti (I117_model22)	103
5.3.2 Modello di rete neurale con 3 nodi nascosti (I117_model23)	104
5.3.3 Le effettive prestazioni delle 2 reti	106
5.4 Reti Neurali con l'aggiunta di dati ritardati di un periodo	107
5.4.1 Rete neurale con 2 nodi nascosti e dati ritardati (I1617_model2)	107
5.4.2 Rete neurale con 3 nodi nascosti e dati ritardati (I1617_model3)	108
5.4.3 Rete neurale con 4 nodi nascosti e dati ritardati (I1617_model22)	110
5.4.4 Rete neurale con 5 nodi nascosti e dati ritardati (I1617_model5)	112
5.4.5 Le effettive prestazioni delle reti con dati ritardati	113

5.5 Reti Neurali con diverse percentuali nella composizione del training set e validation set	116
5.5.1 Trainig set 75%, Validation set 25% (l7525_model)	116
5.6 Risultati in sintesi	118
Appendice 1	120
Appendice 2	122
Appendice 2.1.1	135
Appendice 2.1.2	136
Appendice 2.1.3	137
Appendice 2.2.1	139
Appendice 2.2.2	139
Appendice 2.2.3	140
Appendice 2.3.1	141
Appendice 2.3.2	141
Appendice 2.3.3	142
Appendice 2.3.4	143
Appendice 2.4.1	144
Appendice 2.4.2	144
Appendice 2.4.3	145
Appendice 2.4.4	147
Appendice 2.4.5	148
Appendice 2.4.6	150
Appendice 2.5.1	150
Appendice 2.5.2	152
Bibliografia	155
Z-Score e analisi discriminante:	155
Reti Neurali	156



## Introduzione

Le principali motivazioni che hanno spinto lo sviluppo e la revisione dei modelli di *Credit scoring* è dovuta principalmente a due motivazioni: la prima si basa sull'introduzione dei principi di Basilea II i quali impongono dei requisiti patrimoniali a copertura delle attività creditizie mentre la seconda è dovuta al fatto che, nei primi anni del nuovo millennio, le imprese statunitensi hanno registrato un altissimo aumento del tasso di insolvenza o di bancarotta. Malgrado questo alto tasso sia poi diminuito con il trascorrere degli anni, l'attività di sviluppo di modelli per lo *score* del credito non si è mai arrestata.

Per poter funzionare queste misure, come molte altre, devono poggiarsi su una forte conoscenza del rischio di credito e devono essere utilizzate come strumenti aggiuntivi e non come i soli criteri per prendere delle decisioni in ambito dell'analisi del rischio di credito. Le banche stanno cercando in questi anni di sviluppare dei metodi per la valutazione della PD (*Probability of Default*) e della LGD (*Loss Given Default*) per attività di vario genere. La stima di queste due grandezze è di fondamentale importanza per gli istituti di credito in quanto consente di poter adoperare il cosiddetto approccio dei rating interni avanzato (IRB) previsto ai sensi di Basilea II.

I modelli di *score* del credito sono utilizzati nel mercato americano del credito oramai da 150 anni per poter valutare i rischi legati alle obbligazioni, ai prestiti e ai crediti delle compagnie da parte degli istituti finanziari e di altri generi di creditori. L'evoluzione dei modelli di *score* del credito è sinteticamente descritta di seguito secondo il modello presentato da Altman nel *Corporate Financial Distress and Bankruptcy, Third Edition (Edward I. Altman, Edith Hotchkiss)*, cap. 11:

- i. Modelli qualitativi (metodo molto soggettivo)
- ii. Modelli univariati (misure di contabilità o basate su valori di mercato)
- iii. Modelli Discriminanti Multivariati (misure di contabilità o basate su valori di mercato), modelli logit/probit (modelli lineari o quadratici), modelli non lineari (come per esempio l'analisi partecipativa ricorsiva e reti neurali)<sup>1</sup>
- iv. Sistemi di Intelligenza Artificiale, Expert System (un altro sistema di intelligenza artificiale), Reti Neurali

---

<sup>1</sup> Modelli Discriminanti e Logit in uso:

- Modelli di consumo
- Z-Score (utilizzato per la manifattura)
- Z-Score (utilizzato nel settore industriale)
- Risk Calc, Z"-Score (modello per le imprese non quotate)
- Score EM (modello utilizzato per i mercati emergenti e per il settore industriale)
- Altri (sistemi specifici elaborati dalle banche)

- v. Modello di Option Pricing/Contratti Contingenti, Rischio di Rovina, KMV per il monitoraggio del credito
- vi. Modelli a Rapporti Misti (Blended Ratio)/a Valori di Mercato, Risk Calc di Moody, BondScore, Z-Score

I primi modelli, risalenti al 1850, comprendevano una valutazione qualitativa delle aziende e dei loro proprietari come per esempio le garanzie personali, garanzie reali, esperienza di gestione del proprietario e segmento industriale. Questi prototipi di modelli di *scoring* furono poi integrati, all'inizio del ventesimo secolo, con delle funzioni univariate e da altri dati estrapolati dai bilanci aziendali. A partire da questo periodo iniziarono ad operare anche le agenzie di rating con la nascita di *Moody'* nel 1909 la quale di poco precedette la costituzione del 1916 di *Standard and Poor*.

Lo *Z-score* di Altman (1968) è il primo modello multivariato di *scoring* del credito che combina assieme dei rendiconti finanziari e dei valori estrapolati dal mercato. Questo modello fu utilizzato per classificare le aziende in due categorie: la prima comprende le imprese in difficoltà e/o in bancarotta mentre la seconda riguarda le imprese che non presentano particolari criticità. La tecnica statistica per svolgere questa classificazione si chiama *analisi discriminante*; questa è un particolare processo statistico, tra quelli disponibili, utile per separare o prevedere lo stato di salute futuro di un'azienda.

Nel tempo furono sviluppati altri processi statistici con la medesima finalità basati su:

- Regressione *logit*
- Regressione *probit* (per esempio *Ohlson 1980*)
- Parzionamento ricorsivo (per esempio *Frydman, Altman e Kao 1985*)
- Discriminanti quadratici (per esempio *Altman, Haldeman e Narayanan 1977*)
- Reti neurali (per esempio *Altman, Marco e Varetto*)
- Altre sofisticate tecniche come quelle basate sugli algoritmi genetici (per esempio *McKee e Lensbergn 2002*).

Da sottolineare è come, secondo Altman, molte di queste tecniche utilizzino sostanzialmente i medesimi principi per giungere allo stesso scopo ovvero, partendo da un'analisi puntuale di indici legati all'impresa, si punta a classificare e a predire eventuali stati di difficoltà aziendali. Ci sono delle tecniche più recenti che mirano a classificare la salute di un'impresa ma esse si basano su sistemi di intelligenza artificiale, sull'andamento di derivati o opzioni (per esempio il modello *Moody's/KMV's EDF*) e sistemi ibridi (come *CreditSight's BondScore*).

In questa ricerca verrà inizialmente analizzato lo *Z-Score* formulato da Altman nel *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy* del 1968 introducendo anche alcune delle sue riformulazioni. Per fare questo si è fornita una spiegazione teorica in merito all'analisi discriminante ovvero quella metodologia statistica essenziale per la costruzione di questo indice.



Nella seconda parte del lavoro (quindi nel capitolo 3 e 4) verranno invece introdotte le Reti Neurali Artificiali le quali, come per lo Z-Score, possono essere utilizzate per fornire delle previsioni in merito allo stato di salute di un'impresa. Le Reti Neurali non verranno introdotte nella loro interezza ma, dalla loro iniziale teorizzazione, si giungerà fino all'algoritmo della *Back-propagation* ovvero la regola di costruzione di una rete neurale più usata in ambito finanziario .

## Capitolo 1: Il modello Z-Score di Altman

### 1.2 Il modello Z-score

Lo sviluppo della procedura di valutazione della PD (*Probability of Default*) negli ultimi anni ha subito un notevole incentivo in quanto, a partire dagli anni 2000, è di gran lunga aumentato il numero delle imprese soggette a fallimento. Questo fenomeno è stato dettato soprattutto dalla necessità per gli istituti di credito di coprirsi dal rischio di insolvenza di un'impresa non solvente precedentemente finanziata. Passati i primi anni del 2000 le imprese hanno registrato una diminuzione nel numero di default ma la ricerca e lo sviluppo di nuove tecniche per misurazione del rischio di credito è continuata senza sosta. Tale ricerca si è sempre focalizzata sulla stima della PD e della LGD (*Loss Given Default*).

Il modello Z-Score è un modello sviluppato nel 1968 ma nel corso del tempo ha seguito numerose modifiche le quali permettono di adattarlo in base all'evoluzione del contesto economico e alla tipologia di soggetti analizzati.

In generale per stimare una corretta probabilità di default legata alle attività di credito verso le imprese si deve seguire un processo a tre passi che deve basarsi sull'utilizzo consecutivo di:

1. Modelli per lo score del credito (*Credit Score*)
2. Strumenti che sintetizzino il rischio del mercato di capitali; solitamente si utilizzano i rating dei titoli obbligazionari.
3. L'assegnazione delle PD e delle LGD ai portafogli di crediti.

Il modello Z-score si incentra sul primo step mentre gli altri due solitamente sono esemplificati utilizzando il modello sulle frequenze attese dei default formulato da Moody/KVM. Seguendo questo schema si è in grado di poter stimare la PD.

Queste tecniche sono anche le basi per poter valutare il rischio di mercato associato ad un'attività creditizia (*Value at Risk* o *VaR*). Altman e Hotchkiss nel *Corporate Financial Distress and Bankruptcy* (2006) sostengono che un modello di *scoring*, se basato su dei dati comprensibili e rappresentativi, riesca a rendere il *VaR* sul credito preciso ed utile sia ai fini della determinazione del capitale regolamentare da accantonare per la copertura del rischio sia per la previsione dello stato di dissesto di un'impresa. Se questo non dovesse accadere, nessuna quantità o nessuna struttura per l'analisi di portafoglio risulta essere valida per la stima del rischio di credito

### 1.3 I modelli di *scoring* del credito

Tutti i modelli di *scoring* del rischio di credito che sono utilizzati al giorno d'oggi sono delle variazioni dello stesso tema. Questo tema riguarda la combinazione di una serie di indicatori finanziari quantificabili riportanti le prestazioni dell'azienda con l'aggiunta di qualche variabile che punta a misurare degli elementi specificatamente legati al processo che descrive il credito. Il modello, qualunque esso sia, concerne una gran mole di dati relativi ai default passati per la determinazione della PD e, sebbene ci si concentri esclusivamente su delle variabili quantitative, Altman (2006) afferma che non si dovrebbero sottovalutare l'importanza delle misure qualitative nel processo di determinazione della PD e sostiene che, in ambito bancario, gli elementi qualitativi sintetizzati dal giudizio del responsabile della gestione dei rischi, possono apportare una miglioria nel potere esplicativo del modello di *score* del credito che va dal 30% fino al 50%.

Sempre dal lavoro di Altman e Hotchkiss, *Corporate Financial Distress and Bankruptcy*, 3e (2006) si riporta che dagli anni 60 molti esperti ed accademici hanno iniziato ad operare un processo di esclusione dell'analisi basata su indici finanziari come metodo analitico per studiare le prestazioni dell'impresa; la regola generale che si basava sulla comparazione degli indici aziendali tra varie imprese fu abbandonata verso l'utilizzo di approcci basati su criteri statistici più rigorosi. Questa esclusione dell'analisi degli indici aziendali verso metodologie meramente empiriche sembra indirizzarsi verso un approccio puramente statistico, tralasciando ogni spazio alla teoria economica. Scott nello studio *The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models. Journal of Banking & Finance (September)* (1981) ha studiato le correlazioni tra i modelli strutturati e quelli multivariati, come per esempio lo Z-Score e il modello ZETA arrivando a smentire questa logica meramente statistica dell'analisi del rischio associato al credito. Scott infatti afferma che i modelli teorici e quelli empirici hanno molti punti in comune e quindi devono procedere pari passo nell'analisi della rischiosità di un asset.

### 1.4 Analisi tradizionale degli indici

Una compagnia che versa in condizioni di difficoltà, secondo Altman, è particolarmente adatta per essere sottoposta ad un'analisi con gli indici finanziari

Prima dell'introduzione di misure quantitative, in merito alle performance dell'impresa, le agenzie davano esclusivamente dei giudizi qualitativi sul merito di credito dei mercati. William H. Beaver nel '66 e nel '68 produsse due studi dai titoli *Financial Ratios As Predictors of Failure* e *The Information Content of Annual Earnings Announcements* concernenti l'analisi di alcuni

indici di impresa per la classificazione dello stato di salute delle imprese. Beaver essenzialmente analizzò una serie di indicatori economici per poterne stimare la capacità previsiva in merito allo stato di salute dell'impresa; la sua fu quindi un'analisi univariata la quale, come afferma Altman, aprì la strada verso l'approccio multivariato formulato in precedenza. L'analisi di Beaver scoprì che alcuni di questi indicatori riescono a discriminare una serie di campioni uguali di imprese fallite e non fallite fino a cinque anni prima del fallimento. Beaver tuttavia mise in discussione l'utilizzo di un approccio multivariato che però, nello stesso periodo, fu messo in pratica da Altman con il modello *Z-Score*, evidenziando la non problematicità di questo metodo.

Altman selezionò alcuni indicatori finanziari, adatti per la previsione dello stato di un'impresa in termini di salute finanziaria, per costruire un modello che restituisse con estrema velocità uno *score* di impresa. Si riuscì a descrivere un profilo multivariato dell'impresa basato su diversi pesi associati ad ogni variabile il quale forniva dei risultati chiari, non ambigui o ingannevoli. Gli indici aziendali utilizzati per questo indicatore si riferiscono alla profittabilità, alla liquidità, all'effetto leva, alla solvibilità e anche a misure multidimensionali, come i guadagni e la copertura dei flussi finanziari. Alcuni di questi prevalsero come gli indicatori più significativi ma, anche oggi, attraverso vari studi, non si è in grado di definire quale sia quello più importante (Altman & Hotchkiss, 2006). In sintesi Altman approfondì gli studi di Beaver combinando le diverse misure analizzate per creare un modello predittivo multivariato.

## 1.5 L'analisi discriminante

Nel modello di Altman si è scelto inizialmente di utilizzare la c.d. *analisi discriminante* (MDA) come approccio statistico per la costruzione dello *Z-Score*. Sebbene questa non sia popolare come l'analisi di regressione, l'analisi discriminante è stata utilizzata in svariati settori accademici dopo la sua prima applicazione nelle scienze biologiche negli anni 30 (Altman & Hotchkiss, 2006).

La MDA è una tecnica statistica utilizzata per classificare un'unità statistica dentro a uno dei molti gruppi definiti a priori, in dipendenza delle caratteristiche individuali della singola osservazione. Questa tecnica è utilizzata per classificare o fare previsioni su una variabile dipendente che si presenta in forma qualitativa<sup>2</sup>.

Il primo passo da fare è quello di fissare dei gruppi in cui sia possibile svolgere una classificazione delle nostre variabili; i gruppi imposti quindi devono essere almeno due. Dopo

---

<sup>2</sup> Nel caso del *credit scoring* le caratteristiche qualitative che può assumere la variabile dipendente è "insolvente" o "solvente"

aver fatto questa operazione si deve procedere alla raccolta dei dati identificativi delle unità statistiche che costituiscono il campione; questa tecnica ha il vantaggio di considerare molte caratteristiche comuni tra le imprese come anche le interazioni tra queste variabili considerate. Un approccio multivariato alternativo sviluppato successivamente (Ohlson, 1980) all'analisi discriminante di Altman fu un modello basato su un'analisi logistica. Questo è utilizzato soprattutto dalle aziende di consulenza quando devono offrire assistenza alle banche; in particolare questa tipologia di analisi considera lo specifico contributo di ogni variabile e riesce a stimare direttamente come output una misura della probabilità di default.

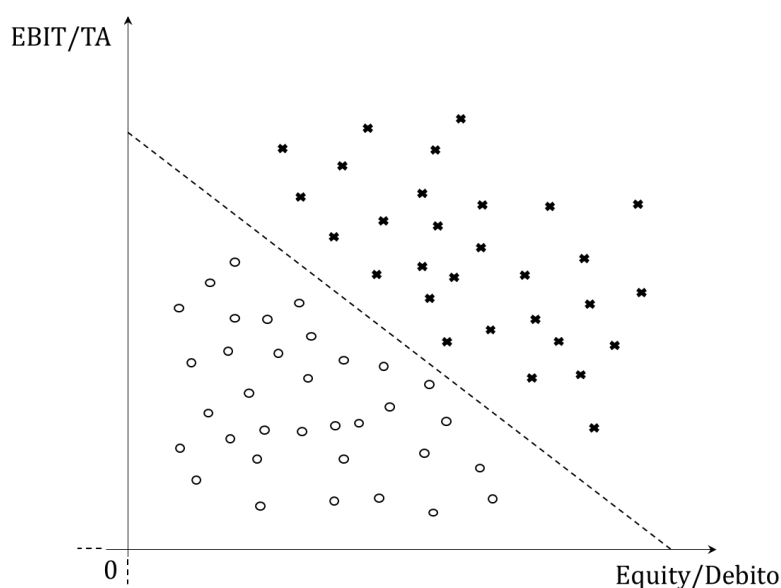
In merito all'analisi discriminante, per descrivere la struttura lineare discriminante tra aziende sane o in dissesto si è scelta, a titolo esemplificativo, una struttura a due variabili; queste sono i guadagni e l'effetto leva.

Forma lineare:  $Z = a_1x_1 + a_2x_2$

dove  $Z$  rappresenta lo score discriminante (Z-score)

$a_1, a_2$  = coefficienti discriminanti o pesi delle variabili;

$x_1, x_2$  = variabili discriminanti rappresentate in questo caso da guadagni ( $EBIT/TA$ ) ed effetto leva ( $Equity/Debito$ )



**Figura 1:** Previsione dello stato di salute di un campione di imprese con l'analisi discriminante. Con la croce (X) vengono denotate le imprese sane mentre con il cerchio (O) le imprese in dissesto.

**Fonte:** Altman & Hotchkiss, *Corporate Financial Distress and Bankruptcy*, 3e (2006)

Per poter svolgere la separazione nei gruppi definiti in precedenza (aziende sane e aziende fallite) si deve procedere con la determinazione della linea tratteggiata (Figura 1) chiamata "linea di separazione" (o *decision boundary*). Per far ciò si devono stimare i valori dei pesi  $a_1$

e  $a_2$  in modo da rendere massima la distanza tra le medie dei gruppi e allo stesso tempo minima la varianza all'interno di ciascun gruppo<sup>3</sup>.

## 1.6 Lo sviluppo del modello Z-Score

### 1.6.1 La selezione del campione

Per la creazione della funzione che avrebbe definito lo *Z-score* Altman prese un campione di 66 imprese composto per la metà da imprese fallite e per l'altra metà da imprese non fallite. Le imprese in bancarotta erano tutte manifatturiere e tutte andarono in default tra il 1946 e il 1965. Prendere in considerazione un periodo di tempo di 20 anni, anche secondo lo stesso autore però non è stata una buona scelta poiché il tasso medio di fallimenti varia nel tempo rendendo quindi l'analisi distorta. Sarebbe stato preferibile esaminare una lista di indicatori al tempo  $t$  in modo da fare delle previsioni riguardanti l'andamento delle imprese nel periodo di tempo successivo  $t + 1$ . I dati a disposizione però a quel tempo erano pochi, soprattutto per quanto riguardava le imprese fallite, e quindi non era possibile svolgere una raccolta di dati sufficientemente ampia e uniforme in termini temporali in modo da porre in essere l'analisi voluta dallo stesso Altman. Si afferma inoltre che, nella selezione del campione, il gruppo delle imprese in bancarotta non è omogeneo a causa di differenze sostanziali tra le imprese; queste differenze sono date dal settore di appartenenza e dalle dimensioni (Altman, 1968).

Oggi, a differenza di allora, il numero dei fallimenti è molto cresciuto e quindi si riesce a ricavare delle informazioni molto più consistenti in termini di numerosità (Altman & Hotchkiss, 2006).

Altman nella costruzione dello Z-Score ha operato un'attenta selezione anche delle imprese manifatturiere non in difficoltà. Questo gruppo è formato da campioni accoppiati (o campioni dipendenti) di imprese manifatturiere scelte su una base casuale stratificata, ovvero divisi in gruppi in base a delle caratteristiche condivise con le corrispondenti imprese fallite come la grandezza e la tipologia di azienda.

Inizialmente si era scelto un range, in termini di attivo di bilancio, che andava da 1 milione di dollari ai 25 milioni di dollari come intervallo in cui comprendere le imprese del campione<sup>4</sup>. Al tempo (ovvero nel 1968), un default di un'impresa con un valore contabile degli attivi di 25 milioni di dollari era considerato molto rilevante (Altman, 2006). I dati raccolti nel campione

---

<sup>3</sup> Per un approfondimento quantitativo sul tema della determinazione dei pesi dell'analisi discriminante si invita a fare riferimento al *Capitolo 2* di questo elaborato.

<sup>4</sup> I valori degli asset del campione avevano una media pari a 6,4 milioni di \$ e un range che andava dai 0,7 milioni di \$ di minimo e 25,9 milioni di \$ di massimo.

delle imprese rivelatesi sane erano degli stessi anni di quelli per le aziende fallite e l'intero campione iniziale (quindi imprese fallite e non) conteneva i dati ricavati dai bilanci di un anno prima la bancarotta. Tyler Shumway nel *Forecasting Bankruptcy More Accurately: A Simple Hazard Model* del 2002 ha criticato questo approccio di analisi statistica ma si è subito dimostrato che gli indici ricavati da un bilancio prima della bancarotta fornisce, ex post, un modello più accurato nei risultati dei test (Altman & Hotchkiss, 2006).

### 1.6.2 La selezione delle variabili e i loro pesi

Nell' primo articolo pubblicato da Edward I. Altman, *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, The Journal of Finance, Vol. 23, No. 4 (Sep., 1968), pp. 589-609 viene esposta la metodologia di selezione delle variabili per la costruzione dello Z-Score. L'autore, dopo aver selezionato le imprese e definiti i gruppi composti rispettivamente da 33 imprese fallite e da 33 sane, ne raccolse i bilanci e le varie scritture contabili. Ci sono moltissime variabili che possono fungere da indicatore dello status dell'impresa; alla fine Altman individuò 22 variabili potenzialmente utili per la valutazione dello stato di salute dell'impresa. Di queste variabili solamente 5 avrebbero svolto la miglior previsione di un possibile fallimento dell'impresa. La funzione discriminante stimata da Altman è stata ottenuta mediante un procedimento iterativo; questo quindi non offre alcuna garanzia sull'ottima capacità discriminante della funzione calcolata (Altman, 2006 e Altman & Hotchkiss, 2006) . La funzione discriminante risulta essere:

$$Z = 1.2 X_1 + 1.4 X_2 + 3.3 X_3 + 0.6 X_4 + 1.0 X_5$$

dove:

$X_1 = \text{Capitale circolante} / \text{Totale delle attività}$

$X_2 = \text{Utile trattenuti} / \text{Totale delle attività}$

$X_3 = \text{EBIT o Margine Operativo Netto} / \text{Totale delle attività}$

$X_4 = \text{Valore di mercato dell'equity} / \text{Valore contabile di tutte le passività}$

$X_5 = \text{Vendite} / \text{Totale delle attività}$

$Z = \text{Indice complessivo o score}$

**Capitale Circolante Netto/Totale delle attività:** questa variabile misura le attività nette liquide della società in relazione al totale della capitalizzazione. Il *Capitale Circolante Netto* è definito dalla differenza che vi è tra attività correnti e passività correnti. La liquidità e la dimensione

dell'impresa sono caratteristiche considerate in questa variabile. In questo indice vengono solitamente inclusi i beni materiali ma esclusi quelli immateriali.

Questo rapporto dà il contributo meno forte alla discriminazione tra i due gruppi di aziende presi a campione.

**Utili trattenuti/Totale delle attività:** gli *utili trattenuti* rappresentano il totale degli utili reinvestiti e/o le perdite di una società maturati durante tutta la sua vita. Questa voce di bilancio può essere chiamata anche *surplus guadagnato*. Gli utili trattenuti possono essere associati al profitto cumulato sull'intera vita della società quindi, all'interno di questo indice, è rappresentata implicitamente anche la durata della vita dell'impresa.

Quando non si distribuiscono dei dividendi si crea un'immagine negativa dell'impresa agli occhi degli azionisti e di altri operatori che vorrebbero entrare nella compagine azionaria dell'impresa; si deve quindi in qualche modo fare una correzione di questi pregiudizi nei conti dell'impresa.

Questo rapporto misura il grado di indebitamento dell'impresa; un *UT/TA* alto testimonia un alto finanziamento delle attività dell'impresa attraverso la trattenuta degli utili maturati; l'impresa in questo caso ricorre in maniera marginale all'indebitamento. Questa variabile quindi evidenzia l'utilizzo di fondi generati internamente (capitale a basso rischio) o l'utilizzo di fondi provenienti da altri operatori di mercato (capitale ad alto rischio).

Questa variabile ha evidenziato un crescente deterioramento del valore delle imprese nel mercato durante gli ultimi 20 anni (Altman, 2006) perciò, nella struttura della funzione che definisce il parametro *Z* si è cercato di ridurre l'impatto di questo fattore in modo da non inficiare troppo i punteggi ottenuti dallo *Z-score*. Questo aspetto sarà trattato più avanti sul tema dell' "equivalenza con i rating obbligazionari".

**Earnings Before Interest and Taxes (EBIT o Margine Operativo Netto)/Totale delle attività:** questa è una misura del grado di produttività degli asset aziendali indipendentemente dal livello di indebitamento o dal livello di tassazione. La sopravvivenza di una società è strettamente legata al grado di guadagni che è in grado di ottenere nel tempo perciò questa variabile è estremamente importante nello studio del rischio di credito della singola impresa. Si è dimostrato come questa misura, sebbene possa essere soggetta a manipolazioni contabili essendo strettamente dipendente dai guadagni ottenuti dall'impresa, si sia però dimostrata estremamente efficace dal punto di vista predittivo limitatamente ai cash flow aziendali<sup>5</sup>.

---

<sup>5</sup> A titolo esplicativo si veda *ZETA model results* in Altman, Haldeman, and Narayanan 1977 e Altman 1993



**Valore di mercato dell'equity/Valore contabile di tutte le passività:** il patrimonio netto della società è misurato con la combinazione del valore di mercato di tutte le azioni in circolazione mentre le passività comprendono sia i debiti correnti che quelli a lungo termine. Questo misura è di estrema importanza poiché misura di quanto le attività dell'impresa possano diminuire in valore<sup>6</sup> prima che le passività eccedano le attività in modo da rendere insolvente l'impresa<sup>7</sup>. Questo rapporto mette in relazione lo *Z-score* con una dimensione legata al valore di mercato che in molti studi viene spesso tralasciata. Si vuole inoltre precisare che si sostituirà al valore di mercato del patrimonio netto il corrispondente valore contabile in modo da ricavare una funzione discriminante per le imprese non quotate (*Z'-Score*) e per le imprese non manifatturiere (*Z''-Score*)

**Vendite/Totale delle attività:** il coefficiente di rotazione del capitale è un rapporto finanziario molto utilizzato in ambito contabile poiché mostra la capacità delle attività aziendali di riuscire a generare delle vendite. Questo rapporto a livello statistico è molto poco significativo infatti, se si svolgesse un test sulla significatività statistica univariata, questo parametro non sarebbe selezionato poiché non significativo ai fini della stima del modello. Il coefficiente di rotazione del capitale tuttavia, a causa delle sue relazioni con le altre variabili del modello, è la variabile che offre il maggiore contributo alla *capacità discriminante complessiva del modello* (Altman, 2006). Sussistono grandi variazioni tra le diverse aziende e tra le varie aree geografiche del *Tasso di rotazione del Capitale Investito (Asset Turnover)* le quali richiedono la specificazione di un modello alternativo (*Z'''*) che non comprenda questa variabile.

Inserendo i dati estratti dai bilanci, si può calcolare il valore *Z* che, per definiti limiti, rende lo stato di salute dell'impresa. Quindi se:

$Z > 2,99$	l'impresa è sana
$1,81 < Z < 2,99$	l'impresa si trova in uno stato di incertezza
$Z < 1,81$	l'impresa è in bancarotta

(da *Z-Scores- A Guide to Failure*, Editor: Douglas R. Carmichael, PhD, CPA Baruch College)

Si può notare come il modello non contenga un termine costante. Come si è giunti a determinare i pesi associati alle varie variabili che rientrano all'interno del modello? Questi dati sono stati sviluppati attraverso un algoritmo del computer. Ovviamente questi "pesi"

---

<sup>6</sup> Il valore delle attività di un'impresa è dato dal valore di mercato dell'equity sommato al debito

<sup>7</sup> Si vuole ricordare che questa comparazione tra attività e passività per valutare il rischio legato al debito dell'impresa è stata trattata per la prima nel modello KMV prima che Merton (1974) mettesse questa relazione dentro ad un modello con struttura opzionale.

cambiarebbero se vi fossero dei campioni differenti o se fossero utilizzate delle nuove variabili (Altman & Hotchkiss, 2006).

Le variabili e le loro medie sono state misurate in un istante finanziario prima del default osservando la relativa *F-statistica*: questo test statistico confronta le differenze tra le medie dei valori degli indici in un gruppo con la variabilità (o spread) dei valori di tali indici all'interno di ogni gruppo. Altman sottolinea che le variabili  $X_1$  fino alla  $X_4$  manifestino delle significative differenze tra un gruppo e l'altro: questo è testimoniato dal fatto che il valore corrispondente della loro *F-statistica* è significativo per un livello di confidenza del 0,001. Questo ci porta rifiutare l'ipotesi nulla  $H_0$  e ciò ci spinge a considerare l'aggiunta di queste variabili come significativa ai fini della specificazione del modello. La variabile  $X_5$  invece non presenta questa caratteristica sui 2 diversi campioni in quanto la *F-statistica* non è significativa dato il livello di confidenza prima indicato. Ciò ci spinge ad accettare l'ipotesi  $H_0$  e quindi l'aggiunta di questa variabile non sembrerebbe significativa ai fini della specificazione del modello (Altman, 1968). La presenza di questa ultima variabile nel modello di Altman verrà motivata in seguito.

Variabile	Media del gruppo delle imprese insolventi	Media del gruppo delle imprese sane	Valore del test F
	$n=33$	$n=33$	
$X_1$	-6,1%	41,4%	32,6
$X_2$	-62,6%	35,5%	58,86
$X_3$	-31,8%	15,3%	26,56
$X_4$	40,1%	247,7%	33,26
$X_5$	150,0%	190,0%	2,84

: significative per un livello di confidenza del 0,001

$$F_{1,60}(0,001) = 12,00$$

$$F_{1,60}(0,01) = 7,00$$

$$F_{1,60}(0,05) = 4,00$$

**Tabella 1:** Medie delle variabili nei due campioni e test F di significatività

**Fonte:** Edward I. Altman, *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, *The Journal of Finance*, Vol. 23, No. 4 (Sep., 1968), pp. 589-609

Altman afferma che, se ci si basasse su un modello univariato per ogni variabile, tutti i rapporti soprariportati rilevarebbero dei valori più alti per le imprese sane inoltre i coefficienti discriminanti del test, facenti parte della funzione per la determinazione dello *Z-Score*, presentano sempre dei segni positivi; questo è quello che ci si aspetta poiché, ad imprese più in salute, si associa uno *score* più elevato. Al contrario si associa uno *score* più contenuto alle imprese che manifestano una pericolosità maggiore (Altman, 1968).

La variabile  $X_5$  risulta essere importante per la determinazione dello *score* di un'impresa in quanto Altman nel 1968 ha dimostrato che questa, nella determinazione della capacità

discriminante di tutte e cinque le variabili, risulta una tra le più efficaci se pesata per la propria deviazione standard.

Sebbene quattro delle cinque variabili usate presentino sostanziali differenze tra i due gruppi, l'*analisi discriminante multivariata* (MDA) è di fondamentale importanza poiché è in grado di separare i due gruppi di imprese considerate (rischiose e non rischiose) utilizzando dei modelli multivariati.

Stimati i coefficienti discriminanti è possibile calcolare lo *score discriminante* per ogni unità del campione e quindi assegnare l'osservazione a uno dei gruppi. Questa procedura si basa sulla comparazione del profilo individuale di ogni impresa con quella dei due gruppi alternativi (imprese in difficoltà e non in difficoltà) (Altman & Hotchkiss, 2006).

### 1.6.3 Test del modello su altri campioni di imprese in difficoltà

Svolgendo consecutivamente tre test per il calcolo del modello *Z-Score*, Altman esaminò 86 imprese in difficoltà dal 1969 al 1975, 110 fallite tra il 1976 e il 1995 e 120 fallite tra il 1997 e il 1999 (Altman & Hotchkiss, 2006). Con questa analisi si scoprì che il modello *Z-Score*, assegnando un punteggio pari a 2,675 di *cut-off*, ha avuto un range di precisione che va dall'82% e il 94%. Ripetendo i test, la precisione del modello *Z-Score* in campioni di imprese in difficoltà ha mostrato una precisione variante tra l'80% e il 90%; questi test sono basati su dati finanziari presi dai bilanci di un periodo finanziario precedente il fallimento. L'errore di seconda specie<sup>8</sup> tuttavia è cresciuto sostanzialmente (del 25%) se il valore *cut-off* fosse fissato pari a 1,81 ovvero pari al livello più basso c.d. della banda di ignoranza. Questo risultato è assai curioso poiché un *cut-off* pari a 1,81 dà un valore di *Z-Score* più realistico rispetto al valore ottenuto fissando il limite a 2,675 sebbene quest'ultimo, come riporta Altman, sia stato quello che nei test originari abbia provocato complessivamente meno errori (Altman & Hotchkiss, 2006).

Vi è poi anche una tecnica alternativa, ma comunque molto importante, per la classificazione delle aziende poiché consente di stimare anche la probabilità di default (PD). Questo si basa sull'utilizzo della *tecnica di equivalenza con i rating dei bond* (BRE<sup>9</sup>) per ogni score associato calcolato per le singole imprese.

---

<sup>8</sup> L'errore nel classificare una data impresa come "in difficoltà" quando in realtà non è sarà soggetta a bancarotta o in default a causa del mancato adempimento delle proprie obbligazioni.

<sup>9</sup> *Bond Rating Equivalent*

### 1.6.4 Ricalibrazione del modello

L'accuratezza del modello *Z-Score* è stata osservata da Altman per quasi quattro decenni a partire dalla data della sua formulazione originale. La precisione del modello ha continuato ad essere accettabile infatti ha registrato un tasso del 80% di default nelle predizioni degli stessi entro un anno dalla data scelta ma è cresciuto anche l'errore di seconda specie. La crescita di questa tipologia di errore è dovuta al fatto che il 25% delle imprese statunitensi hanno una struttura finanziaria molto più simile a quella di aziende in stato di bancarotta che ad entità sane. Un'altra ragione che spiega la consistenza dell'errore di seconda specie è legata al fatto che le imprese americane sono molto più rischiose rispetto al passato; questo fatto si è concretizzato nel deterioramento di due indici principali che costituiscono lo *Z-score* ovvero il rapporto utili non distribuiti sul totale delle attività e il rapporto capitale/debito (Altman & Hotchkiss, 2006).

Se a questa misura si applicasse una trasformazione logaritmica si riesce a migliorare le capacità previsive del modello e si va a ridurre l'errore di seconda specie; un approfondimento di questa discussione è riportata più in dettaglio nello studio di Edward I. Altman e Herbert A. Rijken, *How rating agencies achieve rating stability*, pubblicato il 26 agosto 2004.

### 1.7 Z-Score per le imprese private

Il modello di *Z-Score* originario è stato pensato secondo gli esperti per le imprese quotate infatti la variabile  $X_4$  richiede l'utilizzo del prezzo delle azioni; nelle aziende private quindi questa stima non è possibile se non mediante la quotazione dell'azienda stessa in borsa. Per poter adattare questo modello alle imprese non quotate si potrebbe sostituire al valore di mercato dell'impresa il valore contabile del capitale netto; questa modifica però non è scientificamente valida (Altman & Hotchkiss, 2006).

Al posto di sostituire la variabile  $X_4$  inserendo una nuova variabile proxy che vada a soddisfare le imprese private, Altman preferisce procedere con una nuova stima del modello sostituendo il valore di libro dell'impresa al rispettivo valore di mercato. Ci si aspetta quindi che cambino sia tutti i coefficienti legati alle variabili (non solamente il parametro legato alla nuova variabile) sia il criterio di classificazione sia il punteggio legato al *cut-off*. La nuova formulazione dell'equazione di stima dello *score* risulta essere (Altman & Hotchkiss, 2006).

$$Z' = 0.717X_1 + 0.847X_2 + 3.107X_3 + 0.420X_4 + 0.998X_5$$

dove:

$$X_1 = \text{Capitale circolante} / \text{Totale delle attività}$$

$$X_2 = \text{Utile trattenuto o Utile portato a nuovo} / \text{Totale delle attività}$$

$$X_3 = \text{EBIT o Margine Operativo Netto} / \text{Totale delle attività}$$

$$X_4 = \text{Valore contabile del capitale netto} / \text{Valore contabile di tutte le passività}$$

$$X_5 = \text{Vendite} / \text{Totale delle attività}$$

nel quale se:

$$Z' > 2,90$$

l'impresa è sana

$$1,23 < Z' < 2,90$$

l'impresa si trova in uno stato di incertezza

$$Z' < 1,23$$

l'impresa è in bancarotta

(da *Z-Scores- A Guide to Failure*, Editor, Douglas R. Carmichael, PhD, CPA Baruch College)

Il nuovo modello di *Z-Score* stimato da Altman con la nuova variabile  $X_4$  si presenta quindi in modo diverso rispetto alla funzione stimata originariamente. Il coefficiente associato alla variabile  $X_1$  si è abbassato ad un valore pari a 0,7 mentre originariamente era di 1,2. In generale però il nuovo *Z-Score* per le imprese private si presenta abbastanza simile a quello originario visto per le imprese quotate.

### 1.8 Z-Score per le imprese non manifatturiere o operanti nei mercati emergenti.

Un'ulteriore modifica del modello *Z-Score* prevede l'eliminazione dal modello della variabile  $X_5$  (*Vendite/Totale delle Attività*). Altman e Hotchkiss (2006) riportano che questo elemento viene eliminato dal modello poiché si vuole andare a minimizzare i potenziali effetti specifici del settore industriale; questi infatti si manifestano in maniera più consistente quando l'indice di rotazione del capitale è incluso nel modello.

Questo nuovo modello è stato utilizzato per valutare lo stato di salute delle imprese non manifatturiere e delle società non statunitensi (Altman & Hotchkiss, 2006). Altman, Hartzell e Peck (1995, 1997) in particolare hanno applicato questa nuova riformulazione del modello *Z-Score* (*Z''-Score*) a delle società operanti nei mercati emergenti, soprattutto messicane<sup>10</sup>. In questo caso, per la variabile  $X_4$  si è utilizzato il valore contabile del capitale netto.

La precisione di questo nuovo modello nella classificazione è nei fatti identica a quella risultante dal modello a cinque variabili rivisto (*Z'-Score*). La funzione che descrive questa ulteriore nuova riformulazione è:

<sup>10</sup> Si vuole sottolineare che tutte queste società avevano emesso degli bond in dollari americani

$$Z'' = 3,25 + 6,56 X_1 + 3,26 X_2 + 6,72 X_3 + 1,05 X_4$$

nel quale se:

$Z'' > 5,83$	l'impresa è sana
$4,15 < Z'' < 5,83$	l'impresa si trova in uno stato di incertezza
$Z'' < 4,15$	l'impresa è in uno stato di forte criticità

(da *Z-Score Models' Application to Italian Companies Subject to Extraordinary Administration*, Edward I. Altman, Alessandro Danovi, and Alberto Falini, 2013)

In questo caso se lo  $Z''$ -Score va al di sotto del valore 0 indica uno stato di criticità per l'impresa. Tutti i pesi delle variabili, dalla  $X_1$  alla  $X_4$ , sono differenti rispetto al modello  $Z$ -Score originale e questo rende differenti anche le medie e i punteggi di *cut-off*. In questo modello riferito ai mercati emergenti si è aggiunto un termine costante il quale ci permette di associare il punteggio di *cut-off* del modello con il valore 0 in modo da paragonarlo al giudizio "D" riferito ad uno stato di default dell'obbligazione. Si ritiene che questo modello sia più appropriato per le imprese manifatturiere rispetto all'originale  $Z$ -Score e si aggiunge che lo sviluppo di modelli per le specifiche imprese si presentano più efficienti nella stima di potenziali pericoli per imprese simili (Altman e Hotchkiss, 2006).

In una più recente ricerca dal titolo *Z-Score Models' Application to Italian Companies Subject to Extraordinary Administration* del 2013, condotta dallo stesso Altman assieme ad Alessandro Danovi e Alberto Falini, si testimonia la consistenza dell'indicatore  $Z''$ -Score se applicato alle imprese italiane soggette ad una procedura di Amministrazione straordinaria, come previsto dal *Decreto Legislativo 270/1999* e dal *Decreto Legge 347/2003*, ovvero in grave stato di crisi. Il campione preso in considerazione comprende imprese che abbiano tutti i bilanci pubblicati nel periodo temporale che va dal 2001 e il 2009 e che posseggano più di 200 lavoratori; Altman, in questo lavoro, riporta che l'utilizzo dell'indice  $Z''$ -Score è stato motivato dalla presenza nel campione di una percentuale molto bassa di imprese pubbliche (pari al 5% del totale) mentre risulta essere estremamente elevata la presenza di società manifatturiere. Sintetizzando i risultati di questa ricerca l'indice  $Z''$ -Score risulta essere estremamente preciso se a questo vengono applicati dei dati ad un anno dal default classificando in maniera corretta il 95,5% delle imprese del campione.

## 1.9 Z-Score per le PMI

Un'altra rielaborazione del modello originale Z-Score, formulato da Altman, che merita di essere riportata è quella condotta da Pietro Bottani, Letizia Cipriani e Francescomaria Serao nell'articolo *Analisi del rischio d'insolvenza di una PMI tramite l'utilizzo del modello dello Z-Score* del 2004. In questa ricerca, seguendo il procedimento suggerito da Altman, si è ricostruito uno specifico Z-Score (per semplicità verrà definito come  $Z_{PMI}$ ) atto a prevedere lo stato di salute o di insolvenza di un'impresa dalle piccole o medie dimensioni.

Proprio come aveva fatto Altman nel 1968, la ricostruzione di questo indice è stata fatta con l'utilizzo di un campione di 66 imprese diviso in 33 imprese fallite nel 2002 e in 33 imprese sane. Tutte queste società sono classificate secondo quanto riporta il decreto ministeriale DM18.9.1997 in recepimento della direttiva n. 96/C 213/4. L'equazione ottenuta per la stima dello stato di salute di una società risulta quindi essere:

$$Z_{PMI} = 1,981X_1 + 9,841X_2 + 1,951X_3 + 3,206X_4 + 4,037X_5$$

dove le variabili da inserire sono quelle già viste per il modello Z'-Score. Ciò che cambia sono i pesi associati a ciascuna variabile e il punteggio che definisce le tre zone in cui può essere classificata l'impresa. Se  $Z > 8,105$  l'impresa risulta essere in ottima condizione strutturale; se  $Z < 4,846$  la società fallirà mentre se  $4,846 < Z < 8,105$  l'impresa si trova in una situazione incerta in merito alla sua "salute finanziaria" futura.

Al fine di poter valutare la capacità discriminante di ciascuna variabile gli autori di questa ricerca hanno svolto un *F-Test* il quale, come descritto in precedenza, mette in relazione le differenze tra le medie degli indici in ciascun gruppo rispetto alla media totale con la variabilità degli indici all'interno di ciascun gruppo. Il risultato di questo test ha messo in evidenza che le variabili aventi una forte differenziazione tra i gruppi sono  $X_1$ ,  $X_2$ ,  $X_3$  e  $X_5$  con un livello di significatività a 0,001. La variabile  $X_4$  invece presenta una differenziazione più lieve tra i gruppi considerati essendo significativa al livello dello 0,01.

Bottani, Cipriani e Serao, procedendo con la standardizzazione delle variabili, hanno inoltre misurato la capacità discriminante di queste variabili; tra queste, quelle che risultano contribuire in maniera più sostanziale alla discriminazione all'interno dei gruppi sono rispettivamente le variabili  $X_5$ ,  $X_4$  e  $X_1$ .

L'analisi dei coefficienti di correlazione indica che le variabili sono scarsamente correlate tra loro mentre, procedendo con il *t-test*, il grado di affidabilità della classificazione operata dall'indice  $Z_{PMI}$  è pari al 99,9%.

Anche nella ricerca di Bottani, Cipriani e Serao si evidenzia che con dati più recenti si ottiene una capacità predittiva migliore infatti più i bilanci considerati si avvicinano alla data del fallimento della società e più le variabili considerate sono predittive della situazione di crisi.

### 1.10 L'equivalenza con i rating obbligazionari (*Bond Rating Equivalents*)

La ragione principale per sviluppare un modello per lo *score* del credito è quello di quantificare la probabilità di *default*, dato un certo livello di rischio stimato. Basilea II prevede che gli approcci basati sui rating interni<sup>11</sup> si riferiscano a delle stime ricavate dall'esperienza del mercato bancario o del capitale (Altman & Hotchkiss, 2006). Altman afferma che si deve essere consci che tutte le agenzie di rating<sup>12</sup>, malgrado non siano perfette nel calcolo del rischio di credito, forniscono comunque delle informazioni fondamentali in merito alla probabilità di default di un'impresa. In aggiunta, se si ha a disposizione delle consistenti serie storiche e un numero sufficientemente alto di default di imprese alle quali era stato associato un rating, si può sfruttare questa informazione, confrontando lo *score* del credito stimato con questi rating storici, e derivare quindi la *Probabilità di Default* (PD) inattesa ed attesa oltre alle perdite che si subiranno nel caso si verificasse il *default* (LGD).

Queste stime possono essere finalizzate all'analisi di un periodo di tempo fisso a partire dalla data di stima del rating (ci si riferisce ad un periodo di tempo che va da  $t - 1$  a  $t$ ) o utilizzando una base cumulata delle PD per una stima della probabilità di default su periodo di tempo maggiore (si riferisce quindi alla probabilità che il default avvenga nel periodo  $t$  o prima, ovvero fino al periodo 0). Queste possono essere desunte dai calcoli condotti dalle agenzie di rating con il metodo del "pool statico"<sup>13</sup> o della "coorte dinamica"<sup>14</sup> (Altman & Hotchkiss, 2006). Un'alternativa è quella di utilizzare l'approccio sul tasso di mortalità teorizzato da Altman nel 1989 calcolato annualmente; questo è basato sul default atteso dalla data originale di emissione e dal rating associato.

In relazione alle imprese sulle quali non vi è a disposizione una valutazione da parte di un'agenzia di rating, si può calcolare uno *score* basato su un generico modello a disposizione e collegarlo con il rating equivalente associato ad un'obbligazione; questo quindi ci consente di stimare la PD.

---

<sup>11</sup> Si fa esplicito riferimento sia ai rating interni "di base" sia ai rating interni "avanzati"

<sup>12</sup> Moody's, S&P, Fitch e Dominion Bond Rating Service

<sup>13</sup> Proprio dell'agenzia di rating "Standard and Poors"

<sup>14</sup> Letteralmente "*dynamic cohort*": metodologia tipica dell'agenzia di rating *Moody's*. Per approfondimenti si veda "Christian Bluhm, Ludger Overbeck, Christoph Wagner, *An Introducing of Credit Risk Modelling*, 2003.



**Un esempio** (da Altman & Hotchkiss - *Corporate Financial Distress and Bankruptcy*, 3e, 2006)

Le tabelle sottostanti (Tabella 2 e Tabella 3) riportano i rating obbligazionari equivalenti per vari intervalli di Z-Score e Z"-Score basati sulle medie delle obbligazioni calcolate nelle loro rispettive categorie di rating. Si può osservare, dalla Tabella 2 che nel periodo che va dal 1996 al 2001 le obbligazioni quotate come AAA hanno uno Z-Score medio pari a 6,2 mentre le obbligazioni quotate B hanno uno score medio pari a 1,8. Si vuole ricordare che 1,8 corrispondeva alla banda superiore della "zona di pericolo" nel modello Z-Score originale. Nel grafico si riporta anche che la mediana dei punteggi delle imprese che sono andate in bancarotta è appena sotto il valore 0 ovvero pari a -0,2. Lo score delle imprese in bancarotta è stato derivato da un campione di 244 aziende tra il periodo che va dal 2000 al 2004.

<b>Z-Score medio stimato da S&amp;P Bond Rating</b>			
	<b>Media annua del n° di imprese</b>	<b>Z-Score medio</b>	<b>Deviazione Standard</b>
<b>AAA</b>	66	6,2	2,06
<b>AA</b>	194	4,73	2,36
<b>A</b>	519	3,74	2,29
<b>BBB</b>	530	2,81	1,48
<b>BB</b>	538	2,38	1,85
<b>B</b>	390	1,8	1,91
<b>CCC</b>	10	0,33	1,16
<b>D</b>	244	-0,2	n.a.

**Tabella 2:** Altman & Hotchkiss, *Corporate Financial Distress and Bankruptcy*, 3e, pag. 247, 2006

A questo punto l'analista dovrebbe osservare la PD associata ai giudizi delle società di *rating* in merito alla rischiosità dei bond. Per esempio, ad un rating B, le società di rating associano una PD in un'intorno del 6% mentre, come riporta Altman (2006), la PD media per un titolo con la stessa fascia di rating un anno dopo l'emissione è pari al 2,85%.

Altman afferma che la PD, stimata secondo il modello da lui teorizzato, associata al tasso di mortalità del primo anno, è di molto inferiore rispetto alla PD derivata da un paniere di Moody's o S&P formato da obbligazioni con rating B contenente titoli di diverse età e scadenze. Con ciò si vuole sottolineare che l'analista deve applicare la stima corretta della PD basata sulle caratteristiche di un portafoglio di dimensioni rilevanti di attività creditizie.

**Equivalenza con i rating obbligazionari degli  
Score dei mercati emergenti**

$$Z'' = 3,35 + 6,56X_1 + 3,36X_2 + 6,72X_3 + 1,05X_4$$

<b>Rating equivalente negli U.S.A.</b>	<b>Media degli Score nei mercati emergenti</b>
AAA	8,15
AA+	7,6
AA	7,3
AA-	7
A+	6,85
A	6,65
A-	6,4
BBB+	6,25
BBB	5,85
BBB-	5,65
BB+	5,25
BB	4,95
BB-	4,75
B+	4,5
B	4,15
B-	3,75
CCC+	3,2
CCC	2,5
CCC-	1,75
D	0

**Tabella 3:**Altman e Hotchkiss, *Corporate Financial Distress and Bankruptcy*, 3e, pag. 248, 2006

### 1.11 Stima la Probabilità di Default mediante l'equivalenza con i rating obbligazionari

Come già menzionato prima, il modello *Z-Score* ci dà una misura nel continuo dello stato di salute dell'impresa ma da questo *score* non si riesce a stimare precisamente la probabilità di default di un'impresa. Malgrado vi siano alcune tecniche statistiche che riescano a fornire un criterio di classificazione, come per esempio i modelli *logit* o *probit*, della probabilità di default definendola solitamente in un intervallo tra compreso tra 0 e 1, Altman preferisce il metodo dei "*bond rating equivalent*" (BRE) i quali si basano sull'esperienza di oltre 2000 imprese andate in default nei precedenti 35 anni (Altman, 2006). A sostegno della propria tesi Altman afferma che una specifica funzione logistica non è adeguata poichè non è detto che essa sia quella più precisa per modellare la stima dei default. Altman inoltre continua dicendo che i modelli

*logit* offrono, se applicati a moltissimi campioni presi in esame, o delle magnitudo di default estremamente alte o estremamente basse.

Come visto in precedenza il BRE è un metodo a tre passi; questi sono:

- 1- Il calcolo di uno score del credito su un nuovo o esistente credito in portafoglio
- 2- Mappatura dello score del credito su un'obbligazione che abbia un rating equivalente
- 3- L'utilizzo di un tasso di mortalità per le nuove emissioni, o una probabilità cumulata di default per le emissioni stagionali in modo da specificare una stima precisa delle probabilità di default per certi periodi di tempo nel futuro.

Nelle tabelle specificate prima (Tabella 2 e Tabella 3) si è visto rispettivamente i BRE per i modelli *Z-Score* e per i modelli *Z''-Score*. Questi erano basati su dei campioni molto recenti di *score* medi per le maggiori classi di obbligazioni.

Basandosi su 35 anni di esperienza di default e calibrandoli sul *rating* obbligazionario originale di un'emissione o sul *rating* iniziale a qualche periodo di misurazione<sup>15</sup>, si può osservare la verosomiglianza storica che una determinata emissione con un dato *score* e il BRE hanno di andare in default in un periodo compreso tra 1 e 10 anni dopo il periodo associato al calcolo dello *score* stesso (Altman & Hotchkiss, 2006). Il tasso di mortalità per il periodo che va dal 1971 al 2004 è riportato nella seguente tabella (Tabella 4). Nella tabella successiva (Tabella 5) invece vengono confrontati i tassi di mortalità cumulativi individuati con il criterio di Altman e dai *rating* di agenzia<sup>16</sup>.

Si può osservare che l'emissione di un nuovo titolo con *Z-Score* uguale a 1,8 (Tabella 2) ha un BRE pari a B e una probabilità di default, un anno dopo l'emissione, di 2,85% secondo le stime condotte da Altman, mentre sono pari a 5,81% e 5,71% (Tabella 5) rispettivamente le probabilità di default date da *Moody's* ed *S&P*.

---

<sup>15</sup> Si fa riferimento ai tassi di default cumulati di *Moody's* e *S&P*

<sup>16</sup> Anche in questo caso si fa riferimento ai tassi annuali cumulativi dei default stilati dalle agenzie *Moody's* e *S&P*

		1	2	3	4	5	6	7	8	9	10
<b>Tasso di mortalità ricavati dai rating originari su tutte le obbligazioni aziendali soggette a stima del rating: periodo temporale 1971-2004</b>											
<b>AAA</b>	Marginale	0,00%	0,00%	0,00%	0,00%	0,03%	0,00%	0,00%	0,00%	0,00%	0,00%
	Cumulato	0,00%	0,00%	0,00%	0,00%	0,03%	0,03%	0,03%	0,03%	0,03%	0,03%
<b>AA</b>	Marginale	0,00%	0,00%	0,32%	0,16%	0,03%	0,03%	0,00%	0,00%	0,03%	0,02%
	Cumulato	0,00%	0,00%	0,32%	0,48%	0,51%	0,54%	0,54%	0,54%	0,57%	0,59%
<b>A</b>	Marginale	0,01%	0,10%	0,02%	0,09%	0,06%	0,11%	0,06%	0,21%	0,11%	0,06%
	Cumulato	0,01%	0,11%	0,13%	0,22%	0,28%	0,39%	0,45%	0,65%	0,76%	0,82%
<b>BBB</b>	Marginale	0,36%	3,22%	1,43%	1,28%	0,77%	0,45%	0,20%	0,20%	0,14%	0,40%
	Cumulato	0,36%	3,56%	4,49%	6,16%	6,89%	7,31%	7,50%	7,68%	7,87%	8,18%
<b>BB</b>	Marginale	1,19%	2,48%	4,40%	2,01%	2,51%	1,16%	1,60%	0,88%	1,70%	3,60%
	Cumulato	1,19%	3,64%	7,88%	9,74%	12%	12,93%	14,36%	15,07%	16,52%	19,60%
<b>B</b>	Marginale	2,85%	6,85%	7,40%	8,55%	6%	4,16%	3,72%	2,28%	1,96%	0,86%
	Cumulato	2,85%	9,51%	16,20%	23,37%	27,94%	30,96%	33,46%	34,97%	36,25%	36,80%
<b>CCC</b>	Marginale	7,98%	15,57%	19,55%	12,10%	4,26%	9,45%	5,60%	3,15%	0,00%	4,28%
	Cumulato	7,98%	22,31%	37,50%	45,06%	47,37%	52,35%	55,01%	56,43%	56,43%	58,30%

**Tabella 4:** Rating svolto da S&P al momento dell'emissione: si basa su 1796 emissioni in seguito andate in default  
**Fonte:** Altman & Hotchkiss - *Corporate Financial Distress and Bankruptcy*, 3e, pag.250, 2006

Confronto dei tassi di default cumulati (in percentuale per un massimo di 10 anni)											
		1	2	3	4	5	6	7	8	9	10
AAA/Aaa	Altman	0.00	0.00	0.00	0.00	0.03	0.03	0.03	0.03	0.03	0.03
	Moody's	0.00	0.00	0.00	0.04	0.12	0.21	0.30	0.41	0.52	0.63
	S&P	0.00	0.00	0.03	0.06	0.10	0.17	0.24	0.36	0.41	0.45
AA/Aa	Altman	0.00	0.00	0.32	0.48	0.51	0.54	0.54	0.59	0.57	0.59
	Moody's	0.00	0.00	0.03	0.12	0.20	0.29	0.37	0.47	0.54	0.61
	S&P	0.01	0.04	0.09	0.19	0.30	0.41	0.54	0.64	0.74	0.85
A/A	Altman	0.01	0.11	0.13	0.22	0.28	0.39	0.45	0.65	0.76	0.82
	Moody's	0.02	0.08	0.22	0.36	0.50	0.67	0.85	1.04	1.25	1.48
	S&P	0.04	0.13	0.24	0.40	0.61	0.84	1.11	1.34	1.63	1.94
BBB/Baa	Altman	0.36	3.56	4.49	6.16	6.89	7.31	7.50	7.68	7.87	8.18
	Moody's	0.19	0.54	0.98	1.55	2.08	2.59	3.12	3.65	4.25	4.89
	S&P	0.29	0.81	1.40	2.19	2.99	3.73	4.34	4.95	5.50	6.10
BB/Ba	Altman	1.19	3.64	7.88	9.74	12.00	12.93	14.36	15.07	16.52	19.60
	Moody's	1.22	3.34	5.79	8.27	10.72	12.98	14.81	16.64	18.40	20.11
	S&P	1.20	3.58	6.39	8.97	11.25	13.47	15.25	16.75	18.16	19.20
B/B	Altman	2.85	9.51	16.20	23.37	27.94	30.96	33.46	34.97	36.25	36.80
	Moody's	5.81	12.93	19.51	25.33	30.48	35.10	39.45	42.89	45.89	48.64
	S&P	5.71	12.49	18.09	22.37	25.40	27.77	29.76	31.32	32.54	33.75
CCC/Caa	Altman	7.98	22.31	37.50	45.06	47.37	52.35	55.01	56.43	56.43	58.30
	Moody's	22.43	35.96	46.71	54.19	59.72	64.49	68.06	71.91	74.53	76.77
	S&P	28.83	37.97	43.52	47.44	50.85	52.13	53.39	54.05	55.56	56.45

**Tabella 5:** Altman e Hotchkiss, *Corporate Financial Distress and Bankruptcy*, 3e, pag. 174-175, 2006

Il calcolo delle PD è di fondamentale importanza poiché esse costituiscono il principale input per gli IRB (*Internal Rating Based*) interni od esterni richiesti da Basilea II e per la valutazione di ogni attività creditizia. I flussi di cassa intrinseci nell'emissione dell'obbligazione possono quindi essere aggiustati per la probabilità di default e per il tasso di recupero in modo da poter stimare i flussi di cassa attesi. Il tasso di sconto quindi, nel calcolo del valore attuale, può essere confrontato con il tasso privo di rischio; il primo infatti presenta un premio aggiuntivo basato sulla perdita inattesa delle obbligazioni che si inseriscono in quella specifica classe di rating.

## 1.12 La comparazione dei modelli

Alcuni studi hanno analizzato l'accuratezza delle diverse tecniche statistiche per predire la bancarotta delle imprese: per esempio Altman sottolinea l'importanza della ricerca fatta da Mark E. Zmijewski chiamata *Methodological Issues Related to the Estimation of Financial Distress Prediction Models*, 1984 (Altman & Hotchkiss, 2006). In aggiunta si cita anche lo studio *An empirical comparison of bankruptcy Models* condotto da Charles E. Mossman, Geoffrey G. Bell, L. Mick Swartz, Harry Turtle nel 1998 nel quale si va a comparare 4 diversi tipi di modelli in merito alla loro capacità predittiva sulla salute finanziaria di un'impresa in vari periodi prima del fallimento. Questi, nella loro formulazione, si presentavano molto restrittivi in

quanto il primo presentava solamente rapporti di variabili, il secondo solamente variabili rappresentanti dei flussi di cassa, il terzo esclusivamente variabili esplicitanti i tassi di rendimento mentre l'ultimo delle variabili che rappresentavano le varianze dei rendimenti. In questo studio si è scoperto che ad un anno dal fallimento, nessuno dei 4 modelli può essere escluso senza perdere parte del potere esplicativo. Altman afferma quindi che questo risultato è coerente con la sua argomentazione secondo la quale un modello multivariato è sicuramente migliore rispetto a qualsiasi modello univariato o a singola variabile (Altman & Hotchkiss, 2006).

Nello studio di Mossman si deve sottolineare anche che, mentre le variabili rappresentanti i flussi di cassa dominano da due a tre anni prima della bancarotta, le variabili sotto forma di rapporto si sono mostrate più efficienti nella previsione ad un anno prima della bancarotta.

### **1.13 La stima della *Loss Given Deafault* (Perdite date dal default)**

Molti dei modelli di analisi del rischio di credito e tutti i modelli *VaR* (per esempio la *Credit Metrics*) si basano sull'assunto di indipendenza tra la PD e il tasso di recupero dei debiti in stato di default (Altman & Hotchkiss, 2006). Nello studio *The Link Between Default and Recovery Rates: Theory, Empirical Evidence and Implications* condotto da Altman, Brady, Resti e Sironi (2005) gli autori sono riusciti a dimostrare che l'assunzione sopracitata non è corretta e hanno simulato l'impatto sui requisiti patrimoniali quando si considera una correlazione negativa significativa tra la PD e i tassi di recupero per un certo periodo di tempo. Andando sul particolare, si spiega che in periodi costituiti da alti tassi di default delle obbligazioni, il tasso di recupero è basso (relativamente alla media storica) e le perdite attese risultano essere maggiori. Se si prendesse ad esempio gli anni 2001 e 2002, le capacità di recupero dei titoli obbligazionari (ovvero i prezzi dopo i default) erano rispettivamente del 26,4% e del 25,5%. Il contrario invece accade quando i tassi di default sono relativamente bassi: si fa specifico riferimento agli anni 2003, 2004 quando i tassi di recupero crebbero e si assestarono in un intervallo compreso tra il 45% e il 57%. Hu e Perraudin nel *The Dependence of Recovery Rates and Defaults* (2002) trovarono dei risultati molto simili mentre in uno studio di Frye chiamato *Depressing Recoveries* (ottobre 2000) si testimonia che vi è una notevole influenza macroeconomica nei tassi di recupero.

Questa prociclicità dell'approccio basato sui rating, il quale è peraltro affrontata anche da Basilea II, ha causato una certa preoccupazione nelle banche centrali. A seguito di queste

scoperte dal 2004 il Comitato di Basilea iniziò ad analizzare la “*LGD in downturns*<sup>17</sup>” in modo da valutare gli effetti di questa correlazione. Da allora gli investitori nelle obbligazioni societarie ad alta rischiosità o nelle obbligazioni collateralizzate (*CDO*) devono essere informati del fatto che i possibili recuperi delle perdite possono essere molto inferiori nei periodi temporali caratterizzati da molti default.

Basilea II quindi ha dato la spinta a varie banche centrali e a diverse istituzioni verso la costruzione e la valutazione dei modelli sul rischio di credito i quali si sono concretizzati nelle tecniche di *scoring* del credito, nella stima delle perdite a seguito dei default e negli approcci di portafoglio per affrontare i problemi legati al rischio di credito.

---

<sup>17</sup> Per “*LGD in downturns*” letteralmente si fa riferimento alle perdite attese a seguito del default avvenuto in un momento in cui l’economia è in recessione

## Capitolo 2: Approfondimento quantitativo sull'analisi discriminante e la regressione logistica

### 2.1 Elementi di Analisi statistica discriminante

L'analisi che seguirà è tratta principalmente dalla Tesi di Laurea in Matematica dal titolo *Analisi discriminante, regressione logistica e reti per la stima delle probabilità di fallimento. Applicazioni ed estensioni del metodo Z-score* di Jacopo Iannucci, 2004.

L'analisi discriminante si configura in concreto con la presenza di due classi predefinite  $D_1$  e  $D_0$  e con i vettori di osservazioni  $x_1, x_2, \dots, x_n$  relativi a un numero  $n$  di soggetti di una popolazione. A questo punto si deve costruire una funzione, chiamata *regola d'assegnazione*, la quale ha l'obiettivo di associare ciascun vettore  $x_i$ , sulla base di un coerente criterio discriminatorio, la classe  $D_1$  o  $D_0$ .

Si supponga di dover effettuare uno studio sulla probabilità che un certo numero di individui affidati da una banca, riescano a risanare il proprio debito. Per ognuno di questi soggetti si possono prendere a riferimento due "fattori" come il reddito e i beni che possono fungere da garanzia; questi saranno riassunti in un vettore numerico  $\underline{x}$ . Si ha quindi la necessità di avere un modello classificatore (in questo caso sarà l'analisi discriminante) il quale farà corrispondere a ciascun vettore  $\underline{x}_i$  o la classe

$$D_1 = \{\text{individui che ripagheranno i debiti con la banca}\}$$

o la classe alternativa

$$D_0 = \{\text{individui che non ripagheranno i debiti con la banca}\}.$$

Si parte da un *dataset* costituito da  $N$  esempi di oggetti pre-classificati suddivisi in due sottoinsiemi formati rispettivamente da  $n$  elementi e  $k$  elementi in modo che  $N = n + k$ . Tale suddivisione permette di utilizzare il primo gruppo per creare la regola di classificazione mentre il secondo lo si utilizza per verificare la fondatezza del modello da noi utilizzato. In questo caso il gruppo formato dagli  $n$  elementi prende il nome di *training set* mentre l'altro prende il nome di *validation set*.

L'analisi discriminante è un metodo parametrico dove quindi si deve identificare dei parametri con il quale costruire la soglia di discriminazione per poter in seguito svolgere la classificazione.

L'obiettivo principale dell'analisi discriminante è quello di poter classificare un generico individuo in uno dei gruppi  $D_1, D_2, \dots, D_g$  basandosi su un vettore di informazioni  $\underline{x} = (x_1, x_2, \dots, x_p)$ .



Si supponga di avere 2 gruppi ( $g = 2$ ) che identificano l'insieme di imprese fallite e quello delle imprese solventi e di poter osservare per ogni individuo 3 variabili ( $p = 3$ ), che, in questo specifico esempio sono:  $x_1 =$  stipendio,  $x_2 =$  garanzie e  $x_3 =$  tipologia di prestito richiesto.

La tipologia di analisi di seguito illustrata è l'analisi discriminante di Fischer applicata al caso in cui vi siano due gruppi ( $g = 2$ ) e  $p$  variabili. L'obiettivo è quello di trovare una combinazione lineare  $\underline{\alpha}^t \underline{x}$  delle  $p$  variabili  $\underline{x}$  che riesca a separare i due gruppi di individui. Questo può essere svolto determinando un vettore  $\underline{\alpha}$  in modo da rendere massima la distanza tra le medie dei gruppi e allo stesso tempo minima la varianza all'interno di ciascun gruppo. Il vettore dei pesi  $\underline{\alpha}$  è calcolato come il vettore che massimizza il rapporto

$$J(\underline{\alpha}) = Q_B(\underline{\alpha})/Q_W(\underline{\alpha}) \quad [2.1.1]$$

dove

$$Q_B(\underline{\alpha}) = \underline{\alpha}^t \hat{S}_B \underline{\alpha} \quad [2.1.2]$$

e

$$Q_W(\underline{\alpha}) = \underline{\alpha}^t \hat{S}_W \underline{\alpha}. \quad [2.1.3]$$

Essendo

$$\hat{\underline{x}}_1 = (\hat{x}_{11}, \hat{x}_{12}, \dots, \hat{x}_{1p}) = \left( \frac{1}{n_1} \sum_{k=1}^{n_1} x_{k1}^{(1)}, \frac{1}{n_1} \sum_{k=1}^{n_1} x_{k2}^{(1)}, \dots, \frac{1}{n_1} \sum_{k=1}^{n_1} x_{kp}^{(1)} \right) \quad [2.1.4]$$

$$\hat{\underline{x}}_0 = (\hat{x}_{01}, \hat{x}_{02}, \dots, \hat{x}_{0p}) = \left( \frac{1}{n_0} \sum_{k=1}^{n_0} x_{k1}^{(0)}, \frac{1}{n_0} \sum_{k=1}^{n_0} x_{k2}^{(0)}, \dots, \frac{1}{n_0} \sum_{k=1}^{n_0} x_{kp}^{(0)} \right) \quad [2.1.5]$$

I vettori delle medie con  $\hat{x}_{ij} =$  media campionaria all'interno del  $i$ -esimo gruppo del  $j$ -esimo fattore,  $n = n_1 + n_0$  dove  $n_1$  e  $n_0$  sono le dimensioni campionarie dei due gruppi in cui si vuole suddividere gli  $N$  oggetti.

Le componenti  $Q_B(\underline{\alpha})$  e  $Q_W(\underline{\alpha})$  sono le forme quadratiche associate rispettivamente  $\hat{S}_B$  e  $\hat{S}_W$  con

$$\hat{S}_B = (\hat{\underline{x}}_0 - \hat{\underline{x}}_1)(\hat{\underline{x}}_0 - \hat{\underline{x}}_1)^t \quad [2.1.6]$$

$\hat{S}_B$  è la matrice delle differenze al quadrato fra le medie dei gruppi. Di seguito la componente  $\hat{S}_W$  è così composta:

$$\hat{S}_W = \frac{1}{n_1 + n_0 - 2} [n_1 \hat{S}_{W1} + n_0 \hat{S}_{W0}] \quad [2.1.7]$$

con

$$\hat{S}_{W1} = \sum_{i=1}^{n1} (x_i - \hat{x}_1)(x_i - \hat{x}_1)^t \quad [2.1.8]$$

e

$$\hat{S}_{W0} = \sum_{i=n1+1}^{n0} (x_i - \hat{x}_0)(x_i - \hat{x}_0)^t \quad [2.1.9]$$

Si può quindi dimostrare che la funzione [2.1.1] è massimizzata per  $\underline{\alpha} =$  *massimo autovalore* di

$$\hat{S}_W^{-1} \hat{S}_B$$

Dal quale si ottiene

$$\underline{\alpha} \propto \hat{S}_W^{-1} [(\hat{x}_1 - \hat{x}_0)]$$

Dopo aver svolto il calcolo dei pesi associati ai vari fattori basandosi sui dati contenuti nel *training set*, un nuovo soggetto del quale si dispone il vettore delle osservazioni  $\underline{x} = (x_1, x_2, \dots, x_p)$  è collocato nella classe  $D_1$  se:

$$\underline{\alpha}^t [\underline{x} - \frac{1}{2}(\hat{x}_1 + \hat{x}_0)] > 0 \quad [2.1.10]$$

$\underline{\alpha}$  infatti è descritta dalla seguente equazione:

$$\underline{\alpha} = \hat{S}_W^{-1} (\hat{x}_1 - \hat{x}_0) \quad [2.1.11]$$

mentre la funzione

$$y = \underline{a}^t \underline{x} \quad [2.1.12]$$

è anche definita come la *funzione discriminante lineare* e l'iperpiano

$$\alpha^t [\underline{x} - \frac{1}{2}(\hat{x}_1 + \hat{x}_0)] = 0 \quad [2.1.13]$$

viene definito anche come *iperpiano separatore* o *decision boundary* (Iannucci, 2004).

L'analisi statistica discriminante può essere per esempio utilizzata in ambito medico/clinico; se all'interno di un certo numero di pazienti si volesse discriminare tra quelli che hanno una buona possibilità di contrarre il diabete e tra quelli che non lo contrarranno allora in numero dei gruppi necessari per attuare la discriminazione sono solamente due ( $g = 2$ ) (Iannucci, 2004). Si può ipotizzare che la tipologia di osservazioni da utilizzare per poter effettuare questa discriminazione si basi sulla misurazione del glucosio nel sangue, del peso ed dell'età ( $p = 3$ ); quindi:

$x_1 = \text{concentrazione del glucosio nel sangue}$

$x_2 = \text{peso}$

$x_3 = \text{età}$ .

Generalmente tali variabili, rilevate per ciascun individuo, vengono chiamate fattori.

In generale, se a ciascun gruppo  $D_j$  è associata una certa densità di probabilità che descrive la probabilità che i fattori appartengano ad un certo gruppo  $j$  della forma  $f_j(\underline{x})$  dove  $\underline{x} = (x_1, x_2, \dots, x_p)$ , per le variabili osservate una metodologia per il processo di discriminazione potrebbe basarsi sulla seguente relazione

$$f_{i0}(\underline{x}) = \max\{f_1(\underline{x}), f_2(\underline{x}), \dots, f_g(\underline{x})\} \quad [2.1.14]$$

dove il soggetto che presenta le caratteristiche contenute nel vettore  $\underline{x}$  è collocato nel gruppo  $D_{i0}$ . Per rendere più chiara questa nostra supposizione si può far ricorso ad alcuni esempi riportati nell'appendice (*Esempio 1* ed *Esempio 2*) tratti dalla ricerca di Iannucci dal nome *Analisi discriminante, regressione logistica e reti per la stima delle probabilità di fallimento. Applicazioni ed estensioni del metodo Z-score* del 2004.

## 2.2 Caso gaussiano univariato

Prendendo come punto di riferimento il caso gaussiano è possibile operare un'analisi dettagliata. Si supponga che  $X$  sia una variabile continua e che  $D_0$  e  $D_1$  siano i due gruppi con i quali operare l'analisi discriminante. In  $D_1$  la variabile  $X$  ha una distribuzione normale avente media  $\mu_1$  e varianza  $\sigma_1^2$ , mentre in  $D_0$  la variabile  $X$  ha una distribuzione normale con media  $\mu_0$  e varianza  $\sigma_0^2$ ; si supponga inoltre che  $\mu_1 > \mu_0$  e  $\sigma_1 > \sigma_0$ ; seguendo la regola [2.1.14], un soggetto per il quale si può osservare la realizzazione  $x$  verrà assegnato a  $D_1$  se (Iannucci, 2004):

$$f_1(x) > f_0(x) \quad [2.2.1]$$

Dove  $f_1(x) = f(x | Y = 1)$  e  $f_0(x) = f(x | Y = 0)$  sono i valori delle funzioni di densità condizionate valutate in  $x$  (Elena Stanghellini, *Introduzione ai metodi statistici per il credit scoring*, 2009).

Con semplici calcoli algebrici è possibile dimostrare che vale questa relazione se e soltanto se:

$$\frac{\sigma_1}{\sigma_0} \exp \left\{ -\frac{1}{2} \left[ \frac{(x - \mu_1)^2}{\sigma_1^2} - \frac{(x - \mu_0)^2}{\sigma_0^2} \right] \right\} > 1 \quad [2.2.2]$$

Passando alla funzione logaritmica ed eseguendo dei semplici calcoli algebrici si perviene quindi alla formula:

$$x^2 \left[ \frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right] - 2x \left[ \frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2} \right] + \left[ \frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2} \right] - 2 \ln \frac{\sigma_1}{\sigma_0} < 0 \quad [2.2.3]$$

Se questa disequazione è verificata allora il soggetto per il quale è stato osservato il valore  $x$  sarà assegnato al gruppo  $D_1$  altrimenti viene inserito nel gruppo  $D_0$  (Iannucci, 2004).

Se per ipotesi si avesse  $\sigma_1 = \sigma_0 = \sigma$  la funzione discriminante diventa di tipo lineare e quindi la formula precedente, sempre procedendo attraverso semplici passaggi algebrici diventerebbe:

$$-2x \left[ \frac{\mu_1}{\sigma^2} - \frac{\mu_0}{\sigma^2} \right] + \left[ \frac{\mu_1^2}{\sigma^2} - \frac{\mu_0^2}{\sigma^2} \right] < 0 \quad [2.2.4]$$

Da questo risultato si può quindi supporre che  $\mu_1 > \mu_0$  e quindi dalla disequazione appena descritta si può ricavare la condizione di appartenenza alla classe  $D_0$ :

$$-\frac{2x}{\sigma^2}(\mu_1 - \mu_0) + \frac{1}{\sigma^2}(\mu_1^2 - \mu_0^2) < 0 \quad [2.2.5]$$

ed essendo:

$$(\mu_1^2 - \mu_0^2) = (\mu_1 - \mu_0)(\mu_1 + \mu_0) < 2x(\mu_1 - \mu_0) \quad [2.2.6]$$

risulta quindi:

$$x > \frac{1}{2}(\mu_1 + \mu_0). \quad [2.2.7]$$

Per i valori di  $x$  che risolvono:

$$x^2 \left[ \frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right] - 2x \left[ \frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2} \right] + \left[ \frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_0^2}{\sigma_0^2} \right] - 2 \ln \frac{\sigma_1}{\sigma_0} = 0 \quad [2.2.8]$$

L'insieme di questi valori di  $x$  è chiamato *decision boundary* ed è di tipo quadratico infatti è definito graficamente attraverso una parabola con due radici reali e da una concavità rivolta verso il basso. L'asse  $\mathbb{R}$  viene quindi diviso in due parti in cui una si presenta limitata e connessa mentre la seconda si presenta illimitata e sconnessa.

### 2.3 Caso multivariato

Nel gruppo  $D_1$  al vettore di variabili aleatorie  $\underline{x}$  viene assegnata una distribuzione normale multivariata che presenta vettore di medie  $\underline{\mu}_1$  e matrice di covarianza  $\Sigma$  mentre nell'gruppo  $D_0$  sempre una distribuzione normale multivariata con media  $\underline{\mu}_0$  e uguale matrice di covarianza  $\Sigma$ . Utilizzando sempre la regola [2.1.14], ad un individuo viene assegnata la classe  $D_1$  solamente se vale la relazione (Iannucci, 2004):

$$f_1(\underline{x}) > f_0(\underline{x}) \quad [2.3.1]$$

Dove  $f_1(\underline{x}) = f(\underline{x} | Y = 1)$  e  $f_0(\underline{x}) = f(\underline{x} | Y = 0)$  sono i valori delle funzioni di densità condizionate valutate in  $\underline{x}$

Questa, a sua volta, corrisponde alla seguente notazione (Iannucci, 2004):

$$\exp -\frac{1}{2} \left[ (\underline{x} - \underline{\mu}_1)^t \Sigma^{-1} (\underline{x} - \underline{\mu}_1) - (\underline{x} - \underline{\mu}_0)^t \Sigma^{-1} (\underline{x} - \underline{\mu}_0) \right] > 1 \quad [2.3.2]$$

la quale, se le venisse applicato l'operatore logaritmo, diverrebbe:

$$-\frac{1}{2} \left[ (\underline{x} - \underline{\mu}_1)^t \Sigma^{-1} (\underline{x} - \underline{\mu}_1) - (\underline{x} - \underline{\mu}_0)^t \Sigma^{-1} (\underline{x} - \underline{\mu}_0) \right] > 0 \quad [2.3.3]$$

$$\Leftrightarrow \frac{1}{2} \underline{x}^t (\Sigma^{-1} - \Sigma^{-1}) \underline{x} + (\Sigma^{-1} \underline{\mu}_1 - \Sigma^{-1} \underline{\mu}_0)^t \underline{x} + \frac{1}{2} (\underline{\mu}_0^t \Sigma^{-1} \underline{\mu}_0 - \underline{\mu}_1^t \Sigma^{-1} \underline{\mu}_1) > 0 \quad [2.3.4]$$

Da ciò, sapendo che:

$$\underline{x}^t (\Sigma^{-1} - \Sigma^{-1}) \underline{x} = 0 \quad [2.3.5]$$

si può quindi affermare che la regola di discriminazione ricavata dall'assunto [2.1.14] ci consente di assegnare un individuo con vettore di osservazioni  $\underline{x}$  a  $D_1$  se:

$$\underline{\alpha}^t (\underline{x} - \underline{\mu}) > 0 \quad [2.3.6]$$

dove:

$$\underline{\alpha} = \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_0) \quad [2.3.7]$$

$$\underline{\mu} = \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_0) \quad [2.3.8]$$

Mente l'equazione:

$$\underline{\alpha}^t (\underline{x} - \underline{\mu}) = 0 \quad [2.3.9]$$

va a definire un iperpiano nello spazio  $\mathbb{R}^p$  dei  $p$  fattori il quale prende il nome di *decision boundary* (Iannucci, 2004).

A volte però si ritiene adeguato fare l'assunzione che i membri di alcuni gruppi siano più facilmente osservabili di altri appartenenti a gruppi diversi<sup>18</sup>. Se i gruppi hanno probabilità a priori  $\rho_1, \rho_2, \dots, \rho_g$ , la regola  $f_{i0}(\underline{x}) = \max \{f_1(\underline{x}), f_2(\underline{x}), \dots, f_g(\underline{x})\}$  cambia in modo da collocare un individuo che abbia un certo vettore di osservazioni  $x$  nel gruppo di popolazione per il quale

$$\rho_j f_j(\underline{x}) \quad [2.3.10]$$

è massimo (Iannucci, 2004). Se si hanno a disposizione due gruppi caratterizzate da densità multivariate normali aventi una comune matrice di covarianza, la considerazione delle distribuzioni a priori cambia la regola [2.1.14] in

$$\underline{\alpha}^t(\underline{x} - \underline{\mu}) > \ln \frac{\rho_0}{\rho_1} \quad [2.3.11]$$

Spesso però le probabilità a priori non sono note ma se il campione è estratto casualmente è possibile stimare queste probabilità con le corrispondenti frequenze relative osservate nel campione (Stanghellini, 2009).

## 2.4 Analisi discriminante classica

Nel caso visto in precedenza si sono prese delle variabili distribuite in maniera normale e quindi si presuppone la conoscenza dei primi due momenti della distribuzione di queste variabili. Questo, nella pratica, può comportare alcuni problemi dal momento che non sempre è possibile avere dei dati per poter stimare questi valori ma devono essere ricavati da dei valori campionari (Iannucci, 2004).

Nell'esempio precedente in cui si avevano a disposizione due gruppi ( $D_1$  e  $D_2$ ) e  $p$  fattori osservati che si distribuiscono con delle densità normali multivariate, aventi medie diverse ma identica matrice di covarianza, allora i valori  $\underline{\mu}_1 = (\mu_{11}, \mu_{12}, \dots, \mu_{1p})$ ,  $\underline{\mu}_0 = (\mu_{01}, \mu_{02}, \dots, \mu_{0p})$  e  $\Sigma$  possono essere sostituiti dai corrispondenti valori campionari  $\hat{x}_1, \hat{x}_0$  e  $\hat{S}$  nella regola [2.3.11]. Se si ha a disposizione un certo set di dati formato da  $N = n + k$  individui lo si può quindi scomporre in due parti:

---

<sup>18</sup> Si prende ad esempio, in ambito medico, il caso del raffreddore, una malattia molto più comune rispetto, ad esempio, alla poliomelite. La prima ha quindi una probabilità a priori molto più alta di essere osservata. Questo esempio è preso da *Analisi discriminante, regressione logistica e reti per la stima delle probabilità di fallimento. Applicazioni ed estensioni del metodo Z-score*, tesi di laurea in matematica di Jacopo Iannucci.

1. Un *training set*  $X = X_1 \cup X_0$  dove  $X_1$  (*training set* formato da individui che appartengono alla classe  $D_1$ ) e  $X_0$  (*training set* formato da individui che appartengono alla classe  $D_0$ ) hanno cardinalità rispettivamente  $n_1$  e  $n_0$  dalle quali  $n_1 + n_0 = n$ . Questo *training set* serve per poter stimare il vettore dei pesi  $\underline{\alpha}$  del modello.
2. Un *validation set* che presenta una cardinalità pari a  $k$  il quale ci serve per dimostrare la validità del modello.

Con queste considerazioni si può quindi definire:

$$\hat{\underline{x}}_1 = (\hat{x}_{11}, \hat{x}_{12}, \dots, \hat{x}_{1p}) = \left( \frac{1}{n_1} \sum_{k=1}^{n_1} x_{k1}^{(1)}, \frac{1}{n_1} \sum_{k=1}^{n_1} x_{k2}^{(1)}, \dots, \frac{1}{n_1} \sum_{k=1}^{n_1} x_{kp}^{(1)} \right)$$

$$\hat{\underline{x}}_0 = (\hat{x}_{01}, \hat{x}_{02}, \dots, \hat{x}_{0p}) = \left( \frac{1}{n_0} \sum_{k=1}^{n_0} x_{k1}^{(0)}, \frac{1}{n_0} \sum_{k=1}^{n_0} x_{k2}^{(0)}, \dots, \frac{1}{n_0} \sum_{k=1}^{n_0} x_{kp}^{(0)} \right)$$

Con  $\hat{x}_{ij}$  = *media campionaria all'interno del i – esimo gruppo del j – esimo fattore* mentre la matrice di covarianza  $\hat{S}_w$ , se si avesse scelto due gruppi, è data da:

$$\hat{S}_w = \frac{1}{n_1 + n_0 - 2} [n_1 \hat{S}_{w1} + n_0 \hat{S}_{w0}] \quad [2.4.1]$$

sapendo che:

$$\hat{S}_{w1} = \sum_{i=1}^{n_1} (x_i - \hat{\underline{x}}_1)(x_i - \hat{\underline{x}}_1)^t$$

$$\hat{S}_{w0} = \sum_{i=n_1+1}^{n_1+n_0} (x_i - \hat{\underline{x}}_0)(x_i - \hat{\underline{x}}_0)^t$$

dove  $n_1$  e  $n_0$  rappresentano le dimensioni campionarie dei due gruppi mentre  $\hat{S}_{w1}$  e  $\hat{S}_{w0}$  sono le matrici di covarianza del gruppo<sup>19</sup>. In questo caso la regola di classificazione colloca un individuo per il quale è stato osservato  $\underline{x} = (x_1, x_2, \dots, x_p)$  nel gruppo  $D_1$  se:

$$\underline{\alpha}^t \left[ \underline{x} - \frac{1}{2} (\hat{\underline{x}}_1 + \hat{\underline{x}}_0) \right] > 0. \quad [2.4.2]$$

<sup>19</sup> Per ulteriori chiarimenti si può andare ad osservare Elena Stanghellini, *Introduzione ai metodi statistici per il credit scoring*, 2009



Ciò corrisponde a

$$\underline{\alpha}^t \underline{x} > \frac{1}{2}(\hat{x}_1 + \hat{x}_0) \quad [2.4.3]$$

dove si sa che

$$\underline{\alpha} = \hat{S}_W^{-1}(\hat{x}_1 - \hat{x}_0). \quad [2.4.4]$$

Sapendo che la seguente equazione

$$y = \underline{\alpha}^t \underline{x} \quad [2.4.5]$$

rappresenta la *funzione discriminante lineare*, allora l'iperpiano

$$\underline{\alpha}^t \underline{x} - \frac{1}{2}(\hat{x}_1 + \hat{x}_0) = 0 \quad [2.4.6]$$

viene chiamato *iperpiano separatore* o *decision boundary*.

Trovare una combinazione lineare  $\underline{\alpha}^t \underline{x}$  delle  $p$  variabili  $\underline{x}$  in modo da dividere in modo sensato due campioni di prova fu inizialmente un'idea proposta da Fisher nel 1963 (Iannucci, 2004). Questa si basava sulla determinazione di un vettore  $\underline{\alpha}$  che rendesse massima la distanza tra le medie dei gruppi e allo stesso tempo minima la varianza interna a ciascun gruppo. La soluzione a questo problema è rappresentata dalla funzione [2.4.4]<sup>20</sup>.

## 2.5 Valutazione della correttezza del modello dato dall'analisi discriminante

Elena Stanghellini nel volume *Introduzione ai metodi statistici per il credit scoring*, pubblicato nel 2009 da Springer-Verlag Italia affronta, nel contesto dell'analisi discriminante, il problema della selezione delle variabili la cui capacità discriminante sia elevata per escludere invece quelle la cui influenza risulta essere minima<sup>21</sup>. La procedura per la selezione del modello si

<sup>20</sup> La dimostrazione di questa affermazione la si può trovare nella tesi di laurea in matematica "*Analisi discriminante, regressione logistica e reti per la stima delle probabilità di fallimento. Applicazioni ed estensioni del metodo Z-score*", di Iacopo Iannucci, relatore Prof. Alessandro Ramponi, 2004.

<sup>21</sup> Questo problema si presenta anche quando si ha a che fare con il modello logistico.

differenziano a seconda della natura delle variabili della funzione discriminante; se viene accettata l'ipotesi di normalità allora queste si basano su test statistici. Per poter svolgere una prima analisi ci si può basare su test univariati quale il *T-student* per testare l'uguaglianza tra le medie di due popolazioni.

Se  $\mu_{0j}$  e  $\mu_{1j}$  sono i valori attesi di  $X_j$  nelle due popolazioni e  $\sigma_{0j}^2$  e  $\sigma_{1j}^2$  la varianza nelle due popolazioni. Ponendo  $\sigma_{0j}^2 = \sigma_{1j}^2$  questo test ha come ipotesi nulla  $H_0: \mu_{0j}^2 = \mu_{1j}^2$  contro l'ipotesi alternativa  $H_1: \mu_{0j}^2 \neq \mu_{1j}^2$ . La statistica  $t$  si presenta nella seguente forma:

$$t = \frac{\bar{x}_{1j} - \bar{x}_{0j}}{\sqrt{s_{jj} \left( \frac{1}{n_0} + \frac{1}{n_1} \right)}} \quad [2.5.1]$$

dove  $s_{jj}$  rappresenta il  $j$ -esimo elemento della diagonale principale della matrice di covarianza  $\hat{S}_W$  la quale ha distribuzione T di Student con  $n_0 + n_1 - 2$  gradi di libertà. L'ipotesi nulla  $H_0$  viene rifiutata se il valore osservato di  $t$  è, in modulo, molto elevato. Qualora la numerosità risultasse essere molto elevata, questo test può essere utilizzato anche quando l'ipotesi di normalità non è presente e le varianze delle due popolazioni sono diverse.

L'estensione al caso multivariato del test precedente è nota come test  $T^2$  di Hotelling. L'ipotesi nulla  $H_0: \mu_0 = \mu_1$  supporta l'uguaglianza fra i vettori delle medie in due popolazioni le quali presentano anche identica matrice delle varianze e covarianze. Ovviamente l'ipotesi alternativa è  $H_1: \mu_1 \neq \mu_0$ . Questo test si basa sulla distanza standardizzata fra i vettori delle medie delle due popolazioni chiamata *distanza di Mahalanobis*.

$$\Delta^2 = (\mu_1 - \mu_0)^T \Sigma^{-1} (\mu_1 - \mu_0) \quad [2.5.2]$$

Per poter svolgere la stima campionaria di questa distanza si deve utilizzare la grandezza  $D^2$  la quale risulta essere pari a:

$$D^2 = (\bar{x}_1 - \bar{x}_0)^t \Sigma^{-1} (\bar{x}_1 - \bar{x}_0) \quad [2.5.3]$$

La statistica  $T^2$  è uguale a:

$$T^2 = \frac{n_0 n_1}{n} D^2 \quad [2.5.4]$$

Questa, sotto l'ipotesi  $H_0$ , assume una distribuzione definita di Hotelling. Si può inoltre dimostrare che, mediante una semplice trasformazione monotona, questa statistica si distribuisce secondo una *F di Fisher* con  $p$  e  $n - p - 1$  gradi di libertà (Stanghellini, 2009).

Valori elevati della statistica  $T^2$  (o della *F di Fisher*) portano al rifiuto di  $H_0$  e questa tipologia di test è utilizzata per valutare la capacità discriminante di un vettore di variabili congiuntamente considerate.

Qualora non dovesse avere fondamento l'ipotesi di normalità, ci si può basare su dei criteri empirici i quali hanno il fine di valutare in maniera formale la validità della funzione discriminante; si può quindi ricorrere a diversi procedimenti. Si vuole sottolineare che, se si applica la funzione stimata ai dati originali per verificare quando le singole unità sono valutate in maniera errata, questa presenta degli inconvenienti in quanto offre una stima di errata classificazione molto ottimistica (Stanghellini, 2009).

La prima tipologia di stima si basa sul *criterio di minimizzazione dell'errore di classificazione*, sia esso presente nel campione di derivazione (c.d. *training set*) o nel campione di classificazione (c.d. *validation set*).

Se si volesse per esempio definire uno dei due insiemi sopra riportati ( $D_1$ ) una prima regola elementare è quella di assegnare a questo insieme tutti i valori  $\underline{x}$  tali che:

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} > s \quad [2.5.5]$$

dove  $s$  rappresenta il valore la soglia di separazione tra i due gruppi; questo valore deve essere scelto opportunamente (Stanghellini, 2009). In modo più rigoroso  $D_1$  è quell'insieme delle  $x$  tali che:

$$D_1 = \left\{ \underline{x} \mid \frac{f_1(\underline{x})}{f_0(\underline{x})} > s \right\}. \quad [2.5.6]$$

Il problema ora sussiste nello scegliere il miglior valore di  $s$  in modo da poter ripartire nella maniera più corretta i due insiemi  $D_0$  e  $D_1$ ; si deve quindi tenere conto del rapporto tra i due costi legati all'errata classificazione.

Se  $C$  è il costo risultante da una sbagliata classificazione, il suo valore atteso è

$$E(C) = C(0 \mid 1)P(Y = 1) \int_{D_0} f_1(\underline{x}) \delta \underline{x} + C(1 \mid 0)P(Y = 0) \int_{D_1} f_0(\underline{x}) \delta \underline{x} \quad [2.5.7]$$

quindi  $s$  non è altro che quel valore che va a minimizzare il valore atteso del costo.

Sapendo che:

$$\int_{D_0} f_r(\underline{x}) \delta \underline{x} + \int_{D_1} f_r(\underline{x}) = 1$$

dalla quale risulta che:

$$\int_{D_0} f_r(\underline{x}) \delta \underline{x} = 1 - \int_{D_1} f_r(\underline{x}).$$

L'espressione quindi si semplifica in:

$$E(C) = C(0 | 1)P(Y = 1) + \int_{D_1} \{C(1 | 0)P(Y = 0)f_0(\underline{x}) - C(0 | 1)P(Y = 1)f_1(\underline{x})\} \delta \underline{x}$$

e poiché  $C(0 | 1)P(Y = 1)$  è costante, la minimizzazione del valore atteso del costo avviene solamente se si sceglie  $A_1$  come l'insieme di tutti i valori di  $\underline{x}$  in cui  $C(1 | 0)P(Y = 0)f_0(\underline{x}) - C(0 | 1)P(Y = 1)f_1(\underline{x})$  è negativo, ovvero deve risultare:

$$C(0 | 1)P(Y = 1)f_1(\underline{x}) > C(1 | 0)P(Y = 0)f_0(\underline{x}) \quad [2.5.8]$$

Questa a sua volta può essere vista con nella seguente formulazione mediante semplici trasformazioni algebriche:

$$\frac{f_1(\underline{x})}{f_0(\underline{x})} > \frac{C(1 | 0)P(Y = 0)}{C(0 | 1)P(Y = 1)} \quad [2.5.9]$$

Facendo quindi riferimento all'espressione [2.5.5] la soglia ottimale risulta essere (Stanghellini, 2009):

$$s = \frac{C(1 | 0)P(Y = 0)}{C(0 | 1)P(Y = 1)} \quad [2.5.10]$$

A questo punto, il valore associato alla soglia  $s$  ottimale è determinato dall'espressione appena definita e la regione associata all'insieme  $D_1$  è quindi:

$$D_1 = \left\{ \underline{x} \mid \frac{f_1(\underline{x})}{f_0(\underline{x})} > \frac{C(1 | 0)P(Y = 0)}{C(0 | 1)P(Y = 1)} \right\}. \quad [2.5.11]$$

Se invece si utilizzasse la formula di Bayes, tale espressione diverrebbe:

$$D_1 = \left\{ \underline{x} \mid \frac{P(Y = 1 \mid \underline{x})}{P(Y = 0 \mid \underline{x})} > \frac{C(1 \mid 0)}{C(0 \mid 1)} \right\}. \quad [2.5.12]$$

A queste vi è inoltre una formulazione alternativa che si ottiene applicandovi il logaritmo dei rapporti tra funzioni di densità:

$$D_1 = \left\{ \underline{x} \mid \log \frac{f_1(\underline{x})}{f_0(\underline{x})} > \log \frac{C(1 \mid 0)}{C(0 \mid 1)} + \log \frac{P(Y = 0)}{P(Y = 1)} \right\} \quad [2.5.13]$$

oppure

$$D_1 = \left\{ \underline{x} \mid \log \frac{P(Y = 1 \mid \underline{x})}{P(Y = 0 \mid \underline{x})} > \log \frac{C(1 \mid 0)}{C(0 \mid 1)} \right\} \quad [2.5.14]$$

Nel caso in cui non sussista l'ipotesi di normalità, si deve procedere *con la minimizzazione della frequenza dell' errore* che può essere stimata mediante la *tabella di confusione*.

In concreto si sostituiscono i parametri  $\alpha$  con le loro stime ottenute con i metodi visti precedentemente definendo quindi lo score ottenuto mediante la funzione  $s_i = \hat{\alpha}^T x_i$ . Un soggetto viene assegnato al gruppo  $D_1$  se lo score  $s_i$  è superiore ad  $s$  altrimenti viene classificato in  $D_0$ . Per costruire la tabella di confusione si devono prima riclassificare le unità del campione di validazione nel modo seguente; si costruisce la variabile Flag stimata la quale vale 1 se lo score di una unità è maggiore di  $s$  e 0 altrimenti (Stanghellini, 2009). A questo punto si deve costruire la tabella 2 x 2 denominata di *confusione*.

		Flag stimata		Totale
		1	2	
Flag	1	a	b	a + b
	2	c	d	c + d
Totale		a + c	b + d	1

**Tabella 6:** Tabella di confusione

**Fonte:** Stanghellini, 2009

Sulla diagonale principale di questa tabella ci sono le frequenze relative delle unità classificate in modo corretto. Al di fuori della diagonale ci sono invece le frequenze relative legate agli errori di classificazione. Facendo riferimento alla *tabella 6* la stima della probabilità di errore

la si trova quindi mediante la somma  $b + c$  mentre i rapporti  $b/(a + b)$  e  $c/(c + d)$  sono le stime della probabilità di errore: la prima rappresenta la probabilità di assegnare tra la controparte dei cattivi un buono mentre la seconda è la probabilità di assegnare ai buoni un cattivo. Come introdotto all'inizio di questo capitolo, se le probabilità sono calcolate attraverso lo stesso campione sulla quale è stata effettuata la sottostima queste tenderanno a fornire dei valori ottimistici. Da notare è che se si aumenta la soglia si aumenta la probabilità che le unità del gruppo  $D_2$  siano classificate correttamente ma si diminuisce la probabilità che ciò avvenga per le unità che provengono da  $D_1$ .

Nel caso di due variabili, la funzione discriminante può essere mostrata da una retta quindi un possibile innalzamento (abbassamento) della soglia influisce sull'intercetta di questa retta aumentandola (diminuendola) (Stanghellini, 2009).

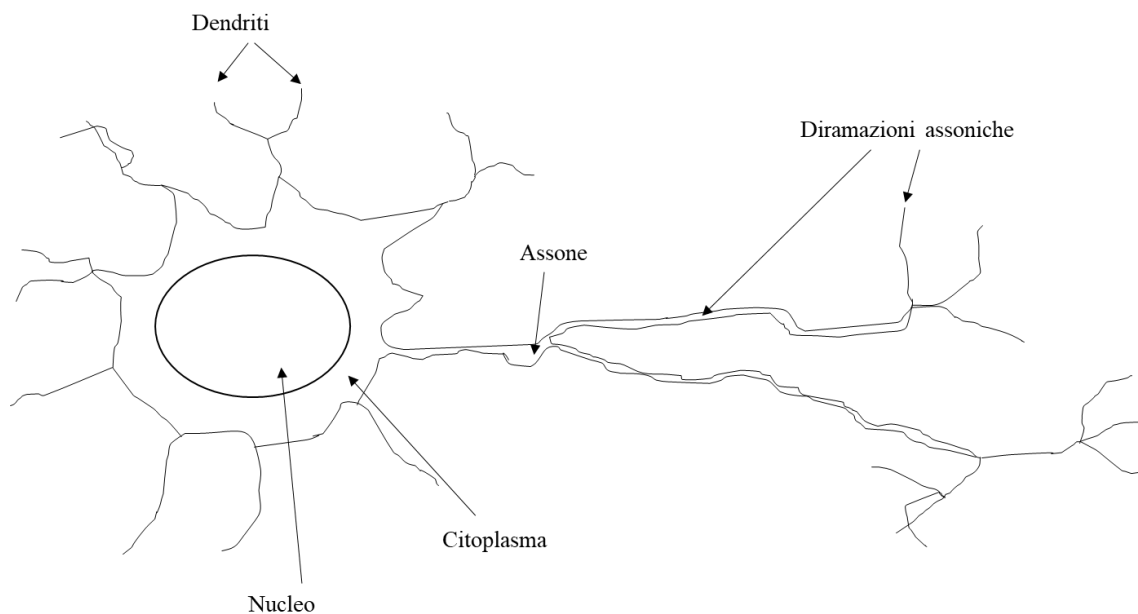
Una seconda tipologia di stima può essere la cosiddetta *leaving one out the method*. Con questa metodologia la funzione discriminante viene calcolata sulla base di  $n - 1$  soggetti. La stima ottenuta viene poi utilizzata per classificare l'individuo non incluso nella creazione della funzione discriminante; si procede quindi ripetendo il processo per ciascun soggetto calcolando al termine la frequenza delle unità mal classificate (Stanghellini, 2009).

## Capitolo 3: Le Reti Neurali

### 3.1 Introduzione e caratteristiche generali delle Reti Neurali Artificiali

In letteratura sono numerosissimi i modelli sviluppati per trattare il rischio di insolvenza. In questa parte della ricerca viene trattato un'ulteriore metodologia che ha preso piede dall'inizio degli anni '90 la quale si basa sull'utilizzo delle reti neurali artificiali.

Le reti neurali sono dei modelli che cercano di imitare la struttura del cervello umano e il suo funzionamento in modo da poter simulare come un generico soggetto prenda decisioni e o compia specifici gesti. Per questa ragione ci si è concentrati sulla forma, le caratteristiche e l'organizzazione dei neuroni oltre che alle relative sinapsi così da poter definire una struttura matematica che, sfruttando la programmazione informatica, riuscisse a replicare l'attività del cervello umano (D'Annunzio, Falavigna, 2004). A livello concreto i modelli neurali artificiali costituiscono una semplificazione della loro controparte biologica malgrado questa approssimazione dipenda dalla tipologia di modello adottata.



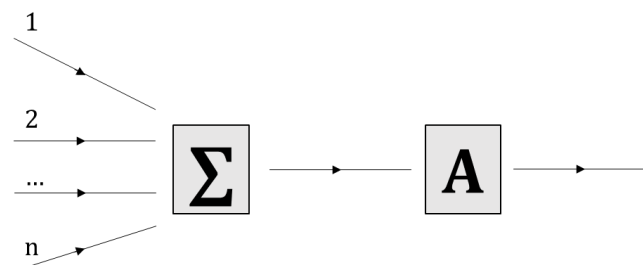
**Figura 2:** Rappresentazione di un neurone biologico e delle sue componenti principali  
**Fonte:** disegno a cura dell'autore tratto da D'Annunzio, Falavigna, 2004

Il neurone viene considerato quindi l'unità computazionale elementare del cervello e in letteratura può essere anche definito come *unità*, *nodo* o *processore*. Questo ha la capacità di poter generare un potenziale elettrico che si propaga lungo l'*assone*<sup>22</sup> quando nel corpo del

<sup>22</sup> Si può quindi considerare l'assone come l'output del neurone

neurone vi è un certo quantitativo di attività elettrica che superi una determinata soglia. Al contrario nel neurone l'input di ingresso è rappresentato da una serie di fibre chiamate *dendriti* le quali sono collegate con gli assoni di altri neuroni in modo da poter ricevere tutti i flussi elettrici in uscita da questi. Il punto di contatto fra un assone di un neurone e il dendrite di un altro neurone si chiama *sinapsi* la quale ha la capacità di modulare l'impulso elettrico che proviene dall'assone. Il potenziale elettrico che viene generato da un neurone e di tipo *tutto-o-nulla* infatti l'impulso viene generato se e solamente se l'attività elettrica di un neurone supera una data soglia. Se ciò non avviene la scarica generata non differisce per intensità da quella presente tra un neurone ed un altro. Il potenziale di energia generata si propaga lungo l'assone per poi giungere alla sinapsi che si collega con il dendrite di un altro neurone. A questo punto però si deve considerare il fatto che il potenziale post-sinaptico sul dendrite dipende dalle caratteristiche biomediche proprie di ogni specifica sinapsi quindi, in presenza del medesimo potenziale pre-sinaptico, due sinapsi differenti danno vita a potenziali post sinaptici differenti. I potenziali post-sinaptici, mediante i dendriti del neurone si propagano e si sommano all'interno del *soma*<sup>23</sup>. Se il risultato della somma di questi potenziali elettrici supera una data soglia il neurone innesca il potenziale elettrico al suo interno che si propagherà attraverso il suo assone; se invece il limite non viene superato non accade nulla (D'Annunzio, Falavigna, 2004).

Lo schema proposto di seguito rappresenta in termini più schematici il funzionamento di un neurone:



**Figura 3:** Diagramma schematico del neurone biologico  
**Fonte:** D'Annunzio, Falavigna, 2004

In questa semplice rappresentazione i dendriti sono rappresentati dalle linee di ingresso che portano intuitivamente verso il soma, raffigurato mediante l'operatore sommatoria  $\Sigma$ ; in questa cella si sommano i segnali (stimolatori) o si sottraggono (inibitori) in modo da generare un altro segnale che perverrà al modulo A. Questo funge da elemento decisionale in quanto trasmette, mediante l'assone, un segnale corredato da particolari caratteristiche; a sua volta questo

<sup>23</sup> Genericamente si può definire "soma" il corpo della cellula nervosa all'interno del quale vi è il nucleo



passerà sui dendriti di un nuovo neurone se, al momento dell'uscita da  $\Sigma$ , il segnale è maggiore rispetto al valore soglia intrinseco nel neurone stesso (D'Annunzio, Falavigna, 2004).

Questo schema è esemplificativo di tutta una serie di caratteristiche e passaggi rappresentabili da un punto di vista informatico. Ciò significa che si è riusciti a riprodurre artificialmente il funzionamento biologico dei neuroni e si è quindi dato vita ai nodi artificiali che consentono la connessione di più neuroni artificiali creando la c.d. *Rete Neurale Artificiale (RNA)* o *Artificial Neural Network (ANN)*.

Il procedimento appena descritto è stato rappresentato dal p.d.v. informatico poiché consente di dare risposta a problemi complessi quale la quantificazione del rischio di insolvenza di un operatore economico. Le reti neurali, presentandosi come degli eccellenti classificatori, non sono utilizzate esclusivamente per condurre delle analisi o previsioni finanziarie ma anche per risolvere dei problemi computazionali di natura completamente diversa da quella economico-finanziaria o biologica (D'Annunzio, Falavigna, 2004).

Considerando una definizione tecnica, le reti neurali artificiali sono rappresentate da una serie di modelli composti da strati di unità elementari di elaborazione chiamati *Processing Elements (PEs)* le quali elaborano le informazioni per mezzo di una funzione non lineare.

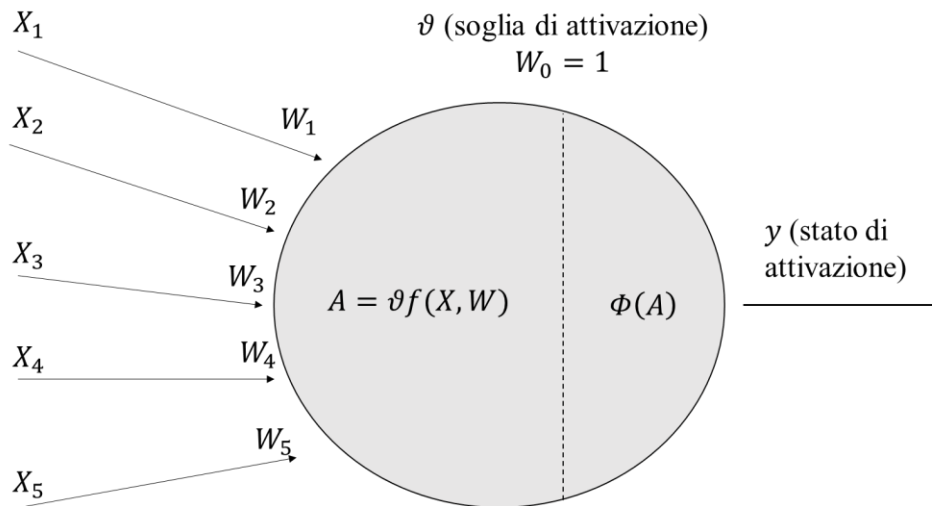
I modelli che descrivono il funzionamento delle reti neurali fanno parte di quella macrocategoria dei sistemi di intelligenza artificiale (IA) con la particolarità che questi, a differenza di quelli tradizionali, non prevedono la presenza di un decisore centrale; la decisione infatti proviene dal risultato ottenuto con il lavoro distribuito svolto da più unità elementari. Non a caso le caratteristiche principali delle reti neurali ne rappresentano anche le differenze fondamentali rispetto a tutti gli altri sistemi di IA. Le reti neurali sono:

1. "Adattive" in quanto apprendono.
2. In grado di generalizzare.
3. Resistenti al rumore poiché riescono a fornire dei risultati consistenti malgrado i segnali in input possano a volte risultare distorti.
4. Resistenti alle lesioni in quanto se vengono in parte danneggiate, il danno subito non ne compromette il funzionamento nel suo complesso.
5. Rappresentabili attraverso una simbologia elementare (D'Annunzio, Falavigna, 2004).

Dal punto di vista empirico vi è la possibilità di individuare le componenti di un nodo e verificarne se nella costruzione e/o implementazione dello stesso sono emersi dei problemi. Il neurone è costituito da degli elementi fondamentali che sono:

- Inputs o stati di attivazione ( $x_i$ )
- Connessioni o pesi o sinapsi ( $w_i$ )
- Stato di attivazione netto ( $A$ )

- Valore soglia o bias ( $\vartheta$ ) che definisce lo stato di attivazione o inibizione di un singolo nodo<sup>24</sup>
- Funzione di attivazione ( $\Phi(A)$ )
- Output o stato di attivazione ( $y$ )



**Figura 4:** Rappresentazione semplificata di un nodo di una rete neurale

**Fonte:** D'Annunzio, Falavigna, 2004. (La simbologia è stata leggermente modificata per renderla coerente con ciò che verrà rappresentato in seguito)

D'Annunzio e Falavigna chiariscono l'efficacia del sistema delle reti neurali spiegando, in linea generale, la sua costruzione in modo da poter essere applicata per la classificazione e/o la previsione di insolvenza di una controparte. Questa fase è quella che prende il nome di *fase di apprendimento* o *training* e avviene attraverso un algoritmo costruito da chi intende determinare la struttura della rete. In questa fase entrano in gioco le leggi di apprendimento che impongono le regole attraverso le quali si possono operare delle variazioni alle connessioni della rete. Queste regole possono essere generalmente di 2 tipi:

- *Non supervisionato*: la rete si avvale di stimoli  $S$  attuati nel mondo reale e, a seguito di questi stimoli, fornisce di rimando delle risposte  $R$ . Il sistema a questo punto si autoregola in modo da fornire una rappresentazione della realtà quanto più veritiera possibile.

La rete ha la capacità di auto-organizzarsi basandosi sulle somiglianze tra i vari esempi cambiando autonomamente le connessioni a seguito delle indicazioni concernenti le varie distanze da tenere tra i vari differenti esempi.

<sup>24</sup> In questo caso si fa riferimento al caso in cui si stia parlando di una *rete a perceptrone*; verrà trattata in seguito. Nelle versioni più recenti il bias è rappresentato dal peso applicato all'input e che assume un valore pari a 1. La funzione, essendo sempre presente questa variabile, risulta essere arricchita di un grado di libertà aggiuntivo.

➤ *Supervisionato*: la rete non interagisce direttamente con l'ambiente esterno; si avvale sempre degli stimoli  $S$  e delle risposte  $R$  tratte dall'ambiente ma queste sono fornite da un intermediario esterno. Questa regola impone a sua volta di seguire dei passaggi operativi precisi.

1. Si deve prefissare il valore che ogni unità di output assumerà se la rete classifica in modo corretto un'osservazione
2. Alle unità di input devono essere presentate alcuni esempi di apprendimento (*training set*) dei quali si possa fornire una rappresentazione in merito all'appartenenza o meno a uno dei gruppi presi in considerazione nello studio.
3. L'ultimo *step* è quello di valutare la risposta fornita dalla rete stessa. Nel caso in cui questa sia corretta si continua la fase di apprendimento con un'altra osservazione altrimenti se la risposta non è soddisfacente, si deve calcolare l'errore tra la risposta stimata e quella teorica e si devono modificare i pesi in modo da ottenere dalla rete la risposta corretta (D'Annunzio, Falavigna, 2004).

In tutti e due i casi il miglioramento della rete avviene mediante i cosiddetti *cicli* o *epoche*; gli esempi del *training set* vengono infatti rielaborati più volte dal modello.

Dal punto di vista matematico la tecnica maggiormente utilizzata è quella denominata *backpropagation* che si basa su una tipologia di apprendimento, da parte della rete, basata sulla minimizzazione dell'errore nella definizione del risultato.

**Esempio** (tratto da *Modelli di analisi e previsione del rischio di insolvenza. Una prospettiva delle metodologie applicate* - Nadia D'Annunzio, Greta Falavigna - 2004 – pag. 26)

Si supponga di avere una stanza con 20 persone e di voler costruire una rete neurale che, sulla base di 4 input, riesca ad individuare precisamente la persona a cui si fa riferimento. I quattro input sono quindi delle caratteristiche oggettive riguardanti: età, colore dei capelli (0 se chiari, 1 se scuri), portatore di occhiali (0 se no, 1 se sì), sesso (0 se maschio, 1 se femmina). Queste variabili indipendenti sono sia qualitative (3) sia quantitative (1); esse possono essere osservate su ognuna delle 20 persone. Dopo aver numerato tutti i soggetti da 1 a 20, selezionando l'individuo n° 2 si sa che esso ha i capelli chiari, è un maschio di 20 anni ed è portatore di occhiali. Il suo vettore di input avrà questi risultati:

	$X1$	$X2$	$X3$	$X4$
<b>1</b>	...	...	...	...
<b>2</b>	20	0	1	1
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
<b>20</b>	...	...	...	...

che quindi da un output descritto dalla seguente tabella.

	<b>Y1</b>	<b>Y2</b>	<b>Y3</b>	<b>...</b>	<b>Y20</b>
<b>1</b>	1	0	0	0	0
<b>2</b>	0	1	0	0	0
<b>3</b>	0	0	1	0	0
<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>
<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>
<b>20</b>	0	0	0	0	1

Si vuole inoltre ricordare che sia le componenti  $x$  che  $y$  devono essere inserite nella rete già all'inizio in quanto sono componenti a noi note. Le incognite del problema sono i pesi  $w_i$  i quali rendono la rete capace di riconoscere l'appartenenza degli input all'individuo  $n^{\circ}$ ; ovviamente ciò avviene dopo la fase di apprendimento (*training*).

La fase di *training* è fondamentale poiché, attraverso il numero dei cicli della rete e attraverso le sue caratteristiche, si è in grado di stimare dei pesi idonei; inizialmente questi vengono scelti casualmente per poi essere corretti sulla base del criterio di minimizzazione dell'errore. Si ricorda che l'errore è la differenza che vi è tra l'output stimato  $y$  e quello teorico  $t$  (D'Annunzio, Falavigna, 2004).

La funzione dell'errore da minimizzare è quindi:

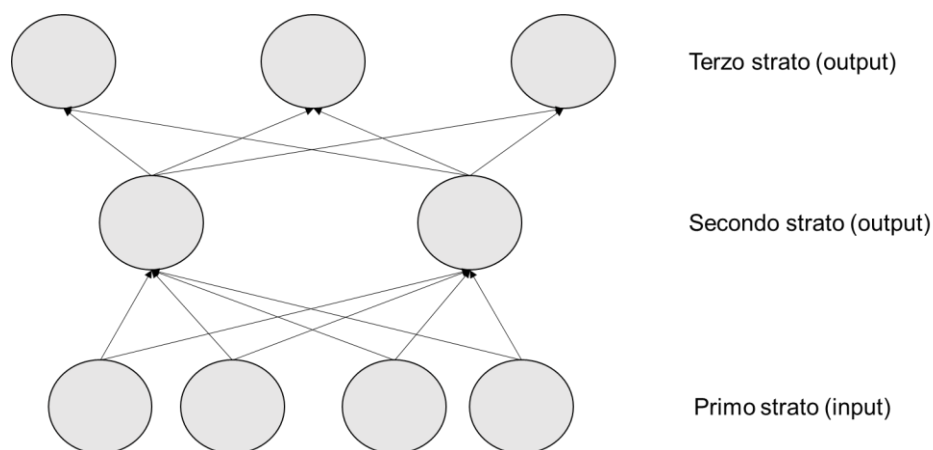
$$\sum_n \sum_k (t_k^n - y_k^n)^2$$

In questa funzione l'indice  $n$  va da 1 al numero totale degli pattern del *training set* mentre l'indice  $k$  descrive i singoli *Processing Elements* di output.

Se si volesse basare questo ragionamento su un contesto economico-finanziario, ovvero sul rischio di insolvenza delle imprese, è necessario definire un set di imprese fallite e uno di aziende sane. A questo punto si può calcolare i pesi discriminanti per poi applicarli ad un set di imprese che non sono fallite ma delle quali se ne vuole conoscere lo stato economico-finanziario futuro. Questo ultimo set preso in considerazione non è un campione attraverso il quale la rete apprende ma costituisce un campione di controllo o *validation set* ovvero un campione atto a testare la corretta organizzazione della rete: in parole povere se la rete ha imparato o meno. In questo set si conosce lo stato di salute delle imprese ma la rete lo ignora e se il modello creato dalla rete stessa riesce a fornire una buona classificazione di queste imprese sarà possibile utilizzare quel sistema di pesi per fare delle previsioni in merito al rischio di insolvenza di ciascuna impresa (D'Annunzio, Falavigna, 2004).

Le reti neurali hanno anche la caratteristica di essere costituite da una serie di parti tutte strettamente interrelate tra loro e se anche una sola di queste viene modificata, ne risente la validità e l'accuratezza del risultato finale (output). Si deve però sottolineare come la rete, malgrado i risultati del test tendano a modificarsi con la sostituzione dei pesi ottimali, offra comunque un'ottima rappresentazione della realtà anche con l'inserimento di un valore errato. La rete ha anche alcuni punti deboli sia dal punto di vista teorico che pratico. Il principale problema riguarda l'eventuale mancanza di apprendimento di qualche aspetto della realtà di difficile esplicitazione<sup>25</sup>. In questo caso è possibile incorrere in un errore di valutazione di questi specifici casi i quali rientrano quindi nella parte "oscura" della realtà.

Un secondo punto critico della rete neurale nasce dalla sua natura multistrato la quale presenta dei livelli "nascosti" poiché intermedi tra gli strati degli input e degli output. Se il problema presenta delle soluzioni che si intersecano, si dovrà introdurre uno strato nascosto in modo che la rete sia in grado di fornire una soluzione corretta.



**Figura 5:** Esempificazione di una rete neurale multistrato  
**Fonte:** D'Annunzio, Falavigna, 2004

Un ulteriore inconveniente della rete è che malgrado possa esservi la presenza di vettori di *inputs* uguali, si possono comunque generare dei risultati diversi; con questo si richiama l'ultimo esempio visto nel quale, sebbene alcuni soggetti possano presentare delle caratteristiche comuni, non è detto che ci si riferisca sempre al medesimo soggetto. In particolare, sempre in riferimento all'esempio della classe, due persone hanno una miriade di altre caratteristiche specifiche non prese in considerazione tali da consentire la distinzione (esempio altezza e peso). Il vettore corretto nel caso visto risulta essere diverso da individuo ad individuo ma la rete non sarà in grado di distinguere soggetti molto simili in quanto le

<sup>25</sup> In *Modelli di analisi e previsione del rischio di insolvenza. Una prospettiva delle metodologie applicate di Nadia D'Annunzio e Greta Falavigna* si fa riferimento ad una particolare situazione aziendale non considerata.

informazioni discriminanti sono poche. Si deve quindi fare in modo che le variabili siano in un numero sufficientemente adeguato affinché lo specifico fenomeno sia individuato e discriminato; ciò viene fatto al fine di non commettere degli errori di classificazione.

### 3.2 La letteratura riguardante le reti neurali

Di grande importanza è l'analisi della letteratura economico-finanziaria in merito ai risultati ottenuti con l'applicazione delle reti neurali. Ciò ci permette di poter evidenziare i punti di forza e i punti deboli di questo sistema di classificazione mettendoli anche in confronto con altre metodologie di discriminazione per l'analisi del rischio di insolvenza.

In un contesto economico le variabili da utilizzare per la creazione della rete sono prevalentemente di tipo quantitativo poiché lo stato di salute aziendale viene rappresentato dalle voci di Stato Patrimoniale e Conto Economico. Con questo si può fare uno specifico riferimento ad Altman (1968) il quale, come indicato nei capitoli precedenti, individuò le principali variabili per la descrizione del rischio di insolvenza in un'impresa<sup>26</sup>. Una delle prime ricerche sull'analisi dell'efficacia delle reti neurali è stato condotto da Chilanti nell'*Analisi e previsione delle insolvenze: un approccio neurale*, Finanza Imprese e Mercati, del 1993 nel quale è stato utilizzato un campione per il *training set* non diviso in maniera equa tra imprese fallite e sane. Questo metodo di procedere ha suscitato molti dubbi in merito alla necessità di adottare un identico numero di imprese per tipologia per la creazione della rete. Perez nello studio *Artificial neural networks and bankruptcy forecasting: a state of the art* del 2006 afferma che, dopo aver analizzato 24 studi, quelli che presentano *inputs* migliori sono quelli che dispongono di un campione 50-50 tra imprese sane e imprese fallite.

Per quanto concerne la funzione di attivazione sempre Chilanti afferma di aver utilizzato una di tipo sigmoide ovvero:

$$f(A) = \frac{1}{1 + e^{-kA}}$$

Turban e Trippi però nella ricerca *Neural Networks in Finance and Investing. Using artificial neural intelligence to improve real-world performance*, come anche Sironi e Marsella nel *La misurazione e la gestione del rischio di credito. Modelli, strumenti e politiche*, Bancaria editrice del 1998 affermano che possano essere utilizzate anche altre forme funzionali. Sironi e

---

<sup>26</sup> Le variabili individuate da Altman sono: Capitale circolante/Totale attivo, Utili ritenuti/Totale attivo, Utili prima di interessi e tasse/Totale attivo, Capitalizzazione di mercato/Debito totale, Fatturato/Totale attivo.

Marsella in particolare affrontano questo problema più approfonditamente proponendo le seguenti tipologie di funzioni:

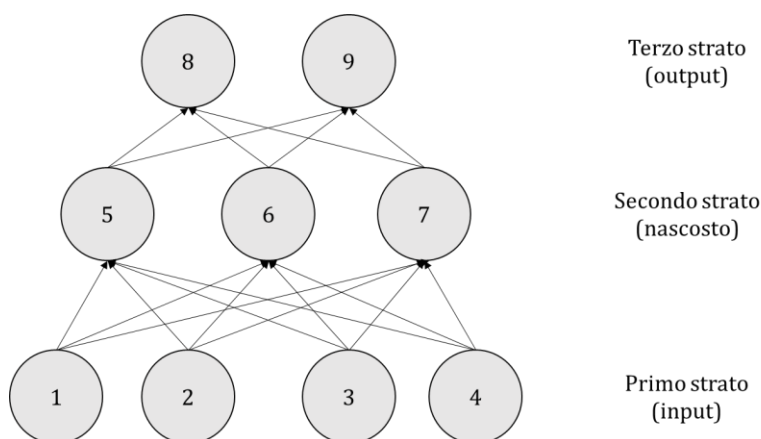
Lineare	$f(x) = \alpha \cdot x$
Sinusoidale	$f(x) = \sin(x)$
Gaussiana	$f(x) = e^{-x^2}$
Logistica	$f(x) = \frac{1}{1 + e^{-x}}$
Logistica simmetrica	$f(x) = \left(\frac{2}{1 + e^{-x}}\right)^{-1}$
Tangente iperbolica	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

**Tabella 7:** questa tabella di funzione è stata proposta da Sironi e Marsella (1998)

Se la funzione fosse lineare si rientra nel caso della regressione multipla lineare con la conseguente perdita degli strati nascosti della rete. Definendo a titolo esemplificativo una rete avente 4 nodi in input, 3 nascosti e 2 di output, si ottengono le seguenti matrici di pesi.

$$A = \begin{bmatrix} w_{51} & w_{61} & w_{71} \\ w_{52} & w_{62} & w_{72} \\ w_{53} & w_{63} & w_{73} \\ w_{54} & w_{64} & w_{74} \end{bmatrix} \quad B = \begin{bmatrix} w_{85} & w_{95} \\ w_{86} & w_{96} \\ w_{87} & w_{97} \end{bmatrix}$$

Nella matrice  $A$ , l'elemento 1,1 (quindi  $w_{51}$ ) stabilisce il legame tra il nodo nascosto 5 e il nodo in input 1; questa tipologia di connessione è definita anche per tutti gli altri elementi in base all'indice espresso. Nella matrice  $B$  invece l'elemento 1,1 (quindi  $w_{85}$ ) è il peso associato all'influenza che il nodo 5 ha sul nodo in output 8.



**Figura 6:** Rete con 4 nodi in input, 3 nascosti e 2 di output.  
**Fonte:** D'Annunzio, Falavigna, 2004.

Questa rete può essere descritta secondo una generica funzione che va da  $\mathbb{R}^4$  ad  $\mathbb{R}^2$ :

$$y = f[B \cdot f(Ax)]$$

I valori che il vettore  $x$  assume in corrispondenza di ognuno dei casi del *training set* corrispondono ai nodi in input mentre i valori associati ai neuroni dello strato nascosto rappresentano gli elementi del vettore risultante da  $f(Ax)$ . Infine i valori dei nodi dello strato in output sono rappresentati dalle componenti del vettore risultato dati da  $f[B \cdot f(Ax)]$ .

Si vuole sottolineare che se la funzione  $f(Ax)$  è lineare essa “scompare” eliminando lo strato nascosto; la funzione generale della rete quindi diventa:

$$y = f(BAx) = f(Cx)$$

Se invece si avesse una funzione non lineare da introdurre nella rete normalmente viene utilizzata una funzione di tipo logistico (o in alternativa sigmoide) poiché offre un output di tipo dicotomico. Questa tipologia di funzione rende il risultato pari a 0 se non si supera una determinata soglia o 1 dopo aver superato tale valore; questo consente di introdurre il discorso sulla forma dell'output il quale, nella quasi totalità degli studi, definisce se un'impresa è sana o fallita, quindi di tipo dicotomico (D'Annunzio e Falavigna, 2004).

Solitamente, nelle letterature analizzate, quella migliore da utilizzare per poter attuare delle previsioni in campo finanziario appare essere quella della *Back-propagation* tuttavia, in origine, la tipologia di rete neurale più utilizzata era la *Multilayer Perceptron (MLP)* implementata da Rosenblatt. Si vuole però sottolineare l'esistenza di molteplici generi di rete che vengono utilizzate in base alla tipologia di fenomeno che si vuole studiare e, malgrado si sia cercato di associare ad ogni problema un dato tipo di rete neurale, anche in ambito economico finanziario vengono utilizzati dei sistemi neurali diversi dalla *Back-propagation* o dal *Multilayer Perceptron*. Nella seguente tabella presa da Sironi e Marsella nel *La misurazione e la gestione del rischio di credito. Modelli, strumenti e politiche* del 1998 viene offerta una semplice sintesi di vari tipi di rete neurale oltre ai rispettivi punti forti e difetti e dei fenomeni che sono in grado di rappresentare.



Modello Neurale	Applicazione principale	Vantaggi	Limiti
Hopfield	Ricostruzioni delle immagini da frammenti	Implementazioni su larga scala	Non apprende e i pesi devono essere attribuiti
Perceptron	Riconoscimento caratteri ed immagini	Esperienza di applicazione	Non riconosce modelli troppo complessi
Multilayer Perceptron	Riconoscimento dei pattern sottostanti	Rete semplice e generalizzabile	Non riconosce modelli troppo complessi
Macchina di Boltzman	Riconoscimento dei pattern sottostanti (radar, sonar)	Semplicità basata sulla funzione di errore	Richiede tempi lunghi di apprendimento
Counter-propagation	Comprensione di immagini e analisi statistica	Più semplice del Perceptron, meno potente della Back-propagation	Elevato numero di connessioni ed elementi da processare
Neocognitron	Riconoscimento di grafia e dei simboli	Rete raffinata per pattern complessi	Elevato numero di connessioni ed elementi da processare
Self-Organizing Map	Mappatura e classificazione di dati	Performance superiore alle tecniche statistiche	Training esteso e dispersivo
Back-propagation	Ampie applicazioni: in finanza previsione e credit scoring	Rete diffusa semplice ed efficace	Training supervisionato senza leggi universali

**Tabella 8:** Tabella riportante varie tipologie di reti neurali artificiali

**Fonte:** Sironi e Marsella, 1998.

Un ulteriore tema da affrontare è quello legato ai vari generi di architettura delle reti stesse. Gran parte della letteratura è propensa ad utilizzare uno strato nascosto (*hidden layer*) all'interno della rete neurale mentre una piccola parte degli studi introduce un secondo strato nascosto; questa scelta, in nessun caso, è stata giustificata (D'Annunzio e Falavigna, 2004). In questo caso si riferiscono agli studi di Altman, Marco e Varetto (1994), De Almeida e Dumonter (1993), Roghupahi, Schkade e Raju (1991). Sempre rimanendo in tema di struttura della rete neurale Hornik, Stinchcombe e White nell'articolo *Multilayer feedforward networks are universal approximators* del 1989 giungono ad un risultato molto significativo in quanto riescono ad approssimare qualsiasi funzione utilizzando una rete MLP (*Multilayer Perceptron*) ovvero multistrato avente un solo strato nascosto. La superiorità delle reti neurali è stata poi confermata da tutta la letteratura in quanto non vi sono delle ipotesi particolari che devono stare alla base nella definizione delle variabili in *inputs*; questo è un risultato molto importante

in quanto si differenzia da moltissime altre tecniche statistiche<sup>27</sup>. In aggiunta a ciò appena detto la rete neurale, per funzionare, non necessita di serie storiche e, se basata su funzioni di attivazione non lineari, può rappresentare in modo più corretto la realtà dei fatti.

### 3.3 Confronto tra il modello a rete neurale e altre metodologie statistiche

I primi a confrontare l'efficacia dei modelli neurali con altre metodologie statistiche furono Odom e Sharda nel 1990 che mediante lo studio *A neural network model for bankruptcy prediction* paragonarono il sistema neurale con l'analisi discriminante formulata nel 1968 da Altman sulla base della loro efficienza ed efficacia. Le reti neurali, al fine di favorirne il confronto, furono utilizzate applicando loro, come *inputs*, le stesse variabili economico finanziarie riportate da Altman nella sua ricerca. Il risultato fu che il sistema di reti neurali risultò decisamente migliore rispetto all'analisi discriminante per quanto riguarda l'accuratezza dei risultati. Altre ricerche confermano questo assunto a partire per esempio da Coast e Fant nel *Recognizing financial distress patterns using a neural network tool* del 1993 fino a Kerling e Poddig nel *Klassifikation von Unternehmen mittels KNN* del 1994; entrambe queste ricerche hanno svolto un confronto sempre tra le reti neurali e l'analisi discriminante multivariata (*MDA*). Al contrario di queste tesi appena formulate, la ricerca *Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)* condotta da Altman, Marco e Varetto nel 1994 afferma e dimostra la superiorità dell'analisi discriminante lineare (*ADL*) rispetto alle reti neurali. Queste ultime, secondo Altman, si presentano di difficile interpretazione dal punto di vista economico-finanziario soprattutto per quanto riguarda l'architettura se la complessità del fenomeno aumenta mentre l'analisi svolta con il metodo *ADL* risulta di più immediata comprensione. Secondo questa ricerca inoltre è difficile individuare quali siano le variabili significative per la rappresentazione del fenomeno quando si deve andare a comporre la rete neurale.

La tesi finale di Altman quindi si basa sul fatto che, secondo i risultati ottenuti, le reti neurali non riescono a dominare i modelli statistici a differenza dell'analisi discriminante lineare.

Un'ulteriore ricerca condotta da Kiviluoto e pubblicata nel 1998 dal titolo *Predicting bankruptcies with the self-organizing map* punta a confrontare le reti neurali *SOM (Self Organizing Map)*<sup>28</sup>, l'analisi discriminante multivariata e il modello di rete *LVQ (Linear Vector*

---

<sup>27</sup> Come già visto in precedenza e come affermato anche da Nadia D'Annunzio Greta Falavigna (2004) nell'analisi discriminante si deve ipotizzare che la distribuzione dei dati sia di tipo normale. Questa premessa può sembrare un forte ostacolo per il proseguimento dell'analisi statistica ma può essere facilmente superato applicando una trasformazione logaritmica ai dati iniziali.

<sup>28</sup> Un approfondimento su questo tipo di rete neurale si trova nell'articolo *Self-Organizing Map* di Kohonen pubblicato nel 1990

*Quantization*)<sup>29</sup>. I risultati riportati testimoniano una maggiore precisione della rete SOM rispetto alle altre due tecniche considerate.

Nel 1995 sono state testate le capacità predittive della rete *back-propagation*, SOM e della *Boltzman Machine*; i risultati e le caratteristiche di questo confronto sono riportate nell'articolo *A comparative study of neural networks in bankruptcy prediction* di Back, Sere, Oosterom e Wezel. Questa ricerca mostra i risultati delle previsioni ad un anno, a due anni e a tre anni dall'evento insolvenza evidenziando come la rete *backpropagation* dia risultati migliori a distanza di un anno dall'evento mentre, se si prendesse un periodo temporale più ampio, nessuno dei tre riuscirebbe ad offrire dei risultati soddisfacenti. Sempre in questo contesto si inserisce Atiya che, nel 2001, sperimenta l'accuratezza delle previsioni, sempre in diversi periodi temporali prima del verificarsi dell'evento. Si giunge quindi alla conclusione che se si considera un periodo temporale ampio, al massimo due anni, le previsioni risultano essere soddisfacenti mentre se si eccede tale arco temporale si perde accuratezza da parte delle reti.

Tempo prima del default	N° corrette (nel campione)	Numerosità nel campione	% corrette	N° corrette (fuori dal campione)	Numerosità fuori dal campione	% corrette (fuori dal campione)
<b>5 mesi o meno</b>	35	38	<b>92,11</b>	56	65	<b>86,15</b>
<b>6-12 mesi</b>	43	51	<b>84,31</b>	44	54	<b>81,48</b>
<b>12-18 mesi</b>	33	37	<b>89,19</b>	47	63	<b>74,60</b>
<b>18-24 mesi</b>	33	37	<b>89,19</b>	25	32	<b>78,13</b>
<b>più di 24 mesi</b>	19	25	<b>76,00</b>	28	42	<b>66,67</b>
<b>Default totali</b>	163	188	<b>86,70</b>	200	256	<b>78,13</b>
<b>Controparti solventi</b>	276	303	<b>91,09</b>	372	413	<b>90,07</b>
<b>Totale</b>	439	491	<b>89,41</b>	572	669	<b>85,50</b>

**Tabella 9:** Precisione delle reti neurali in base a inputs di diversa natura.

**Fonte:** Atiya, 2001.

Si vuole sottolineare come questi risultati siano stati ottenuti utilizzando non solamente degli indicatori di bilancio ma anche dei dati estratti direttamente dal mercato. Come per lo Z-Score di Altman infatti, utilizzando solamente dei dati finanziari la previsione risulta essere peggiore in quanto i documenti contabili riportano dei valori aziendali ritardati mentre le valutazioni del mercato azionario incorporano delle valutazioni qualitative altamente previsive come lo stadio di maturità dell'attività o l'immagine aziendale (D'Annunzio e Falavigna, 2004).

Un ulteriore confronto tra le reti neurali ed altri modelli statistici è stata condotta da Tam e Kiang nel *Predicting bank failures: a neural network approach* e nel *Managerial applications of neural networks: the case of bank failure predictions* rispettivamente nel 1990 e nel 1992. Dopo

<sup>29</sup> Metodologia che costruisce una rete supervisionata nella quale i vettori in input del set di training sono suddivisi in classi; il target della rete è proprio la classe di appartenenza di ogni vettore di input. La rete deve essere costruita dopo una fase di training nella quale si devono analizzare i vettori in input allo scopo di classificarli correttamente.

aver paragonato tra loro vari sistemi come regressioni, MDA, *k-Nearest Neighbour*, *Logit* e *ID3*<sup>30</sup> si giunge alla conclusione che le reti neurali danno dei risultati migliori rispetto a tutti gli altri metodi analizzati se i dati in *inputs* sono basati su valori entro un anno dal fallimento dell'impresa. Qualora invece si estendesse il periodo temporale dei dati in input a due anni dal fallimento, le previsioni sono più precise se ci si basasse sulla MDA. Tam e Kiang (1990 e 1992) inoltre evidenziano come, nei loro studi, la rete neurale con uno strato nascosto risulti essere più performante rispetto a quella che ne è priva.

Da sottolineare è anche la ricerca *Toward global optimization of neural networks: a comparison of the genetic algorithm and backpropagation* svolta da Sexton, Dorsey e Johndon la quale paragona la tecnica della *backpropagation* con l'algoritmo generico basato su una funzione di apprendimento attraverso il metodo Monte Carlo. Il lavoro svolto evidenzia che, quando il fenomeno da rappresentare è articolato, l'algoritmo *backpropagation* non riesce sempre a dare delle soluzioni performanti<sup>31</sup> ma risulta comunque più semplice ed intuitivo. È stato tuttavia dimostrato come l'algoritmo genetico sia una metodologia di ricerca che riesce a focalizzarsi sull'area in cui è presente la soluzione migliore quando si passa da un insieme di soluzioni ad un altro. Ciò permette di semplificare progressivamente anche l'insieme dei parametri associati alle variabili e consente di ottenere un ottimo globale e non locale, rendendo la rete efficiente nel trovare una soluzione che rappresenti ogni realtà considerata.

Un altro studio da sottolineare è quello svolto da Hill, Marquez, O'Connor e Remus chiamato *Artificial neural network models for forecasting and decision making* del 1993 il quale confronta le reti neurali con i modelli ARMA (*Autoregressive Mobile Average*). I risultati ottenuti testimoniano la superiorità delle reti neurali ma solamente al ricorrere di determinate condizioni; gli autori infatti supportano la dimostrazione della validità a livello teorico della *backpropagation* ma a livello pratico è di difficile implementazione. Sul piano empirico infatti si predilige adottare un sistema diverso come un *ID3* o un algoritmo genetico.

La superiorità delle reti neurali rispetto ai modelli autoregressivi a media mobile è testimoniato anche da Fenga nella ricerca *L'utilizzo di reti neurali per la stima anticipata e la previsione di serie storiche stagionali. Un'analisi comparativa con modelli della classe ARMA su serie ISTAT delle presenze turistiche in Italia*. Anche in questo articolo si evidenzia come la rete che offre dei risultati migliori sia quella basata su una funzione sigmoide avente al suo interno uno strato nascosto.

Per concludere il capitolo si presenta nella seguente tabella, presa dalla ricerca *Modelli di analisi e previsione del rischio di insolvenza. Una prospettiva delle metodologie applicate di*

---

<sup>30</sup> *ID3* è una particolare tecnica predittiva che serve per la costruzione degli alberi decisionali

<sup>31</sup> I risultati non sempre ottimali sono dati dal fatto che la *backpropagation* può a volte trovare degli ottimi locali e non globali

Nadia D'Annunzio e Greta Falavigna, che riporta i pregi e i difetti dei modelli maggiormente utilizzati in letteratura per l'analisi del rischio legato al default di un'impresa.

	<b>Pregi</b>	<b>Difetti</b>
<b>Alberi decisionali</b>	<ol style="list-style-type: none"> <li>1. Impegno finanziario richiesto è minore di quello della MDA</li> <li>2. Risultato trasparente dal punto di vista interpretativo</li> <li>3. Non dipendenza da ipotesi metodologiche tipiche</li> <li>4. dell'analisi discriminante e della statistica parametrica.</li> </ol>	<ol style="list-style-type: none"> <li>1. L'ottimizzazione avviene a livello locale e non globale</li> <li>2. Il risultato è di tipo dicotomico</li> </ol>
<b>Algoritmi genetici</b>	<ol style="list-style-type: none"> <li>1. Si ottengono attraverso funzioni lineari ottimizzate che non dipendono da ipotesi di normalità</li> <li>2. Si può gestire il segno degli indicatori che entrano nella funzione</li> </ol>	Risultati meno precisi che MDA
<b>Analisi discriminante multivariata</b>	<ol style="list-style-type: none"> <li>1. Dà un risultato continuo</li> <li>2. Dà risultati migliori sia degli AD che degli AG</li> </ol>	<ol style="list-style-type: none"> <li>1. Ipotesi di normalità dei dati finanziari</li> <li>2. Le matrici di varianza e covarianza devono essere uguali</li> </ol>
<b>Reti neurali</b>	<ol style="list-style-type: none"> <li>1. Capacità di apprendere dalla realtà e di generalizzare</li> <li>2. Capacità di adattarsi a situazioni complesse e dinamiche</li> <li>3. Capacità di dare poco peso agli outliers</li> <li>4. Velocità computazionale</li> </ol>	<ol style="list-style-type: none"> <li>1. Fase di apprendimento lungo</li> <li>2. Grande quantità di dati</li> <li>3. Overfitting<sup>32</sup></li> </ol>

**Tabella 10:** Pregi e difetti di diversi modelli utilizzati per l'analisi del rischio di insolvenza  
**Fonte:** D'Annunzio e Falavigna, 2004.

### 3.4 Le reti neurali ibride

Le reti neurali ibride sono considerate dei modelli complessi e sono state trattate per la prima volta nello studio *Neural network forecasting of short, noisy time series* di Foster, Collopy e

<sup>32</sup> D'Annunzio e Falavigna asseriscono che l'overfitting è un particolare fenomeno che può presentarsi nella fase di apprendimento della rete neurale; ovvero quanto vengono presentati i casi del *training set* al fine di minimizzare gli errori di classificazione. Questo fatto fa in modo che la rete si specializzi solamente nel classificare i casi presentati in fase di apprendimento.

Ungar del 1992. Questi inizialmente confrontarono le reti neurali con un semplice modello costruito su una regressione lineare ed in un secondo momento combinarono le due tecniche ottenendo delle stime più soddisfacenti rispetto ai due sistemi (reti neurali e regressione semplice) distinti.

Il più grande lavoro in tema delle reti neurali ibride è però da attribuire a Lee, Han e Kwon i quali, nella ricerca *Hybrid neural network models for bankruptcy predictions* del 1996, paragonarono l'efficacia dei risultati delle reti neurali con altre procedure statistiche. Essi evidenziarono la forte capacità previsiva dei modelli ibridi paragonando 3 modelli: il primo ottenuto dall'unione di sistemi a reti neurali, il secondo ottenuto mediante l'analisi discriminante multivariate e il terzo era un ID3. Il risultato di questa ricerca mostra che l'ibrido ad offrire dei risultati migliori è stato il *SOFM<sup>33</sup>(MDA)-assisted NN* nel quale i modelli ID3 e MDA vengono utilizzati in una prima fase di classificazione dei dati e per la selezione di altre variabili significative per la costruzione della rete neurale.

In analogia con la ricerca appena vista, lo studio *A comparison of corporate failure models in Australia: Hybrid neural networks, logit models and discriminant analysis* di Yim e Mitchell analizza se i modelli statistici tradizionali possano essere superati in favore di due particolari reti ibride MLP. Questi due modelli ibridi sono stati costruiti secondo dei particolari criteri:

- Si è fatto inizialmente ricorso ai modelli statistici per la selezione delle variabili *inputs*.
- Le variabili *inputs* rilevate, in termini probabilistici, sono state poi introdotte nella ANN.

Yim e Mitchell hanno motivato quindi l'introduzione dei modelli statistici generali, all'interno delle reti neurali, riprendendo il problema dell'*overfitting*; questo problema di specializzazione della rete si presenta se le variabili in *inputs* sono elevate. Sostanzialmente in questa ricerca vengono introdotti tre generi di modelli ibridi:

- I modelli *Logit* e dell'Analisi Discriminante vengono utilizzati nella fase di selezione delle variabili. I modelli facenti parte di questa categoria prendono il nome di ANN-Logit e ANN-DA.
- I modelli Logit e dell'Analisi Discriminante stimano la probabilità di fallimento dell'impresa (*Probability of Default*) la quale viene poi introdotta come input della rete. Questi modelli vengono nominati ANN-Plogit e ANN-PDA.
- I modelli Logit e dell'Analisi Discriminante vengono utilizzati per la selezione delle variabili e per la stima della Probabilità di Default dell'impresa le quali vengono poi inserite come *inputs* della rete. I modelli costruiti con questo processo vengono chiamati ANN-Logit-Plogit, ANN-Logit-PDA, ANN-DA-PDA e ANN-DA-Plogit.

---

<sup>33</sup> SOFM (Self Organizing Feature Map) è una tipologia di rete neurale ottenuta dalla fusione tra un modello SOM (Self Organizing Map) e LVQ (Linear Vector Quantization). (Fonte: Modelli di analisi e previsione del rischio di insolvenza. Una prospettiva delle metodologie applicate di Nadia D'Annunzio e Greta Falavigna)

I risultati di questa ricerca sono illustrati nelle 2 tabelle sottostanti le quali evidenziano come i modelli ibridi siano migliori in termini previsivi rispetto alle altre procedure statistiche. La tabella 11 si riferisce ai risultati ottenuti applicando i vari modelli ai dati del *training set* mentre la tabella 12 riporta la precisione di stima dei modelli se applicati a dati diversi da quelli del *training set*.

Model	1 year before failure		2 year before failure	
	Non failed firms	Failed firms	Non failed firms	Failed firms
	correctly classified (%)	correctly classified (%)	correctly classified (%)	correctly classified (%)
<b>Analisi Discriminante</b>	86	75	86,3	60
<b>Logit</b>	91	80	91,2	55
<b>ANN</b>	94	80	95	65
<b>ANN-DA</b>	98	75	96,2	65
<b>ANN-PDA</b>	96	80	95	75
<b>ANN-DA-PDA</b>	93	75	96,2	65
<b>ANN-Logit</b>	98	70	96,2	65
<b>ANN-Plogit</b>	96	85	95	75
<b>ANN-Logit-Plogit</b>	93	85	97,5	65
<b>ANN-DA-Plogit</b>	93	85	97,5	65
<b>ANN-Logit</b>	91	80	95	65

**Tabella 11:** Tipologie di modelli ibridi creati e precisione delle stime su dati applicati al *training set*.  
**Fonte:** Yim e Mitchell, 2002.

Secondo la ricerca di Yim e Mitchell quindi i risultati più performanti, come anche esplicitato dalla tabella sopra, sono generati dai modelli ANN-Plogit, ANN-Logit-Plogit e ANN-DA-Plogit se applicati a dati riguardanti il *training set*. Questi modelli ibridi dominano i sistemi previsivi più elementari. Anche se si utilizza invece un campione di dati non proveniente dal set di *training*, i modelli ibridi e in generale le reti neurali offrono dei risultati migliori rispetto a delle semplici modelli statistici (vedi *tabella 12*).

Da questa ultima ricerca analizzata, Nadia D'Annunzio e Greta Falavigna (2004) affermano che le reti ibride saranno i sistemi di previsione più utilizzati rispetto a tutti gli altri modelli poiché, combinando assieme diverse metodologie di analisi, si riesce a superare i punti deboli di queste, valorizzandone allo stesso tempo i loro pregi. Uno dei limiti però che continuano a presentare le reti neurali resta le difficoltà inerenti l'analisi di significatività, in particolare nella

costruzione degli intervalli di confidenza; cosa che al contrario risulta essere normale nei modelli statistico-econometrici.

Model	1 year before failure	
	Non failed firms	Failed firms
	correctly classified (%)	correctly classified (%)
<b>Analisi Discriminante</b>	86	60
<b>Logit</b>	89	60
<b>ANN</b>	94	50
<b>ANN-DA</b>	92	50
<b>Ibrido (ANN-PDA)</b>	94	50
<b>Ibrido (ANN-DA-PDA)</b>	92	60
<b>ANN-Logit</b>	94	50
<b>Ibrido (ANN-Plogit)</b>	89	60
<b>Ibrido (ANN-Logit-Plogit)</b>	89	60
<b>Ibrido (ANN-DA-Plogit)</b>	94	80

**Tabella 12:** Risultati nel caso in cui non vengano utilizzati dati appartenenti al training set

**Fonte:** Yim e Mitchell, 2002.

Sempre secondo Nadia D'Annunzio e Greta Falavigna la soluzione risulta essere quella di applicare gli *algoritmi genetici* alle reti neurali; nella ricerca appena richiamata si fa anche una breve descrizione di questa prima procedura da adottare. Si inizia generando casualmente un determinato numero di reti neurali descritte dal relativo vettore dei pesi, in seguito si procede incrociando, in modo casuale, a due a due le varie reti con l'obiettivo di creare una popolazione finita di individui. Alla fine di questo processo rimarranno solamente quei sistemi incrociati che meglio di tutti gli altri riescono a fornire delle soluzioni consistenti in merito alla previsione dell'insolvenza. Questa procedura si presenta di difficile implementazione utilizzando semplicemente un *personal computer* in quanto ci si deve dotare di una elevata capacità computazionale altrimenti non si giungerebbe ad una soluzione in breve tempo.

La seconda soluzione fornita da D'Annunzio e Greta per la creazione di modelli ibridi è quella che si basa sulla logica *fuzzy* la quale assegna dei valori numerabili a variabili qualitative<sup>34</sup>.

<sup>34</sup> Si riporta ad esempio la capacità imprenditoriale o in generale si fa riferimento a delle caratteristiche soggettive del soggetto da affidare o affidato



## Capitolo 4: Descrizione quantitativa del neurone artificiale

Come definito in precedenza il neurone è costituito da un insieme di sinapsi che fungono da collegamento con i terminali di altri neuroni, da una funzione di attivazione e da una soglia. In letteratura si usa non fare alcune distinzioni tra le connessioni sinaptiche e, nella quasi totalità dei modelli, si tende non prendere in considerazione i possibili ritardi di trasmissione dovuti a caratteristiche qualitative come, ad esempio, la lunghezza (Dario Floreano, Claudio Mattiussi, *Manuale sulle Reti Neurali*, Il Mulino, 2002).

Definito come  $w$  il peso attribuito alla sinapsi, l'effetto di un segnale  $x$  sul neurone postsinaptico è uguale a  $w \cdot x$ ; a questo punto si può definire il *potenziale di attivazione* o *input netto* ( $A_i$ ) di un neurone  $i$ -esimo come la somma dei prodotti fra tutti i segnali in ingresso ( $x_j$ ) e i valori dei pesi corrispondenti alle varie sinapsi ( $w_{ij}$ )

$$A_i = \sum_j^N w_{ij} x_j \quad [4.1]$$

dove  $i$  rappresenta quale sia il neurone ricevente,  $j$  si riferisce al neurone emittente il segnale che riceverà l'unità  $i$  mentre  $N$  si riferisce al numero totale di neuroni.

I pesi  $w_{ij}$  delle sinapsi possono essere positivi o negativi, anche continui; questi, come si vedrà, vengono modificati durante la fase di apprendimento per poter meglio rappresentare il fenomeno oggetto di analisi (Floreano, Mattiussi, 2002).

A questo punto si deve sottrarre il valore della soglia  $\vartheta_i$  appartenente allo specifico neurone ottenendo:

$$A_i = \sum_j^N w_{ij} x_j - \vartheta_i \quad [4.2]$$

La risposta data dal neurone ( $y_i$ ) è uguale alla *funzione di attivazione*  $\Phi(A)$  applicata al potenziale di attivazione:

$$y_i = \Phi(A_i) = \Phi\left(\sum_j^N w_{ij} x_j - \vartheta_i\right) \quad [4.3]$$

Sono varie le funzioni di attivazione applicate alle reti neurali e solamente alcune verranno trattate in seguito.

Essendo le reti neurali composte da più neuroni i quali sono a loro volta costituiti da numerose connessioni sinaptiche, risulta più agevole l'utilizzo di un sistema in notazione vettoriale. Il potenziale di attivazione di un intero strato di neuroni  $A^T = [A_1, A_2, \dots, A_m]$ , essendo il potenziale di attivazione di un singolo nodo funzione lineare degli input in ingresso, può essere semplicemente riscritto secondo la seguente notazione

$$A = \mathbf{W} \cdot \mathbf{x}$$

Questa uguaglianza definisce un prodotto tra la matrice  $\mathbf{W}(m \times n) = \{w_{11}, w_{12}, \dots, w_{1n}, w_{21}, w_{22}, \dots, w_{mn}\}$  e un vettore  $\mathbf{x}^T = \{x_1, x_2, \dots, x_n\}$  dove la prima riporta tutte le connessioni sinaptiche (le righe  $m$  corrispondono al numero dei neuroni riceventi mentre le colonne  $n$  corrispondono ai segnali di ingresso) mentre la seconda rappresenta i segnali di ingresso.

#### 4.1 La funzione di attivazione

La funzione di attivazione è fondamentale per poter determinare quale sia l'output che un singolo neurone è in grado di generare.

Una delle prime formulazioni risale al 1943 dalla ricerca *A logical calculus of the Ideas Immanent in Nervous Activity* condotta da McCulloch e Pitts i quali teorizzarono una risposta del neurone basata su una funzione a gradino, mantenendo la soglia del neurone ( $\vartheta$ ) fuori dal calcolo del potenziale di attivazione.

$$\Phi(A) = \begin{cases} 1 & \text{se } A > \vartheta \\ 0 & \text{altrimenti} \end{cases} \quad [4.1.1]$$

Altrimenti l'output del neurone può essere bipolare:

$$\Phi(A) = \begin{cases} 1 & \text{se } A > \vartheta \\ -1 & \text{altrimenti} \end{cases} \quad [4.1.2]$$

Queste due realizzazioni del possibile output del neurone definiscono, attraverso due stati discreti, la possibile risposta; il neurone può quindi trasmettere solamente un bit di

informazione. Se invece si utilizzasse una funzione di attivazione che sia anche continua e lineare l'informazione trasmessa può essere maggiore (Floreano, Mattiussi, 2002). Definendo  $k$  come una costante, una generica funzione di attivazione potrebbe essere

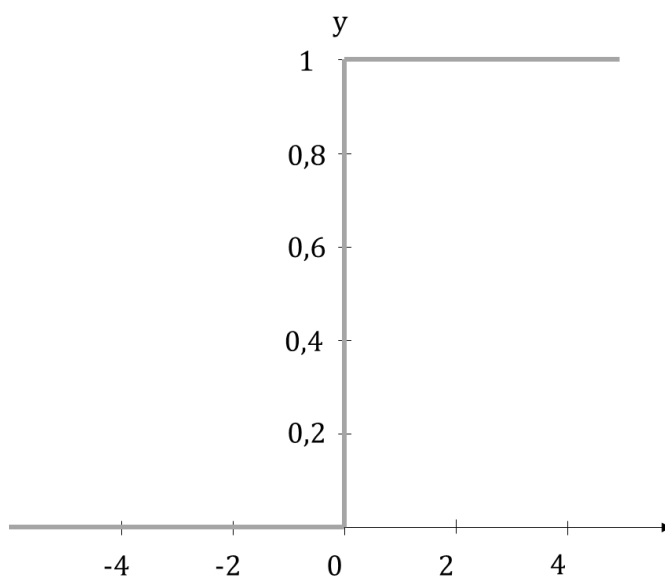
$$\Phi(A) = kA \quad [4.1.3]$$

Le funzioni continue presentano delle particolarità in quanto nel sistema neurale permettono al neurone di trasmettere dei segnali secondo diversi gradi di intensità i quali possono essere opportunamente sfruttati dai neuroni riceventi.

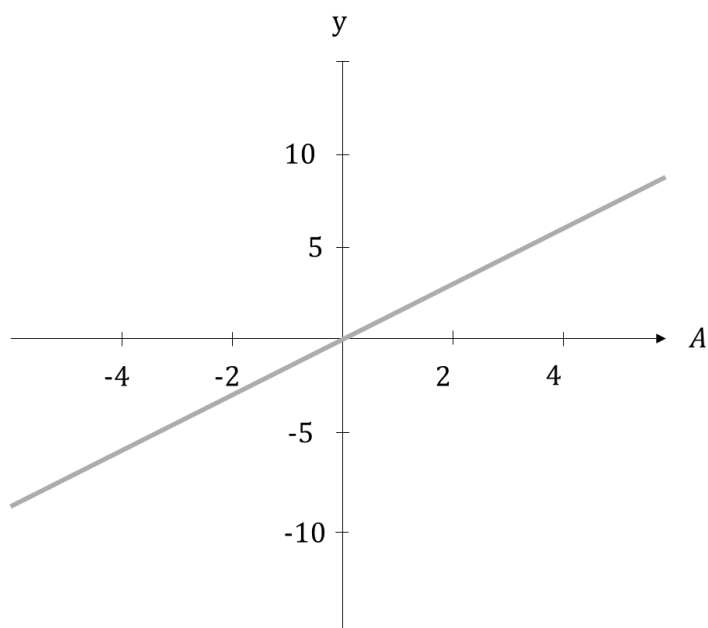
Alle funzioni continue si aggiunge anche qualche tipologia non-lineare; tra queste è bene ricordare quella più comune, ovvero la funzione *sigmoide* o *logistica*.

$$\Phi(A) = \frac{1}{1 + e^{-kA}} \quad [4.1.4]$$

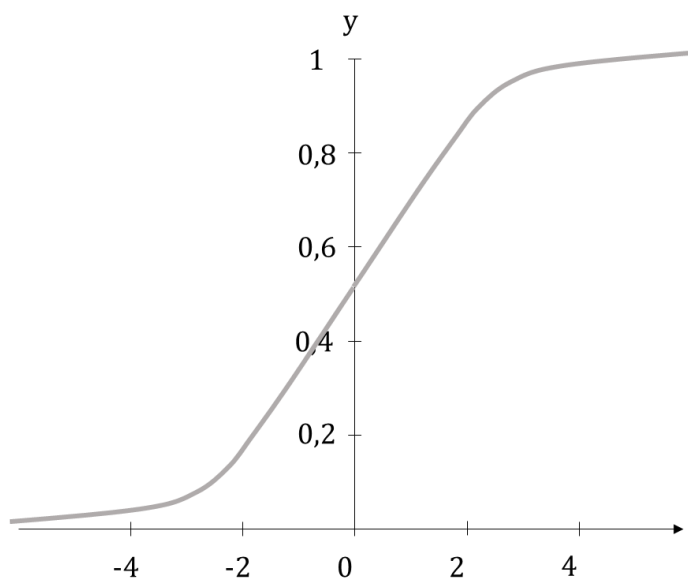
In questo caso  $k$  è una costante che determina l'inclinazione della curva e se  $k \rightarrow \infty$  la funzione sigmoide approssima la funzione a gradino. In questa funzione possono essere individuati due asintoti orizzontali rappresentati dalle rette  $y = 0$  e  $y = 1$ .



**Figura 7:** Funzione di attivazione a gradino con  $\vartheta = 0$   
**Fonte:** Floreano, Mattiussi, 2002



**Figura 8:** Funzione di attivazione continua  
**Fonte:** Floreano, Mattiussi, 2002



**Figura 9:** Funzione di attivazione sigmoide  
**Fonte:** Floreano, Mattiussi, 2002

In quasi tutti i sistemi di reti neurali tutte le unità, ad eccezione di quelle in ingresso, utilizzano la medesima funzione di attivazione per poter calcolare il proprio segnale d'uscita.

## 4.2 Apprendimento

La risposta di una rete neurale è data dai valori sinaptici di connessione che vi sono tra i nodi. A priori è difficile individuare i valori associati a tali sinapsi in una rete neurale che deve esibire un certo comportamento: questi eventi sono rari e ristretti a casi estremamente semplici (Floreano, Mattiussi, 2002). Le reti neurali, similmente alle reti biologiche, modificano i propri valori sinaptici mediante una presentazione, anche ripetuta, di vari esempi; questa è una caratteristica estremamente importante poiché, proprio come nella realtà, esse riescono ad apprendere dall'esperienza passata.

In letteratura si distinguono due modalità di apprendimento:

- *Apprendimento supervisionato (o apprendimento con insegnante)*: i valori sinaptici vengono modificati utilizzando una misura dell'errore tra la rete neurale e la risposta desiderata per ogni vettore di input. L'apprendimento viene eseguito su un set di coppie di pattern dei quali uno è composto da un *set di informazioni in input  $x$*  mentre l'altro è formato *dall'insieme delle risposte osservate  $t$* . L'apprendimento supervisionato deve comprendere anche una serie di algoritmi che si limitino ad offrire una stima della bontà di risposta della rete neurale piuttosto che la risposta corretta associata ad ogni pattern di addestramento. Questa tipologia di algoritmi quindi sono definiti tramite il cosiddetto *apprendimento per rinforzo*.
- *Apprendimento per auto-organizzazione*: in questa tipologia di organizzazione della rete neurale non vi è una risposta ottimale desiderata dall'ambiente esterno ma vi è l'introduzione di alcune semplici regole di "plasticità sinaptica" (Floreano, Mattiussi, 2002) le quali vanno ad auto-organizzare le reti neurali mentre sono sottoposte all'analisi di alcuni pattern di input. Le reti neurali che riescono ad operare secondo questo modello di apprendimento possono estrarre autonomamente delle informazioni dall'ambiente esterno, a classificare una larga serie di pattern e per tale motivo, in alcuni casi, tali modelli artificiali si presentano con una struttura interna molto simile a quella biologica.

La distinzione tra questi due modelli di apprendimento non si presenta tuttavia molto marcata. È, per esempio, possibile utilizzare degli algoritmi per l'*apprendimento supervisionato* senza sapere quale sia l'effettiva risposta corretta. In altri casi invece, per l'*apprendimento per auto-organizzazione*, è possibile che l'informazione in ingresso debba essere formulata in modo preciso con delle dettagliate assunzioni in merito ai pattern di ingresso.

Tutti gli algoritmi di apprendimento presentano degli elementi in comune:

1. I valori dei pesi sinaptici vengono inizialmente presi in maniera casuale all'interno di un certo campo di variazione solitamente molto piccolo, ad esempio  $[-0,1; 0,1]$ , oppure possono essere tutti fissati pari a 0.

2. L'apprendimento consiste nella presentazione di una serie di vettori chiamati anche *pattern di apprendimento*. Come visto prima nel caso di apprendimento supervisionato, i pattern sono composti da due vettori (vettore di ingresso e vettore afferente alla risposta desiderata) mentre nell'apprendimento per auto-organizzazione vi è un solo vettore nel pattern di apprendimento (quello di ingresso).

La modifica di valori sinaptici  $\Delta w_{ij}$  appartenenti ad una data rete neurale artificiale viene calcolata dopo ogni presentazione di un generico pattern (*apprendimento per cicli*) oppure dopo che tutti i pattern di addestramento sono stati presentati alla rete artificiale (*apprendimento per epoche*). Con entrambe queste tipologie di apprendimento, il nuovo valore associato ad un peso sinaptico dopo un ciclo/epoca è calcolato addizionando la modifica ricavata  $\Delta w_{ij}$  alla configurazione sinaptica presente prima dell'inizio del ciclo/epoca  $w_{ij}^{t-1}$ :

$$w_{ij}^t = w_{ij}^{t-1} + \Delta w_{ij}^t \quad [4.2.1]$$

Sostanzialmente l'utilizzo degli algoritmi di apprendimento riguarda solamente il calcolo di  $\Delta w_{ij}^t$  quindi il processo di apprendimento punta ad introdurre nuove conoscenze  $\Delta w_{ij}^t$  a quelle già consolidate in precedenza  $w_{ij}^{t-1}$ .

3. Al fine di non perdere le conoscenze acquisite in precedenza con la presentazione di nuovi pattern di input, l'apprendimento avviene in modo *ricorsivo* e *graduale*. Ciò implica innanzitutto che un medesimo pattern di apprendimento viene presentato più volte e che solamente una piccola parte della modifica sinaptica viene effettivamente aggiunta ai valori sinaptici. La velocità di apprendimento viene regolata da una costante  $\eta$  chiamata *tasso di apprendimento* la quale definisce la porzione di modifica che viene data ai valori sinaptici; essa può assumere valori anche maggiori di 1 ma in questo caso si può incorrere in potenziali elementi di instabilità (Floreano, Mattiussi, 2002). L'equazione generale di apprendimento diventa quindi

$$w_{ij}^t = w_{ij}^{t-1} + \eta \Delta w_{ij}^t \quad 0 < \eta < 1 \quad [4.2.2]$$

4. Terminata la fase di apprendimento i valori delle sinapsi vengono congelati in modo da poter studiare la risposta che la rete rende applicandole dei *pattern di test*.<sup>35</sup> Questa fase

---

<sup>35</sup> La distinzione tra fase di apprendimento e di test non si utilizza agli algoritmi di apprendimento operanti secondo la *Teoria della Risonanza Adattiva*; qui le reti neurali attivano in modo autonomo il processo di apprendimento quando gli vengono applicati dei pattern sconosciuti.

di test comporta la presentazione di vettori di ingresso e il calcolo dell'attivazione dei nodi della rete senza però che i pesi sinaptici vengano modificati: in questo modo si può analizzare:

- La capacità della rete nel generalizzare anche se sottoposta a nuovi stimoli
- Il tipo di soluzione trovata dalla rete per adempiere ad un dato compito.

La capacità di apprendere quindi è uno dei maggiori punti di forza delle reti neurali in quanto permette di risolvere dei problemi non giungendo necessariamente ad una soluzione analitica ma sottoponendo la rete neurale ad una serie di esempi. Questo spiega il perché le reti neurali sono utilizzate per risolvere delle problematiche che richiedono l'applicazione di trasformazioni non lineari le quali sarebbero di difficile analisi con dei metodi analitici i quali, in questi casi, sono impossibilitati o di non facile applicazione nel fornire delle soluzioni ottimali (Floreano, Mattiussi, 2002).

### 4.3 Il neurone artificiale: analisi vettoriale

Considerando che la risposta data una semplice unità lineare è:

$$y = a \left( \sum_i^N w_i x_i \right) \quad a = 1 \quad [4.3.1]$$

questa relazione può essere vista come il prodotto interno tra il vettore  $\mathbf{w}$ , rappresentante i valori sinaptici, e il vettore  $\mathbf{x}$  degli input:

$$y = \mathbf{w} \cdot \mathbf{x} \quad [4.3.2]$$

Il valore scalare  $y$  corrisponde alla risposta dell'unità, ovvero una misura della somiglianza tra il vettore di input e il vettore dei valori sinaptici. Sapendo infatti che la norma di un vettore ne stabilisce la sua lunghezza

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \quad [4.3.3]$$

e che l'angolo  $\vartheta$  (più precisamente il coseno) formato dai due vettori  $\mathbf{w}$  e  $\mathbf{x}$  è dato da:

$$\cos(\vartheta) = \frac{\mathbf{w} \cdot \mathbf{x}}{\|\mathbf{w}\| \|\mathbf{x}\|} \quad 0 \leq \vartheta \leq \pi \quad [4.3.4]$$

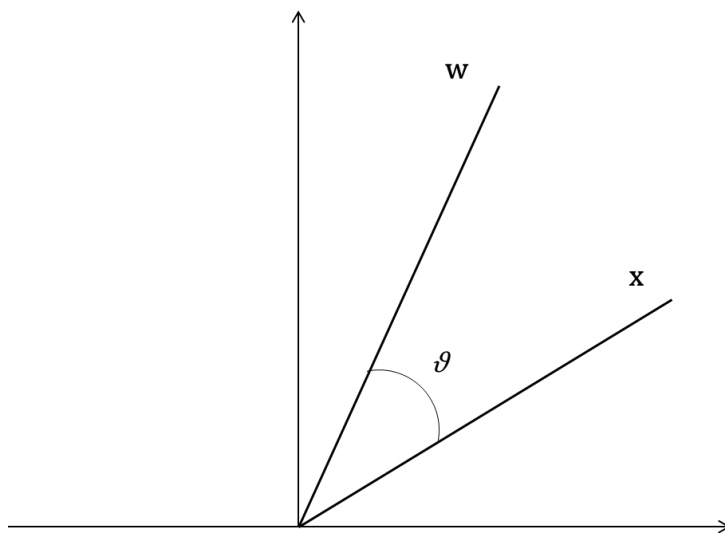
attraverso dei semplici calcoli algebrici, si giunge a definire il prodotto dato dalla [4.3.2] come

$$\mathbf{w} \cdot \mathbf{x} = \|\mathbf{w}\| \|\mathbf{x}\| \cos(\vartheta) \quad [4.3.5]$$

Con questo si vuole affermare che se si immagina di traslare nello spazio i due vettori, mantenendo inalterata a loro lunghezza, allora il loro prodotto interno è proporzionale al coseno dell'angolo  $\vartheta$  formatosi tra i due vettori nello spazio. Per questo se:

$$\begin{aligned} \vartheta = 0^\circ &\rightarrow \cos(\vartheta) = 1 \\ \vartheta = 90^\circ &\rightarrow \cos(\vartheta) = 0 \\ \vartheta = 180^\circ &\rightarrow \cos(\vartheta) = -1 \end{aligned}$$

Quindi la risposta data dal neurone (il prodotto) sarà tanto maggiore quanto più piccola è la distanza angolare che vi è tra il vettore degli input e il vettore sinaptico all'interno del medesimo quadrante. Se la risposta di un'unità risulta essere nulla allora i due vettori formano, tra di loro, un angolo di  $90^\circ$  (c.d. ortogonali); se invece la distanza risulta essere maggiore di  $90^\circ$  l'unità assume valori negativi.



**Figura 10:** rappresentazione vettoriale di pesi sinaptici e pattern di input appartenenti ad un'unità con due sinapsi  
**Fonte:** Floreano, Mattiussi, 2002

Un generico neurone che utilizza una funzione di attivazione a gradino  $\Phi(x) \in \{0,1\}$ <sup>36</sup> e con una soglia fissata sul valore zero evidenzia perciò se il vettore degli input si trova ad una distanza minore (la risposta risulta essere uguale ad 1) o maggiore (risposta uguale ad 0) di

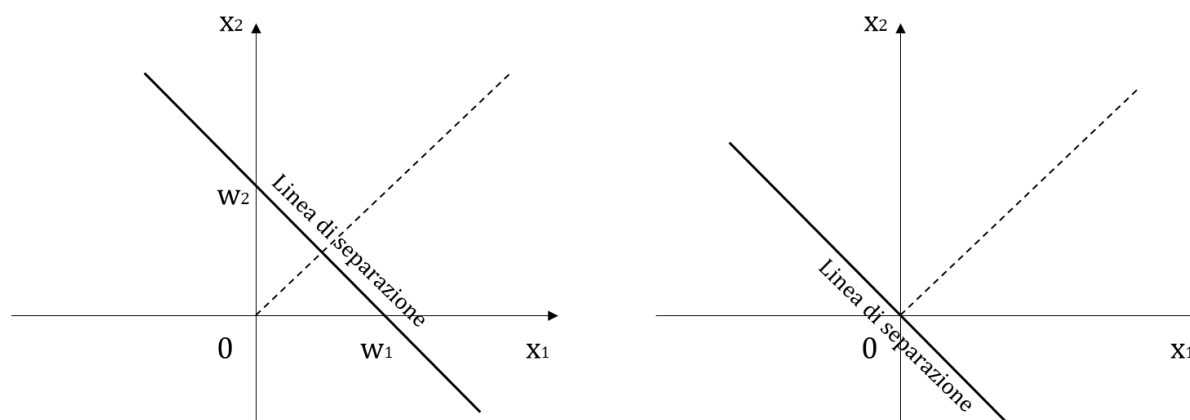
<sup>36</sup> Questa tipologia di unità è chiamata neurone di *McCulloch e Pitts* (Floreano, Mattiussi, 2002).



un angolo retto dal proprio vettore sinaptico. Da sottolineare è che in una rete con molte unità, per osservare quale di questa possiede un vettore sinaptico simile al pattern di input, ci si deve basare sul livello di attivazione solamente se i vettori sinaptici sono stati normalizzati (Floreano, Mattiussi, 2002).

Generalmente una rete neurale artificiale è utilizzata per poter classificare dei pattern. Nel caso appena visto si sta trattando un neurone binario, ovvero un'unità che opera una classificazione imparando ad attivarsi solamente se vi è un determinato gruppo di pattern in ingresso ma rimanendo inattiva in tutti gli altri casi. Per apprendere si deve sviluppare un vettore di valori sinaptici atto a tracciare una linea di demarcazione nello spazio dell'input in modo tale che tutti i pattern di un gruppo si trovino da una parte della linea mentre tutti gli altri nel lato opposto. Se in un dato nodo sussistono solamente due connessioni in ingresso allora è opportuno parlare di *linea di separazione*, se ve ne sono tre si deve fare riferimento ad un *piano* mentre se vi sono più di tre connessioni si fa riferimento ad un *iperpiano di separazione* nello spazio dimensionale dell'input. Se si presenta una di queste tre opzioni per la classificazione dei gruppi di input in due categorie allora tale separazione può essere definita *lineare*. Questo si presenta come un elemento importante anche per la scelta dell'architettura relativa alla rete neurale artificiale; per esempio tutte le reti neurali che presentano solamente uno strato di sinapsi non riescono a classificare in modo adeguatamente corretto i vettori di input che non possano essere linearmente separabili (Floreano, Mattiussi, 2002).

Sia che si parli di linea, di piano o di iperpiano, questi sono tutti perpendicolari rispetto ai pesi sinaptici del singolo nodo. Nei grafici sottostanti si riporta la linea di separazione di uno spazio di input la quale viene caratterizzata da un neurone con due connessioni di ingresso poste uguali a 1. Nel grafico di sinistra la soglia  $\vartheta$  è posta uguale a 1 mentre in quello di destra è uguale a 0. In entrambi i grafici si può individuare: la linea di separazione (linea grossa) e la linea del vettore sinaptico (linea tratteggiata).



**Figura 11:** Linea di separazione dello spazio di input in due neuroni con due connessioni d'ingresso. Nella figura a sinistra la soglia  $\vartheta = 1$  mentre in quello a destra la soglia  $\vartheta = 0$ . La linea nera evidenzia la linea di separazione mentre la linea tratteggiata la direzione del vettore sinaptico.

**Fonte:** Floreano, Mattiussi, 2002.

Quindi un nodo ad attivazione bipolare il quale abbia due connessioni in ingresso, rende una risposta (o output) pari a:

$$y = \Phi(A) = \Phi\left(\sum_j w_j x_j - \vartheta\right) \quad [4.3.6]$$

mentre la linea di separazione che individua due regioni nelle quali in una  $A > 0$  e nell'altra  $A < 0$  è definibile secondo la relazione:

$$w_1 x_1 + w_2 x_2 - \vartheta = 0 \quad [4.3.7]$$

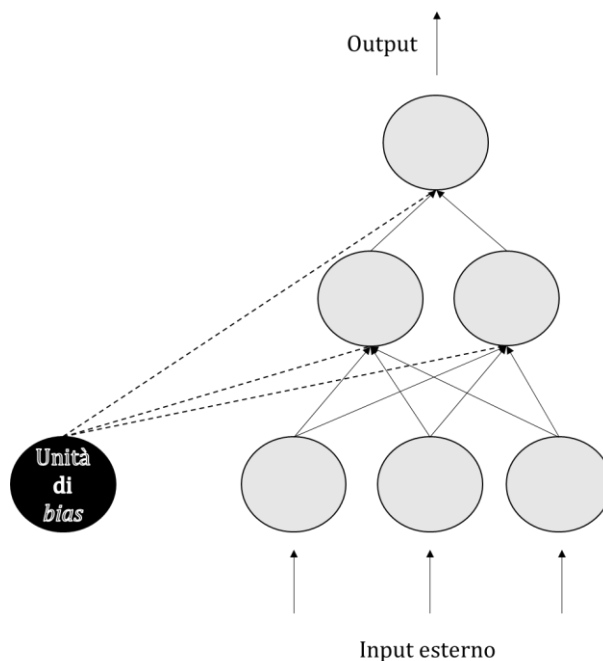
con  $\vartheta$  pari al valore della soglia. Procedendo algebricamente si ricava l'equazione:

$$x_2 = \frac{\vartheta}{w_2} - \frac{w_1}{w_2} x_1 \quad [4.3.8]$$

In questa relazione la soglia e i pesi sinaptici sono determinati dall'algoritmo di apprendimento. In  $A > 0$  tutti i vettori di informazioni in input provocheranno un output pari a 1 mentre tutti gli altri pattern, classificati nella regione  $A < 0$ , avranno come output  $-1$ .

Dall'ultimo grafico riportato si può notare che se  $\vartheta = 0$  la soglia non è presente poichè la linea di separazione passa per l'origine dello spazio dell'input quindi il neurone potrebbe avere delle difficoltà nella separazione tra diverse distribuzioni di pattern. Impostando allora una casuale configurazione sinaptica iniziale e una serie di pattern di input per svolgere l'apprendimento, si deve muovere il vettore sinaptico in modo che la linea di separazione classifichi correttamente i pattern nei due gruppi.

La soglia  $\vartheta$  può quindi essere definita come un peso sinaptico di valore  $\vartheta$  legata ad un'unità aggiuntiva in input che abbia un'attivazione costante pari a  $-1$ . In questo caso il peso legato alla sinapsi prende il nome di *bias* ( $w_0$ ) mentre il neurone aggiuntivo prende il nome di *unità di bias* ( $x_0$ ). Ogni unità all'interno della rete neurale possiede una connessione di *bias* mentre le unità di input non la posseggono in quanto la loro attivazione dipende dal corrispondente vettore di ingresso.



**Figura 12:** Riproduzione di una rete neurale con relativa unità di bias in nero. Ogni unità possiede una connessione di bias mentre le unità in input ne sono prive poichè vengono attivate dall'elemento corrispondente del vettore di ingresso  
**Fonte:** Floreano, Mattiussi, 2002

Con questo per poter attivare il neurone si utilizza la seguente notazione:

$$y = \Phi \left( \sum_{j=0} w_j x_j \right) \quad [4.3.9]$$

Dove la sommatoria viene fatta partire da zero e  $x_0 = -1$ . La soglia quindi viene vista come un ulteriore peso sinaptico derivante da un neurone sempre attivo; ciò implica che, a seguito di un processo di apprendimento, essa può essere modificata come per gli altri pesi sinaptici della rete (Floreano, Mattiussi, 2002).

#### 4.4 L'analisi di una rete neurale e gli strumenti per la sua valutazione

Il periodo di addestramento di una rete neurale dipende essenzialmente dal paradigma di apprendimento utilizzato.

1. L'apprendimento supervisionato: in questa tipologia di apprendimento si fa ricorso *all'errore medio fra la risposta desiderata e la risposta fornita dalla rete neurale* per ciascun pattern di addestramento. Si termina la fase di addestramento, con la registrazione dei valori sinaptici, quando l'errore raggiunge un valore minimo prefissato; si passa quindi alla

fase di test. Per impostare il livello minimo di errore ci si deve basare su svariati fattori come il grado di precisione che si vuole ottenere, la funzione di attivazione utilizzata, il numero di unità di output e la distribuzione dei pattern nello spazio. Si deve inoltre considerare che vi è un *trade-off* tra la precisione nella fase di apprendimento dei pattern e la capacità di riuscire a generalizzare ai nuovi pattern di test.

2. L'apprendimento per auto-organizzazione: in questo caso non vi è una misura dell'errore ma per alcuni di questi algoritmi può essere definita una *funzione di energia* (o *funzione oggettiva*) dei parametri della rete. Questa funzione ci restituisce una misura continua in merito alla convergenza della rete verso una soluzione. In alternativa possono essere utilizzati anche altre metodologie come l'analisi di alcuni indici. Un primo esempio di indice può essere la misura della variazione delle sinapsi: si ritiene infatti di aver raggiunto una soluzione quando i valori delle sinapsi hanno una certa stabilità. Un secondo indice che può essere preso in considerazione riguarda la variazione dello stato di attivazione dei nodi della rete per ogni singolo pattern.

Se invece si volesse valutare la qualità della risposta di una rete neurale dopo la fase di addestramento ci si deve essenzialmente domandare quale sia la sua finalità di impiego.<sup>37</sup>

Utilizzare un modello neurale comporta il vantaggio di poter studiare le dinamiche di sviluppo dello stesso; questo permette di capire al meglio il funzionamento del modello finale risultante ma anche di analizzare quantitativamente i risultati di molti esperimenti. Floreano e Mattiussi nel *Manuale sulle Reti Neurali* (2002) riportano che questi esperimenti possono essere quantificati anche se svolti in ambito psicologico verso soggetti animali o umani; si può quindi comprendere in che modo questi soggetti imparano.

A differenza del sistema neurale presente nei soggetti umani, la valutazione di una Rete Neurale Artificiale non può avvenire mediante una semplice "interrogazione" quindi si ritiene opportuno pianificare a priori la struttura e la metodologia simulativa; questo è fatto in modo da rappresentare al meglio la situazione sperimentale che si ha intenzione di analizzare. Una ulteriore divergenza si basa sul fatto che nella sperimentazione su soggetti umani od animali, le prestazioni possono essere misurate per esempio calcolando il tempo di risposta del soggetto esaminato; con le reti neurali, non si può procedere in questo modo poiché tali strutture, come risposta, restituiscono solamente l'attivazione o meno dei nodi di uscita.

Malgrado le differenze che sussistono tra soggetti umani e le reti neurali risulta essere più produttivo individuare dei confronti sulle proprietà delle prestazioni. Floreano e Mattiussi (2002) a questo proposito riportano l'esempio delle reti neurali che utilizzano il meccanismo di apprendimento supervisionato: in questo confronto tra reti si deve tenere conto di quali siano

---

<sup>37</sup> In merito all'approfondimento di questo argomento si faccia riferimento Dario Floreano e Claudio Mattiussi nel *Manuale sulle Reti Neurali*, 2002, pag. 57.

i pattern in ingresso che provocano i gli errori più marcati, se vi è sistematicamente una risposta errata per dati tipi di pattern oppure, in caso di errore, quale sia la risposta della rete. In generale infatti la misura di errore della rete neurale è sintomo del grado di difficoltà nella elaborazione del pattern in ingresso; questa motivazione ci consente di poter paragonare l'errore della rete neurale<sup>38</sup> con le percentuali delle risposte corrette e dei relativi tempi dei soggetti analizzati. In questo contesto quindi non devono essere prese in considerazione le scale di misura ma esclusivamente l'andamento delle prestazioni sotto i vari vincoli imposti dall'esperimento.

Un'altra tipologia di valutazione concerne l'analisi interna della rete ottenuta dopo la fase di apprendimento. Prestando attenzione alle caratteristiche di attivazione dei nodi interni e alla configurazione dei valori sinaptici si può giungere al tipo di soluzioni utilizzate dalla rete per risolvere un dato problema. Se si volesse, ad esempio, analizzare una rete *feedforward*<sup>39</sup> avente un solo strato di sinapsi, la visualizzazione dei campi consiste nella proiezione su un piano dei valori sinaptici associati a ciascun neurone; a valori assoluti più alti corrisponde un'associazione più forte la quale, a sua volta può significare una specializzazione del neurone per quanto riguarda una data risposta; questo fa in modo che si possa comprendere quali siano gli stimoli in input che in maniera più efficace riescano ad attivarlo. Questa metodologia però si presenta molto complessa nel caso in cui gli input da introdurre nella rete abbiano una dimensionalità molto elevata (Floreato, Mattiussi, 2002).

Un metodo alternativo per l'analisi dei valori sinaptici punta a estrapolare le caratteristiche dei campi recettivi sulla base della misurazione dell'attività dei nodi stessi. Questa capacità di attivazione di un nodo si calcola mediante la presentazione di molteplici stimoli nel suo campo recettivo in modo da poter stimare delle curve di risposta che vanno a definire la selettività del nodo stesso; ciò permette quindi di poter delineare le caratteristiche dello stimolo e risulta molto utile per l'analisi dei nodi interni alle reti multistrato.

L'analisi di come avviene l'attivazione di un singolo nodo non ci dice nulla in merito sul come l'intera rete organizza internamente i vari stimoli in ingresso. Per poter giungere a quest'informazione si deve procedere mediante un'analisi per gruppi (o *cluster analysis*) in merito all'attivazione di tutti i nodi interni, misurata per ciascun pattern in ingresso. Al termine di questa analisi si ottiene una suddivisione degli input (o delle somiglianze percepite dalla rete) in funzione dei nodi attivati interni.

---

<sup>38</sup> L'errore della rete neurale può essere per esempio la somma delle discrepanze quadratiche medie fra il vettore di output della rete e il vettore della risposta corretta, la distanza di Hamming o altre in base al modello utilizzato.

<sup>39</sup> La rete *feedforward* fu uno tra i primi modelli di rete neurali messi a punto; si caratterizza per la sua semplicità in quanto le informazioni si muovono esclusivamente in una sola direzione ovvero dalle unità di input verso quelle degli strati di output. Per questo motivo questo sistema non crea cicli e si differenzia quindi dalle reti neurali ricorrenti. La rete *feedforward* è anche chiamata *rete neurale con flusso in avanti*.

La valutazione quindi punta a sottolineare le qualità della modello neurale in funzione delle finalità di ricerca la quale non sempre si limita a voler riprodurre dei dati sperimentali esistenti ma anche a fornire delle nuove predizioni.

#### 4.5 L'apprendimento della rete neurale: la regola di Hebb

Tutti gli algoritmi di apprendimento sono stati teorizzati e costituiti basandosi sulla regola di modifica sinaptica teorizzata da Donald Hebb nella ricerca *The Organization of Behaviour* del 1949<sup>40</sup> o su delle sue riformulazioni. La regola hebbiana, secondo Floreano e Mattiussi 2002, per quanto semplice costituisce la base su cui poter costruire dei modelli neurali più complessi. Hebb afferma che se due neuroni, collegati tra di loro, sono attivi contemporaneamente allora l'efficacia sinaptica viene rafforzata e quindi la sola attivazione di un dato nodo comporta l'attivazione anche dell'altro. La forza di tale connessione viene rafforzata ogni volta che i due neuroni saranno attivi contemporaneamente.

Se si prende ad esempio una rete neurale avente nodi di uscita caratterizzati da una funzione di attivazione binaria e se gli si presenta un pattern di ingresso  $\mathbf{x}$  e uno di uscita  $\mathbf{y}$ , la modifica dei pesi sinaptici risulta essere:

$$\Delta w_{ij} = \eta y_i x_j \quad [4.5.1]$$

con  $\eta$  pari al tasso di apprendimento. Da ciò la matrice finale dei valori sinaptici sarà uguale a tutte le variazioni calcolate per ciascuna coppia di pattern  $\mu$  sommate:

$$w_{ij} = \sum_{\mu} \Delta w_{ij}^{\mu} \quad [4.5.2]$$

Questa formula vale se si ipotizza che la fase di apprendimento inizi imponendo la matrice dei pesi  $w_{ij} = 0$  e fissando  $\eta = 1$

Dopo che ogni coppia è stata appresa, se si presenta un pattern di input, anche incompleto, automaticamente la rete va ad identificare il corrispondente pattern in uscita.

La regola di Hebb tuttavia presenta dei punti deboli in quanto non riesce ad apprendere da input che presentano degli elementi in comune ma che richiedono delle risposte diverse associate ai vari neuroni di output. In questo scenario infatti i neuroni in uscita attivati sono

---

<sup>40</sup> La traduzione in italiano di questa opera è stata fatta nel 1975 con il titolo *L'organizzazione del comportamento. Una teoria neuropsicologica*, Milano, Angeli.

molteplici e quindi offrono una risposta mista e non univoca<sup>41</sup>. In sintesi la regola di Hebb permette di apprendere esclusivamente in presenza di pattern in ingresso ortogonali i quali devono quindi offrire un prodotto dei singoli componenti che sia pari a 0 (Floreato, Mattiussi, 2002). Viene definita "interferenza" la produzione di risposte miste a causa della sovrapposizione dei pattern.

#### 4.5.1 La regola postsinaptica, presinaptica e della covarianza

Visto che l'apprendimento hebbiano può presentare alcune problematiche si deve procedere con il ridurre l'efficacia sinaptica. La regola *Stent-Singer* o *regola postsinaptica* (si veda *A Physiological Mechanism for Hebb's Postulate of Learning* di Stent, 1973) infatti asserisce che il valore della connessione sinaptica sia incrementato ogni volta che l'unità presinaptica e postsinaptica sono entrambe attive nel medesimo istante; allo stesso tempo però questa connessione viene indebolita se, al momento dell'attivazione dell'unità postsinaptica, l'unità presinaptica rimane inattiva. Da ciò si perviene al seguente risultato:

$$\Delta w_{ij} = \eta [y_i x_j + (x_j - 1) y_i] \quad [4.5.3]$$

La parte  $y_i x_j$  è quella modifica sinaptica legata all'apprendimento in cui entrambe le unità (presinaptica e postsinaptica) si presentano attive; la parte  $(x_j - 1) y_i$  costituisce quella parte di modifica sinaptica in cui l'apprendimento è dato dalla non attivazione del nodo presinaptico mentre il nodo postsinaptico si è attivato. Ipotizzando che il nodo postsinaptico sia sempre attivo, se:

- Il nodo presinaptico è attivo ( $x_j = 1$ ) la funzione appena descritta si trasforma nell'equazione della variazione dei pesi sinaptici vista per la regola di Hebb.
- Se il nodo presinaptico non è attivo ( $x_j = 0$ ) allora la modifica dei pesi diventa funzione della sola risposta  $-y_i$  ponderata per il tasso di apprendimento  $\eta$ .

La *regola postsinaptica* riesce a superare il presentarsi del fenomeno dell'interferenza ma anche questa metodologia di apprendimento presenta un punto critico; infatti se si introducessero nella rete troppi pattern di input, parzialmente sovrapposti ai quali deve essere

---

<sup>41</sup> Il fenomeno secondo il quale vengono prodotte delle risposte miste, dovute alla sovrapposizione di pattern in ingresso, è definita "interferenza".

associato un medesimo pattern di output, la regola postsinaptica non riesce ad apprendere in modo corretto poiché si creano troppe sinapsi inibitorie<sup>42</sup> (Floreano, Mattiussi, 2002).

La *regola presinaptica* si propone diametralmente opposta alla regola postsinaptica. Il valore della connessione sinaptica cresce se l'unità presinaptica e postsinaptica sono entrambe attive ma decresce se l'unità postsinaptica è inattiva nel momento in cui l'unità presinaptica è attiva. Tale relazione può quindi essere descritta come segue:

$$\Delta w_{ij} = \eta[y_i x_j + (y_j - 1)x_j] \quad [4.5.4]$$

Questa regola funziona meglio rispetto alla regola di *Stent-Singer* quando ad uno stesso pattern di uscita vengono associati diversi pattern di input parzialmente sovrapposti (Floreano, Mattiussi, 2002).

L'ultima regola di apprendimento basata sulla teoria di Hebb si presenta come la combinazione delle precedenti due appena trattate (postsinaptica e presinaptica) e viene chiamata *regola della covarianza* o di *Hopfield* (si veda *Neural networks and physical systems with emergent collective computational abilities* di Hopfield, 1982). Questa norma prevede che la connessione venga rafforzata se l'unità presinaptica e postsinaptica si presentano nello stesso stato; quindi o entrambe attive o entrambe inattive mentre, se dovessero presentarsi in due stati diversi, allora la connessione si indebolisce. Si deve sottolineare che se la regola della covarianza viene applicata ad una rete i cui nodi utilizzano una funzione di attivazione bipolare allora, tale regola di apprendimento, assume la stessa forma della regola di Hebb. Per questo motivo questa norma prende il nome di *regola di Hebb estesa* e presenta la peculiarità di poter svolgere dei compiti di classificazione o di associazione complessi.

Le capacità di apprendimento della regola postsinaptica, presinaptica e della covarianza sono state confrontate in uno studio dal titolo *Optimal Plasticity from Matrix Memories: What goes up Must Come down* di Willshaw e Dayan del 1990; in questa ricerca si evidenzia che il metodo di apprendimento ottimale, ovvero quello che offre il minor numero di errori, dipende dalla proporzione di unità attive  $p$  nei pattern di input e dalla proporzione di unità attive  $r$  nei pattern di output per ciascun pattern di addestramento. Per questo motivo Willshaw e Dayan, in relazione allo stato dell'unità presinaptica e postsinaptica, sono riusciti a stimare le quantità ottimali di modifica per ogni regola di apprendimento.

---

<sup>42</sup> da sottolineare è che durante il processo di apprendimento una sinapsi può invertire il suo segno in seguito all'accumulazione delle modifiche. Tale caratteristica è comune quasi in tutte le reti neurali artificiali ma non nei sistemi nervosi biologici.



Per quanto riguarda la regola postsinaptica vale la relazione:

$$\Delta w_{++}^+ = 1 - p, \quad \Delta w_{+-}^- = p \quad [4.5.5]$$

in queste due relazioni gli apici indicano se la modifica è positiva o negativa mentre i pedici indicano in che stato si trovano, in ordine, le unità postsinaptiche e presinaptiche.

Se invece si prendesse in considerazione la regola presinaptica risulta invece:

$$\Delta w_{++}^+ = 1 - r, \quad \Delta w_{-+}^- = r \quad [4.5.6]$$

mentre per la regola della covarianza si utilizza la combinazione della regola postsinaptica e presinaptica:

$$\Delta w_{++}^+ = (1 - p)(1 - r), \quad \Delta w_{--}^- = pr, \quad \Delta w_{-+}^- = (1 - p)r, \quad \Delta w_{+-}^- = (1 - r)p, \quad [4.5.7]$$

In pratica queste relazioni ci mostrano quale sia il potenziale di miglioramento relativo ai pesi sinaptici ma, malgrado queste, le regole di apprendimento hebbiano presentano delle problematiche relativamente alla tipologia di associazioni che sono in grado di apprendere. Esse infatti sono sempre soggette al fenomeno dell'interferenza quanto i pattern di input non sono linearmente indipendenti (Floreano, Mattiussi, 2002).

#### 4.6 L'apprendimento supervisionato e il perceptrone semplice

L'*apprendimento supervisionato* presenta un insegnante esterno (ovvero non compreso nella rete neurale) atto a dare informazioni per la correzione della risposta data alla rete neurale. L'informazione può consistere nella risposta desiderata dalla rete per ciascun pattern di apprendimento (*apprendimento con insegnante classico*) o nella stima della bontà della risposta della rete (*apprendimento critico o per rinforzo*).

L'apprendimento hebbiano può anche presentare la caratteristica di essere supervisionato nel caso in cui le reti siano predisposte in modo da associare a ciascun pattern di input un dato pattern di output. Tuttavia, secondo Dario Floreano e Claudio Mattiussi (2002), se si combinasse l'apprendimento hebbiano con quello supervisionato si darebbe luogo ad un caso abbastanza anomalo in quanto non si richiederebbe il calcolo, durante la fase di addestramento, dell'attivazione dei nodi di output.

Rosenblatt nel *Principles of Neurodynamics* (1962) per la prima volta teorizzò la possibilità di adattare i pesi sinaptici di una rete neurale artificiale sulla base dell'errore generato tra la

risposta dalla rete e la risposta osservata. Le reti neurali che sorsero presero il nome di *perceptroni di Rosenblatt* ed erano caratterizzate da due strati di connessioni nelle quali le sinapsi del primo strato erano fissate (e quindi non modificabili) in modo casuale<sup>43</sup>. Le sinapsi del secondo strato dovevano invece superare un processo di apprendimento; per questo motivo, per *perceptroni* si intende una rete neurale nella quale vi sia un unico strato di connessioni unidirezionali rivolte verso i nodi di output a partire dai nodi di entrata. In letteratura sono stati poi trovati molti tipi di perceptroni differenziati in base alla tipologia di nodi utilizzati; quelli che utilizzano nodi binari, bipolar o bipolar con una zona di indecisione sono anche detti perceptroni semplici in quanto utilizzano una semplice funzione a gradino.

#### 4.7 Regola di apprendimento con un unico nodo di uscita ad attivazione bipolare

Avendo  $P$  coppie di addestramento, ognuna con un vettore di ingresso bipolare  $\mathbf{x}_p$ , una risposta desiderata  $t_p$  bipolare, e una rete neurale con  $N$  nodi di ingresso allora l'output per ciascun pattern di ingresso della rete può essere rappresentato dalla seguente funzione

$$y = \begin{cases} 1 & \text{se } \sum_{i=0}^N w_i x_i \\ -1 & \text{altrimenti} \end{cases} \quad [4.7.1]$$

Il risultato ottenuto da un certo pattern di input deve poi essere confrontato con la risposta desiderata  $t_p$ . Nel caso in cui la rete dia una risposta identica alla risposta desiderata allora non è necessario modificare i valori sinaptici mentre se sussiste una divergenza tra le due (ovvero un errore) i pesi sinaptici devono essere modificati in funzione della risposta desiderata. Tale modifica dei pesi viene calcolata secondo la funzione:

$$\Delta w_i = \eta t x_i \quad [4.7.2]$$

La variazione dei pesi sinaptici deve quindi essere sommata ai valori delle sinapsi precedenti secondo la formula già vista in precedenza:

$$w_i^t = w_i^{t-1} + \Delta w_i^t \quad [4.7.3]$$

---

<sup>43</sup> Costituiscono quindi un modo alternativo per riportare l'attivazione dei recettori sui nodi di input della rete neurale.

I valori iniziali delle connessioni sinaptiche possono essere impostati a piacimento; solitamente o si pongono uguali a 0, o si assumono piccoli valori casuali o si possono porre uguali ad uno qualsiasi dei pattern di addestramento<sup>44</sup>. Il valore del tasso di addestramento  $\eta$  solitamente è ipotizzato essere pari a 1 ma secondo Floreano e Mattiussi (2002) può anche essere fissato a  $1/||x||$  per ciascun pattern che provoca un cambiamento dei pesi sinaptici in modo che il vettore di modifica dei pesi sinaptici sia di lunghezza pari a 1.

A questo punto le modifiche dei pesi sinaptici dei perceptron avvengono solamente quando sorgono degli errori delle risposte della rete al momento della presentazione dei pattern in ingresso: si richiede quindi la presentazione ripetuta dei pattern di addestramento. Procedendo in questo modo le modifiche sinaptiche decrescono poiché la rete impara a fornire progressivamente meno risposte scorrette. Quando gli errori di risposta per ogni pattern vengono totalmente annullati, cessa anche la fase di apprendimento. Da questa metodologia di "allenamento" della rete neurale si possono quindi evincere due proprietà dei perceptron:

1. Gradualità dell'apprendimento: questa proprietà è rispettata presentando più volte le coppie di addestramento
2. La regola di apprendimento è sempre in grado di trovare il vettore (o i vettori), se esiste, dei pesi sinaptici ottimali per soddisfare la trasformazione dei pattern di input ed output; questo processo avviene in un numero finito di cicli. Tale condizione può essere rispettata esclusivamente se i pattern sono linearmente separabili.

La regola di apprendimento e le due proprietà appena viste del perceptrone possono essere applicate anche ad input binari ma in questo caso le modifiche vengono fatte solamente nei pesi sinaptici corrispondenti ad unità di ingresso attive le quali hanno provocato una risposta scorretta. Con questo Floreano e Mattiussi (2002) affermano che i perceptron sono molto più potenti del meccanismo di apprendimento hebbiano poiché, per esempio, secondo quanto riportato da Fausett nel *Fundamentals of Neural Networks* del 1994 la regola di Hebb non riesce ad apprendere la funzione AND quando si utilizzano input binari mentre un perceptrone trova la soluzione (vettore dei pesi sinaptici) in 10 epoche<sup>45</sup>.

Confrontando la regola di Hebb con quella del perceptrone si possono trovare anche alcuni elementi in comune; quello più importante è sicuramente legato al fatto che entrambi modificano le proprie sinapsi in funzione di un evento pre e postsinaptico. Per quanto concerne l'algoritmo di apprendimento si vuole sottolineare come questo possa essere applicato anche alle reti neurali costituite da più di un nodo di uscita poiché, siccome ogni nodo è indipendente dagli altri, la rete stessa può essere scomposta in tanti piccoli perceptron strutturati con un solo nodo di output, medesimo vettore di input ma diverse connessioni sinaptiche.

---

<sup>44</sup> Si veda Minsky e Papert 1988 per un eventuale approfondimento

<sup>45</sup> Per "epoca" si intende la presentazione di tutte le coppie di addestramento.

#### 4.8 Percettrone semplice: dimostrazione sulla convergenza

In questo capitolo si riporta la dimostrazione in forma vettoriale contenuta nel *Perceptrons. Expanded Edition* di Minsky e Papert del 1988 in cui si mostra che un percettrone semplice converge verso la soluzione in una serie finita di presentazioni dei pattern di addestramento se esiste un vettore di pesi sinaptici che costituisce tale soluzione.

Avendo  $P$  coppie di addestramento si procede inizialmente scomponendo il gruppo di addestramento in due sottoinsiemi in funzione del tipo di risposta desiderata.

$$F^+ = \{\mathbf{x} \text{ tale che } t = +1\} \quad [4.8.1]$$

$$F^- = \{\mathbf{x} \text{ tale che } t = -1\} \quad [4.8.2]$$

In questo modo l'esistenza di una soluzione è legata al fatto che vi sia un vettore dei pesi sinaptici  $\mathbf{w}^*$  per cui

$$\mathbf{x} \cdot \mathbf{w}^* > 0 \quad \text{se } \mathbf{x} \in F^+ \quad [4.8.3]$$

e

$$\mathbf{x} \cdot \mathbf{w}^* < 0 \quad \text{se } \mathbf{x} \in F^- \quad [4.8.4]$$

A questo punto si deve creare un nuovo insieme

$$F = F^+ \cup F^- \quad [4.8.5]$$

nel quale

$$-F^- = \{-\mathbf{x}, \text{ tale che } \mathbf{x} \in F^-\} \quad [4.8.6]$$

Sostanzialmente sono stati invertiti i segni degli elementi dei pattern di input che corrispondevano alla risposta desiderata  $-1$ . Ora quindi l'esistenza della soluzione  $\mathbf{w}^*$  è data da una sola condizione:

$$\mathbf{x} \cdot \mathbf{w}^* > 0 \quad \text{se } \mathbf{x} \in F \quad [4.8.7]$$

la quale restituisce tutte le risposte desiderate del nuovo insieme rendendole pari a +1. A questo punto la regola di apprendimento impone che se un vettore in input provoca un errore nella risposta, ovvero

$$\mathbf{x} \cdot \mathbf{w} \leq 0, \quad [4.8.8]$$

indicando con  $k$  l'istante temporale, la modifica dei pesi sinaptici avviene nel seguente modo:

$$\mathbf{w}(k) = \mathbf{w}(k - 1) + \eta t \mathbf{x} \quad [4.8.9]$$

Dal momento che nel nuovo insieme  $F$  risulta che  $t = +1$ , assumendo  $\eta = 1$ , la regola di apprendimento può essere riscritta come segue:

$$\mathbf{w}(k) = \mathbf{w}(k - 1) + \mathbf{x} \quad [4.8.10]$$

Ipotizzando quindi di sottoporre tutti i pattern di addestramento (in modo ciclico) alla rete neurale e di tenere conto solamente degli istanti in cui si incorre in un errore della rete in quanto va a modificare i pesi sinaptici, si avrà:

$$\begin{aligned} \mathbf{w}(1) &= \mathbf{w}(0) + \mathbf{x} \\ \mathbf{w}(2) &= \mathbf{w}(1) + \mathbf{x} \\ \mathbf{w}(3) &= \mathbf{w}(2) + \mathbf{x} \\ \mathbf{w}(k) &= \mathbf{w}(k - 1) + \mathbf{x}. \end{aligned}$$

L'intero processo di modifica quindi può essere rappresentato secondo l'equazione

$$\mathbf{w}(k) = \mathbf{w}(0) + \mathbf{x} + \mathbf{x} + \dots + \mathbf{x} = \mathbf{w}(0) + k\mathbf{x} \quad [4.8.11]$$

La dimostrazione necessita ora di dimostrare che  $k$  non cresce all'infinito quindi, considerando l'esistenza di un vettore  $\mathbf{w}^*$  tale per cui vale la [4.8.7], si prenda uno scalare  $m$  in modo tale che:

$$m = \min_p \{\mathbf{x} \cdot \mathbf{w}^*\}. \quad [4.8.12]$$

Questa relazione rappresenta il potenziale di attivazione minimo tra tutti i pattern  $P$  di addestramento secondo il vettore desiderato  $\mathbf{w}^*$ . Dal momento che quindi  $\mathbf{x}(p) \cdot \mathbf{w}^* \geq m$  per tutte le  $P$  coppie di addestramento, risulta quindi che:

$$\mathbf{w}(k) \cdot \mathbf{w}^* \geq \mathbf{w}(0) \cdot \mathbf{w}^* + km. \text{ }^{46} \quad [4.8.13]$$

A questo punto, sapendo che la disuguaglianza di *Cauchy-Schwartz* ci dice che prendendo due vettori qualsiasi  $\mathbf{a}$  e  $\mathbf{b}$  risulta:

$$(\mathbf{a} \cdot \mathbf{b})^2 \leq \|\mathbf{a}\|^2 \|\mathbf{b}\|^2$$

si può introdurre la seguente disuguaglianza:

$$(\mathbf{w}(k) \cdot \mathbf{w}^*)^2 \leq \|\mathbf{w}(k)\|^2 \|\mathbf{w}^*\|^2 \quad [4.8.14]$$

dalla quale, mediante dei semplici calcoli algebrici si perviene a:

$$\|\mathbf{w}(k)\|^2 \geq \frac{[\mathbf{w}(k) \cdot \mathbf{w}^*]^2}{\|\mathbf{w}^*\|^2} \quad [4.8.15]$$

Questa relazione, basandoci sulla [4.8.13], risulta verificata anche se si opera la seguente sostituzione al numeratore:

$$\|\mathbf{w}(k)\|^2 \geq \frac{[\mathbf{w}(0) \cdot \mathbf{w}^* + km]^2}{\|\mathbf{w}^*\|^2} \quad [4.8.16]$$

Ricordando che, in un dato istante  $k$ , i pesi sinaptici sono determinati dall'equazione [4.8.10], essi vengono modificati solamente in presenza di un errore, tale errore è determinato da

$$\mathbf{w}(k-1) \cdot \mathbf{x} \leq 0. \quad [4.8.17]$$

A questo punto, mettendo assieme i risultati ottenuti, si può dire che

$$\|\mathbf{w}(k)\|^2 = \|\mathbf{w}(k-1)\|^2 + \|\mathbf{x}\|^2 + 2\mathbf{x} \cdot \mathbf{w}(k-1) \quad [4.8.18]$$

$$\leq \|\mathbf{w}(k-1)\|^2 + \|\mathbf{x}\|^2 \quad [4.8.19]$$

---

<sup>46</sup> Questa relazione può essere dimostrata a ritroso; essendo  $\mathbf{w}(k) = \mathbf{w}(0) + k\mathbf{x}$  allora risulta

$$[\mathbf{w}(0) + k\mathbf{x}] \cdot \mathbf{w}^* \geq \mathbf{w}(0) \cdot \mathbf{w}^* + km$$

$$\mathbf{w}(0) \cdot \mathbf{w}^* + k\mathbf{x} \cdot \mathbf{w}^* \geq \mathbf{w}(0) \cdot \mathbf{w}^* + km$$

a questo punto, semplificando i fattori comuni si giunge alla condizione esplicitata in precedenza ovvero:

$$\mathbf{x} \cdot \mathbf{w}^* \geq m$$

e procedendo a ritroso si ottiene

$$\begin{aligned}
 \|\mathbf{w}(k)\|^2 &\leq \|\mathbf{w}(k-1)\|^2 + \|\mathbf{x}\|^2 \\
 &\leq \|\mathbf{w}(k-2)\|^2 + \|\mathbf{x}\|^2 + \|\mathbf{x}\|^2 \\
 &\leq \|\mathbf{w}(k-1)\|^2 + \|\mathbf{x}\|^2 + \|\mathbf{x}\|^2 + \|\mathbf{x}\|^2 \\
 &\vdots \\
 &\leq \|\mathbf{w}(0)\|^2 + k\|\mathbf{x}\|^2
 \end{aligned} \tag{4.8.20}$$

Considerando ora un valore scalare  $M$  espressione della lunghezza quadratica in assoluto più grande di tutti i vettori di addestramento  $P$

$$M = \max_p(\|\mathbf{x}\|^2) \tag{4.8.21}$$

si può riscrivere la disuguaglianza [4.8.20] come:

$$\|\mathbf{w}(k)\|^2 \leq \|\mathbf{w}(0)\|^2 + kM. \tag{4.8.22}$$

A questo punto se si mettono assieme la relazione espressa dalla [4.8.22] con la [4.8.16] si ottiene il margine di variazione del vettore dei pesi sinaptici

$$\frac{[\mathbf{w}(0) \cdot \mathbf{w}^* + km]^2}{\|\mathbf{w}^*\|^2} \leq \|\mathbf{w}(k)\|^2 \leq \|\mathbf{w}(0)\|^2 + kM \tag{4.8.23}$$

In questo caso si può notare che il limite inferiore è quadraticamente dipendente dal parametro  $k$  mentre il limite superiore è linearmente dipendente con  $k$ ; ciò ci dice che questa relazione non può essere soddisfatta per  $k$  eccessivamente grande quindi il numero di correzioni esplicitate indirettamente da questo parametro deve essere finito.

Al fine di poter stimare in via teorica il numero di iterazioni necessarie per raggiungere una soluzione si può fare riferimento al caso in cui i pesi iniziali della rete siano tutti 0 (quindi  $\mathbf{w}(0) = 0$ ) e in cui il vettore desiderato possieda lunghezza 1 ( $\|\mathbf{w}^*\| = 1$ ).

Adoperando le necessarie sostituzioni nella relazione [4.8.23] questa risulta essere;

$$(km)^2 \leq kM. \tag{4.8.24}$$

Per questo il numero massimo di modifiche sinaptiche  $k$  è definito da:

$$k \leq \frac{M}{m^2} \quad [4.8.25]$$

Come detto in precedenza, il valore  $k$  è meramente teorico in quanto  $m$  non è calcolabile, essendo il vettore dei pesi desiderati  $\mathbf{w}^*$  sconosciuto. Questa ultima relazione però ci dice che, in generale, un perceptrone raggiungerà una soluzione (se esiste) in termini vettoriali in un numero finito di iterazioni ma tale valore non è a priori stimabile.

## 4.9 La regola delta

La regola delta è una ulteriore regola di apprendimento la quale si presenta molto simile al meccanismo di apprendimento dei perceptron semplici appena trattato ma, mentre quest'ultimo è caratterizzato da unità con attivazione a gradino, si differenzia per il fatto che essa può essere utilizzata per quei neuroni di output che presentano una funzione di attivazione continua e differenziabile. La continuità e la differenziabilità di queste unità consente la descrizione delle performance di una rete neurale e in particolare dei suoi errori mediante una funzione continua  $E_w$ . La differenziabilità della funzione di errore può essere fatta rispetto alla matrice dei pesi  $\mathbf{W}$ ; ciò ci consente di poter osservare il cambiamento dell'errore della rete rispetto ad una variazione dei pesi sinaptici (Floreato e Mattiussi, 2002).

### 4.9.1 La regola delta applicata ad unità lineari

Prendendo ad esempio una rete neurale di tipo *feedforward* con unità di output ad attivazione lineare si sa che la sua risposta è definita da

$$y_i = \sum_{j=0} w_{ij} x_j \quad [4.9.1]$$

Considerando ora un gruppo di pattern di addestramento formato da  $M$  coppie le quali sono a loro volta composte da un vettore di input  $\mathbf{x}^u$  e da un vettore di output desiderato  $\mathbf{t}^u$ , si vuole stimare il valore ottimale dei pesi sinaptici tali per cui

$$y_i^u = t_i^u \quad \forall i, u \quad [4.9.2]$$



Le performance generali della rete possono essere descritte tramite la funzione di errore (o funzione di costo)  $E_W$  la quale, per far sì che il modello risulti ottimale, deve essere ridotta il più possibile:

$$E_W = \frac{1}{2} \sum_u \sum_i (t_i^u - y_i^u)^2 \quad [4.9.3]$$

Analiticamente la funzione appena vista rappresenta lo scarto quadratico medio tra la risposta desiderata e la risposta ricavata da tutte le unità della rete sommate su tutte le unità e su tutte le coppie di addestramento<sup>47</sup>. Dal momento che si stanno impiegando delle unità lineari la funzione di errore può anche essere vista nel seguente modo

$$E_W = \frac{1}{2} \sum_u \sum_i \left( t_i^u - \sum_{j=0} w_{ij} x_j^u \right)^2 \quad [4.9.4]$$

Tanto più le risposte stimate della rete sono vicine a quelle desiderate tanto più questa funzione di costo restituisce dei valori piccoli. Tale funzione  $E_W$  dipende da un unico parametro, ovvero dal valore associato alle connessioni sinaptiche contenute nella matrice  $\mathbf{W}$  infatti la direzione di massima crescita della funzione (gradiente) è descritta dal vettore delle derivate parziali di  $E_W$  rispetto a ciascun peso sinaptico  $w_{ij}$ .

Poiché il valore di  $E_W$ , in pratica lo scarto quadratico medio tra la risposta desiderata e la risposta ottenuta, deve essere minimizzato, i pesi sinaptici devono essere modificati in via opposta rispetto al gradiente di  $E_W$  quindi

$$\Delta w_{ij} = - \frac{\partial E}{\partial w_{ij}} \quad [4.9.5]$$

Sapendo inoltre che ogni unità di risposta è indipendente dalle altre, se si risolve la derivata per ogni singola unità si ottiene la seguente equazione

$$\Delta w_{ij} = \sum_u (t_i^u - y_i^u) x_j^u \quad [4.9.6]$$

---

<sup>47</sup> La frazione serve unicamente a semplificare alcuni calcoli matematici nei passaggi successivi

la quale ci permette di asserire che la variazione dei pesi sinaptici per ciascuna coppia di apprendimento  $u$ , prendendo un dato tasso di apprendimento  $\eta$ , è il risultato della differenza tra la risposta desiderata e l'output della rete moltiplicata per l'attività presinaptica.

$$\Delta w_{ij} = \eta(t_i - y_i)x_j. \quad [4.9.7]$$

Come detto in precedenza l'apprendimento può essere condotto sia per cicli che per epoche e quando vengono utilizzate delle unità ad attivazione lineare queste due modalità di apprendimento sono equivalenti (Floreano e Mattiussi, 2002).

Uguualmente alla regola del perceptrone, anche la regola delta impone che i pattern di addestramento siano presentati più volte alla rete neurale; la differenza più marcata si può individuare nel fatto che quest'ultima opera anche con funzioni continue. La condizione sufficiente ma non necessaria affinché la regola delta, applicata a unità continue, possa arrivare ad una soluzione sta nel nell'indipendenza lineare dei pattern di ingresso. Questa regola infatti a volte riesce a giungere ad una soluzione anche quando i pattern in ingresso sono linearmente dipendenti, purchè siano linearmente separabili; in questo caso, l'ottenimento di una risposta soddisfacente dipende da come vengono formulati i vettori di risposta desiderata (Floreano e Mattiussi, 2002).

La regola di apprendimento delta, possedendo un algoritmo che si basa sulla differenza  $\delta$  tra la risposta desiderata e la risposta ottenuta, viene brevemente indicata, dal punto di vista matematico, con

$$\Delta w_{ij} = \eta \delta_i x_j \quad [4.9.8]$$

ed essa può essere anche definita: *regola di Widrow-Hoff*<sup>48</sup>, *regola ADALINE*<sup>49</sup> o *regola LMS*<sup>50</sup>. Dal momento che la regola di apprendimento delta si basa sulla stima dei pesi sinaptici che minimizzino la funzione di costo  $E_W$ , si può intuitivamente vedere la funzione di errore come la superficie di uno spazio a  $n$  dimensioni dove  $n$  corrisponde ai pesi sinaptici della rete e ciascuna dimensione descrive lo spazio di potenziale variazione di ciascuna sinapsi<sup>51</sup>. Dal momento che la superficie è caratterizzata da massimi e minimi, se i pattern in ingresso sono linearmente indipendenti, la funzione di errore  $E_W$  per unità lineari presenta forma a scodella con un minimo globale in  $E_W = 0$ . Assumendo che il tasso di apprendimento sia

---

<sup>48</sup> Widrow-Hoff: dal nome degli autori della ricerca *Adaptive Switching Circuits*, 1960.

<sup>49</sup> ADALINE: in inglese corrisponde a *Adaptive Linear Neuron*.

<sup>50</sup> LMS: è l'acronimo in inglese per *Least Mean Square* per evidenziare lo scarto quadratico medio utilizzato nella funzione di errore.

<sup>51</sup> Da sottolineare che questa rappresentazione è semplicistica in relazione con reti neurali più complesse che presentano più connessioni o funzioni di attivazione non lineari.

sufficientemente contenuto, la modifica continua dei pesi sinaptici porta a discendere le pareti della superficie fino ad arrivare al punto di minimo ricercato. Se invece i pattern di addestramento non sono linearmente indipendenti la regola delta riesce a discendere le pareti della superficie ma il minimo non è detto sia uguale a 0 se solamente alcune associazioni vengono apprese o se alcune associazioni presentano delle unità di output scorrette. In questo caso il sistema neurale riesce a trovare solamente una soluzione parziale per i pattern di addestramento (Floreano e Mattiussi, 2002).

La velocità di movimento sulla superficie dell'errore è dettata dal tasso di apprendimento  $\eta$ : se tale indice è troppo grande si rischia di spostarsi da una parte all'altra della superficie senza mai trovare il livello minimo di errore ricercato; se  $\eta$  invece è troppo piccolo si giunge a trovare la soluzione al nostro problema (solamente nel caso di unità lineari) ma lo spostamento lungo la superficie rischia di essere troppo lento quindi può essere necessario presentare molte volte le coppie di addestramento alla rete neurale. Windrow, Winter e Baxter nello studio *Learning Phenomena in Layered Neural Networks* del 1987 indicano di assumere il valore del tasso di apprendimento in base alla relazione  $0,1 \leq N\eta \leq 1,0$  nella quale  $N$  rappresenta il numero di unità di input per ogni unità di output. Un'altra ricerca di Hecht-Nielsen del 1990 dal nome *Neurocomputing* stabilisce che i margini siano ben più ampi come ad esempio  $0,1 \leq \eta \leq 10,0$ .

Il *metodo della pseudo inversione* prevede di poter stimare il valore esatto dei pesi sinaptici in un solo passaggio; questo metodo è però applicabile solamente se le unità sono lineari. Se si introduce la *matrice di sovrapposizione dei pattern di input*  $\mathbf{Q}$ , dove  $N$  rappresenta il numero di elementi all'interno di ciascun pattern di input, questa risulta essere pari a

$$\mathbf{Q}_{uv} = \frac{1}{N} \sum_j x_j^u x_j^v \quad [4.9.9]$$

la sua inversa  $\mathbf{Q}^{-1}$  esiste solamente se i pattern di input sono linearmente indipendenti ovvero se  $\mathbf{Q}$  è non singolare. I pesi sinaptici sono quindi calcolabili secondo la formula

$$w_{ij} = \frac{1}{N} \sum_u \sum_v t_i^u (\mathbf{Q}^{-1})_{uv} x_j^v \quad [4.9.10]$$

Questa formula non è applicabile se i pattern sono linearmente dipendenti ma se questi rimangono linearmente separabili allora è applicabile il metodo di discesa del gradiente visto in precedenza (si veda *Introduction to the theory of Neural Computation*, Hertz, Krogh e Palmer 1991, pag. 102) .

#### 4.9.2 La regola delta applicata ad unità continue non-lineari

Avendo una funzione continua, differenziabile e non lineare  $y_i = \Phi(A_i)$  del potenziale di attivazione  $A_i$ , rappresentante l'attivazione dei nodi di uscita, la funzione di errore viene rappresentata secondo la seguente notazione:

$$E_W = \frac{1}{2} \sum_u \sum_i \left[ t_i^u - \Phi \left( \sum_{j=0} w_{ij} x_j \right) \right]^2 \quad [4.9.11]$$

Mentre la regola di modifica sinaptica, ottenuta secondo la regola della discesa del gradiente della funzione rispetto ai pesi sinaptici, diventa

$$\Delta w_{ij} = - \frac{\partial E}{\partial w_{ij}} = \sum_u (t_i^u - y_i^u) \dot{\Phi}(A_i^u) x_j^u \quad [4.9.12]$$

con  $\dot{\Phi}(A_i)$  pari alla derivata prima della funzione di attivazione<sup>52</sup>. La modifica dei pesi sinaptici associata ad ogni coppia di addestramento è rappresentata dalla seguente equazione:

$$\Delta w_{ij} = \eta (t_i - y_i) \dot{\Phi}(A_i) x_j \quad [4.9.13]$$

Qualora la funzione di attivazione assumesse la forma:

- I. Sigmoide essa offre dei valori tra 0 e 1 poichè è definita dall'equazione  $y_i = \Phi(A_i) = (1 + e^{-\beta A_i})^{-1}$ . In questo caso il calcolo della derivata si presenta alquanto semplice ed è esprimibile direttamente basandosi all'output dell'unità stessa ovvero  $\dot{\Phi}(A_i) = \beta y_i (1 - y_i)$
- II. Tangente iperbolica essa dà dei valori compresi nell'intervallo  $-1$  e  $1$  ed è rappresentata dall'equazione  $y_i = \Phi(A_i) = \tanh(A_i) = (1 - e^{-2\beta A_i}) / (1 + e^{-2\beta A_i})$  la quale fornisce una derivata prima uguale a  $\dot{\Phi}(A_i) = \beta (1 - y_i^2)$

Le due funzioni hanno degli asintoti orizzontali i quali descrivono un cambiamento repentino se i valori di  $y_i$  sono intermedi<sup>53</sup>. I pesi sinaptici di unità che possiedono quindi valori intermedi vengono modificati in modo maggiore dalla regola delta poichè possiedono una derivata prima maggiore mentre le unità che offrono risposte vicine agli estremi, avendo derivata molto vicina

<sup>52</sup> Si noti che la derivata presenta un punto sopra il simbolo  $\Phi$ .

<sup>53</sup> Nel caso della funzione sigmoide un valore intermedio si colloca quindi in un'intorno a  $y = 0,5$  mentre per la funzione tangente iperbolica in un intorno a  $y = 0$ .

allo 0, non vedono grosse mutazioni dei propri pesi sinaptici a seguito di tale regola di apprendimento. In parole povere le unità più “incerte” hanno dei pesi sinaptici che vengono maggiormente modificati rispetto a quelli delle unità che invece presentano più “certe”.

Hertz, Krogh e Palmer nella ricerca *Introduction to the Theory of Neural Computation* affermano che se si utilizzano delle unità continue non lineari, l'ottenimento di una soluzione è legato all'indipendenza lineare dei pattern in ingresso ma, a differenza delle unità lineari, se i pattern si presentano linearmente dipendenti allora si ottiene un numero maggiore di soluzioni parziali. Si può inoltre notare che utilizzando delle generali funzione asintotiche come quelle viste in precedenza (funzione sigmoide e tangente iperbolica), la soluzione della rete si avvicina ma non riesce a raggiungere il livello di errore nel minimo assoluto  $E_W = 0$  nel caso in cui le soluzioni desiderate si attestino agli estremi<sup>54</sup>. A differenza delle funzioni continue lineari nelle quali la superficie della funzione di errore assume una forma a scodella, nelle funzioni continue non-lineari la forma risulta essere più articolata definendo diversi minimi locali. La presenza di questi punti nella superficie in cui gli errori sono relativamente piccoli testimonia che vi sono vari possibili set di valori sinaptici applicabili alla rete ma questi tuttavia non rappresentano la vera risposta desiderata.

Nelle reti multistrato, essendo queste non caratterizzate dalle condizioni di indipendenza e separabilità lineare dei pattern in input (a differenza delle reti con un solo strato di connessioni), si rivela particolarmente utile l'applicazione delle unità continue non-lineari. Le reti multistrato riescono quindi a fornire una risposta esatta poiché vi sono uno o più strati di unità interne che rielaborano i pattern in input permettendo alle unità in uscita di dare l'output desiderato. Risulta perciò comune applicare delle unità ad attivazione non lineare alle reti multistrato anche perché una trasformazione lineare di una trasformazione lineare è uguale ad un'unica trasformazione lineare. Una rete multistrato che utilizzi quindi dei nodi ad attivazione lineare, sarebbe uguale ad una rete con un solo strato di sinapsi ad unità lineari e sarebbe perciò sottoposta alle medesime condizioni di indipendenza lineare dei pattern (Floreato e Mattiussi, 2002).

#### 4.10 Back-propagation

Il metodo di *Back-propagation* è un algoritmo applicabile alle reti neurali che presentano architetture molto diverse e con un qualsiasi numero di strati di connessioni; è un'estensione della regola delta per reti neurali multistrato e per questo motivo può essere definito anche come *regola delta generalizzata*. La *Back-propagation* rappresenta uno dei modelli più

---

<sup>54</sup> Quindi [0,1] per la funzione sigmoide e [-1, +1] per la funzione tangente iperbolica

importanti del “connessionismo” (Floreato e Mattiussi, 2002) poichè ha un funzionamento estremamente semplice, è applicabile ad una vasta serie di problemi ed è estremamente potente come metodo di classificazione.

*Back-propagation* è simile alla regola delta poichè i pesi sinaptici vengono modificati in funzione quanto grande è la differenza che vi è tra la risposta fornita dalla rete e la risposta desiderata. Si consideri per esempio un modello di rete semplice come quello *feedforward* avente due strati di connessioni con unità di input indicate con  $x_k$ , unità interne nascoste denotate con  $h_j$  e unità di output associate al simbolo  $y_i$ . Si assegni il valore della risposta corretta per ogni unità di output il simbolo  $t_i^u$  nel quale  $u$  rappresenta un dato pattern di ingresso e si denoti con  $w_{ij}$  le connessioni sinaptiche associate a ciascuna unità di output provenienti dalle unità nascoste le quali, a loro volta ricevono le connessioni sinaptiche  $v_{jk}$  da tutte le unità di input. Si ipotizzi inoltre che tutti i nodi posseggano un’attivazione continua non lineare, per esempio la funzione sigmoide, ad eccezione delle unità in ingresso le quali posseggono un valore pari al pattern in input. Le unità possono avere valori binari, bipolari o continui.

A questo punto, prendendo un pattern in input  $\mathbf{x}^u$  si può procedere inizialmente con la stima dell’attivazione delle unità interne  $h_j$  mediante la seguente equazione

$$h_j^u = \Phi \left( \sum_{k=0} v_{jk} x_k^u \right) \quad [4.10.1]$$

e successivamente si può condurre anche il calcolo dell’attivazione dei nodi di output  $y_i$

$$y_i^u = \Phi \left( \sum_{j=0} w_{ij} h_j^u \right) \quad [4.10.2]$$

In entrambi questi casi i nodi hanno una connessione con l’unità di *bias* (corrispondente al pedice 0) la quale ha valore costante  $-1$ .

Quest’ultimo risultato è utile per poter calcolare  $E_W$  ovvero quanto sia grande la divergenza che vi è tra risposta desiderata  $t_i^u$  e la risposta stimata dal modello  $y_i^u$ . La formula per il calcolo dell’errore è uguale a quella già vista per la regola delta

$$E_W = \frac{1}{2} \sum_u \sum_i (t_i^u - y_i^u)^2 \quad [4.10.3]$$

la quale, se viene espansa, mostra la propria dipendenza dai due gruppi di connessioni sinaptiche ovvero  $w_{ij}$  e  $v_{jk}$

$$E_W = \frac{1}{2} \sum_u \sum_i \left[ t_i^u - \Phi \left( \sum_j w_{ij} \left( \sum_k v_{jk} x_k^u \right) \right) \right]^2 \quad [4.10.4]$$

Calcolato l'errore, per quanto concerne la regola di modifica sinaptica, si può utilizzare il metodo della discesa del gradiente. Se ci si concentrasse sulle connessioni tra le unità nascoste e le unità di uscita, descritte dalla matrice  $w_i$  si ottiene una funzione simile alla regola delta

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = \eta \sum_u (t_i^u - y_i^u) \Phi(A_i^u) h_j^u \quad [4.10.5]$$

la quale, imponendo  $\delta_i^u = (t_i^u - y_i^u) \Phi(A_i^u)$ , diventa

$$\Delta w_{ij} = \eta \sum_u \delta_i^u h_j^u \quad [4.10.6]$$

Per stimare invece come vengono modificate le connessioni dello strato inferiore ci si deve basare sulla *regola di derivazione delle funzioni composte* (Floreato e Mattiussi, 2002).

Questa regola può esplicitarsi in due procedimenti distinti dove:

- i. Il primo considera la variazione dell'attivazione delle unità interne rispetto alla variazione dell'attivazione delle unità di input
- ii. Il secondo invece mette in relazione la variazione dell'attivazione delle unità interne con la variazione dei pesi sinaptici  $v_{jk}$  i quali rendono un segnale da un input costante.

Di seguito si riporta il secondo metodo partendo dall'equazione che definisce la variazione dei pesi  $v_{jk}$

$$\Delta v_{jk} = -\eta \frac{\partial E}{\partial v_{jk}} = \sum_u \frac{\partial E}{\partial h_j^u} \frac{\partial h_j^u}{\partial v_{jk}} \quad [4.10.7]$$

dalla quale, svolgendo le derivate parziali sopraindicate, si ottiene

$$\Delta v_{jk} = \eta \sum_u \sum_i (t_i^u - y_i^u) \Phi(A_i^u) w_{ij} \Phi(A_j^u) x_k^u \quad [4.10.8]$$

Da questa equazione si giunge ad individuare l'operatore sopra definito  $\delta_i^u$  per le unità di output quindi si può procedere con una semplificazione

$$\Delta v_{jk} = \eta \sum_u \sum_i \delta_i^u w_{ij} \Phi(A_j^u) x_k^u \quad [4.10.9]$$

Da evidenziare è come i pesi sinaptici dello strato inferiore, per ogni unità interna  $h_j$ , siano pari alla somma dei prodotti tra i delta  $\delta_i^u$  appartenenti allo strato superiore di unità a cui l'unità nascosta è collegata e i relativi pesi sinaptici  $w_{ij}$ . Con questo si vuole implicitamente affermare che le variazioni (o errori) rappresentate dai *delta* dello strato superiore vengono propagati all'indietro e da ciò si può intuire il perché questo algoritmo prenda il nome di *Back-propagation*. La propagazione avviene per mezzo di quelle stesse connessioni sinaptiche per poi essere sommati in ciascuna unità inferiore dalla quale l'unità di output provocante l'errore riceve il segnale.

A questo punto, ipotizzando che i delta delle unità nascoste  $\delta_j^u$  siano pari a

$$\delta_j^u = \Phi(A_j^u) \sum_i w_{ij} \delta_i^u \quad [4.10.10]$$

La modifica dei pesi sinaptici  $v_{jk}$  si può riscrivere in modo simile all'equazione che definisce la variazione dei pesi sinaptici  $w_{ij}$

$$\Delta v_{jk} = \eta \sum_u \delta_j^u x_k^u \quad [4.10.11]$$

I delta quindi di un generico strato di unità sono ottenuti in base ai delta dello strato superiore; ciò fa in modo che l'algoritmo della *Back-propagation* possa essere usata per la costruzione di reti aventi un numero qualsiasi di strati a sua volta costituiti da un numero qualsiasi di nodi (Floreano e Mattiussi, 2002).

In sintesi questa regola di apprendimento si concretizza nella presentazione dei pattern di input, propagando l'attivazione in tutti gli strati fino al raggiungimento dei nodi di uscita. A questo punto i pesi sinaptici  $w_{mn}$  vengono corretti secondo l'equazione:

$$\Delta w_{mn} = \eta \sum_u \delta_m^u x_n^u \quad [4.10.12]$$



dove  $m$  indica la specifica unità dello strato superiore,  $n$  è l'indice delle unità dello strato inferiore e

$$\delta_m^u = (t_m^u - y_m^u) \dot{\Phi}(A_m^u) \quad [4.10.13]$$

nel caso in cui  $m$  corrisponda all'unità di uscita. Se invece  $m$  corrisponde ad una generica unità interna vale la seguente relazione

$$\delta_m^u = \dot{\Phi}(A_m^u) \sum_s w_{sm} \delta_s^u \quad [4.10.14]$$

Nella quale  $s$  è l'indice dell'unità appartenente allo strato superiore rispetto a  $m$ .

La regola della *Back-propagation* si presenta quindi molto potente e flessibile essendo in grado di trovare una soluzione anche se sottoposto ad input non linearmente separabili. Questo algoritmo inoltre si è molto diffuso in ambito applicativo poiché il suo utilizzo è semplice e conformabile a qualsiasi struttura di rete neurale (Floreato e Mattiussi, 2002).

Uno delle problematiche che può avere questo algoritmo è che la funzione di errore  $E_W$ , alla quale, come per la regola delta, viene applicato il metodo di discesa del gradiente, solitamente ha un andamento complesso. Tale complessità è data dal fatto che nel calcolo dell'attivazione dei nodi si eseguono in sequenza una serie di operazioni non-lineari che generano una serie di soluzioni parziali (minimi locali). I minimi locali rappresentano un problema per l'algoritmo di *Back-propagation* poiché questo ne può rimanere intrappolato senza riuscire più ad uscirne.

Una seconda problematica di questo algoritmo sta sempre nella metodo della discesa del gradiente ma nel caso in cui ci si trovi in una superficie piatta della funzione di errore. In questo caso le derivate prima dell'attivazione del nodo postsinaptico tende a 0 facendo in modo che i pesi sinaptici vengano modificati pochissimo e rendendo perciò lenta la convergenza verso una soluzione.

Un ultimo caso che può indurre l'algoritmo di *Back-propagation* in errore si può notare quando la soluzione corrisponde ad un minimo su una superficie molto ripida e stretta; in questo caso può accadere che la rete, nel percorso di minimizzazione dell'errore, balzi fuori da questo "avvallamento" proseguendo in direzione opposta rispetto alla soluzione.

## Capitolo 5: Confronto tra Z-Score e Reti Neurali

In questo ultimo capitolo si procede, mediante il software *R*, ad un'analisi dello Z-Score e delle Reti Neurali Artificiali svolgendo un semplice test statistico per poter osservare quale dei due offra una stima migliore dei default se applicato a delle S.r.l. italiane.

Il campione di imprese (e le relative variabili di nostro interesse) è stato scaricato inserendo gli opportuni filtri dalla banca dati AIDA, una piattaforma a disposizione per gli studenti dell'Università Ca' Foscari di Venezia. Il set di aziende comprende in totale 400 s.r.l. con 200 società fallite nel 2019 e con 200 società ancora in salute.

Considerando che Altman ha dato varie stime dello Z-Score, l'analisi svolta prende sia la parte di s.r.l. ad oggi ancora solventi sia la parte di società fallite, riprendendone i dati di bilancio associati alle variabili esplicative del *coefficiente Z* registrati a chiusura dell'anno 2017. Questi stessi dati verranno poi utilizzati anche per la stima di varie reti neurali con lo scopo di poter confrontare alla fine i risultati ottenuti con quelli risultanti dal modello *Z'-Score* e *Z''-Score*. I dati registrati nel 2017 e 2016 sono riportati nell'appendice **Appendice 2**, *tabella 26* e i comandi di *R* e i relativi risultati sono esplicitati di seguito.

### 5.1 Z'-Score

#### 5.1.1 Z'-Score con il campione di sole imprese solventi

Nell'**appendice 2.1.1** sono riportati i comandi svolti in *R* per il calcolo dello Z'-Score applicato alle sole 200 imprese solventi. Questo indice è stato calcolato inserendo i dati del 2017 nelle relative variabili secondo la formula:

$$Z' = 0.717X_1 + 0.847X_2 + 3.107X_3 + 0.420X_4 + 0.998X_5$$

dove:

$X_1 = \text{Capitale circolante}/\text{Totale delle attività}$

$X_2 = \text{Utili trattenuti o Utili portati a nuovo}/\text{Totale delle attività}$

$X_3 = \text{EBIT o Margine Operativo Netto}/\text{Totale delle attività}$

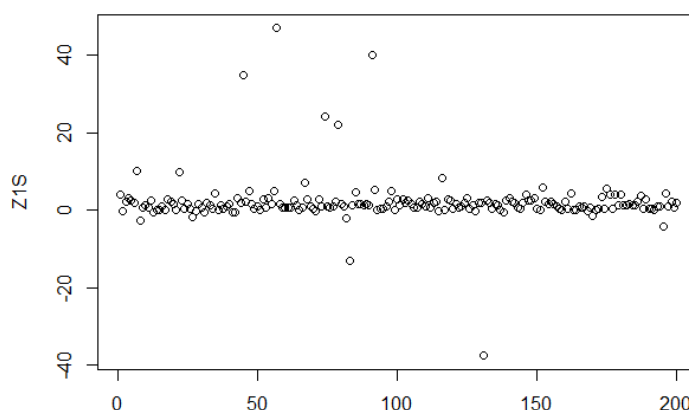
$X_4 = \text{Valore contabile del capitale netto}/\text{Valore contabile di tutte le passività}$

$X_5 = \text{Vendite}/\text{Totale delle attività}$

nel quale se:

$Z' > 2,90$	l'impresa è sana
$1,23 < Z' < 2,90$	l'impresa si trova in uno stato di incertezza
$Z' < 1,23$	l'impresa è in bancarotta

Nella Figura 13 e Tabella 13 sono riportati tutti i risultati ( $Z1S$ ) ottenuti



**Figura 13:** Grafico riportante tutti gli  $Z'$ -Score del campione di 200 imprese solventi

$Z'$ -Score	Stato impresa	Risposte stimate (%)	Risposte osservate (%)
$Z' > 2,90$	Solventi	17	
$1,23 < Z' < 2,90$	Solventi, ma situazione incerta	33,5	100
<b>Totale imprese solventi</b>		50,5	100
$Z' < 1,23$	Fallita	49,5	0

**Tabella 13:** Risultati ottenuti con il modello  $Z'$ -Score applicato ad un campione di imprese solventi

I risultati non sono soddisfacenti infatti la risposta desiderata dovrebbe darci una percentuale di imprese solventi pari al 100% ma applicando lo  $Z'$ -Score solamente il 17% delle società risulta essere inequivocabilmente "sano" (quindi con  $Z'$ -Score maggiore di 2,90) mentre si prevede che le imprese che andranno in bancarotta saranno pari al 49,5% malgrado nella realtà siano solventi. Da sottolineare è che vi è una zona grigia in cui l'indice non riesce con precisione a dire se i soggetti appartenenti a questo insieme, in futuro, saranno solventi o meno; questa zona è costituita dal 33,5% delle imprese del campione le quali, pur avendo una

situazione incerta, possono essere considerate tra le aziende sane. In parole povere la percentuale di imprese che si prevede rimangano solvienti è pari al 50,5%.

### 5.1.2. Z'-Score con il campione di sole imprese insolventi

Nell'appendice 2.1.2 sono riportati i comandi svolti in R per il calcolo dello Z'-Score applicato alle sole 200 imprese insolventi secondo la formula vista nel capitolo precedente.

Nella Figura 14 e nella tabella 14 sono riportati tutti i risultati ( $Z1F$ ) ottenuti.

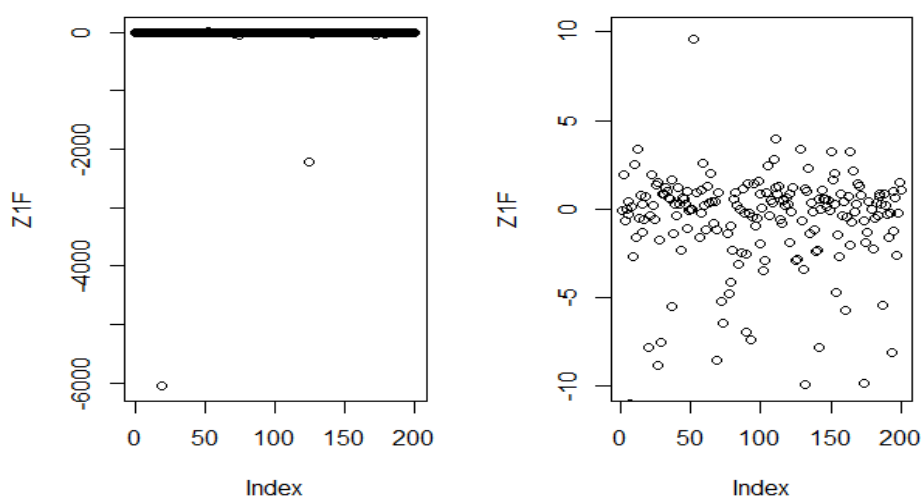


Figura 14: Grafico riportante tutti gli Z'-Score del campione di 200 imprese non solvienti

Z'-Score	Stato impresa	Risposte stimate (%)	Risposte osservate (%)
$Z' > 2,90$	Solvienti	3	
$1,23 < Z' < 2,90$	Solvienti, ma situazione incerta	12,5	0
<b>Totale imprese solvienti</b>		50,5	0
$Z' < 1,23$	Fallita	84,5	100

Tabella 14: Risultati ottenuti con il modello Z'-Score applicato ad un campione di imprese insolventi

In questo caso si può notare come le previsioni risultino essere decisamente migliori; le imprese classificate correttamente, ovvero quelle che si prevede siano in stato di insolvenza dall'indice Z'-Score, sono l' 84,5% e quindi la stima delle risposte corrette è in percentuale

molto maggiore rispetto al caso in cui il campione comprenda esclusivamente imprese solventi (50,5%).

In questo ultimo caso, le risposte sbagliate sono pari a 15,5% e sono quelle che comprendono le imprese giudicate solventi (3%) più l'insieme di soggetti appartenenti alla zona di ignoranza (12,5%). La percentuale di errore è quindi molto minore di 49,5%, ovvero la percentuale di risposte errate date dall'esperimento visto in precedenza nel quale tutto il campione di soggetti risultava essere solvente.

### 5.1.3 Risultati complessivi Z'-Score

Il grafico seguente (Tabella 15) riassume i risultati ottenuti mediante lo Z'-Score stimato in precedenza applicandolo al campione aggregato di imprese (quindi imprese fallite e sane). La diagonale principale della tabella esplicita la percentuale di risposte corrette sia nel caso in cui l'impresa solvente venga classificata come tale, sia nel caso in cui un'impresa insolvente risulti essere effettivamente in bancarotta. Da sottolineare è anche come, malgrado il campione sia diviso equamente tra imprese sane e non, il modello Z'-Score tenda a sovrastimare la percentuale di imprese che effettivamente falliscono (67%) e quindi a sottostimare quelle che effettivamente rimangono sane (33%).

		Risposte stimate		
		Aziende solventi	Aziende fallite	
Risposte osservate	Aziende solventi	25,25%	24,75%	50%
	Aziende fallite	7,75%	42,25%	50%
		33,00%	67,00%	100%

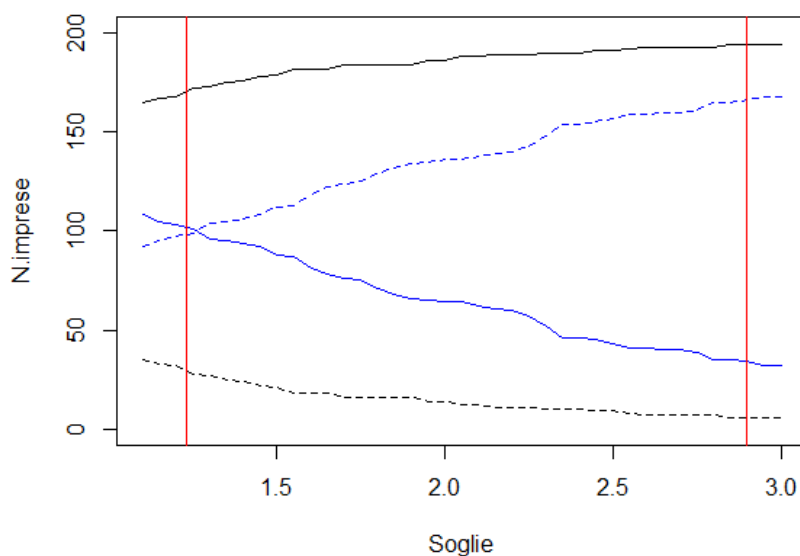
**Tabella 15:** Tabella a doppia entrata dove sulla diagonale principale sono esplicitate la percentuale di risposte corrette e sulla secondaria quelle errate.

La percentuale di risposte corrette totali sul campione di 400 imprese composto da 200 imprese solventi e da 200 non solventi è del **67,5%**.

Dopo aver stimato la precisione dello Z'-Score si è voluto andare ad analizzare l'andamento delle prestazioni del modello ipotizzando diverse soglie, soprattutto se inserite all'interno c.d. *banda di ignoranza* stimata da Altman. Ciò ci permette di vedere come varia la precisione del modello e se il valore associato alla soglia è quello più adeguato tra tutti gli altri possibili valori contenuti nella "zona grigia" definita da Altman.

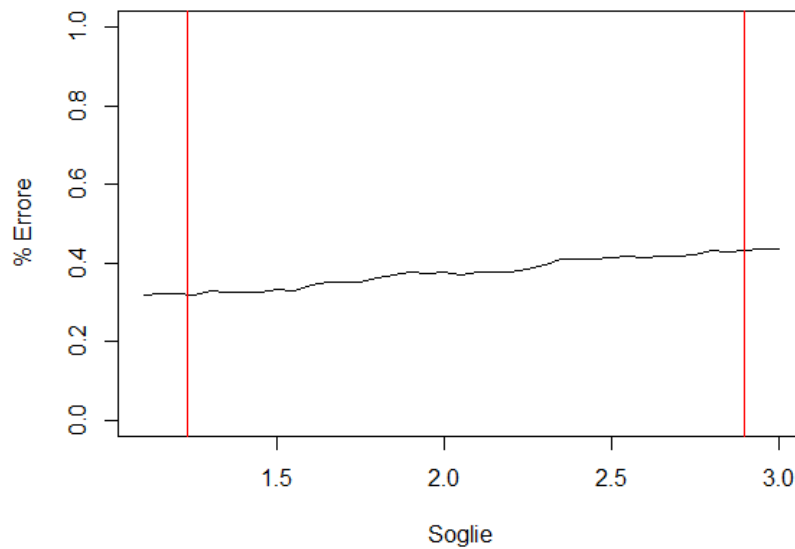
Siccome la banda di ignoranza è compresa in un intervallo di Z'-Score compreso tra 1,23 e 2,90 si è analizzato l'andamento delle prestazioni dello Z'-Score costruendo un vettore di valori,

incrementati l'uno dall'altro di 0,05 ma compresi in un intervallo tra 1,10 e 3. Il seguente grafico (*Figura 15*) identifica mediante la linea continua blu e la linea continua nera rispettivamente il totale delle imprese solventi e classificate, dallo Z'-Score, come tali e il totale delle imprese in default e classificate come effettivamente insolventi. Le linee tratteggiate identificano, di riflesso, le due tipologie di errori; il tratteggio blu identifica le imprese solventi ma classificate come insolventi (errore di I specie) mentre il tratteggio nero denota le imprese insolventi ma classificate come sane (errore di II specie). I due segmenti rossi invece denotano i valori che delimitano la zona di incertezza dello Z'-Score.



**Figura 15:** Rappresentazione della zona di ignoranza (area compresa tra le linee rosse), l'errore di I specie (linea tratteggiata blu), l'errore di II specie (linea tratteggiata nera), le imprese solventi classificate come tali (linea continua blu) e le imprese insolventi classificate correttamente (linea nera).

Dal grafico si può notare come più si alza il valore della soglia più diminuisce l'errore di seconda specie, ovvero l'errore più grave. Tale decremento però non è molto rilevante se paragonato all'incremento dell'errore di I specie all'aumentare del valore della soglia. Se infatti si andasse ad analizzare esclusivamente la percentuale delle risposte errate, al variare del valore soglia (si veda *Figura 16*), si nota che, al decrescere della soglia, la percentuale di errori totali (errori di I specie ed errori di II specie denotati dalla linea blu) diminuisce; ciò fa preferire un valore di separazione anche più basso di quello stimato da Altman per lo Z'-Score in quanto, applicando una soglia pari a 1,10, si ottiene una percentuale di errore di classificazione pari al 31,75% mentre, utilizzando il valore soglia stimato da Altman per lo Z'-Score, si ha un errore di classificazione totale pari a 32,5%.



**Figura 16:** Andamento dell'errore totale (errore di I specie ed errore di II specie) al variare della soglia.

Questa analisi non solo ci conferma la correttezza, nell'utilizzo dello  $Z'$ -Score, nell'adottare il valore soglia minimo rispetto agli altri compresi nella banda di ignoranza ma ci spingerebbe ad utilizzare una soglia ancora più piccola rispetto a quella calcolata da Altman in quanto non ottimale nella minimizzazione dell'errore<sup>55</sup>.

## 5.2 $Z''$ -Score

Siccome il risultato dato dallo  $Z'$ -Score si presenta altamente impreciso si è provato ad applicare anche lo  $Z''$ -Score che, teoricamente, dovrebbe funzionare in maniera più precisa per le imprese non manifatturiere o appartenenti ai mercati emergenti; i risultati sono riportati di seguito.

### 5.2.1. $Z''$ -Score con il campione di sole imprese solventi

Nell'**appendice 2.2.1** sono riportati i comandi svolti in *R* per il calcolo dello  $Z''$ -Score applicato alle sole 200 imprese solventi. Questo indice è stato calcolato inserendo i dati del 2017 nelle relative variabili secondo la formula:

<sup>55</sup> Gli *script* in *R* di questa ultima analisi in merito all'errore al variare della soglia sono presentati nell'*appendice 2.1.3*

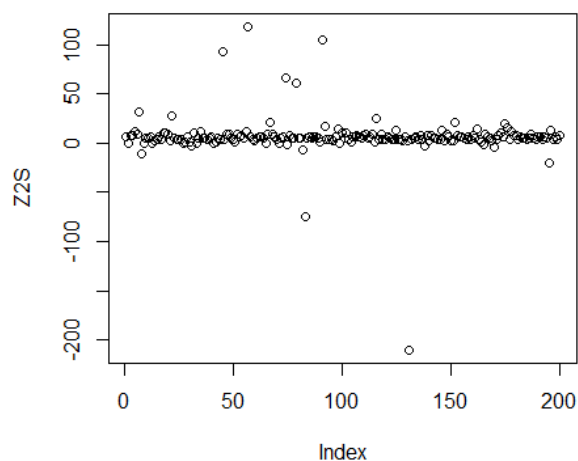
$$Z'' = 3,25 + 6,56 X_1 + 3,26 X_2 + 6,72 X_3 + 1,05 X_4$$

dove le variabili, tranne  $X_5$  che è stata eliminata, sono le stesse dello Z'-Score.

Questa relazione ci dice che se:

$Z'' > 5,83$	l'impresa è sana
$4,15 < Z'' < 5,83$	l'impresa si trova in uno stato di incertezza
$Z'' < 4,15$	l'impresa è in bancarotta

Nella Figura 17 e nella tabella 16 sono riportati tutti i risultati ( $Z2S$ ) ottenuti.



**Figura 17:** Grafico riportante tutti gli Z''-Score del campione di 200 imprese solventi

Z'-Score	Stato impresa	Risposte stimate (%)	Risposte osservate (%)
$Z'' > 5,83$	Solventi	44,5	
$4,15 < Z'' < 5,83$	Solventi, ma situazione incerta	21	100
<b>Totale imprese solventi</b>		65,5	100
$Z'' < 4,15$	Fallita	34,5	0

**Tabella 16:** Risultati ottenuti con il modello Z'-Score applicato ad un campione di imprese solventi

Con questo indice la risposta stimata si avvicina in modo molto sensibile a quella desiderata in quanto la percentuale di imprese che si stima saranno solventi, quindi le risposte corrette, sale al 44,5% mentre la percentuale di risposte sbagliate, che vede le imprese del campione



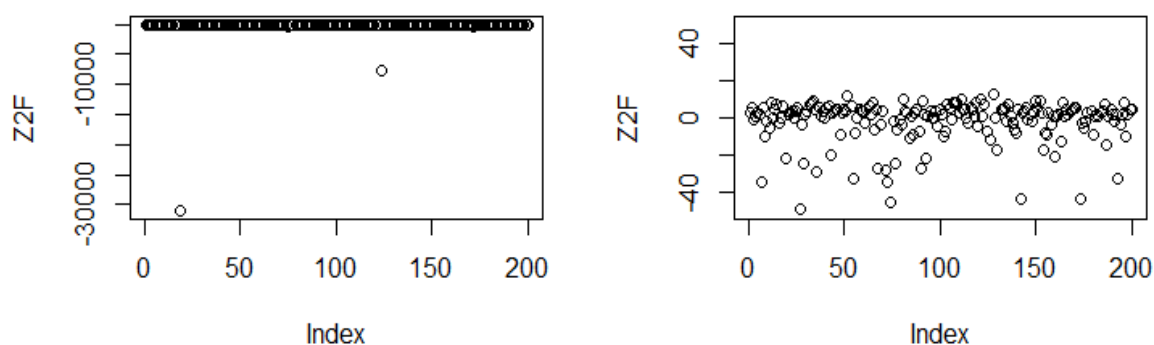
come insolventi, scende al 34,5%. La percentuale di soggetti che fa parte della zona grigia si attesta al 21% quindi, se gli si considera tra le imprese solventi, la percentuale di risposte corrette sale al 65,5%.

Se si prendessero le parti di un istituto di credito che deve erogare dei finanziamenti questo errore potrebbe risultare grave in quanto non si finanzierebbero delle controparti (pari al 34,5%) poiché, secondo lo  $Z''$ -Score, si stima essere inadempienti malgrado nella realtà questo non avviene. Questo errore di stima comporterebbe la perdita, da parte della banca di una buona parte della propria clientela sana dal punto di vista finanziario. Un errore ancora più grave si presenterebbe invece se la banca erogasse degli affidamenti ad imprese che prevede essere solventi malgrado nella realtà vadano in default. Questa tipologia di errore viene stimata con lo stesso metodo utilizzato prima ma, invece di prendere un campione di imprese solventi, si prende un campione composto esclusivamente da imprese fallite.

### 5.2.2. $Z''$ -Score con il campione di sole imprese insolventi

Nell'appendice 2.2.2 sono riportati i comandi svolti in R per il calcolo dello  $Z''$ -Score applicato alle sole 200 imprese non solventi. Questo indice è stato calcolato inserendo i dati del 2017 nelle relative variabili secondo la formula vista nel capitolo precedente.

Nella Figura 18 e nella tabella 17 sono riportati tutti i risultati ( $Z2F$ ) ottenuti.



**Figura 18:** Grafico riportante tutti gli  $Z''$ -Score del campione di 200 imprese non solventi

Si può osservare che le stime in merito alla correttezza di classificazione del campione di imprese non solventi torna ad essere lievemente distorta. Le imprese classificate correttamente, ovvero quelle che l'indice prevede che falliscano, sono pari al 72% mentre le

risposte sbagliate sono pari al 28% ovvero la percentuale che lo  $Z''$ -Score ritiene essere solventi (14%) più quelle appartenenti alla zona di ignoranza (14%).

$Z'$ -Score	Stato impresa	Risposte stimate (%)	Risposte osservate (%)
$Z'' > 5,83$	Solventi	14	
$4,15 < Z'' < 5,83$	Solventi, ma situazione incerta	14	0
	<b>Totale imprese solventi</b>	28	0
$Z'' < 4,15$	Fallita	72	100

**Tabella 17:** Risultati ottenuti con il modello  $Z''$ -Score applicato ad un campione di imprese non solventi

### 5.2.3 Risultati complessivi $Z''$ -Score

Il grafico seguente (Tabella 18) riassume i risultati ottenuti mediante lo  $Z''$ -Score stimato in precedenza applicandolo al campione aggregato di imprese (quindi imprese fallite e sane). La diagonale principale della tabella esplicita la percentuale di risposte corrette sia nel caso in cui l'impresa solvente venga classificata come tale, sia nel caso in cui un'impresa insolvente risulti essere effettivamente in bancarotta.

		Risposte stimate		
		Aziende solventi	Aziende fallite	
Risposte osservate	Aziende solventi	62,5%/2= <b>32,75%</b>	37,5%/2= <b>17,25%</b>	50%
	Aziende fallite	26,5%/2= <b>14%</b>	73,5%/2= <b>36%</b>	50%
		46,75%	53,25%	100%

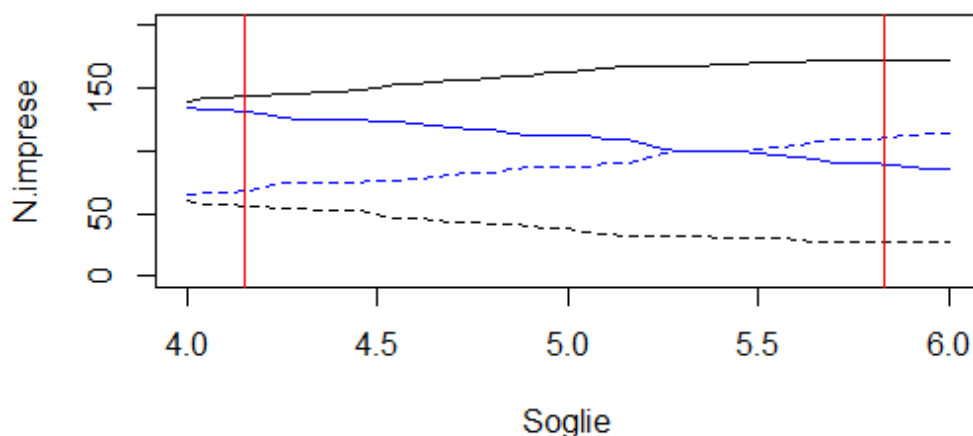
**Tabella 18:** Tabella a doppia entrata dove sulla diagonale principale sono esplicitate la percentuale di risposte corrette e sulla secondaria quelle errate date dallo  $Z''$ -Score.

La percentuale di risposte corrette dello  $Z''$ -Score sul campione di 400 imprese composto da 200 imprese solventi e da 200 non solventi è del **68,75%**.

Come per lo  $Z'$ -Score, anche con lo  $Z''$ -Score si deve dimostrare che il valore soglia scelto sia quello più adatto rispetto ai valori costituenti la banda di ignoranza.

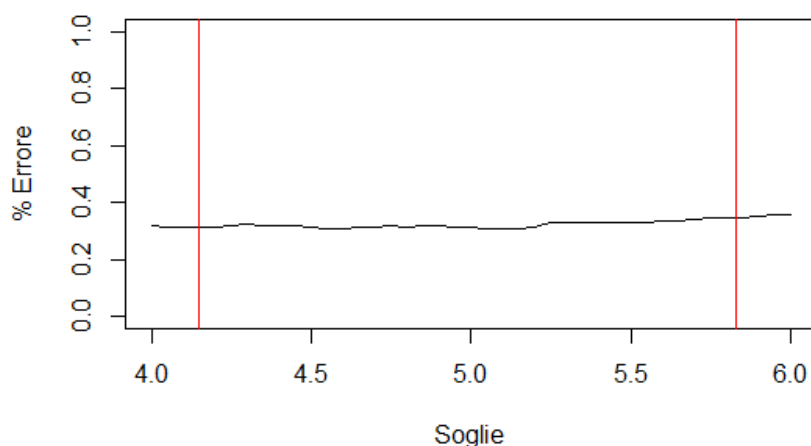
Dalla figura 19 si può vedere come, anche in questo caso, gli errori di II specie (linea tratteggiata nera) decrescano all'aumentare della soglia mentre, a tale incremento corrisponda un più sostanziale aumento degli errori di I specie (linea tratteggiata blu). Questo ci

spingerebbe a scegliere un valore soglia più alto in modo da limitare la sbagliata classificazione di imprese che andranno in default ma che il modello stima come solventi.



**Figura 19:** Rappresentazione della zona di ignoranza (area compresa tra le linee rosse), l'errore di I specie (linea tratteggiata blu), l'errore di II specie (linea tratteggiata nera), le imprese solventi classificate come tali (linea continua blu) e le imprese insolventi classificate correttamente (linea nera).

Andando però a vedere la percentuale totale di imprese classificate in modo errato (*figura 20*) si può vedere come questo valore venga leggermente minimizzato per valori della soglia corrispondenti alla parte bassa della zona di ignoranza. Si passa infatti da una percentuale di errore pari al 35,5% per un valore soglia fissato a 6 fino ad arrivare ad una percentuale di errore pari al 31,75% per un valore soglia di 4; quest'ultima è la medesima percentuale di errore che si è stimata utilizzando il valore minimo della banda di ignoranza stimata da Altman per lo Z"-Score (*soglia = 4,15*).



**Figura 20:** Andamento dell'errore totale (errore di I specie ed errore di II specie) al variare della soglia.

Questo risultato quindi può testimoniare un buon grado di correttezza nell'utilizzo, all'interno dei nostri test, del valore soglia stimato da Altman per lo  $Z'$ -Score poiché si presenta come quel punto in cui la funzione raggiunge l'errore minimo nel range di soglie analizzato<sup>56</sup>.

### 5.3 Reti neurali

Al fine di poter applicare la rete neurale al nostro campione si devono inizialmente fare alcune premesse. Il modello di rete neurale scelto è quello che presenta una funzione di attivazione sigmoide la quale, a sua volta, entra nell'algoritmo di *back-propagation* scelto per la determinazione della struttura della rete stessa.

La funzione sigmoide è una tra le funzioni di attivazioni più utilizzate poiché limita gli output ad essere compresi tra 0 e 1 ed è semplice da implementare a causa della immediata possibilità di calcolo delle derivate parziali in riferimento ad un processo di minimizzazione dell'errore. La scelta invece dell'algoritmo di *backpropagation* è legata al fatto che in letteratura esso è sempre associato all'utilizzo delle reti neurali in ambito finanziario (D'Annunzio, Falavigna, 2004). Durante il processo di calcolo il programma ha avuto non pochi problemi e molte volte non riusciva a raggiungere un minimo globale; questo fenomeno è dato dal fatto che, impostando un tasso di apprendimento eccessivamente alto, per l'algoritmo risultava difficile poter discendere la curva di errore. Il test ha quindi necessitato di varie ripetizioni durante le quali si è abbassato progressivamente il tasso di apprendimento fino a quando non si sono raggiunti dei risultati soddisfacenti. Di seguito vengono riportati i risultati associati alle reti neurali stimati avendo a disposizione gli stessi indici di bilancio individuati da Altman e utilizzati in precedenza per lo  $Z'$ -Score e  $Z''$ -Score; l'utilizzo delle stesse variabili nelle reti neurali è motivato da una più efficiente comparazione in termini di prestazioni con gli  $Z$ -Score.

In riferimento alla struttura della rete neurale sono state considerate due differenti configurazioni: una con due nodi nascosti e una con tre.

Il campione di dati utilizzato è lo stesso adottato in precedenza per le stime dei vari  $Z$ -Score solamente che, in questo caso, non si è fatta distinzione tra le 200 imprese solventi ed le 200 imprese non solventi. Si è proceduto inizialmente aggregando i due campioni distinti in un unico campione di 400 imprese; da questo poi sono state estratte casualmente 200 imprese in modo da creare due campioni di ampiezza uguale; uno è stato utilizzato come *training set* mentre il secondo come *validation set*. Entrambi i campioni (si fa riferimento al *training set* e al *validation set*), inoltre, sono stati composti per la metà da imprese fallite e per l'altra metà

---

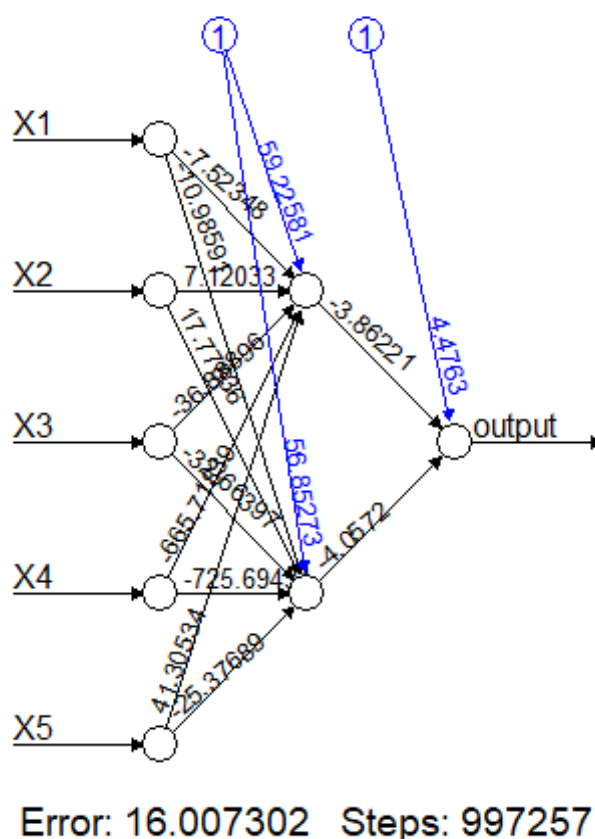
<sup>56</sup> Gli *script* in *R* di questa ultima analisi in merito all'errore al variare della soglia sono presentati nell'*appendice 2.2.3*

da imprese sane. Inizialmente viene introdotto il campione complessivo composto da 400 imprese; in un secondo momento vengono trattati i dati in modo da normalizzarli e infine, prima di stimare il modello, vengono creati questi due campioni. Si vuole sottolineare che, da qui in avanti, i dati sono predisposti in modo da creare una uguale suddivisione tra imprese fallite e non fallite all'interno del *validation* e del *training set*, qualunque siano le loro percentuali sulla totalità dei dati presi in considerazione.

### 5.3.1 Modello di rete neurale con 2 nodi nascosti (I117\_model22)

I dati per la creazione di questo modello sono riportati nell'appendice 2.3.1. La rete neurale con 2 nodi nascosti stimata verrà chiamata I117\_model22. Le operazioni svolte in R con i relativi risultati sono riportate nell'appendice 2.3.2.

Il modello di rete neurale risultante è rappresentato nella Figura 21 mentre i pesi di ciascuna sinapsi sono riportati nella Tabella 19.



**Figura 21:** rappresentazione della rete neurale con 5 variabili input, 2 unità nascoste, 1 unità di *bias* per ogni livello (blu), i pesi associati alle connessioni, la stima degli errori quadratici (*Error*) e il numero di passaggi (*Steps*) necessari per la sua costruzione

<b>Nodi in input</b>	<b>Nodo nascosto 1</b>	<b>Nodo nascosto 2</b>
<b>Bias</b>	59.225812	56.85273
<b>Input 1</b>	-7.523483	-10.98591
<b>Input 2</b>	7.120334	17.77836
<b>Input 3</b>	-36.888960	-32.66397
<b>Input 4</b>	-665.718393	-725.69434
<b>Input 5</b>	41.305344	-25.37689

	<b>Nodo di output</b>
<b>Nodo nascosto bias</b>	4.476305
<b>Nodo nascosto 1</b>	-3.862208
<b>Nodo nascosto 2</b>	-4.057195

**Tabella 19:** Tabelle riportanti i pesi di ciascun nodo. La prima si riferisce ai pesi associati alle sinapsi che collegano i nodi in input a quelli nascosti mentre la seconda descrive i pesi dai nodi nascosti al nodo di output.

Il modello stimato risulta avere una somma dei quadrati degli errori pari a 16,007 (questo indice è riportato sotto la figura della rete neurale alla dicitura "Error") mentre l'indice di correlazione tra i dati stimati e quelli reali è pari a 0,505004.

La rete neurale stima che, sul campione di validazione, ci siano 122 imprese su 200 che resteranno solventi.

### 5.3.2 Modello di rete neurale con 3 nodi nascosti (I117\_model23)

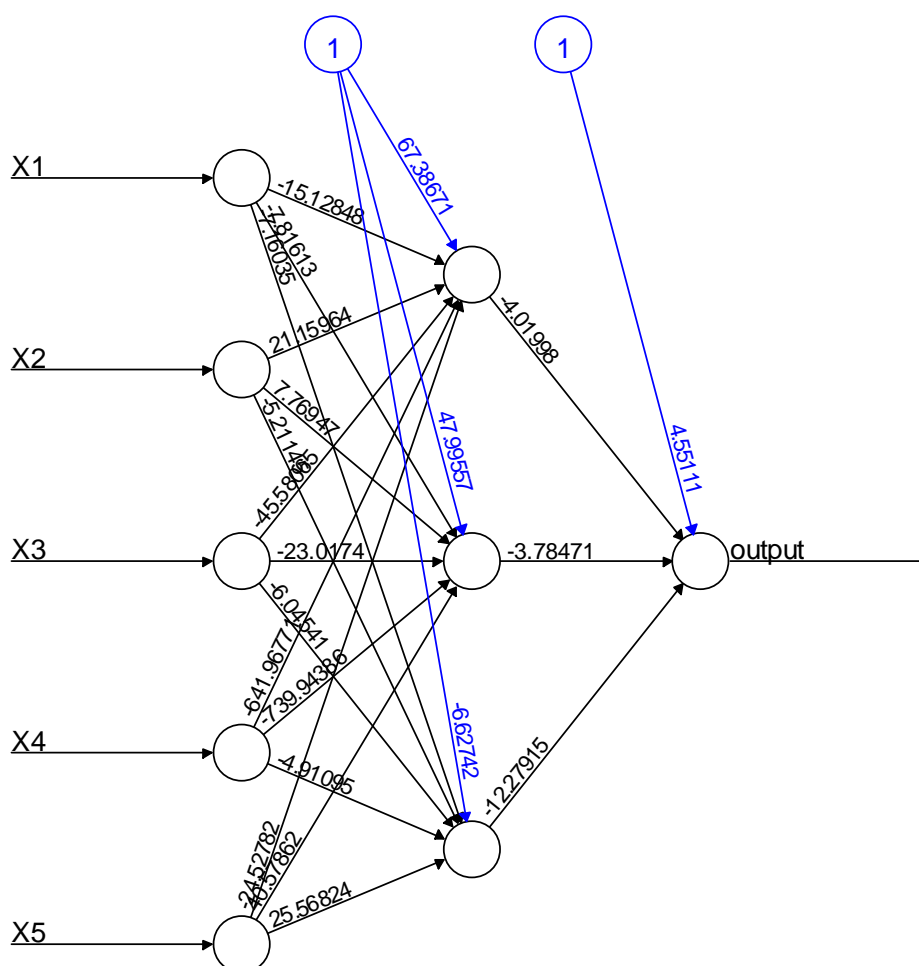
I dati per la creazione di questo modello sono riportati nell'appendice 2.3.1. La rete neurale con 3 nodi nascosti stimata verrà chiamata I117\_model23. Le operazioni svolte in R con i relativi risultati sono riportate nell'appendice 2.3.3.

Il modello di rete neurale risultante è rappresentato nella Figura 22 mentre i pesi di ciascuna sinapsi sono riportati nella Tabella 20.

<b>Nodi in input</b>	<b>Nodo nascosto 1</b>	<b>Nodo nascosto 2</b>	<b>Nodo nascosto 3</b>
<b>Bias</b>	67.38671	47.995571	-6.627423
<b>Input 1</b>	-15.12848	-7.816131	-7.160352
<b>Input 2</b>	21.15964	7.769468	-5.211451
<b>Input 3</b>	-45.58065	-23.017401	-6.045414
<b>Input 4</b>	-641.96771	-739.943864	-4.910949
<b>Input 5</b>	-24.52782	40.578620	25.568237

	Nodo di output
Nodo nascosto bias	4.551112
Nodo nascosto 1	-4.019979
Nodo nascosto 2	-3.784709
Nodo nascosto 3	-12.279149

**Tabella 20:** Tabelle riportanti i pesi di ciascun nodo. La prima si riferisce ai pesi associati alle sinapsi che collegano i nodi in input a quelli nascosti mentre la seconda descrive i pesi dai nodi nascosti al nodo di output.



Error: 15.803698 Steps: 987242

**Figura 22:** Rappresentazione della rete neurale con 5 variabili input, 3 unità nascoste, 1 unità di *bias* per ogni livello (blu), i pesi associati alle connessioni, la stima degli errori quadratici (*Error*) e il numero di passaggi (*Steps*) necessari per la sua costruzione

Il modello stimato a 3 nodi nascosti risulta avere una somma dei quadrati degli errori pari a 15,804 (questo indice è riportato sotto la figura della rete neurale alla dicitura "Error") mentre l'indice di correlazione tra i dati stimati e quelli reali è pari a 0,5111887.

La rete neurale stima che sul campione di validazione ci siano 123 su 200 imprese che resteranno solventi.

### 5.3.3 Le effettive prestazioni delle 2 reti

Entrambi i modelli stimano in maniera corretta 83 imprese come solventi mentre nella realtà sono 100. La differenza sta tra le imprese non solventi classificate correttamente; il modello con 2 nodi nascosti ne stima 61 su 100 mentre il modello con 3 nodi nascosti 60 su 100.

		Risposte stimate		
		Aziende solventi	Aziende fallite	
Risposte osservate	Aziende solventi	$83/(100*2)=$ <b>41,5%</b>	$(200-122-61)/(100*2)=$ <b>8,5%</b>	50%
	Aziende fallite	$(122-83)/(100*2)=$ <b>19,5%</b>	$61/(100*2)=$ <b>30,5%</b>	50%
		61%	39%	100%

**Tabella 21:** Tabella a doppia entrata dove sulla diagonale principale sono esplicitate la percentuale di risposte corrette e sulla secondaria quelle errate date dalla rete neurale con 5 variabili in input e 2 nodi nascosti.

La percentuale di risposte corrette della rete neurale a 2 nodi nascosti sul campione di 200 imprese composto da 100 imprese solventi e da 100 non solventi è del **72%**

		Risposte stimate		
		Aziende solventi	Aziende fallite	
Risposte osservate	Aziende solventi	$83/(100*2)=$ <b>41,5%</b>	$(200-123-60)/(100*2)=$ <b>8,5%</b>	50%
	Aziende fallite	$(123-83)/(100*2)=$ <b>20%</b>	$60/(100*2)=$ <b>30%</b>	50%
		61,5%	38,5%	100%

**Tabella 22:** Tabella a doppia entrata dove sulla diagonale principale sono esplicitate la percentuale di risposte corrette e sulla secondaria quelle errate date dalla rete neurale con 5 variabili in input e 3 nodi nascosti.

La percentuale di risposte corrette della rete neurale a 3 nodi nascosti sul campione di 200 imprese composto da 100 imprese solventi e da 100 non solventi è del **71,5%**

Da ciò si può notare come i modelli costruiti sulle reti neurali siano più precisi rispetto agli Z-Score di Altman visti in precedenza.

I risultati trovati mediante il software *R* sono riportati nell'appendice 2.3.4



## 5.4 Reti Neurali con l'aggiunta di dati ritardati di un periodo

Di seguito sono state stimate altre reti neurali con la stessa metodologia appena vista; a differenza di prima però si utilizzano anche i dati ritardati di un periodo (quindi riferiti al 2016) delle stesse imprese in aggiunta a quelli già utilizzati. Le variabili da inserire nel modello quindi non sono più 5 ma 10. In aggiunta alle reti con 2 e 3 nodi nascosti si sono stimate anche delle reti con 4 e 5 nodi nascosti per vedere se possono offrire delle stime migliori.

I dati sono riportati nell'Appendice 2.4.1.

### 5.4.1 Rete neurale con 2 nodi nascosti e dati ritardati (I1617\_model2)

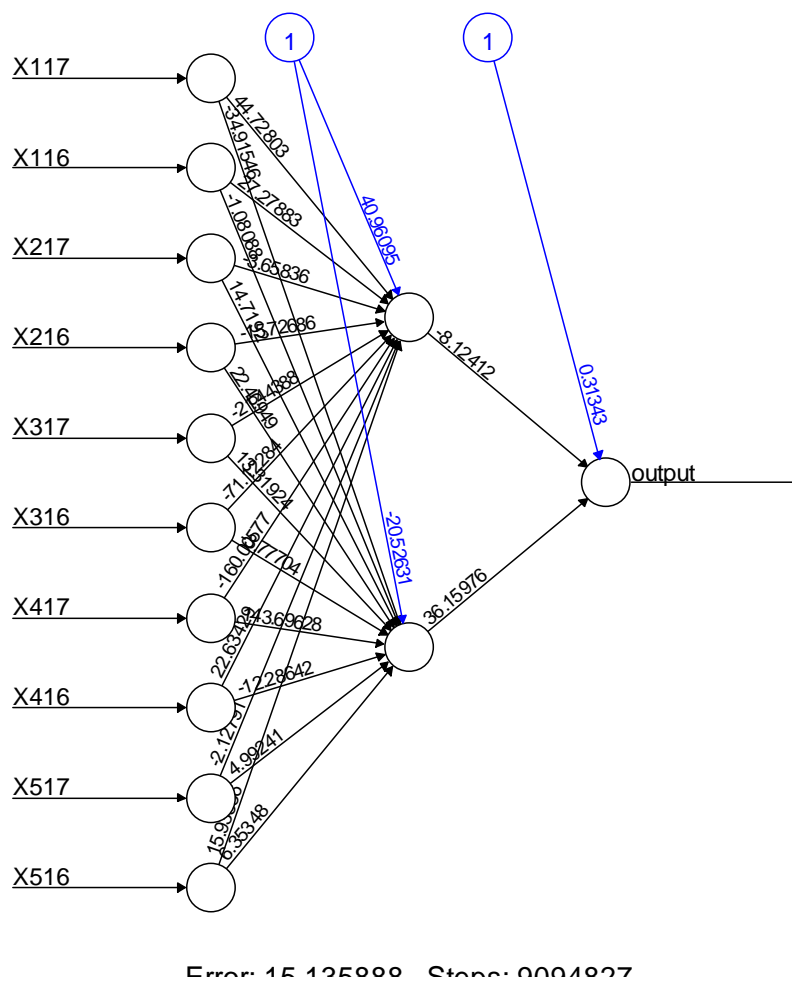
La rete neurale con 2 nodi nascosti stimata verrà chiamata I1617\_model2. Le operazioni svolte in R con i relativi risultati sono riportate nell'appendice 2.4.2.

Il modello di rete neurale risultante è rappresentato nella Figura 23 mentre i pesi di ciascuna sinapsi sono riportati nella Tabella 23.

Nodi in input	Nodo nascosto 1	Nodo nascosto 2
<b>Bias</b>	40.960953	-20.526305
<b>Input 1</b>	44.728029	-34.915464
<b>Input 2</b>	27.278825	-1.080879
<b>Input 3</b>	-3.658360	14.719201
<b>Input 4</b>	-15.726856	22.469486
<b>Input 5</b>	-21.443884	13.319239
<b>Input 6</b>	-71.122837	-3.777040
<b>Input 7</b>	-160.005770	143.696280
<b>Input 8</b>	22.634289	-12.286415
<b>Input 9</b>	-2.127908	4.992410
<b>Input 10</b>	15.935359	6.353479

	Nodo di output
<b>Nodo nascosto bias</b>	0.3134277
<b>Nodo nascosto 1</b>	-8.1241217
<b>Nodo nascosto 2</b>	36.1597578

**Tabella 23:** Tabelle riportanti i pesi di ciascun nodo. La prima si riferisce ai pesi associati alle sinapsi che collegano i nodi in input a quelli nascosti mentre la seconda descrive i pesi dai nodi nascosti al nodo di output.



**Figura 23:** Rappresentazione della rete neurale con 10 variabili input, 2 unità nascoste, 1 unità di *bias* per ogni livello (blu), i pesi associati alle connessioni, la stima degli errori quadratici (*Error*) e il numero di passaggi (*Steps*) necessari per la sua costruzione

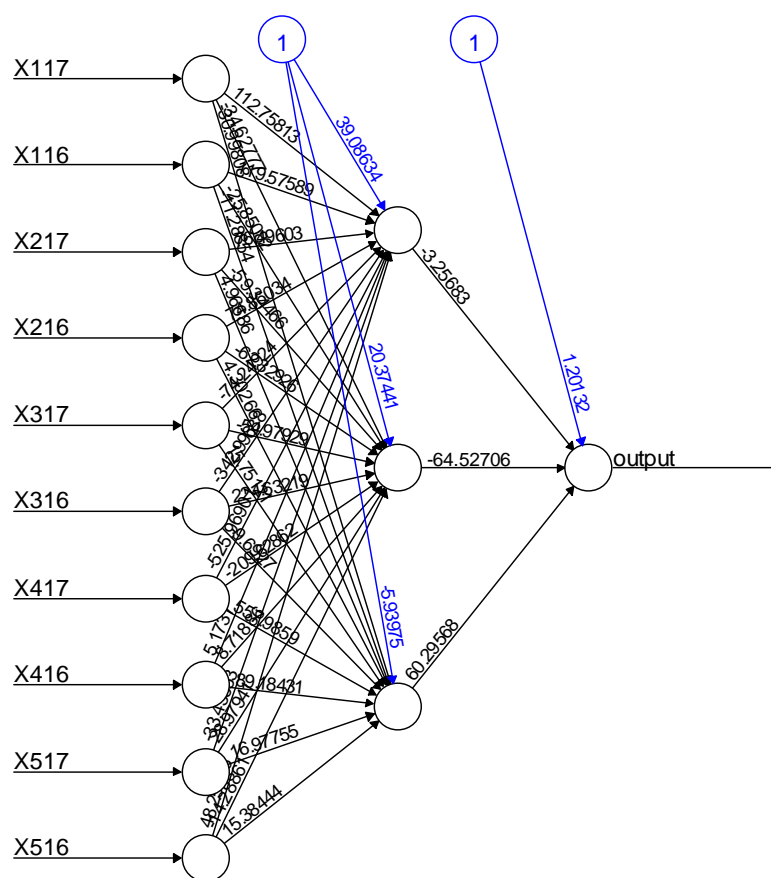
Il modello stimato con 2 nodi nascosti e con l'aggiunta di dati ritardati di un periodo risulta avere una somma dei quadrati degli errori pari a 15,136 (questo indice è riportato sotto la figura della rete neurale alla dicitura "Error") mentre l'indice di correlazione tra i dati stimati e quelli reali è pari a 0,4757701.

La rete neurale stima che sul campione di validazione ci siano 85 su 200 imprese che resteranno solventi.

#### 5.4.2 Rete neurale con 3 nodi nascosti e dati ritardati (I1617\_model3)

La rete neurale con 3 nodi nascosti stimata verrà chiamata I1617\_model3. Le operazioni svolte in R con i relativi risultati sono riportate nell'appendice 2.4.3.

Il modello di rete neurale risultante è rappresentato nella Figura 24 mentre i pesi di ciascuna sinapsi sono riportati nella Tabella 24.



Error: 12.821624 Steps: 2078024

**Figura 24:** Rappresentazione della rete neurale con 10 variabili input, 3 unità nascoste, 1 unità di *bias* per ogni livello (blu), i pesi associati alle connessioni, la stima degli errori quadratici (*Error*) e il numero di passaggi (*Steps*) necessari per la sua costruzione.

Nodi in input	Nodo nascosto 1	Nodo nascosto 2	Nodo nascosto 3
<b>Bias</b>	39.086337	20.374410	-5.939750
<b>Input 1</b>	112.758130	-34.627767	-30.998080
<b>Input 2</b>	119.575886	-25.850823	-11.288536
<b>Input 3</b>	76.496031	-59.344658	4.966863
<b>Input 4</b>	74.860340	-62.329258	4.302658
<b>Input 5</b>	-74.249238	-54.979292	15.751210
<b>Input 6</b>	-342.996875	224.632186	-9.692697
<b>Input 7</b>	-525.969014	-209.628622	557.985897
<b>Input 8</b>	5.173703	8.718858	89.184309
<b>Input 9</b>	-33.436529	28.979406	16.977545
<b>Input 10</b>	48.250961	-14.288605	15.384439

	<b>Nodo di output</b>
<b>Nodo nascosto bias</b>	1.201324
<b>Nodo nascosto 1</b>	-3.256830
<b>Nodo nascosto 2</b>	-64.527058
<b>Nodo nascosto 3</b>	60.295678

**Tabella 24:** Tabelle riportanti i pesi di ciascun nodo. La prima si riferisce ai pesi associati alle sinapsi che collegano i nodi in input a quelli nascosti mentre la seconda descrive i pesi dai nodi nascosti al nodo di output.

Il modello stimato con 3 nodi nascosti e con l'aggiunta di dati ritardati di un periodo risulta avere una somma dei quadrati degli errori pari a 13,834634 (questo indice è riportato sotto la figura della rete neurale alla dicitura "Error") mentre l'indice di correlazione tra i dati stimati e quelli reali è pari a 0,4471527.

La rete neurale stima che sul campione di validazione ci siano 83 imprese su 200 che resteranno solventi.

#### 5.4.3 Rete neurale con 4 nodi nascosti e dati ritardati (I1617\_model22)

La rete neurale con 4 nodi nascosti stimata verrà chiamata I1617\_model22. Le operazioni svolte in R con i relativi risultati sono riportate nell'appendice 2.4.4.

Il modello di rete neurale risultante è rappresentato nella Figura 25 mentre i pesi di ciascuna sinapsi sono riportati nella Tabella 25.

Il modello stimato con 4 nodi nascosti e con l'aggiunta di dati ritardati di un periodo risulta avere una somma dei quadrati degli errori pari a 13,570817 (questo indice è riportato sotto la figura della rete neurale alla dicitura "Error") mentre l'indice di correlazione tra i dati stimati e quelli reali è pari a 0,4418523.

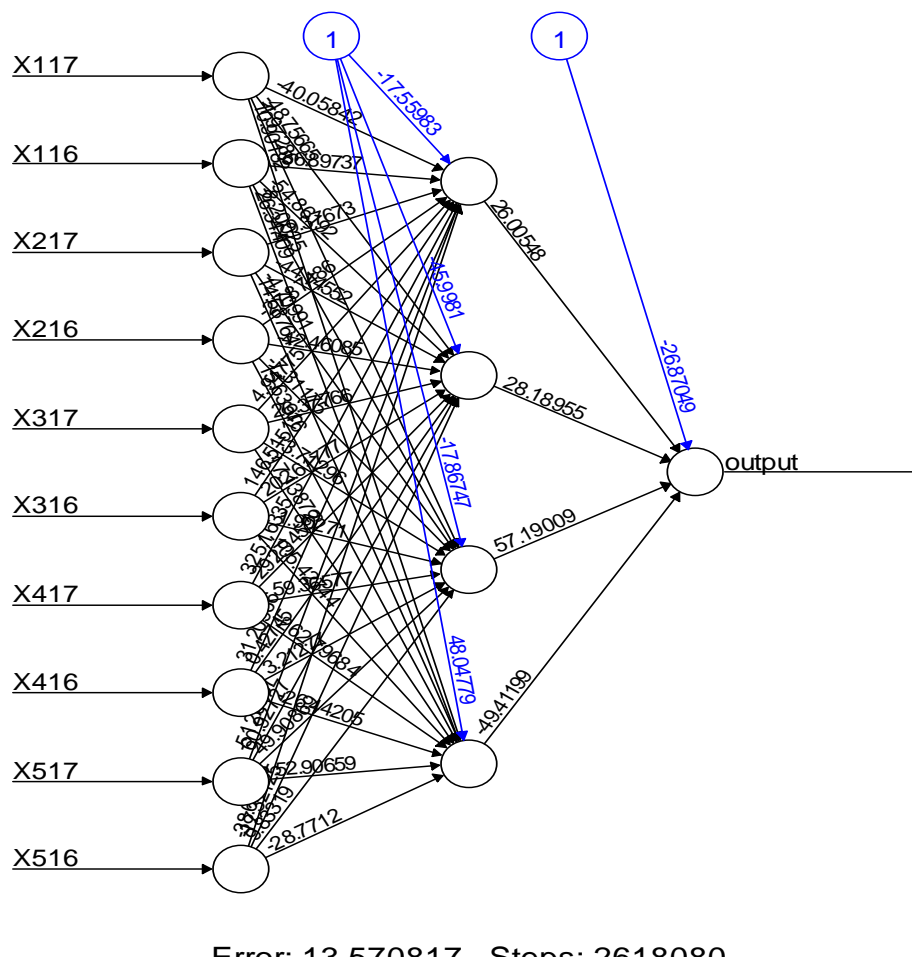
<b>Nodi di input</b>	<b>Nodo nascosto 1</b>	<b>Nodo nascosto 2</b>	<b>Nodo nascosto 3</b>	<b>Nodo nascosto 4</b>
<b>Bias</b>	-17.559831	-45.9981011	-17.867466	48.04779
<b>Input 1</b>	-40.058417	-48.7566485	-1.972861	10.90182
<b>Input 2</b>	-36.897368	-54.8619214	-8.200246	26.34409
<b>Input 3</b>	-29.376733	-44.1455153	-4.709905	14.68764
<b>Input 4</b>	-27.874862	-42.4608458	-7.311745	15.63916
<b>Input 5</b>	4.947747	26.3776610	13.712962	-17.13872
<b>Input 6</b>	146.515154	202.6177737	1.902714	-105.42744
<b>Input 7</b>	325.163347	292.9459890	59.365769	-262.79684
<b>Input 8</b>	31.209665	0.4274478	3.212771	-26.44205
<b>Input 9</b>	-51.263374	60.9272664	49.908687	52.90659

<b>Input 10</b>	-38.621734	-17.3212532	9.833192	-28.77120
-----------------	------------	-------------	----------	-----------

	<b>Nodo di output</b>
<b>Nodo nascosto bias</b>	-26.87049
<b>Nodo nascosto 1</b>	26.00548
<b>Nodo nascosto 2</b>	28.18955
<b>Nodo nascosto 3</b>	57.19009
<b>Nodo nascosto 4</b>	-49.41199

**Tabella 25:** Tabelle riportanti i pesi di ciascun nodo. La prima si riferisce ai pesi associati alle sinapsi che collegano i nodi in input a quelli nascosti mentre la seconda descrive i pesi dai nodi nascosti al nodo di output.

La rete neurale stima che sul campione di validazione ci siano 82 su 200 imprese che resteranno solventi.



**Figura 25:** Rappresentazione della rete neurale con 10 variabili input, 4 unità nascoste, 1 unità di *bias* per ogni livello (blu), i pesi associati alle connessioni, la stima degli errori quadratici (*Error*) e il numero di passaggi (*Steps*) necessari per la sua costruzione.

#### 5.4.4 Rete neurale con 5 nodi nascosti e dati ritardati (I1617\_model5)

La rete neurale con 5 nodi nascosti stimata verrà chiamata I1617\_model5. Le operazioni svolte in R con i relativi risultati sono riportate nell'appendice 2.4.5.

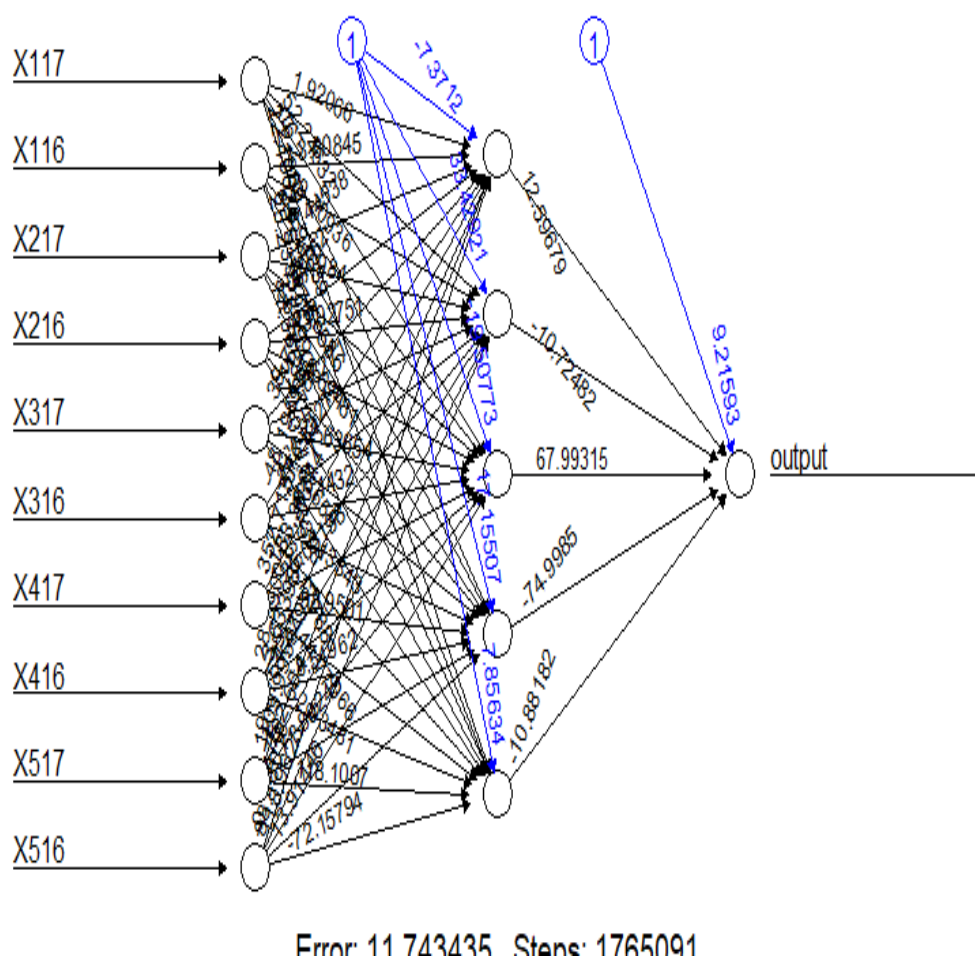
Il modello di rete neurale risultante è rappresentato nella Figura 26 mentre i pesi di ciascuna sinapsi sono riportati nella Tabella 26.

Nodi di input	Nodo nascosto 1	Nodo nascosto 2	Nodo nascosto 3	Nodo nascosto 4	Nodo nascosto 5
<b>Bias</b>	-7.3711960	33.429207	-19.50773	17.155070	7.856339
<b>Input 1</b>	1.9206560	52.246317	-16.20000	1.219090	22.406237
<b>Input 2</b>	-3.6084498	59.709365	-16.94839	6.751864	25.239588
<b>Input 3</b>	-0.4453755	37.983996	-17.56586	3.699917	17.368305
<b>Input 4</b>	-2.4475247	43.927509	-16.62481	5.974321	17.102114
<b>Input 5</b>	39.7533606	-9.882713	-12.63654	-3.686585	-22.433980
<b>Input 6</b>	-48.1961279	-216.812792	45.14320	-52.133431	-69.177877
<b>Input 7</b>	352.7156998	-183.393441	25.03138	-98.950096	-144.770657
<b>Input 8</b>	28.3900127	3.633983	4.27921	-8.549620	-2.934808
<b>Input 9</b>	19.8757546	-109.995907	102.65400	71.802495	118.100698
<b>Input 10</b>	-0.1847462	88.130158	118.96230	13.914462	-72.157937

	Nodo di output
<b>Nodo nascosto bias</b>	9.215927
<b>Nodo nascosto 1</b>	12.596788
<b>Nodo nascosto 2</b>	-10.724823
<b>Nodo nascosto 3</b>	67.993153
<b>Nodo nascosto 4</b>	-74.998502
<b>Nodo nascosto 5</b>	-10.881821

**Tabella 26:** Tabelle riportanti i pesi di ciascun nodo. La prima si riferisce ai pesi associati alle sinapsi che collegano i nodi in input a quelli nascosti mentre la seconda descrive i pesi dai nodi nascosti al nodo di output.

Il modello stimato con 5 nodi nascosti e con l'aggiunta di dati ritardati di un periodo risulta avere una somma dei quadrati degli errori pari a 11,743435 (questo indice è riportato sotto la figura della rete neurale alla dicitura "Error") mentre l'indice di correlazione tra i dati stimati e quelli reali è pari a 0,4369027.



**Figura 26:** Rappresentazione della rete neurale con 10 variabili input, 5 unità nascoste, 1 unità di *bias* per ogni livello (blu), i pesi associati alle connessioni, la stima degli errori quadratici (*Error*) e il numero di passaggi (*Steps*) necessari per la sua costruzione.

La rete neurale stima che sul campione di validazione ci siano 76 su 200 imprese che resteranno solventi.

#### 5.4.5 Le effettive prestazioni delle reti con dati ritardati

Il modello I1617\_model2 stima correttamente 63 imprese effettivamente solventi su 100 e 78 imprese effettivamente non solventi su 100. La rete neurale I1617\_model3 invece si presenta meno precisa infatti essa prevede correttamente 60 imprese solventi su 100 e 77 imprese non solventi su 100. Il modello I1617\_model22 presenta risultati molto simili al modello che possiede 3 nodi nascosti infatti è solamente meno preciso nel prevedere le imprese che si rileveranno solventi stimandone 59 su 100. Il modello I1617\_model5, al contrario, è il peggiore tra i 4 modelli nella stima delle imprese che si pensa saranno solventi poichè rileva che solamente 57 imprese su 100 si dimostreranno effettivamente adempienti; tale modello però

è il migliore nella stima delle imprese effettivamente insolventi stimando correttamente 81 imprese su 100.

I comandi di *R* e gli output del modello sono riportati nell'appendice 2.4.6 invece nella tabella 27, tabella 28, tabella 29 e tabella 30 sono riassunti i risultati rispettivamente del modello I1617\_model2, I1617\_model3, I1617\_model22 e I1617\_model5.

		Risposte stimate		
		Aziende solventi	Aziende fallite	
Risposte osservate	Aziende solventi	$63/(100*2)=$ <b>31,5%</b>	$(200-85-78)/(100*2)=$ <b>18,5%</b>	50%
	Aziende fallite	$(85-63)/(100*2)=$ <b>11%</b>	$78/(100*2)=$ <b>39%</b>	50%
		42,5%	57,5%	100%

**Tabella 27:** Tabella a doppia entrata dove sulla diagonale principale sono esplicitate la percentuale di risposte corrette e sulla secondaria quelle errate date dalla rete neurale con 10 variabili in input e 2 nodi nascosti.

La percentuale di risposte corrette della rete neurale con 10 variabili in input e 2 nodi nascosti su un campione di validazione di 200 imprese, composto da 100 società solventi e da 100 non solventi, è del **70,5%**

		Risposte stimate		
		Aziende solventi	Aziende fallite	
Risposte osservate	Aziende solventi	$60/(100*2)=$ <b>30%</b>	$(200-83-77)/(100*2)=$ <b>20%</b>	50%
	Aziende fallite	$(83-60)/(100*2)=$ <b>11,5%</b>	$77/(100*2)=$ <b>38,5%</b>	50%
		41,5%	58,5%	100%

**Tabella 28:** Tabella a doppia entrata dove sulla diagonale principale sono esplicitate la percentuale di risposte corrette e sulla secondaria quelle errate date dalla rete neurale con 10 variabili in input e 3 nodi nascosti.

La percentuale di risposte corrette della rete neurale con 10 variabili in input e 3 nodi nascosti su un campione di validazione di 200 imprese, composto da 100 imprese solventi e da 100 non solventi, è del **68,5%**



		Risposte stimate		
		Aziende solventi	Aziende fallite	
Risposte osservate	Aziende solventi	$59/(100*2)=$ <b>29,5%</b>	$(200-82-77)/(100*2)=$ <b>20,5%</b>	50%
	Aziende fallite	$(82-59)/(100*2)=$ <b>11,5%</b>	$77/(100*2)=$ <b>38,5%</b>	50%
		41,5%	58,5%	100%

**Tabella 29:** Tabella a doppia entrata dove sulla diagonale principale sono esplicitate la percentuale di risposte corrette e sulla secondaria quelle errate date dalla rete neurale con 10 variabili in input e 4 nodi nascosti.

La percentuale di risposte corrette della rete neurale a 4 nodi nascosti su un campione di validazione di 200 imprese composto da 100 imprese solventi e da 100 non solventi è del **68%**

		Risposte stimate		
		Aziende solventi	Aziende fallite	
Risposte osservate	Aziende solventi	$57/(100*2)=$ <b>28,5%</b>	$(200-76-81)/(100*2)=$ <b>21,5%</b>	50%
	Aziende fallite	$(76-57)/(100*2)=$ <b>9,5%</b>	$81/(100*2)=$ <b>40,5%</b>	50%
		38%	62%	100%

**Tabella 30:** Tabella a doppia entrata dove sulla diagonale principale sono esplicitate la percentuale di risposte corrette e sulla secondaria quelle errate date dalla rete neurale con 10 variabili in input e 5 nodi nascosti.

La percentuale di risposte corrette della rete neurale con 10 variabili in input e 5 nodi nascosti su un campione di validazione di 200 imprese, composto da 100 imprese solventi e da 100 non solventi, è del **69%**.

Riassumendo i risultati di questi test svolti anche con i dati ritardati di un periodo, si può affermare che l'aggiunta di queste informazioni non riesca a migliorare la rete neurale in termini previsivi in nessuno dei 4 nuovi modelli stimati ma anzi, a volte, ne distorce le previsioni.

## 5.5 Reti Neurali con diverse percentuali nella composizione del *training set* e *validation set*

In questo ultimo paragrafo dedicato all'utilizzo delle reti neurali si useranno gli stessi dati usati per la stima dello Z-Score avendo precedentemente verificato che l'aggiunta di informazioni ritardate legate alle variabili prese in considerazioni non contribuisce al miglioramento del modello. Fino ad ora, nella costruzione delle reti, il campione di 400 imprese è stato diviso a metà in modo da creare un *training set* e un *validation set* uguali in termini di numerosità campionaria i quali, a loro volta, sono divisi in ugual proporzione tra imprese fallite e sane.

Ora invece si procederà con la stima di un modello basato su dati presi dalle medesime imprese finora trattate ma si adotterà una diversa numerosità tra *training set* e *validation set*. Si utilizzerà la rete neurale con 2 nodi poiché rappresenta il tipo di struttura neurale artificiale che, in questo elaborato, ha fornito la migliore percentuale di correttezza in termini previsionali sullo stato di salute delle imprese facenti parte il *validation set*.

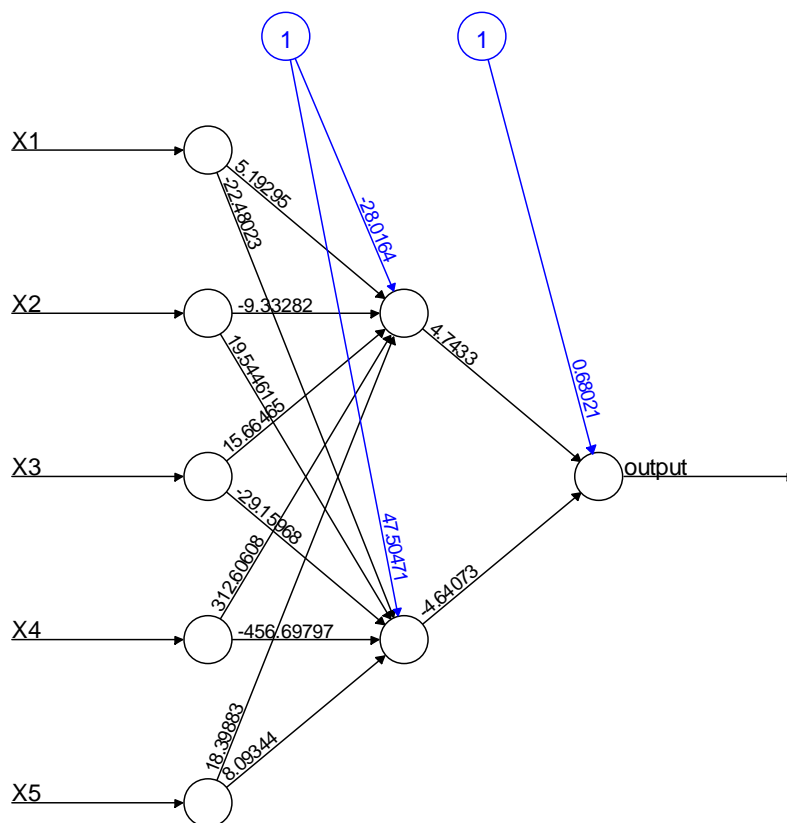
### 5.5.1 Trainig set 75%, Validation set 25% (I7525\_model)

Si farà la stima del modello utilizzando un *training set* pari al 75% del totale delle imprese il quale verrà poi testato con un *validation set* pari al 25% delle imprese totali. In entrambi questi campioni le imprese sono divise equamente tra imprese solventi ed insolventi. Questo modello verrà chiamato I7525\_model. I dati per la creazione di questa rete neurale e i relativi script in R sono riportati nell'appendice 5.5.1 mentre la struttura del modello e i pesi associati ad ogni sinapsi sono riportati rispettivamente nella tabella 31 e nella figura 27.

Nodi di input	Nodo nascosto 1	Nodo nascosto 2
<b>Bias</b>	-28.016399	47.50471
<b>Input 1</b>	5.192952	-22.48023
<b>Input 2</b>	-9.332819	19.54461
<b>Input 3</b>	15.664648	-29.15968
<b>Input 4</b>	312.606077	-456.69797
<b>Input 5</b>	18.398831	8.09344

	Nodo di output
Nodo nascosto bias	0.6802075
Nodo nascosto 1	4.7432955
Nodo nascosto 2	-4.6407260

**Tabella 31:** Tabelle riportanti i pesi di ciascun nodo. La prima si riferisce ai pesi associati alle sinapsi che collegano i nodi in input a quelli nascosti mentre la seconda descrive i pesi dai nodi nascosti al nodo di output.



Error: 25.400279 Steps: 2739275

**Figura 27:** Rappresentazione della rete neurale con 5 variabili input, 2 unità nascoste, 1 unità di *bias* per ogni livello (blu), i pesi associati alle connessioni, la stima degli errori quadratici (*Error*) e il numero di passaggi (*Steps*) necessari per la sua costruzione.

Il modello stimato con 2 nodi nascosti, creato con un *training set* pari al 75% del numero delle imprese considerate nel campione totale originale, risulta avere una somma dei quadrati degli errori pari a 25,400279 (questo indice è riportato sotto la figura della rete neurale alla dicitura "Error") mentre l'indice di correlazione tra i dati stimati e quelli reali è pari a 0,5541945.

La rete neurale stima che sul campione di validazione ci siano 44 su 100 imprese che resteranno solventi.

Il modello, stima correttamente 36 imprese effettivamente solventi su 50 e 42 imprese effettivamente non solventi su 50.

		Risposte stimate		
		Aziende solventi	Aziende fallite	
Risposte osservate	Aziende solventi	$36/(50*2)*100=$ <b>36%</b>	$(100-44-42)/(50*2)*100=$ <b>14%</b>	50%
	Aziende fallite	$(44-36)/(50*2)*100=$ <b>8%</b>	$42/(50*2)*100=$ <b>42%</b>	50%
		44%	56%	100%

**Tabella 32:** Tabella a doppia entrata dove sulla diagonale principale sono esplicitate la percentuale di risposte corrette e sulla secondaria quelle errate date dalla rete neurale costruita sul 75% del campione totale con 5 variabili in input e 2 nodi nascosti.

La percentuale di risposte corrette della rete neurale a 2 nodi nascosti su un campione di validazione di 100 imprese composto da 50 imprese solventi e da 50 non solventi è del **78%**.

## 5.6 Risultati in sintesi

Nella *tabella 33* vengono presentati tutte statistiche più significative concernenti tutti i modelli stimati in questo lavoro.

Le percentuali di imprese che sono state classificate come fallite malgrado le osservazioni ne testimoniassero la buona salute sono state riportate sotto la colonna “% errore di I specie” mentre, al contrario, le percentuali di imprese che sono state stimate essere solventi malgrado si rivelino essere inadempienti sono state riportate nella colonna “% errore di II specie”. Si vuole evidenziare che al fine della concessione di un affidamento, per una banca, l’errore di II specie risulta essere di fondamentale importanza poiché minimizzare questo errore significa aumentare la possibilità di poter rientrare dell’intero credito concesso ai propri clienti. In misura minore anche l’errore di I specie è dannoso poiché implica la perdita, da parte della banca, di una parte della clientela sana la quale non ha potuto usufruire del credito della banca stessa essendo stata, erratamente, classificata come insolvente.

Dai risultati si deduce che complessivamente, il modello che divide in maniera più corretta le imprese fallite da quelle sane è *I7525\_model* ovvero la rete neurale costruita sul *training set* con il 75% delle imprese totali del campione. Questo modello è inoltre molto efficiente nella minimizzazione dell’errore di II specie (ovvero quello più grave) essendo superato, per poco, solamente dallo Z'-Score di Altman.

Metodo	% risposte corrette	% errore di I specie	% errore di II specie	% risposte sbagliate
<b>Z'-Score</b>	67,5	24,75	7,75	32,5
<b>Z"-Score</b>	68,75	17,25	14	31,25
<b>I117_model22</b>	72	8,5	19,5	28
<b>I117_model23</b>	71,5	8,5	20	28,5
<b>I1617_model2</b>	70,5	18,5	11	29,5
<b>I1617_model3</b>	68,5	20	11,5	31,5
<b>I1617_model22</b>	68	20,5	11,5	32
<b>I1617_model5</b>	69	21,5	9,5	31
<b>I7525_model</b>	78	14	8	22

**Tabella 33:** Sintesi dei risultati ottenuti per ogni modello

Per concludere questo lavoro si può dire che, sulla base del campione di imprese selezionate, le reti neurali si pongono come un modello molto efficace nella previsione del rischio di insolvenza. La loro efficacia, data dalla loro capacità di apprendimento, risulta essere molto più forte rispetto al modello Z-Score di Altman il quale offre comunque delle stime discrete.

## Appendice 1

### Esempio 1

Si ipotizzi di poter osservare solamente una variabile  $X$  (quindi  $p = 1$ ) la quale può assumere solamente due valori  $\{0,1\}$  e il soggetto che deve essere sottoposto a classificazione può essere collocato esclusivamente in due gruppi:  $D_0$  e  $D_1$ . In  $D_1$  si supponga di avere:

$$P(X = 0) = P(X = 1) = \frac{1}{2}$$

mentre in  $D_0$  la variabile presenta le seguenti probabilità

$$P(X = 0) = \frac{1}{4};$$

$$P(X = 1) = 1 - \frac{1}{4} = \frac{3}{4}$$

Riprendendo la regola [2.1.14], all'insieme  $D_1$  viene associato l'individuo che presenta  $x = 0$  mentre all'insieme  $D_0$  viene associato l'individuo che presenta  $x = 1$  (Iannucci, 2004). Questo è spiegato dal fatto che un individuo generico, se esso presenta come variabile  $x$  valore 0 esso ha molta più probabilità di essere associato all'insieme  $D_1$  che a  $D_0$ . Al contrario, se si prendesse un secondo soggetto che presenta la variabile  $x$  pari a 1 esso, sarà associato con buona probabilità, all'insieme  $D_0$ .

### Esempio 2

Si supponga di voler suddividere un campione di soggetti secondo una mera caratteristica geografico-spaziale. Prendendo in considerazione il caso italiano, si potrebbero fare una distinzione tra nord, centro e sud; si possono quindi individuare tre gruppi ovvero  $D_N, D_C, D_S$ . Si ipotizzi inoltre di voler fare questa classificazione basandosi su un singolo fattore, per esempio la concentrazione di colesterolo nel sangue supponendo che nelle 3 zone geografiche vi siano naturalmente 3 diverse abitudini nutrizionali e che quindi, in assenza di malattie epatiche, vadano ad influenzare tale concentrazione. Sulla base di precedenti studi sono state calcolate le seguenti probabilità discrete associate alla concentrazione di colesterolo nel sangue nelle 3 aree geografiche precedentemente considerate in cui si è suddivisa l'Italia:

i. In  $D_N$ :

$$P(X < 1) = \frac{2}{10} ; P(1 < X < 2) = \frac{3}{10} ; P(X > 2) = \frac{5}{10}$$

ii. In  $D_C$ :

$$P(X > 1) = \frac{1}{10}; P(1 < X < 2) = \frac{7}{10}; P(X > 2) = \frac{2}{10}$$

iii. In  $D_S$ :

$$P(X < 1) = \frac{6}{10}; P(1 < X < 2) = \frac{3}{10}; P(X > 2) = \frac{1}{10}$$

In queste tre funzioni  $X = \text{tasso di colesterolo}$  e si ha a disposizione un solo fattore e 3 gruppi per svolgere l'analisi discriminante ( $p = 1$  e  $g = 3$ ). Secondo la regola [2.1.14] un individuo con  $X > 2$  verrà inserito nel gruppo  $D_N$ , se presenta  $1 < X < 2$  allora il soggetto verrà collocato nel gruppo  $D_C$  mentre se presenta  $X < 1$  il soggetto preso in esame verrà collocato nel gruppo  $D_S$  (Iannucci, 2004).

Il prossimo capitolo analizzerà la generalizzazione di tale tecnica di classificazione.

## Appendice 2

Di seguito si riporta la tabella con il campione di tutte e 400 le imprese selezionate; questo campione è composto da 200 imprese solventi e da 200 fallite.

Ragione sociale	X1 % 2017	X1 % 2016	X2 % 2017	X2 % 2016	X3 % 2017	X3 % 2016	X4 % 2017	X4 % 2016	X5 % 2017	X5 % 2016	(1=solvente, 0=non solvente)
SPACEREP ROMA S.R.L.	29	20	0	0	14	9	41	27	309	272	1
HOTEL CIMA SRL	-68	-74	10	9	2	3	29	28	0	0	1
GULLI INSURANCE & CONSULTING S.R.L.	29	26	0	0	23	26	48	38	114	97	1
L.S. ENGINEERING S.R.L.	50	59	0	0	6	5	58	108	236	210	1
FUNNEL SOCIETA' A RESPONSABILITA' LIMITATA SEMPLIFICATA	69	66	27	36	27	15	83	81	40	45	1
MEBO S.R.L.	58	34	0	0	17	6	52	39	62	52	1
BETA SRL	79	80	0	0	1	7	2.202	1.885	12	14	1
SOCIETA' IMPIANTI DI RISALITA SESTOLA S.I.R.S. SOCIETA' PER AZIONI IN LIQUIDAZIONE	-117	-123	-177	-174	-4	-12	-38	-37	0	0	1
F.A.T. AUTO S.R.L.	-28	-25	-35	-27	-5	-5	-32	-26	145	148	1
CALLEGARO FRANCESCO ASSICURAZIONI S.R.L.	22	18	0	0	8	11	23	29	79	98	1
COSTRUZIONI S.R.L. DI GIOVANNI STOCHINO	40	33	0	0	-1	-1	1	1	42	58	1
NAUTICA COMPARATO S.R.L.	31	24	0	0	4	1	63	44	182	107	1
G.S. AUTO S.R.L.	-31	-37	-20	-23	-1	0	-17	-19	0	0	1
EDILISA S.R.L.	-17	-18	0	0	0	1	47	47	7	8	1
IMMOBILIARE ARANCI S.R.L.	10	-25	0	0	-12	2	15	5	26	7	1
ESTENSIONE S.R.L.	29	30	33	32	1	1	71	68	5	16	1
PA.IA. S.R.L.	3	3	-2	-3	0	1	5	6	2	2	1
PARISE S.R.L.	15	25	0	0	4	10	583	328	18	23	1
L'EUROCOMES SRL	62	74	0	0	0	1	275	497	51	54	1
TECNOFIL - S.R.L.	58	57	32	0	3	-2	137	133	15	13	1
LA CASA DELL'ALBUM S.R.L.	-19	-14	0	0	1	2	6	2	18	27	1
RIO - S.R.L.	21	23	50	52	10	10	2.022	1.808	31	30	1
ACCIAI SPECIALI TERNI S.P.A.	7	-4	0	0	7	0	61	52	179	193	1
DA GIOVANNI S.R.L. UNINOMINALE	5	6	0	0	3	4	1	1	29	23	1
NOVATION TECH S.P.A.	5	5	-8	-13	3	5	16	18	130	160	1
PRATHOTELS S.R.L.	0	-8	0	0	-1	-5	78	56	5	5	1
NUOVA SPERANZA - PICCOLO ELVIS SOCIETA' COOPERATIVA SOCIALE	17	23	0	0	-63	-8	11	10	0	55	1



GRIECO S.R.L.	-31	-32	-1	0	1	0	5	5	0	0	1
D.I. GUSTO SRL	8	26	32	31	9	2	70	59	69	70	1
CARLUCCI TRASPORTI SOCIETA' COOPERATIVA	-27	-17	10	23	-14	-10	-7	13	134	129	1
PUNTOIMMOBILIARE SOCIETA' A RESPONSABILITA' LIMITATA	-89	-87	1	1	1	0	5	5	4	3	1
SALA GIOCHI PONTE ORLANDO S.R.L.	77	84	0	0	5	6	172	260	37	38	1
OFFICINEPONTE - COOPERATIVA SOCIALE	-23	-10	-4	0	-21	-3	-23	-4	207	218	1
AGO FILO E FANTASIA SRL	32	34	0	0	-2	-2	1	0	3	3	1
BIO EDIL GROUP SOCIETA' A RESPONSABILITA' LIMITATA	63	63	41	52	23	8	171	173	199	0	1
SEMPLIFICATA L.I.G.E. COSTRUZIONI S.R.L.	44	51	0	0	-21	6	85	115	0	47	1
F. MAIORANA RAPPRESENTANZE S.R.L.	21	46	0	0	5	-1	24	50	86	102	1
SIMAR S.R.L.	2	2	0	-2	3	3	13	12	15	59	1
UMBRIA DIGITALE S.C. A R.L.	20	29	0	0	0	1	60	83	44	48	1
MAX S.R.L.	0	0	-20	-18	0	0	424	476	0	0	1
FER.VAL. S.R.L.	-26	-6	0	0	-30	-35	-18	-32	45	95	1
HB SERVIZI S.R.L.	-12	-6	0	0	-28	-56	2	20	33	127	1
COMO CONTROLLO S.R.L.	29	36	0	9	3	-6	19	13	268	256	1
MOKA EXPRESS S.R.L.	3	16	0	0	2	2	11	11	162	139	1
IL ROSETO SRL	98	90	-10	-6	4	1	7.896	859	76	31	1
FOR.MA. DESIGN SOCIETA' COOPERATIVA	11	7	0	0	3	13	7	10	202	154	1
LOYAL COMPANY SOCIETA' A RESPONSABILITA' LIMITATA	57	48	0	0	14	13	79	57	380	374	1
NUOVA ELLE EMME S.R.L.	64	55	19	17	1	1	44	36	76	76	1
ZEROSEI IMMOBILIARE S.R.L.	17	16	0	0	1	0	19	18	4	17	1
SERVICES LAZIO S.R.L.	-8	-10	4	0	5	8	10	7	65	70	1
VIGNE DEL VULTURE SOCIETA' AGRICOLA A RESPONSABILITA' LIMITATA	-27	-23	0	0	1	-3	25	23	10	13	1
SARECA S.R.L.	55	46	0	0	10	12	138	104	162	155	1
GEORESIDENZIALE - SOCIETA' A RESPONSABILITA' LIMITATA	48	58	-1	0	2	2	2	1	12	2	1
PESCHERIA JOLLY S.R.L.	52	52	0	0	5	4	110	111	214	231	1
VAIAN MARMI 2 S.R.L.	16	19	15	12	5	10	36	40	113	115	1
STUDIO SORGENTE S.R.L.	37	43	0	0	29	34	368	131	216	232	1
CHIANESE S.R.L.	12	10	10	0	60	71	10.450	1.013	100	90	1
ONYVA' COOPERATIVA SOCIALE	25	32	0	-8	3	14	75	104	105	93	1

TREND - S.R.L.	17	27	0	-2	-8	21	12	42	69	200	1
EMME S.R.L.	-17	-19	0	0	1	1	9	11	83	134	1
ADM CONSULT S.R.L.	22	18	9	8	1	3	18	19	41	42	1
FELPI S.R.L.	44	37	0	0	1	1	52	57	0	0	1
PARAFARMACIA SIBILLA SRL	16	2	15	13	8	1	55	37	175	149	1
KAKTUS S.R.L.	22	18	11	8	4	5	40	32	55	61	1
DOKDO - SOCIETA' A RESPONSABILITA' LIMITATA	0	38	0	0	-48	2	23	20	141	168	1
C & Z SRL	96	96	0	0	0	-3	0	-2	0	0	1
FITNESSEDANZA.IT - S.S.D.R.L.	45	57	83	49	4	33	1.119	653	111	99	1
OTTICA & ORTOPEDICA SRL	43	34	33	32	11	10	93	84	148	162	1
IMMOBILIARE RONCO S.R.L.	31	31	3	-1	11	9	12	5	16	17	1
COLLIARD S.R.L.	-7	-9	0	0	-4	-4	72	49	30	32	1
RM IMMOBILIARE SRL	-51	-52	0	0	4	5	11	7	0	0	1
ITALPELLI S.R.L.	14	13	0	0	6	8	12	9	239	164	1
MAIL WASH S.R.L.	79	65	-55	-44	2	-17	-9	-11	67	72	1
G.S. IMPIANTI S.R.L.	70	72	-5	-2	-2	-3	5.637	77.273	0	0	1
SAPORI DI BIO S.R.L.	-58	-64	-23	-14	1	-7	10	8	139	57	1
KYTERION SRL	68	69	0	0	-2	1	47	50	7	72	1
SUITEBOX S.R.L.	39	34	0	0	-13	-2	50	45	72	0	1
EDOSS PROJECT S.R.L.	3	38	5	0	15	22	58	54	143	137	1
GEFIM GESTIONI FINANZIARIE IMMOBILIARI E MOBILIARI SRL	48	41	63	67	8	10	4.898	4.517	16	17	1
C. MEDICA S.R.L.	15	54	19	17	4	8	51	54	103	105	1
C.M. COSTRUZIONE MECCANICA DI STERLUTI TARCISIO S.R.L.	11	11	0	0	2	3	139	132	12	10	1
PINDARO - SOCIETA' A RESPONSABILITA' LIMITATA	-70	2	-5	0	-75	0	-41	2	98	0	1
MALIBU' CAFE' SRL SEMPLIFICATA	-682	-216	-393	0	-280	-169	-93	-66	424	184	1
GENTILI CUCINE S.R.L.	13	17	1	1	3	4	23	21	108	101	1
GREENWORLD SRL	18	15	0	0	21	37	86	58	353	398	1
S.G. CARAMELLE STELLA S.R.L.	-1	-25	0	0	15	4	42	23	83	83	1
IMMOBILIARE SIRIO S.P.A.	-13	-17	5	4	3	2	318	306	9	9	1
ASPAN SERVIZI S.R.L.	34	32	0	0	5	3	28	24	60	61	1
ASOLMEC SRL	22	19	0	0	7	13	10	19	108	135	1
ITALY ADSL S.R.L.	-40	-20	0	0	5	3	13	9	138	129	1
IEMMEGI S.R.L.	32	29	0	0	2	2	9.446	9.588	6	6	1
MA.G.I.C.A. DI GIUSEPPE TRIPALDI S.R.L.	21	31	18	26	6	-13	1.099	1.388	0	0	1
INJ ONE S.R.L.	3	4	0	0	0	0	1	1	0	0	1
AGRICOLA LENTI SOCIETA' COOPERATIVA	-5	7	0	0	2	2	35	40	27	29	1

F.Q.R. S.R.L.	-4	-6	0	0	3	2	26	22	7	7	1
MEDICA S.P.A.	-11	-3	0	0	2	2	24	24	79	75	1
KFL S.R.L.	21	24	2	2	17	16	51	47	128	110	1
RESTYLING IMMOBILIARE MILANO S.R.L.	65	52	0	0	60	33	263	292	156	144	1
ASCOM SERVIZI SRL	13	5	-130	-101	5	4	23	14	65	69	1
GIORGETTI S.R.L.	55	51	7	6	2	3	291	262	89	87	1
VALIANI CARROZZERIE INDUSTRIALI S.R.L.	3	12	0	0	6	5	98	107	67	68	1
M&M CONSULTING S.R.L.	44	89	0	0	17	61	262	861	84	132	1
CORTESI SRL	5	15	-2	0	3	0	2	2	171	178	1
DEAN SOCIETA' COOPERATIVA	-21	-19	-33	-6	3	-27	-27	-32	300	335	1
REBIRTH S.R.L.	24	24	0	0	-7	3	177	167	80	82	1
RE.GI.N. IMMOBILIARE S.R.L.	35	35	0	0	0	1	63	64	0	0	1
TRALCIO S.R.L.	81	81	-4	-3	-1	-1	9	10	3	3	1
MAPLE SRL	35	53	1	1	5	3	38	50	166	268	1
CAVUTO A. & G. S.R.L.	25	16	0	0	6	7	30	27	110	109	1
GIEMME - SOCIETA' A RESPONSABILITA' LIMITATA	38	26	0	0	0	3	184	167	3	4	1
NUOVA ALVA S.R.L.	2	-3	0	0	9	10	468	386	66	63	1
ESSEELLEEMME S.R.L.	5	5	3	3	0	1	144	142	7	7	1
G. R. PAINTING S.R.L.	32	29	0	0	7	-18	54	62	131	109	1
AZIENDA TRASPORTI TURISTICI FUSINA - S.R.L.	47	75	0	0	3	7	230	1.251	91	56	1
A. C. CHIEVO - VERONA S.R.L.	-35	-32	0	0	2	3	5	5	2	2	1
OSM BIOMEDICAL S.R.L.	58	90	56	54	40	35	1.294	1.292	81	100	1
ENERGIA SAN MICHELE S.R.L.	16	17	-1	-1	0	0	2	2	0	0	1
IMMOBILIARE GUERRA S.R.L.	33	7	0	0	38	24	72	23	106	121	1
SANABELLA GROUP SRL	-2	-1	0	0	3	3	4	3	222	197	1
TWISTER S.R.L.	-6	-3	3	3	0	2	61	57	6	6	1
MIKO SRL	52	50	0	0	1	1	2	1	125	7	1
EUROPA SERVIZI SOCIETA' COOPERATIVA SOCIALE	24	27	0	0	0	3	24	30	45	59	1
GOLDEN UNION S.R.L.	6	4	6	5	1	1	8	6	87	77	1
EDILTRANS S.R.L.	11	16	-7	-7	5	4	9	6	160	99	1
HOBAMA S.R.L.	82	72	0	0	19	-3	306	184	54	30	1
ALSTEA S.R.L.	2	3	-2	0	1	-2	8	8	12	10	1
AGOLAB STUDIO S.R.L.	3	18	0	0	2	5	8	8	115	182	1
DINO INVESTMENTS S.R.L.	8	18	-7	-11	-7	-31	-6	6	0	0	1
IDECO S.R.L.	29	31	0	0	1	-1	173	344	82	81	1
S & C SRL	-14	-10	0	0	1	2	3	3	182	250	1
TECNOLOGIE INDUSTRIALI S.R.L. IN LIQUIDAZIONE	-1.810	-1.808	-2.865	-2.863	-1	-1	-95	-95	0	0	1
C.S.I. S.R.L.	7	4	0	0	5	3	8	5	227	230	1
EFFEDI S.R.L.	5	5	0	0	14	8	20	11	136	111	1

STONE S.R.L.	34	9	0	0	4	6	10	9	6	0	1
LA VIGNA - SOCIETA' COOPERATIVA SOCIALE ABBREVIABILE, OVE POSSIBILE, IN COOP. LA VIGNA	4	13	0	0	4	7	13	15	139	157	1
IMPRESA FONTANA LUIGI SRL	53	58	0	0	4	3	125	140	16	29	1
RENIT CPV S.R.L.	18	17	0	0	-8	-10	-3	-5	13	6	1
LA LUSTIGNANA S.R.L. SOCIETA' AGRICOLA	-92	-92	0	0	-1	-1	7	7	0	0	1
SMITHS MEDICAL ITALIA SRL	54	49	14	8	8	9	59	51	133	145	1
TRANSITEX ITALIA S.R.L.	4	5	4	0	-11	5	2	5	347	392	1
TRE G S.R.L.	41	37	0	0	15	8	59	38	126	138	1
MULTIPET S.R.L.	20	17	0	-18	-3	-3	83	71	150	203	1
HAR S.R.L.	11	10	0	0	-2	-3	128	142	0	0	1
CAIOLA ECOHOUSE SOCIETA' A RESPONSABILITA' LIMITATA SEMPLIFICATA	0	-8	0	0	4	5	4	4	29	180	1
MR. PIZZAMI S.R.L.	8	-7	0	0	3	7	24	34	162	204	1
NAUTICA F.LLI TULLI - SOCIETA' A RESPONSABILITA' LIMITATA	54	47	0	0	8	16	525	358	118	122	1
HOTEL RISTORANTE RENZO VALBRUNA S.R.L.	-11	-21	2	0	0	6	15	15	232	216	1
CENTRO SPORTIVO SESSO SOCIETA' COOPERATIVA	25	20	-3	-4	5	1	420	594	40	38	1
A.TECH S.R.L.	17	11	0	0	29	20	37	22	195	278	1
ISIDE SRL	-17	-14	0	0	-1	2	91	96	4	27	1
SOLFERINO S.R.L.	-14	-47	0	0	-1	-5	21	25	6	4	1
TOSCANA AERO SERVICE S.R.L.	76	46	-28	-31	17	6	1.193	2.028	10	11	1
NUOVA TERR-LEGNO S.R.L.	25	21	0	0	4	2	245	196	93	86	1
CIRO EXPORT S.R.L.	28	19	13	13	13	-1	48	30	65	126	1
M.A.R.I.DI.ST. S.R.L.	19	15	0	28	1	10	201	135	106	117	1
SILOR S.R.L.	17	21	0	0	5	5	10	10	127	134	1
PLATI SOCIETA' A RESPONSABILITA' LIMITATA SEMPLIFICATA IN BREVE PLATI S.R.L.S.	5	0	0	0	2	3	1	1	69	92	1
IMMOBILIARE LA MILANINA S.R.L.	46	7	0	0	-2	-4	-2	-2	4	4	1
ARENA COSTRUZIONI S.R.L.	9	10	0	0	-1	-1	6	7	0	0	1
DOLCE NETTARE DI SICILIA SOCIETA' COOPERATIVA	43	49	0	0	2	2	184	302	112	93	1
IMMOFINA S.R.L. SOCIETA' DI INTERMEDIAZIONE IMMOBILIARE E FINANZIARIA	82	67	-182	-271	26	-26	44	22	43	0	1
AIRIS JET S.R.L.	15	17	1	1	0	0	918	540	20	16	1

JONICA ZUCCHERI S.R.L.	8	17	0	0	0	0	-1	0	8	11	1
GRETA S.R.L.	-34	-33	0	0	1	3	31	48	7	9	1
SOCIETA' COOPERATIVA GOLFO DI PALMAS	-44	-12	-4	-3	-18	-5	9	89	181	252	1
ARIS S.R.L.	96	97	0	0	0	-1	0	0	0	0	1
PRO LIGHT G.M.B.H.	36	39	0	0	3	1	71	73	42	36	1
FARINELLI S.R.L.	76	66	-111	-91	0	-15	22	24	0	17	1
ACCOPPIATURA SAMUEL SRL	-8	-3	0	0	4	4	16	15	59	54	1
SAN ROCCO SOCIETA' COOPERATIVA	-82	-74	0	0	-31	-34	3	3	3	1	1
S.I.S.GE. 12 S.R.L.	1	1	0	0	-1	-1	1	1	0	0	1
VILLA MORLACCHI S.R.L.	95	95	-30	-26	1	0	-22	-20	2	2	1
A. RAYMOND ITALIANA - S.R.L.	47	38	1	1	24	20	218	181	133	124	1
COS.APP. S.R.L.	42	51	0	0	0	4	5	8	2	22	1
T.T.B. S.R.L.	87	88	0	0	7	4	999	956	41	39	1
I.M.T. SOCIETA' A RESPONSABILITA' LIMITATA SEMPLIFICATA	78	73	13	19	30	1	424	358	54	320	1
TALOS IMMOBILIARE S.R.L.	5	3	7	4	4	4	16	13	9	9	1
IMMOBILIARE D ORSENIGO S.R.L.	1	1	0	0	4	5	832	645	26	25	1
AGI-CASA S.R.L.	48	48	15	13	4	3	68	43	34	12	1
FISIOCENTER - SOCIETA' A RESPONSABILITA' LIMITATA	18	22	8	19	32	34	87	91	240	180	1
ELETTRICI GROUP S.R.L.	13	10	1	1	3	4	12	11	105	118	1
TCM GROUP SRL	1	15	0	0	-6	-1	316	1.013	0	2	1
CAMA 1 SPA	19	27	0	0	5	8	23	22	102	141	1
YELLOW S.R.L.	29	33	0	0	-4	-2	30	35	102	84	1
GE.VI. S.R.L.	8	7	5	4	2	2	13	13	94	86	1
SUPERMARKET MARIA 2 S.R.L.	18	17	0	0	1	-1	21	19	181	141	1
LEGOR GROUP S.P.A.	42	46	0	0	17	15	184	158	196	181	1
MG REAL ESTATE S.R.L.	14	6	5	3	2	3	8	7	7	9	1
CISETTE SOCIETA' CONSORTILE A RESPONSABILITA' LIMITATA	13	14	0	0	-2	-2	3	2	275	283	1
MOREIRA SOCIETA' A RESPONSABILITA' LIMITATA IN SIGLA: MOREIRA S.R.L.	48	50	0	0	-2	-1	31	34	0	0	1
IN.ED. S.R.L.	45	45	-6	-4	0	0	44	48	0	6	1
GIALLO S.R.L.	0	0	-8	-7	-1	-1	54	57	0	0	1
VALFRIDO SOCIETA A RESPONSABILITA LIMITATA	28	29	0	-2	-2	-1	208	213	2	22	1
ANITA SERVICE S.R.L.	10	7	0	0	4	4	145	138	7	8	1
F D S S.R.L. - FORNITURA DISPOSITIVI SICUREZZA	-214	-188	-225	-186	-15	-16	-70	-65	0	0	1

ANSEL TECHNOLOGY S.R.L.	62	59	0	0	56	45	188	178	139	127	1
ITAC SRL	4	16	0	0	2	4	9	37	70	135	1
BI-CLIMA S.R.L.	39	28	12	9	9	6	24	18	150	137	1
LIGHT SOLUTIONS S.R.L.	8	7	0	-7	4	9	8	8	56	103	1
BASSANI ALESSANDRO S.R.L.	30	30	0	0	5	2	218	205	56	50	1
BRAVO PRODUZIONI TELEVISIVE S.R.L. IN LIQUIDAZIONE	-2,7849	-2,5832	0	-42,787	-1,8272	0,80958	0,19834	1,9822	0,31053	2,52113	0
TOMMY S.R.L.	22,4675	18,0442	0	0	2,94295	3,51528	3,49811	3,19579	167,82	188,621	0
C.F.B. S.R.L.	-45,765	-44,862	-30,799	-28,757	-0,9333	-0,6733	-16,397	-15,79	0,19377	0,61757	0
CREAZIONI BARDULOS S.R.L.	-13,023	12,0864	0	0	-24,335	2,89023	-19,107	3,22634	88,9967	70,9046	0
MARETOUR S.R.L.	-15,531	-28,874	-0,6033	-0,6777	2,92023	3,12726	3,35548	2,69777	43,7837	43,6287	0
UNITED TECHNOLOGIES S.R.L. O UNITECH S.R.L.	-10,156	-6,9877	-18,513	-10,414	-2,2287	-0,7211	5,1224	9,28244	0	0	0
BP SERVICE S.R.L. IN LIQUIDAZIONE	-259,18	14,6064	7,11851	1,30405	-294,79	2,00405	-78,039	4,58877	29,596	135,026	0
ZEN 1 S.R.L.	39,4199	44,8203	0	0	-7,8022	1,18493	19,9617	32,4815	0	19,4871	0
PANORAMA SRL	-97,24	-95,69	-197,29	-194,13	-1,2267	0,86503	-65,319	-65,142	0	0	0
PRINCE S.R.L. IN LIQUIDAZIONE	-54,557	-15,891	2,08447	0	-24,596	5,09639	-14,542	12,8296	373,854	336,531	0
C.T.S. SPA - S.R.L. IN LIQUIDAZIONE	-32,116	-19,012	-145,61	0	-18,241	-133,46	-56,922	-52,547	67,4618	160,136	0
TLC SERVICE S.R.L.	45,4389	28,0022	0	0	19,75	8,94555	73,914	42,4506	215,2	281,625	0
REV-AVIATION S.R.L.	-35,844	-20,13	0	0	-16,712	3,75358	0,83411	24,5258	30,1352	53,9036	0
PROJECT SRL	20,0439	16,3896	0	0	-8,6418	2,12248	79,1107	88,1481	57,4047	76,9564	0
SIMARLUNO S.R.L.	68,9775	64,0206	0	0	-7,3394	-2,4501	1,19216	10,102	0,0894	0,08066	0
IMMOBILIARE ARIANNA - S.R.L. - IN LIQUIDAZIONE	-19,746	-23,314	-124,04	-115,17	4,12107	-7,5001	-61,223	-61,623	4,40115	13,6388	0
FATTORIA DI COMPIGNANO S.R.L.	-17,163	-27,952	-52,589	-38,249	-0,662	-8,052	-11,14	-5,9249	2,56652	1,56676	0
ERVI S.R.L.	48,5544	44,6044	0,43869	0,38805	1,46493	1,12526	1,08406	1,55354	32,7189	34,4988	0
BRANA' SOCIETA' A RESPONSABILITA' LIMITATA	-204251	-284439	-540741	-751603	-450	-1327,8	-99,981	-99,986	0	0	0
IMMOBILIARE ELENA SRL - IN LIQUIDAZIONE	69,4428	72,1259	-766,07	-542,23	-2,9041	-221,6	-400,06	-405,66	0	1,18083	0
LA STELLA DEI MARI S.R.L.	-8,0029	1,11844	0	0	-12,241	-3,5895	1,71685	1,16381	4,94222	9,8124	0
FIMAT ITALIA S.R.L.	-2,0121	-3,8222	-3,2612	0	-2,006	-1,9409	2,28871	6,28973	201,755	265,573	0
BIORIGEN S.R.L.	1,02087	14,1962	0	0	1,88198	8,38641	23,2511	34,5314	0,89866	13,246	0
ANGELERI D.F.G. ORTOFRUTTICOLI S.R.L.	-24,22	2,12066	0	0	-24,726	-20,446	-14,116	7,5906	40,4168	45,5417	0
ALWA S.R.L.	0,59668	-4,7371	-5,3421	-3,734	0,94271	0,21054	5,60611	3,64566	133,613	121,173	0
WINNER L. & G. S.R.L.	20,159	26,3819	0	0	4,2623	17,2415	23,0993	35,5373	116,55	106,187	0
COPRA SERVICE - SOCIETA' A RESPONSABILITA' LIMITATA IN LIQUIDAZIONE	-521,1	-475,86	-521,22	-466,2	-8,6356	-17,586	-83,9	-82,635	0	0	0
GESTIONE AEROPORTO ARBATAX DI TORTOLI - GE.AR.TO. S.R.L. - IN FALLIMENTO	-26,658	-11,487	-97,379	-70,916	-18,361	-25,688	-43,585	-35,868	0	0	0

STEFANELLI ENGINEERING S.R.L.	-171,46	-15,419	0	0	-232,22	1,96357	-95,593	4,43291	130,407	71,2537	0
POLITRANS SRL	-21,275	-20,821	0	0	-1,5932	2,98493	5,30312	11,6449	102,316	138,262	0
FERRALLUMINIO TOFFOLI S.R.L.	1,56034	9,38506	0	0	2,70969	2,00863	2,48153	2,68243	77,959	119,526	0
CHEIRON S.R.L.	53,7808	56,9323	0	0	-3,3013	0,54374	69,9758	55,2606	65,2022	53,7513	0
IMMOBILIARE GIARDINO S.R.L.	56,6798	48,1925	0	0	1,99107	2,18143	104,773	93,3625	10,4301	2,49338	0
IMMOBILIARE RONCO CENTRO S.R.L.	99,474	95,5824	-0,8426	-0,6915	-6,5104	-11,875	-3,0369	0,77272	16,742	26,5401	0
ITMONSTER S.R.L.	28,7892	26,3939	-41,62	-10,577	-2,3864	-34,214	37,1478	38,9976	65,6384	23,8968	0
NEW INTER SERVICE SRL IN LIQUIDAZIONE	-317,19	-265,36	-312,72	-254,81	-10,922	-37,993	-76,524	-79,555	10,7338	117,608	0
AG&CO S.R.L.	21,3322	31,2023	0	0	6,35883	10,3842	12,1909	19,5841	129,227	158,906	0
IMPRESL S.R.L.	25,4856	53,8985	0	0	-56,465	-9,8812	-27,108	7,96033	28,6034	8,06454	0
APE S.R.L.	-5,7437	-5,7257	0	0	-0,018	-0,0237	81,6822	81,742	0	0	0
SPORT LEGEND S.R.L. IN LIQUIDAZIONE	-51,295	-50,987	0	0	-2,1429	4,40875	11,4048	14,3882	0	0	0
ECO.M. SRL	18,2102	15,2924	6,03157	7,63657	4,68927	0,80248	10,3012	10,3418	85,7963	79,5514	0
ONE SRL IN FALLIMENTO	53,2592	54,1555	0	0	-3,2806	1,6646	-0,916	2,77891	0	0	0
DAMERI DESIGN S.R.L.	-198,22	-76,932	-132,45	-28,605	-76,47	-35,137	-86,269	-48,66	294,004	50,4521	0
DANA S.R.L.	-10,58	-5,0238	-11,929	-8,0616	2,37275	-1,4288	4,27747	3,71842	72,2079	61,9759	0
NOVAMUSA SOCIETA' A RESPONSABILITA' LIMITATA	2,22941	3,97685	-0,2383	0	0,79385	-4,299	6,74299	5,4334	51,1986	56,296	0
ARIETE GROUP S.R.L.	43,1483	40,7906	-39,504	0	0,33109	-32,731	10,152	11,2326	29,1769	71,7134	0
SUGHERIFICIO BONINO S.R.L.	13,9133	12,5189	0	0	2,71962	3,20617	20,5954	24,1738	69,8384	68,945	0
ELGIU S.R.L.	-82,225	-5,8492	0	0	-89,736	-1,9003	-50,851	10,4315	252,311	281,605	0
LE CINQUE VELE S.R.L.	-13,137	-13,134	-0,3865	0	0,00462	-0,3384	0,4879	0,49747	0,18536	0,26418	0
MARCOROSSI S.R.L. IN LIQUIDAZIONE	29,3122	37,7868	0	0	-32,441	-4,0605	169,013	239,224	7,2541	7,45672	0
LE QUERCE S.R.L.	-1,0758	-1,7088	-0,3065	0	-0,0678	-0,1184	0,91436	1,0709	1,09618	0,942	0
A & B SRL - IN FALLIMENTO	61,3941	57,4202	14,8473	0	25,2651	33,3646	223,917	270,679	731,133	924,174	0
CARPENTERIE CAMPANE S.R.L. IN SCIoglimento	-612,86	-552,48	-804,47	-724,23	-1,2827	-3,827	-97,304	-96,033	0	0	0
F R S.R.L.	52,0597	62,679	0	0	-3,7291	5,14504	6,57727	16,5866	68,4917	129,458	0
TSS - SAFETY SYSTEMS SOCIETA' A RESPONSABILITA' LIMITATA	-126,32	20,2463	-356,55	0	-218,6	-122,01	-84,817	-53,78	0	0	0
GENTILI TRASPORTI S.R.L. IN FALLIMENTO	-82,97	-13,197	-24,099	-14,9	-72,138	-3,7696	-49,466	0,92404	165,946	105,831	0
RAFFAELE GARZIA S.R.L. - IN FALLIMENTO	-47,893	-26,252	-15,252	0	-5,739	-9,5588	24,5548	40,8298	33,2039	43,664	0
GIOR '97 S.R.L. IN LIQUIDAZIONE	18,1493	14,3813	0	0	1,48409	1,02014	2,6977	2,44372	93,8447	107,371	0
GTS SRL SEMPLIFICATA	-11,838	5,08595	2,88727	1,37107	7,89753	2,69069	9,90902	3,20531	237,292	212,83	0
SANTE' DIRECT S.R.L.	2,02294	-1,6562	-3,7307	-10,006	3,7787	-3,9544	19,4267	17,6781	0	0	0
FERREMILIA S.R.L.	-57,371	-57,127	-70,774	-70,466	-0,2311	-0,2924	-37,465	-37,368	0	0	0
ECOVAL S.R.L.	10,2714	-1,0535	0	0	10,4562	-12,549	10,4902	5,71852	88,4018	97,5985	0
I.C.A. COSTRUZIONI S.R.L.	67,9515	-19,002	-20,496	-6,1272	-9,4074	-7,5127	-19,128	-8,466	39,3022	72,2307	0

SICIR - S.R.L.	-3,4373	25,0926	14,8255	15,2686	-29,927	3,64975	-6,9781	40,0399	287,816	121,79	0
GIO.STE. INVEST S.R.L.	74,5036	75,6288	0,57744	0,55657	-3,1641	-0,0466	-1,991	1,15282	0	0	0
LES SACHES SRL	-79,522	-72,543	-72,881	-96,835	-19,496	9,23973	-44,163	-41,405	115,524	217,531	0
P.C.E. PRODUZIONE CABLAGGI ELETTRICI S.R.L. - IN LIQUIDAZIONE	13,8788	9,06346	24,4281	20,4145	-5,7127	8,31501	27,2697	37,3387	23,3848	39,4174	0
GIACHI FIRENZE S.R.L.	-192,04	25,3682	0	0	-254,22	2,09691	-75,735	9,79269	103,818	62,3072	0
B.G.M. DESIGN S.R.L. - IN LIQUIDAZIONE	-82,218	-53,784	0	0	-22,273	-2,1017	-15,54	1,0914	20,781	24,2628	0
EDILNOVA S.R.L.	-0,1121	-0,1348	-0,605	-0,5621	3,02736	5,0396	1,17948	1,55248	86,1954	130,402	0
IMG S.R.L.	-1062,9	-830,33	-1636	-954,29	38,5692	-357,3	-94,246	-92,769	0	0	0
INGROSSO FORNITURE ELETTRICHE S.R.L. ENUNCIABILE ANCHE I.F.E. S.R.L.	-321,77	-300,18	-299,59	-276,7	-1,3799	-3,9425	-71,188	-69,705	0	0	0
AREA VIGILANZA PRATO S.R.L. IN LIQUIDAZIONE	-340,05	-376,84	-382,16	-240,11	-29,429	-130,17	-78,66	-76,467	45,9978	50,5495	0
SOGIT S.R.L. IN LIQUIDAZIONE	-220,25	23,2487	0	0	-494,21	10,0838	-92,233	5,98025	200,888	165,922	0
F.LLI CONTI PATERNO' S.R.L. IN LIQUIDAZIONE	-1860	-1214,7	-3255,1	-2115,9	-54,842	5,56073	-97,004	-95,378	0	0	0
S.I.G.E. S.R.L. IN LIQUIDAZIONE	-14,889	19,1015	0	0	-61,835	1,66504	-31,213	22,5023	80,5831	68,2171	0
LA VELOCE	-270,65	-270,65	-295,99	-295,99	0	0	-73,02	-73,02	0	0	0
AGOSTINO GALLO IMMOBILIARE S.R.L. IN LIQUIDAZIONE	-5,2956	52,416	-11,498	-4,695	-121,29	-0,4312	-54,714	0,50061	0	0	0
COSTRUZIONI IRPINE S.R.L. IN LIQUIDAZIONE	-22,798	-21,682	-68,815	-71,296	-0,9572	2,9177	-36,597	-36,037	0	0,06025	0
MINERVA COSTRUZIONI S.R.L.	11,5398	46,768	-24,787	-0,0518	-92,726	-7,0465	-45,663	14,2121	91,3783	50,6744	0
VIRGINIA 2005 S.R.L.	97,5679	98,2846	0	0	-7,429	-13,299	26,3366	39,8959	0	0	0
CO.E.M.I. S.R.L.	1,86931	-6,3007	1,86698	2,10714	2,30479	2,6053	4,91138	4,45673	81,5377	65,2694	0
IMA LAVORAZIONE LAMIERE - S.R.L. IN LIQUIDAZIONE	-11,814	-5,7078	-6,0117	0	-1,0298	-10,563	40,6507	43,8947	21,9346	29,2098	0
GRAFICHE CORRA' S.R.L. IN LIQUIDAZIONE	-114,44	-77,816	-137,99	0	-28,581	-89,75	-59,38	-42,789	4,06069	86,2037	0
MANIFATTURA TESSILMAGLIA S.R.L. IN LIQUIDAZIONE	-9,2819	21,5168	0	0	-27,745	5,6103	-11,369	17,929	94,7187	108,159	0
RICCELLI - SOCIETA' A RESPONSABILITA' LIMITATA IN LIQUIDAZIONE	-117,22	-69,233	-49,994	-30,72	-32,531	-5,8796	-45,443	-27,194	3,22691	10,751	0
STENAGO SOCIETA' A RESPONSABILITA' LIMITATA	-15,096	-14,794	0	0	3,82447	4,44782	18,4928	19,9271	107,364	117,09	0
MERCURY FAST S.P.A.	73,0115	38,2383	0	0	-55,14	-46,526	15,8372	35,3517	86,692	47,9957	0
ITALIA COSTRUZIONI S.R.L.	-85,201	-6,4307	-14,464	0	-66,183	0,1254	-44,197	-3,1933	42,0879	106,609	0
SISTEMA CABLAGGI S.R.L. IN LIQUIDAZIONE	-205,78	16,0018	0	0	-240,65	-15,671	-67,398	7,89997	231,44	118,893	0
MULTI SERVICE MOIRAGO S.R.L. IN LIQUIDAZIONE	60,9432	-20,236	-12,194	-17,573	25,2244	2,77388	12,422	-12,664	30,1639	31,0879	0



PIZETA GROUP S.R.L. IN FALLIMENTO	-15,132	-12,318	-19,53	0	1,03433	-13,019	-18,069	4,50214	13,1824	9,13615	0
PADOVA S.R.L.	-164,25	-0,3839	-37,21	0,08639	-181,85	-11,229	-68,106	-9,0633	8,51179	57,2483	0
EDILIZIA SESTRI LEVANTE S.R.L.	-10,762	-10,617	-38,647	0	-0,262	-38,452	-9,7387	-9,5907	0,11501	0	0
SINE S.R.L. IN FALLIMENTO	5,25977	8,39013	0	0	0,9252	1,03677	4,2991	9,31535	135,389	28,496	0
TS LAB S.R.L.C.R.	-22,604	6,02229	15,7428	16,1782	-30,958	-3,7818	-11,071	16,5137	11,5445	0	0
G.M.A. GENERALE MECCATRONICA APPLICATA S.R.L. IN LIQUIDAZIONE	-29,443	-18,662	-3,3621	1,24741	-23,664	-8,1633	-15,223	10,6386	55,8137	59,2048	0
LISTO PARQUET S.R.L. IN LIQUIDAZIONE	-8,2887	-18,97	0	0	2,92627	5,68133	22,1783	22,1388	142,257	182,081	0
ENNENOVE IMMOBILIARE S.R.L.	-52,603	-44,035	-87,478	-59,076	-16,962	-17,555	-66,387	-46,14	0	0	0
LISA S.R.L.	-2,7768	-13,339	0	0	11,901	5,10693	15,0462	15,821	46,3829	98,6604	0
VIFIM S.R.L.	-28,32	-32,59	0	0	-2,0948	-2,0519	77,313	79,4485	1,08164	1,15779	0
OFFICINE LEONCINI E C. S.R.L. IN LIQUIDAZIONE	-82,109	18,0229	27,3067	17,6902	-117,31	-4,2998	-47,673	36,9008	76,0364	81,2609	0
LAVAREX S.R.L.	-65,574	-27,462	-47,902	-13,091	-60,837	-11,429	-52,117	-12,077	9,15437	44,1347	0
DVG S.R.L.	58,9544	59,0058	0	0	0,39805	0,5495	2,82284	3,04611	52,9021	69,7926	0
SOFY S.R.L.	32,5341	3,16757	0,64471	0	1,98483	2,91074	4,81079	6,31441	210,694	392,323	0
TOW S.R.L.	-75,244	-90,465	0	0	5,05381	-4,4702	6,7677	-2,8706	0	0	0
VERBENA PROMOZIONI IMMOBILIARI SRL	84,1125	87,5582	-16,467	-16,024	0,62759	-0,3548	1,23217	1,73959	0	0	0
EUROBUNKER S.R.L. - IMP. COSTIERO DI MESSINA	86,2973	86,8606	-9,5889	-8,7564	-0,7247	-0,809	-30,866	-29,031	0	0	0
JOLLY LAMPADARI S.R.L.	63,0462	68,9073	-157,43	-81,305	92,903	-38,526	-26,479	-53,615	90,044	99,0622	0
3 G. GAMES S.R.L. IN LIQUIDAZIONE - IN FALLIMENTO	-22,689	-26,641	0	0	5,73538	-11,772	18,7736	2,74995	114,298	125,11	0
SISSI SRL	38,8823	35,8529	0	0	43,5665	27,6777	103,458	66,9363	191,944	189,229	0
C.B.S. S.R.L.	22,7351	21,064	0	0	-3,945	-1,7511	2,732	6,02431	80,4666	37,3225	0
MDSOLAR S.R.L.	16,1673	29,9207	0	0	-4,2298	3,61611	7,31572	16,8485	124,529	134,155	0
E.A.A. - ELETTRONICA APPLICATA ALL'AUTOMAZIONE - SOCIETA' A RESPO NSABILITA' LIMITATA O, IN FORMA ABBREVIATA, E.A.A. - ELETTRONICA APPLICATA ALL'A UTOMAZIONE - S.R.L.	-7,3039	-3,3469	0	-10,865	-38,5	-6,2252	-30,808	-5,2964	82,9196	75,8152	0
NET SYSTEM S.R.L. - IN LIQUIDAZIONE	-51,925	-42,94	-82,96	0	1,6473	-101,38	-32,293	-25,967	32,9703	70,6303	0
RISTOBEST S.R.L. IN LIQUIDAZIONE	28,4321	74,8704	-11,946	0	-2,2995	-9,7576	44,3603	75,0267	28,0866	146,618	0
SOGRESAL S.R.L.	14,6187	14,6424	0	0	-0,0237	-0,025	17,2732	17,3058	0	0	0
LAMAPLAST 2000 S.R.L. IN LIQUIDAZIONE	-24,531	-14,758	-19,586	-16,744	-0,2252	2,90237	11,0114	13,4852	96,3308	125,112	0
METROPIU' S.R.L. - SOCIETA' IN LIQUIDAZIONE	94,8972	-24,987	-26,063	-21,023	-3,0495	-2,5468	-22,368	-20,048	0	0	0
BOTTEGA VICENTINA S.R.L.	-51,634	3,09185	0	0	-54,949	6,9634	-33,941	6,65767	37,4445	113,453	0

IL GRILLO TURISMO & BENESSERE S.R.L. (IN SIGLA IL GRILLO S.R.L. ) - IN FALLIMENTO	89,8865	89,9185	-0,1452	-0,1132	-0,032	-0,032	53,7431	53,8188	0	0	0
MERIDIONALE COSTRUZIONI S.R.L.	-31,585	-23,568	0	-12,302	-3,2787	0,99583	3,80315	6,72083	12,8421	26,4986	0
ANTIUM S.R.L.	54,1182	73,8821	1,60882	0,58744	7,81603	2,34702	18,4618	12,202	48,9708	19,5755	0
NOVA SIRI SOCIETA' CONSORTILE A RESPONSABILITA' LIMITATA	-58861	0,09448	0	0	-57799	0,63673	-100,22	0,07507	217,721	8,80143	0
CO.EL - SOCIETA' A RESPONSABILITA' LIMITATA IN LIQUIDAZIONE	-50,216	4,32185	-63,572	65,0142	-60,307	-96,854	-79,91	-30,958	24,3605	208,396	0
COIMPRO S.R.L.	-64,468	38,2858	-1,8581	-2,5345	-142,05	1,53348	-84,063	0,45278	245,068	141,273	0
SO.VE.LA. SOCIETA' VENETA LAVANDERIE S.R.L. SOCIETA' UNIPERSONALE IN LIQUIDAZIONE	-2025,4	-94,883	-1509,8	1,91847	-284,34	-141,73	-95,295	-70,213	224,967	143,455	0
B & B INTERNATIONAL S.R.L.	80,6081	65,5194	15,1486	8,93649	9,95568	21,8575	297,949	162,491	111,69	128,179	0
GEI INFORM S.R.L.	-43,014	-21,445	-16,14	-4,3399	-9,2873	-4,2521	20,059	127,846	0	0	0
SPECIAL MILLS SOCIETA' A RESPONSABILITA' LIMITATA	-186,49	-60,637	-60,862	0	-88,424	-33,279	-59,01	-23,685	143,945	261,346	0
NAUSICA S.R.L. IN LIQUIDAZIONE	-807,9	23,7992	0	1,74972	-204,76	-19,577	-88,251	6,14468	262,089	171,042	0
CARTOLAND S.R.L.	12,3987	9,65871	0	0	0,53587	3,02828	12,0603	10,5598	95,8684	74,4559	0
SOC. BIO PUGLIA SOCIETA' A RESPONSABILITA' LIMITATA , IN SIGLA SOC. BIO PUGLIA S.R.L.	7,30876	7,21009	0	0	2,81913	-5,5091	11,9135	13,0266	82,666	216,377	0
PRS MONDO MOTO S.R.L. IN LIQUIDAZIONE	30,0206	9,47153	-17,308	-6,3221	3,48941	-13,278	3,82248	3,82099	208,529	208,861	0
ORIONE SPA IN LIQUIDAZIONE	-47,069	-20,182	-5,0908	-3,8938	-26,949	-0,2476	-47,304	-35,166	0,5808	0,64532	0
MONSUMMANO 2006 S.R.L.	58,4897	59,5711	0	0	-0,7083	-0,2937	0,9968	1,0936	0	0	0
CBC EDILIZIA SRL - IN LIQUIDAZIONE	-18,225	-4,582	-13,927	3,84121	-7,0407	-17,469	-19,49	-10,724	36,6975	108,399	0
LA PANDA S.R.L. IN LIQUIDAZIONE	-65,421	-62,23	-43,678	-17,837	-6,3671	-10,707	-32,122	-26,38	1,18006	1,31761	0
MARCO LAGATTOLLA S.P.A.	-33,951	34,3276	-0,4711	2,8548	-103,25	-1,615	-30,728	75,2597	117,261	77,4446	0
MPA GROUP SOCIETA' COOPERATIVA COSTRUZIONI GENERALI IN LIQUIDAZIONE	-101,57	-63,92	-73,19	-24,475	-25,946	-26,593	-49,63	-33,696	2,45237	106,468	0
IMP.EL. - POLI S.R.L.	12,3435	4,22448	0	-4,4786	3,36505	5,43989	1,30584	1,45881	41,8797	70,2745	0
STARAUTO S.R.L. IN LIQUIDAZIONE	-449,93	-440,09	-512,87	-507,33	3,66635	0,46042	-92,918	-94,34	5,87792	5,84041	0
CASA E BOTTEGA - S.R.L. IN FALLIMENTO	-11,155	10,2281	0	0	-2,6209	-0,6022	11,9232	16,6984	12,3765	25,0117	0
F.B. SERVICES S.R.L. - IN FALLIMENTO	-3,0986	-8,5672	0	0	2,78299	2,70828	23,9764	21,2051	91,1436	89,9612	0

FRANCO & C. S.R.L. IN SCIOGLIMENTO	30,8678	25,3648	-0,5313	0	-0,3205	5,51582	10,5353	10,067	30,7687	68,9416	0
EUROITALY ROMA SRL IN LIQUIDAZIONE	-9,6879	13,7802	-1,5567	-0,8779	-50,355	0,83947	-37,246	2,7461	238,587	152,226	0
F124 - SOCIETA' A RESPONSABILITA' LIMITATA	-29,41	3,68614	-0,4438	-0,4363	0,453	-0,0381	1,75957	1,97932	33,8365	3,9025	0
CO.PR.AL. - S.R.L.	-15,069	22,114	0	0	-53,63	4,05031	-35,105	21,2033	240,607	206,95	0
TOP SEA FISH S.R.L.	-13,981	-13,597	1,49164	2,72523	-1,4658	-0,1622	4,47732	7,98914	3,83669	173,061	0
AGRUMARIA TIRRENICA S.R.L.	61,4554	57,5036	7,22981	0,63714	18,3687	7,99535	26,6077	8,87578	208,411	140,637	0
THE LEGEND S.R.L.	13,1133	4,31361	0	0	9,49591	0,45444	38,5811	27,5642	112,861	119,19	0
ORTOPEDIA E BENESSERE SEMPIONE S.R.L.	37,5595	33,2201	0	-21,406	-0,2304	-1,0542	334,643	194,688	34,2967	45,1876	0
BIO SAC S.R.L. IN LIQUIDAZIONE	-25,797	-36,59	14,7143	18,7992	-0,4566	-5,0612	68,9211	59,0825	6,0405	22,122	0
BESANA S.R.L.	-140,09	-34,098	0	0	-161,57	-5,0683	-66,193	-3,2758	160,394	78,9641	0
IONE S.R.L. IN LIQUIDAZIONE	-65,285	-3,083	-66,027	0	-57,883	-37,811	-86,011	-33,935	170,967	130,715	0
UNICITY S.R.L. IN LIQUIDAZIONE	-107,18	-45,551	-71,8	0,25125	-34,912	-47,196	-52,246	-32,094	3,26891	15,9456	0
FER.F.IM. S.R.L.	-21,5	-26,004	0	0	3,14604	3,36322	23,7626	36,8984	82,368	105,998	0
DIANA 2 - S.R.L.	-106,89	-35,968	0	0	0,88992	3,79127	-37,31	21,8618	50,9297	159,578	0
FIEMME LEGNO CARPENTERIA S.R.L.	-31,83	-34,007	0,00286	0,0026	1,42115	-0,7442	3,28239	-1,2774	62,6922	60,0004	0
SUOLIFICIO NUOVA LINEA S.R.L.	-172,82	-39,117	-128,35	0,99407	-115,06	-77,375	-52,947	3,70824	37,5124	22,8666	0
VILTEX S.R.L. IN LIQUIDAZIONE	-20,282	-21,171	-23,164	-22,657	-0,461	-0,3329	-17,326	-16,917	0,01707	0,76615	0
TIPITALIA - S.R.L.	1,85759	14,9835	0	0	-21,616	1,44782	-20,306	11,1137	144,616	136,286	0
CAMPIONMAX S.R.L.	-164,53	-62,282	-121,99	0,84415	-1,8964	-98,651	-65,202	-57,157	49,4018	60,4353	0
BIPACK SRL	58,3082	51,0263	-8,4749	-8,9949	7,08629	7,00997	90,0739	54,9701	231,837	169,699	0
CASA LIBELLA S.R.L. - IN LIQUIDAZIONE	-21,841	-5,8567	-9,9436	-4,7847	-14,339	1,48257	-11,257	1,99508	0	37,8503	0
IVAC IMPIANTI S.R.L. IN LIQUIDAZIONE	13,3611	6,36813	0	0	6,0609	6,23065	12,0144	9,2117	180,203	177,203	0
FRULLOTTO S.R.L.	-14,858	-7,2678	-29,502	0	-2,2298	-25,691	57,8353	52,5245	1,44128	128,426	0
G & F DI GUERRIERO GIUSEPPE S.R.L.	6,90755	5,9444	0	0	0,51412	0,98676	8,90291	8,64431	17,3172	34,0544	0
NUOVA IRIDE S.R.L.	6,3706	1,17887	-0,1879	-18,343	4,6192	-3,6779	12,6809	1443,14	119,311	0	0
ROITZ LOGISTIC S.R.L.	23,5495	2,3866	0	0	2,073	11,6395	5,73327	9,5937	104,912	260,323	0
MSV ITALIA S.R.L.	39,6159	41,3499	7,81993	10,7929	-9,9504	-0,3805	9,96847	30,5451	73,3599	83,7482	0
SICUREZZA STRADALE ITALIA S.R.L. IN LIQUIDAZIONE, IN BREVE S.S.I. S.R.L. IN LIQUIDAZIONE	-3153,9	-2666,7	-3300,9	-2249,5	-55,943	-431,41	-101,59	-96,287	0	0	0
FRIGERIO CIRILLO & FIGLI S.R.L.	-399,94	-104,16	-387,79	-134,36	-107,04	-0,3788	-79,957	-50,317	0,67187	0,17573	0
PROSCIUTTIFICIO VALDINIEVOLE S.R.L. --- IN LIQUIDAZIONE ---	-94,114	-92,591	-1,9278	-1,3112	-1,3617	-0,0961	-1,543	0,10522	4,63071	4,38154	0
FINEDIL S.R.L. IN LIQUIDAZIONE	-44,741	-44,701	-158,58	-158,51	-0,0272	-0,0272	-48,254	-48,239	0	0	0

EMMETRE S.R.L. IN LIQUIDAZIONE	-17,835	11,9092	-69,242	0	-12,933	-55,998	-45,36	-36,271	1,83191	1,14888	0
VMC TECHNOLOGY S.R.L. IN LIQUIDAZIONE	-13,579	-30,45	0	0	2,85356	3,2047	3,43922	4,0325	43,7115	84,3126	0
BLUE AIR SOCIETA' A RESPONSABILITA' LIMITATA IN LIQUIDAZIONE	-1673,8	-1554,6	-1810,2	0	-11,374	-1706,1	-94,488	-94,134	0	0	0
ST SOLUZIONI EDITORIALI SRL	10,6615	13,2373	-0,0862	0	-3,309	3,74618	0,16662	3,71682	0	3,49099	0
I.S.A.M. S.R.L. - IN LIQUIDAZIONE	-96,938	-108,12	-170,13	-176,03	4,84379	-2,747	-67,826	-69,783	1,66219	2,79472	0
EDM EURODATA MONITOR S.R.L. IN BREVE EDM S.R.L. IN LIQUIDAZIONE	-22,61	-22,118	-32,242	-28,116	-0,3008	-3,9617	-19,462	-19,114	0	0	0
EDICART SOCIETA' A RESPONSABILITA' LIMITATA ED IN BREVE: EDICART S.R.L.	2,01638	14,9848	-32,111	-30,797	-13,55	3,92142	11,753	27,2517	102,316	134,782	0
IGERA S.R.L. - INDUSTRIA GENESI ECOLOGICA RICERCA AGRICOLA	-28,821	-28,756	-12,193	-11,907	-0,065	-0,2838	-6,2934	-6,2248	0	0	0
WELFARE COMPANY S.R.L.	-3,3906	-5,833	0	0	-4,2543	-13,542	0,79979	-3,3003	84,4094	103,451	0
TRANSER S.R.L.	-2,244	-4,9191	0,50036	0,7156	-2,7289	2,29849	15,1118	30,983	91,6401	146,948	0
CENTRO TURISTICO VACANZE EUROPA IL CAVALIERE SOCIETA A RESPONSABILITA LIMITATA	59,4688	57,4658	0,95905	2,56228	-5,141	-1,5252	-0,9283	6,59702	0	0	0
GIMMI S.R.L.	-94,521	32,6547	0	2,22897	-168,83	-37,744	-48,262	45,2584	70,9119	127,331	0
SACITRANS SRL	-44,612	-42,98	-2,5775	-3,458	3,66387	5,93419	16,573	13,3646	99,1493	92,0644	0
SANTELO S.R.L. - IN LIQUIDAZIONE	18,0123	11,0364	0	0,5491	0,21016	-1,1591	5,26725	2,98602	2,53828	98,1226	0
NUOVA CREMONA S.R.L. - IN LIQUIDAZIONE (IN FALLIMENTO)	-23,919	-13,12	0	-2,4029	-4,1189	-0,7535	-2,2239	1,8789	0	0	0
NORDWALL INTERNATIONAL S.R.L.	9,99342	30,8503	0	0	-81,775	-6,987	-0,0698	39,1595	89,3203	89,0067	0
E.L.I. VOLTURNO S.R.L.	-7,067	15,6979	0	0	-11,49	-1,093	2,59393	16,8433	14,9353	45,6331	0
RESEARCHLIFE - SOCIETA' CONSORTILE A RESPONSABILITA' LIMITATA IN LIQUIDAZIONE O ANCHE IN FORMA ABBREVIATA RESEARCHLIFE S.C.R.L. IN LIQUIDAZIONE	-157,21	-149,14	-748,96	-669,75	-3,2604	-55,828	-117,15	-115,9	0	0	0
EXTRA S.R.L. IN LIQUIDAZIONE	-73,078	-57,953	0	0	-19,118	-5,098	-43,498	-30,829	5,80283	6,21231	0
LOGOS INFORMATICA - S.R.L.-	-10,243	9,62064	-3,9873	-8,6227	-7,1332	-2,8702	-4,0892	6,30887	137,141	241,735	0
KARY S.R.L.	72,7977	72,7977	0	0	0	0	29,0812	29,0812	0	0	0
BIESSECI S.R.L. IN LIQUIDAZIONE	-83,374	10,1329	-12,311	-0,161	-104,68	-4,3298	-51,105	7,47092	156,095	79,051	0
M.D.R. SRL	-13,789	-8,1049	-8,9051	-7,7559	-0,1322	-0,2153	-12,144	-11,204	0	0	0

NEW EDIL COSTRUZIONI S.R.L. - IN LIQUIDAZIONE	5,81639	-3,1856	0	0	8,28053	5,25348	15,7178	9,71126	114,882	72,4939	0
LA NUOVA SIPOM S.R.L.	13,9529	5,96743	0	0	3,28407	2,99619	9,93137	14,1904	83,7338	117,415	0

**Tabella 34:** Campione totale delle imprese selezionate, 200 sono solventi e 200 non solventi

**Fonte:** AIDA

## Appendice 2.1.1

```
> #Dati imprese solventi non quotate
> IS17 <- read_excel("dati/Imprese solventi/IS17.xlsx")
> View(IS17)
> str(IS17)
Classes 'tbl_df', 'tbl' and 'data.frame':    200 obs. of  5 variables:
 $ X1
 %
2017: num  29 -68 29 50 69 58 79 -117 -28 22 ...
 $ X2
 %
2017: num  0 10 0 0 27 0 0 -177 -35 0 ...
 $ X3
 %
2017: num  14 2 23 6 27 17 1 -4 -5 8 ...
 $ X4
 %
2017: num  41 29 48 58 83 ...
 $ X5
 %
2017: num  309 0 114 236 40 62 12 0 145 79 ...
> AZS=IS17
> dim(AZS)
[1] 200 5
> #Calcolo dello Z'-score
> Z1=(0.717*AZS[,1]/100)+(0.847*AZS[,2]/100)+(3.107*AZS[,3]/100)+(0.420*AZS[,4]/100)+(0.998*AZS[,5]/100)
> Z1S=as.numeric(unlist(Z1))
> write.table(Z1,file="mydata.csv",quote=F,sep=";",dec="," ,na="NA",row.names=T,col.names=T)
> print(Z1S)
 [1]  3.89893 -0.21892  2.26186  3.14380
 [5]  2.31011  1.78121  9.96566 -2.62196
 [9]  0.66014  1.29132  0.67909  2.42751
[13] -0.49414  0.14537  0.02134  0.86661
[17]  0.04553  2.86007  2.10852  1.50521
[21]  0.09968  9.68655  2.31030  0.42268
[25]  1.42590  0.34643 -1.78932 -0.17867
[29]  1.59065  0.76405 -0.53767  1.79910
[33]  1.11800  0.20144  4.21781  0.02001
[37]  1.26500  0.31185  0.83452  1.61140
[41] -0.74502 -0.61826  3.05558  1.74661
[45] 34.66392  2.21744  4.96787  1.59416
[49]  0.27268  0.82257  0.04228  2.90141
[53]  0.52599  3.12591  1.67606  4.86760
[57] 46.92294  1.63536  0.61235  0.77532
[61]  0.74982  0.56495  2.46783  1.09209
[65]  0.01242  0.68832  6.95752  2.79723
[69]  0.79953  0.42733 -0.19519  2.72242
[73]  0.79358 24.07281  0.84962  0.69268
[77]  0.80428  2.20065 21.85761  1.63490
[81]  0.84457 -2.06866 -13.07733  1.36933
[85]  4.66567  1.46362  1.46777  1.11553
[89]  1.49507  1.30039 40.02466  5.10525
[93]  0.02571  0.44275  0.24359  0.87249
[97]  2.18734  4.99173 -0.10724  2.62620
[101] 1.28819  2.78239  1.82710  2.54373
[105] 1.49639  0.51555  0.58356  2.23105
[109] 1.58947  1.07520  2.91825  0.73592
[113] 1.98111  2.30438 -0.14785  8.37616
[117] 0.11465  2.77755  2.31123  0.29847
[121] 1.65981  0.72198  1.02677  1.80953
[125] 3.00239  0.18183  1.26495 -0.24462
[129] 1.78396  1.75965 -37.67432  2.50460
[133] 1.91211  0.46994  1.59478  1.18897
[137] -0.00236 -0.66131  2.32946  3.19225
[141] 2.26530  1.89579  0.55433  0.43050
[145] 1.86813  4.01838  2.31643  2.47239
[149] 3.12442  0.26916  0.01663  5.94635
[153] 2.26067  1.56508  2.06938  1.58670
[157] 0.79081  0.29920  0.05866  2.26101
```

```

[161] 0.46816 4.17122 0.13300 -0.01265
[165] 0.93556 0.68832 1.06869 -0.30285
[169] 0.72294 -1.50857 -0.01970 0.38568
[173] 3.33408 0.34210 5.44626 3.92119
[177] 0.37644 3.88533 1.22041 3.95166
[181] 1.29319 1.14795 1.40614 1.22761
[185] 1.15457 2.05471 3.55821 0.30833
[189] 2.78817 0.41222 0.45663 0.12797
[193] 1.03218 0.87484 -4.20018 4.36128
[197] 0.82722 2.25870 0.77412 1.84493
> AZS$z1<-z1s
> plot(z1s)

> t1=length(z1s)
> t1
[1] 200

> #Aziende classificate correttamente
> t1s=z1s[z1s>2.90]
> s1=length(t1s)
> c1=s1/t1
> c1
[1] 0.17
> #Aziende classificate in maniera errata
> t1b=z1s[z1s<1.23]
> ns1=length(t1b)
> w1=ns1/t1
> w1
[1] 0.495
> #Aziende nella zona di ignoranza
> u1=1-w1-c1
> u1
[1] 0.335

```

## Appendice 2.1.2

```

> #Dati imprese fallite non quotate
> IF17 <- read_excel("Dati/Imprese fallite/IF17.xlsx")
> View(IF17)
> str(IF17)
Classes 'tbl_df', 'tbl' and 'data.frame':    200 obs. of  5 variables:
 $ X1
 %
2017: num  -3 22 -46 -13 -16 -10 -259 39 -97 -55 ...
 $ X2
 %
2017: num  0 0 -31 0 -1 -19 7 0 -197 2 ...
 $ X3
 %
2017: num  -2 3 -1 -24 3 -2 -295 -8 -1 -25 ...
 $ X4
 %
2017: num  0 3 -16 -19 3 5 -78 20 -65 -15 ...
 $ X5
 %
2017: num  0 168 0 89 44 0 30 0 0 374 ...
> AZF=IF17
> dim(AZF)
[1] 200 5
> #calcolo dello Z'-score
> z1F=(0.717*AZF[,1]/100)+(0.847*AZF[,2]/100)+(3.107*AZF[,3]/100)+(0.420*AZF[,4]/100)+(0.998*AZF[,5]/100)
> z1F=as.numeric(unlist(z1F))
> write.table(z1F,file="mydataF.csv",quote=F,sep=";",dec=".",na="NA",row.names=T,col.names=T)
> print(z1F)
 [1] -0.08365 1.94019 -0.69066 -0.03047
 [5] 0.42174 -0.27377 -10.99159 0.11507
 [9] -2.66815 2.51536 -1.59606 3.40055
[13] -0.48271 0.76443 0.28144 -1.28568
[17] -0.61813 0.71594 -6058.95744 -7.76650
[21] -0.37190 1.92247 0.17589 -0.60843
[25] 1.35841 1.53194 -8.78087 -1.75924
[29] -7.54011 0.82625 0.89439 1.23667
[33] 1.01163 0.64093 0.60413 -5.47939
[37] 1.67481 -1.38465 0.30138 -0.38161
[41] 1.23551 0.28260 -2.32610 0.61699
[45] 0.58379 0.30093 0.98039 -1.08348
[49] -0.09321 -0.00665 0.00701 9.57735
[53] -11.64356 0.95660 -11.08854 -1.58455
[57] -0.22329 1.11085 2.59519 0.18454

```

```

[61] -1.16546 1.30264 0.34795 2.01828
[65] 0.44461 -0.80936 0.46018 -8.54970
[69] -1.12910 0.94722 -20.66170 -5.17901
[73] -6.44709 -15.30640 -43.02230 -1.35571
[77] -4.75679 -4.11949 -0.93581 -2.30024
[81] 0.59437 0.93278 0.22383 -3.09515
[85] -0.03259 -2.44676 1.16019 -0.24998
[89] -2.54429 -6.94091 1.46228 -0.19174
[93] -7.33979 -0.45120 1.43102 -0.91900
[97] -0.48314 1.54541 -1.92229 0.87341
[101] 0.07048 -3.43756 -2.90363 0.96457
[105] 2.43400 -0.35300 0.50203 0.37065
[109] 2.80043 1.23903 3.99547 0.85163
[113] 1.26734 -0.56378 -0.81877 0.50122
[117] 0.17895 0.65563 0.27532 -1.85523
[121] 0.87210 -0.17611 1.21730 -2216.09266
[125] -2.86126 -2.79546 -34.28633 3.38788
[129] -0.63946 -3.39513 -9.91755 1.12559
[133] 1.02214 2.26694 -1.40565 0.38899
[137] -0.17567 -1.14957 -2.40653 -2.34751
[141] 0.60261 -7.77805 -0.00192 1.08068
[145] 0.56938 0.58768 0.13979 0.47285
[149] -0.06626 3.24516 1.66438 2.01878
[153] 0.29031 -4.71754 -1.48175 -2.65294
[157] 0.85463 -0.38254 0.44297 -5.74098
[161] -0.40961 0.69390 -2.06251 3.25895
[165] -0.72362 2.12643 -0.17021 0.28872
[169] 1.44059 1.30732 0.81440 -52.74197
[173] -9.80487 -0.68049 -1.87098 -1.28644
[177] 0.44455 -28.06985 -0.01434 -2.24568
[181] -0.51575 0.37668 -0.33477 0.69673
[185] 0.88208 0.27195 -5.42500 0.83564
[189] 0.18000 -0.30476 -1.58782 -0.22966
[193] -8.05433 -1.23446 1.02739 0.64521
[197] -2.61642 -0.22701 1.50648 1.07391

```

```

> AZF$z1<-Z1F
plot(Z1F)
plot(Z1F,ylim=c(-10,10))

> t1F=length(Z1F)
> t1F
[1] 200

> #Aziende classificate erratamente
> t1ss=Z1F[Z1F>2.90]
> n1f=length(t1ss)
> w1f=n1f/t1F
> w1f
[1] 0.03
> #Aziende classificate in maniera corretta
> t1f=Z1F[Z1F<1.23]
> f1f=length(t1f)
> c1f=f1f/t1F
> c1f
[1] 0.845
> #Aziende nella zona di ignoranza
> u1f=1-w1f-c1f
> u1f
[1] 0.125

```

## Appendice 2.1.3

```

> sz1=c(seq(1.10,3,0.05))
> sz1
[1] 1.10 1.15 1.20 1.25 1.30 1.35 1.40 1.45 1.50 1.55 1.60 1.65
[13] 1.70 1.75 1.80 1.85 1.90 1.95 2.00 2.05 2.10 2.15 2.20 2.25
[25] 2.30 2.35 2.40 2.45 2.50 2.55 2.60 2.65 2.70 2.75 2.80 2.85
[37] 2.90 2.95 3.00
> length(sz1)
[1] 39
> scz1<-c(length(Z1S[Z1S>1.10]), length(Z1S[Z1S>1.15]), length(Z1S[Z1S>1.20]),
+ length(Z1S[Z1S>1.25]), length(Z1S[Z1S>1.30]), length(Z1S[Z1S>1.35]),
+ length(Z1S[Z1S>1.40]), length(Z1S[Z1S>1.45]), length(Z1S[Z1S>1.50]),
+ length(Z1S[Z1S>1.55]), length(Z1S[Z1S>1.60]), length(Z1S[Z1S>1.65]),
+ length(Z1S[Z1S>1.70]), length(Z1S[Z1S>1.75]), length(Z1S[Z1S>1.80]),
+ length(Z1S[Z1S>1.85]), length(Z1S[Z1S>1.90]), length(Z1S[Z1S>1.95]),
+ length(Z1S[Z1S>2]), length(Z1S[Z1S>2.05]), length(Z1S[Z1S>2.10]),
+ length(Z1S[Z1S>2.15]), length(Z1S[Z1S>2.20]), length(Z1S[Z1S>2.25]),
+ length(Z1S[Z1S>2.30]), length(Z1S[Z1S>2.35]), length(Z1S[Z1S>2.40]),
+ length(Z1S[Z1S>2.45]), length(Z1S[Z1S>2.50]), length(Z1S[Z1S>2.55]),

```

```

+ length(Z1S[Z1S>2.60]),length(Z1S[Z1S>2.65]),length(Z1S[Z1S>2.70]),
+ length(Z1S[Z1S>2.75]),length(Z1S[Z1S>2.80]),length(Z1S[Z1S>2.85]),
+ length(Z1S[Z1S>2.90]),length(Z1S[Z1S>2.95]),length(Z1S[Z1S>3]))
> scZ1
[1] 108 105 103 101 96 95 94 92 88 87 81 78 76 75 71
[16] 68 66 65 64 64 62 61 60 57 52 46 46 45 43 41
[31] 41 40 40 39 35 35 34 32 32
> length(scZ1)
[1] 39
> plot(sz1,scZ1,type="l")
> ssZ1=200-scZ1
> ssZ1
[1] 92 95 97 99 104 105 106 108 112 113 119 122 124 125 129
[16] 132 134 135 136 136 138 139 140 143 148 154 154 155 157 159
[31] 159 160 160 161 165 165 166 168 168
> fcZ1<-c(length(Z1F[Z1F<1.10]),length(Z1F[Z1F<1.15]),length(Z1F[Z1F<1.20]),
+ length(Z1F[Z1F<1.25]),length(Z1F[Z1F<1.30]),length(Z1F[Z1F<1.35]),
+ length(Z1F[Z1F<1.40]),length(Z1F[Z1F<1.45]),length(Z1F[Z1F<1.50]),
+ length(Z1F[Z1F<1.55]),length(Z1F[Z1F<1.60]),length(Z1F[Z1F<1.65]),
+ length(Z1F[Z1F<1.70]),length(Z1F[Z1F<1.75]),length(Z1F[Z1F<1.80]),
+ length(Z1F[Z1F<1.85]),length(Z1F[Z1F<1.90]),length(Z1F[Z1F<1.95]),
+ length(Z1F[Z1F<2.00]),length(Z1F[Z1F<2.05]),length(Z1F[Z1F<2.10]),
+ length(Z1F[Z1F<2.15]),length(Z1F[Z1F<2.20]),length(Z1F[Z1F<2.25]),
+ length(Z1F[Z1F<2.30]),length(Z1F[Z1F<2.35]),length(Z1F[Z1F<2.40]),
+ length(Z1F[Z1F<2.45]),length(Z1F[Z1F<2.50]),length(Z1F[Z1F<2.55]),
+ length(Z1F[Z1F<2.60]),length(Z1F[Z1F<2.65]),length(Z1F[Z1F<2.70]),
+ length(Z1F[Z1F<2.75]),length(Z1F[Z1F<2.80]),length(Z1F[Z1F<2.85]),
+ length(Z1F[Z1F<2.90]),length(Z1F[Z1F<2.95]),length(Z1F[Z1F<3]))
> fcZ1
[1] 165 167 168 172 173 175 176 178 179 182 182 182 184 184 184
[16] 184 184 186 186 188 188 189 189 189 190 190 190 191 191 192
[31] 193 193 193 193 193 194 194 194 194
> length(fcZ1)
[1] 39
> fsZ1=200-fcZ1
> fsZ1
[1] 35 33 32 28 27 25 24 22 21 18 18 18 16 16 16 16 16 14 14 12
[21] 12 11 11 11 10 10 10 9 9 8 7 7 7 7 7 6 6 6 6 6
> (plot(sz1,fcZ1,type="l",ylim = c(0, max(200)),
+ xlab="Soglie", ylab="N. imprese")
+ +lines(sz1,scZ1, type="l", col="blue")
+ +lines(sz1,fsZ1, type="l", lty="dashed")
+ +lines(sz1,ssZ1, type="l", lty="dashed", col="blue")
+ + abline(v=1.23,col="red")
+ + abline(v=2.90,col="red"))
integer(0)
> SCZ1=scZ1/(200*2)
> FCZ1=fcZ1/(200*2)
> CTZ1=SCZ1+FCZ1
> SSZ1=ssZ1/(200*2)
> FSZ1=fsZ1/(200*2)
> STZ1=SSZ1+FSZ1
> (plot(sz1,CTZ1,type="l",ylim = c(0, max(1) ))
+ +lines(sz1,STZ1, type="l", col="blue")
+ + abline(v=1.23,col="red")
+ + abline(v=2.90,col="red"))
integer(0)
> (plot(sz1,STZ1,type="l",ylim = c(0, max(1) ),
+ xlab="Soglie", ylab="% Errore" )
+ + abline(v=1.23,col="red")
+ + abline(v=2.90,col="red"))
integer(0)
> SCZ1/SSZ1
[1] 1.1739130 1.1052632 1.0618557 1.0202020 0.9230769 0.9047619
[7] 0.8867925 0.8518519 0.7857143 0.7699115 0.6806723 0.6393443
[13] 0.6129032 0.6000000 0.5503876 0.5151515 0.4925373 0.4814815
[19] 0.4705882 0.4705882 0.4492754 0.4388489 0.4285714 0.3986014
[25] 0.3513514 0.2987013 0.2987013 0.2903226 0.2738854 0.2578616
[31] 0.2578616 0.2500000 0.2500000 0.2422360 0.2121212 0.2121212
[37] 0.2048193 0.1904762 0.1904762
> FCZ1/FSZ1
[1] 4.714286 5.060606 5.250000 6.142857 6.407407 7.000000
[7] 7.333333 8.090909 8.523810 10.111111 10.111111 10.111111
[13] 11.500000 11.500000 11.500000 11.500000 11.500000 13.285714
[19] 13.285714 15.666667 15.666667 17.181818 17.181818 17.181818
[25] 19.000000 19.000000 19.000000 21.222222 21.222222 24.000000
[31] 27.571429 27.571429 27.571429 27.571429 27.571429 32.333333
[37] 32.333333 32.333333 32.333333

```



## Appendice 2.2.1

```

> #Calcolo dello Z"-score
> Z2=(6.56*AZS[,1]/100)+(3.26*AZS[,2]/100)+(6.72*AZS[,3]/100)+(1.05*AZS[,4]/100)
> Z2S=as.numeric(unlist(Z2))
> AZS$Z2<-Z2S
> plot(Z2S)

> t2=length(Z2S)
> t2
[1] 200
> #Aziende classificate correttamente
> t2s=Z2S[Z2S>2.60]
> s2=length(t2s)
> c2=s2/t2
> c2
[1] 0.445
> #Aziende classificate in maniera errata
> t2b=Z2S[Z2S<1.10]
> ns2=length(t2b)
> w2=ns2/t2
> w2
[1] 0.345
> #Aziende nella zona di ignoranza
> u2=1-w2-c2
> u2
[1] 0.21

```

## Appendice 2.2.2

```

> #Dati imprese fallite non quotate
> IF17 <- read_excel("C:/Users/Utente/Desktop/Progetto per tesi/Dati/Imprese fallite/IF17.xlsx")
> view(IF17)
> str(IF17)
Classes 'tbl_df', 'tbl' and 'data.frame':      200 obs. of  5 variables:
 $ X1
 %
2017: num  -3 22 -46 -13 -16 -10 -259 39 -97 -55 ...
 $ X2
 %
2017: num  0 0 -31 0 -1 -19 7 0 -197 2 ...
 $ X3
 %
2017: num  -2 3 -1 -24 3 -2 -295 -8 -1 -25 ...
 $ X4
 %
2017: num  0 3 -16 -19 3 5 -78 20 -65 -15 ...
 $ X5
 %
2017: num  0 168 0 89 44 0 30 0 0 374 ...
> AZF=IF17
> dim(AZF)
[1] 200 5

> #Calcolo dello Z"-score
> Z2F=(6.56*AZS[,1]/100)+(3.26*AZS[,2]/100)+(6.72*AZS[,3]/100)+(1.05*AZS[,4]/100)
> Z2F=as.numeric(unlist(Z2F))
> AZF$Z2<-Z2F
> plot(Z2F)

> t2F=length(Z2F)
> t2F
[1] 200

> #Aziende classificate in maniera corretta
> t2f=Z2F[Z2F<1.10]
> f2f=length(t2f)
> c2f=f2f/t2F
> c2f
[1] 0.72
> #Aziende nella zona di ignoranza
> u2f=1-w2f-c2f
> u2f
[1] 0.14
> #Aziende classificate erratamente
> t2ss=Z2F[Z2F>2.60]
> nf2=length(t2ss)

```

```
> w2f=nf2/t2F
> w2f
[1] 0.14
```

## Appendice 2.2.3

```
> sz2=c(seq(4,6,0.05))
> sz2
[1] 4.00 4.05 4.10 4.15 4.20 4.25 4.30 4.35 4.40 4.45 4.50 4.55
[13] 4.60 4.65 4.70 4.75 4.80 4.85 4.90 4.95 5.00 5.05 5.10 5.15
[25] 5.20 5.25 5.30 5.35 5.40 5.45 5.50 5.55 5.60 5.65 5.70 5.75
[37] 5.80 5.85 5.90 5.95 6.00
> length(sz2)
[1] 41
> scz2<-c(length(Z2S[Z2S>4.00]),length(Z2S[Z2S>4.05]),length(Z2S[Z2S>4.10]),
+ length(Z2S[Z2S>4.15]),length(Z2S[Z2S>4.20]),length(Z2S[Z2S>4.25]),
+ length(Z2S[Z2S>4.30]),length(Z2S[Z2S>4.35]),length(Z2S[Z2S>4.40]),
+ length(Z2S[Z2S>4.45]),length(Z2S[Z2S>4.50]),length(Z2S[Z2S>4.55]),
+ length(Z2S[Z2S>4.60]),length(Z2S[Z2S>4.65]),length(Z2S[Z2S>4.70]),
+ length(Z2S[Z2S>4.75]),length(Z2S[Z2S>4.80]),length(Z2S[Z2S>4.85]),
+ length(Z2S[Z2S>4.90]),length(Z2S[Z2S>4.95]),length(Z2S[Z2S>5.00]),
+ length(Z2S[Z2S>5.05]),length(Z2S[Z2S>5.10]),length(Z2S[Z2S>5.15]),
+ length(Z2S[Z2S>5.20]),length(Z2S[Z2S>5.25]),length(Z2S[Z2S>5.30]),
+ length(Z2S[Z2S>5.35]),length(Z2S[Z2S>5.40]),length(Z2S[Z2S>5.45]),
+ length(Z2S[Z2S>5.50]),length(Z2S[Z2S>5.55]),length(Z2S[Z2S>5.60]),
+ length(Z2S[Z2S>5.65]),length(Z2S[Z2S>5.70]),length(Z2S[Z2S>5.75]),
+ length(Z2S[Z2S>5.80]),length(Z2S[Z2S>5.85]),length(Z2S[Z2S>5.90]),
+ length(Z2S[Z2S>5.95]),length(Z2S[Z2S>6.00]))
> scz2
[1] 134 133 133 131 130 126 125 125 125 125 124 123 122 120 119
[16] 117 117 114 113 112 112 112 110 110 106 101 100 100 100 100
[31] 98 97 95 93 91 90 90 89 87 86 86
> length(scz2)
[1] 41
> plot(sz2,scz2,type="l")
> ssz2=200-scz2
> ssz2
[1] 66 67 67 69 70 74 75 75 75 75 76 77 78 80 81
[16] 83 83 86 87 88 88 88 90 90 94 99 100 100 100 100
[31] 102 103 105 107 109 110 110 111 113 114 114
> fcz2<-c(length(Z2F[Z2F<4.00]),length(Z2F[Z2F<4.05]),length(Z2F[Z2F<4.10]),
+ length(Z2F[Z2F<4.15]),length(Z2F[Z2F<4.20]),length(Z2F[Z2F<4.25]),
+ length(Z2F[Z2F<4.30]),length(Z2F[Z2F<4.35]),length(Z2F[Z2F<4.40]),
+ length(Z2F[Z2F<4.45]),length(Z2F[Z2F<4.50]),length(Z2F[Z2F<4.55]),
+ length(Z2F[Z2F<4.60]),length(Z2F[Z2F<4.65]),length(Z2F[Z2F<4.70]),
+ length(Z2F[Z2F<4.75]),length(Z2F[Z2F<4.80]),length(Z2F[Z2F<4.85]),
+ length(Z2F[Z2F<4.90]),length(Z2F[Z2F<4.95]),length(Z2F[Z2F<5.00]),
+ length(Z2F[Z2F<5.05]),length(Z2F[Z2F<5.10]),length(Z2F[Z2F<5.15]),
+ length(Z2F[Z2F<5.20]),length(Z2F[Z2F<5.25]),length(Z2F[Z2F<5.30]),
+ length(Z2F[Z2F<5.35]),length(Z2F[Z2F<5.40]),length(Z2F[Z2F<5.45]),
+ length(Z2F[Z2F<5.50]),length(Z2F[Z2F<5.55]),length(Z2F[Z2F<5.60]),
+ length(Z2F[Z2F<5.65]),length(Z2F[Z2F<5.70]),length(Z2F[Z2F<5.75]),
+ length(Z2F[Z2F<5.80]),length(Z2F[Z2F<5.85]),length(Z2F[Z2F<5.90]),
+ length(Z2F[Z2F<5.95]),length(Z2F[Z2F<6.00]))
> fcz2
[1] 139 142 142 144 144 146 146 147 147 148 150 153 154 155 156
[16] 156 158 159 160 162 162 164 166 167 168 168 168 168 169 169
[31] 170 170 171 172 172 172 172 172 172 172 172
> length(fcz2)
[1] 41
> fsz2=200-fcz2
> fsz2
[1] 61 58 58 56 56 54 54 53 53 52 50 47 46 45 44 44 42 41 40 38
[21] 38 36 34 33 32 32 32 32 31 31 30 30 29 28 28 28 28 28 28 28
[41] 28
> (plot(sz2,fcz2,type="l",ylim = c(0, max(200) ),
+ xlab="Soglie", ylab="N. imprese")
+ +lines(sz2,scz2, type="l", col="blue")
+ +lines(sz2,fsz2, type="l", lty="dashed")
+ +lines(sz2,ssz2, type="l", lty="dashed",col="blue")
+ + abline(v=4.15,col="red")
+ + abline(v=5.83,col="red"))
integer(0)
> SCZ2=scz2/(200*2)
> FCZ2=fcz2/(200*2)
> CTZ2=SCZ2+FCZ2
> SSZ2=ssz2/(200*2)
> FSZ2=fsz2/(200*2)
> STZ2=SSZ2+FSZ2
> (plot(ssz2,CTZ2,type="l",ylim = c(0, max(1) ))
```

```

+ +lines(sz2,STZ2, type="l", col="blue")
+ + abline(v=4.15,col="red")
+ + abline(v=5.83,col="red")
integer(0)
> plot(sz2,STZ2,type="l",ylim = c(0, max(1) ),
+ xlab="Soglie", ylab="% Errore" )
+ + abline(v=4.15,col="red")
+ + abline(v=5.83,col="red")
integer(0)

```

## Appendice 2.3.1

```

> library(readxl)
> I117 <- read_excel("Dati/Campioni 50-50 completo 2016/Campione definitivo.xlsx",sheet = "Training set")
> str(I117)
Classes 'tbl_df', 'tbl' and 'data.frame':    400 obs. of  6 variables:
 $ x1
 %
2017: num  25 -11 52 20 -1 69 35 -157 28 45 ...
 $ x2
 %
2017: num  0 2 0 0 0 27 1 -749 13 83 ...
 $ x3
 %
2017: num  4 0 -4 -9 15 27 5 -3 13 4 ...
 $ x4
 %
2017: num  245 15 7 79 42 ...
 $ x5
 %
2017: num  93 232 68 57 83 40 166 0 65 111 ...
 $ Output
      : num  100 100 0 0 100 100 100 0 100 100 ...
> dim(I117)
[1] 400  6
> #Normalizzazione delle variabili
> normalize1<-function(I117){return((I117-min(I117))/(max(I117)-min(I117)))}
> I117_norm<-as.data.frame(lapply(I117, normalize1))
> summary(I117_norm)
 x1.....2017      x2.....2017      x3.....2017      x4.....2017
Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
1st Qu.:0.9994   1st Qu.:0.9998   1st Qu.:0.9983   1st Qu.:0.03592
Median :0.9995   Median :0.9998   Median :0.9984   Median :0.03779
Mean   :0.9961   Mean   :0.9972   Mean   :0.9957   Mean   :0.05075
3rd Qu.:0.9997   3rd Qu.:0.9998   3rd Qu.:0.9985   3rd Qu.:0.04129
Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
 x5.....2017      Output
Min.   :0.00000   Min.   :0.0
1st Qu.:0.00513   1st Qu.:0.0
Median :0.06224   Median :0.5
Mean   :0.10094   Mean   :0.5
3rd Qu.:0.15458   3rd Qu.:1.0
Max.   :1.00000   Max.   :1.0
> I117_train<-I117_norm[1:200,]
> show(I117_train)
> colnames(I117_train)<- c("x1","x2","x3","x4","x5","output")
> colnames(I117_train)
[1] "x1" "x2" "x3" "x4" "x5" "output"
> I117_test<-I117_norm[201:400,]
> show(I117_test)
> dim(I117_test)
[1] 200  6
> colnames(I117_test)<- c("x1","x2","x3","x4","x5","output")
> colnames(I117_test)
[1] "x1" "x2" "x3" "x4" "x5" "output"

```

## Appendice 2.3.2

```

> # Modello 2 con backpropagation classica
> sigmoid<- function(I117_train) {1.0 / (1.0 + exp(-I117_train))}
> library(neuralnet)
> I117_model22<-neuralnet(output~x1+x2+x3+x4+x5,
+                          hidden=2,data=I117_train,learningrate = 0.2,
+                          learningrate.factor = list(minus = 0.1,plus = 1.0),

```

```

+           learningrate.limit=list(minus=0.1,plus=1.0),
+           threshold =0.05,stepmax=10000000,algorithm="backprop",
+           act.fct=sigmoid,linear.output=FALSE,likelihood=TRUE)
> plot(I117_model22)

> I117_model22
$weights
$weights[[1]]
$weights[[1]][[1]]
      [,1]      [,2]
[1,]  59.225812  56.85273
[2,]  -7.523483 -10.98591
[3,]   7.120334  17.77836
[4,] -36.888960 -32.66397
[5,] -665.718393 -725.69434
[6,]  41.305344 -25.37689

$weights[[1]][[2]]
      [,1]
[1,]  4.476305
[2,] -3.862208
[3,] -4.057195

$result.matrix
      [,1]
error          1.600730e+01
reached.threshold 3.856678e-02
steps          9.972570e+05
aic            6.201460e+01
bic            1.114894e+02
Intercept.to.l1ayhid1 5.922581e+01
X1.to.l1ayhid1 -7.523483e+00
X2.to.l1ayhid1  7.120334e+00
X3.to.l1ayhid1 -3.688896e+01
X4.to.l1ayhid1 -6.657184e+02
X5.to.l1ayhid1  4.130534e+01
Intercept.to.l1ayhid2 5.685273e+01
X1.to.l1ayhid2 -1.098591e+01
X2.to.l1ayhid2  1.777836e+01
X3.to.l1ayhid2 -3.266397e+01
X4.to.l1ayhid2 -7.256943e+02
X5.to.l1ayhid2 -2.537689e+01
Intercept.to.output  4.476305e+00
l1ayhid1.to.output -3.862208e+00
l1ayhid2.to.output -4.057195e+00

> #Valutazione dell'attendibilità del modello 2
> I117_risultati22<-compute(I117_model22,I117_test)
> I117_risultati22

> SI22=I117_risultati22$net.result[I117_risultati22$net.result>0.5]
> length(SI22)
[1] 122
> prev_attend22<-I117_risultati22$net.result
> cor(prev_attend22,I117_test$output)
      [,1]
[1,] 0.505004

```

### Appendice 2.3.3

```

> # Modello 3 con backpropagation classica
> sigmoid<- function(I117_train) {1.0 / (1.0 + exp(-I117_train))}
> I117_model23<-neuralnet(output~X1+X2+X3+X4+X5,
+ hidden=3,data=I117_train,learningrate = 0.2,
+ learningrate.factor = list(minus = 0.1,plus = 1.0),
+ learningrate.limit=list(minus=0.1,plus=1.0),
+ threshold =0.05,stepmax=10000000,algorithm="backprop",
+ act.fct=sigmoid,linear.output=FALSE,likelihood=TRUE)
> plot(I117_model23)

> I117_model23
$weights
$weights[[1]]
$weights[[1]][[1]]
      [,1]      [,2]      [,3]
[1,]  67.38671  47.995571 -6.627423
[2,] -15.12848  -7.816131 -7.160352
[3,]  21.15964   7.769468 -5.211451
[4,] -45.58065  -23.017401 -6.045414
[5,] -641.96771 -739.943864 -4.910949
[6,] -24.52782  40.578620  25.568237

```

```

$weights[[1]][[2]]
      [,1]
[1,]  4.551112
[2,] -4.019979
[3,] -3.784709
[4,] -12.279149

$result.matrix
      [,1]
error          1.580370e+01
reached.threshold 3.264566e-02
steps          9.872420e+05
aic            7.560740e+01
bic           1.481704e+02
Intercept.to.1layhid1 6.738671e+01
X1.to.1layhid1 -1.512848e+01
X2.to.1layhid1  2.115964e+01
X3.to.1layhid1 -4.558065e+01
X4.to.1layhid1 -6.419677e+02
X5.to.1layhid1 -2.452782e+01
Intercept.to.1layhid2 4.799557e+01
X1.to.1layhid2 -7.816131e+00
X2.to.1layhid2  7.769468e+00
X3.to.1layhid2 -2.301740e+01
X4.to.1layhid2 -7.399439e+02
X5.to.1layhid2  4.057862e+01
Intercept.to.1layhid3 -6.627423e+00
X1.to.1layhid3 -7.160352e+00
X2.to.1layhid3 -5.211451e+00
X3.to.1layhid3 -6.045414e+00
X4.to.1layhid3 -4.910949e+00
X5.to.1layhid3  2.556824e+01
Intercept.to.output  4.551112e+00
1layhid1.to.output -4.019979e+00
1layhid2.to.output -3.784709e+00
1layhid3.to.output -1.227915e+01

> #valutazione dell'attendibilità del modello 3
> I117_risultati23<-compute(I117_model23,I117_test)
> I117_risultati23

> SI23=I117_risultati23$net.result[I117_risultati23$net.result>0.5]
> length(SI23)
[1] 123
> prev_attend23<-I117_risultati23$net.result
> #ora si deve misurare la correlazione tra il valore previsto
> #e quello reale"
> cor(prev_attend23,I117_test$output)
      [,1]
[1,] 0.5111887

```

## Appendice 2.3.4

```

> #Modello 2
> #Imprese solventi classificate correttamente
> I117_risultatiS22<-compute(I117_model22,IS117)
> SIS22=I117_risultatiS22$net.result[I117_risultatiS22$net.result>0.5]
> length(SIS22)
[1] 83
> #Imprese non solventi classificate correttamente
> I117_risultatiF22<-compute(I117_model22,IF117)
> SIF22=I117_risultatiF22$net.result[I117_risultatiF22$net.result<0.5]
> length(SIF22)
[1] 61

> #Modello 3
> #Imprese solventi classificate correttamente
> I117_risultatiS23<-compute(I117_model23,IS117)
> SIS23=I117_risultatiS23$net.result[I117_risultatiS23$net.result>0.5]
> length(SIS23)
[1] 83
> #Imprese non solventi classificate correttamente
> I117_risultatiF23<-compute(I117_model23,IF117)
> SIF23=I117_risultatiF23$net.result[I117_risultatiF23$net.result<0.5]
> length(SIF23)
[1] 60

```

## Appendice 2.4.1

```

> library(readxl)
> I1617<-Campione_50_50_casuale <- read_excel("Dati/Campioni 50-50 completo 2017 2016/Campione
definitivo.xlsx")
> View(I1617)
> #Normalizzazione delle variabili
> normalize1617<-function(I1617){return((I1617-min(I1617))/(max(I1617)-min(I1617)))}
> I1617_norm<-as.data.frame(lapply(I1617, normalize1617))
> summary(I1617_norm)
  X1.....2017      X1.....2016      X2.....2017
Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
1st Qu.:0.9994    1st Qu.:0.9996    1st Qu.:0.9998
Median :0.9995    Median :0.9997    Median :0.9998
Mean   :0.9961    Mean   :0.9971    Mean   :0.9972
3rd Qu.:0.9997    3rd Qu.:0.9998    3rd Qu.:0.9998
Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
  X2.....2016      X3.....2017      X3.....2016
Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
1st Qu.:0.9999    1st Qu.:0.9983    1st Qu.:0.9581
Median :0.9999    Median :0.9984    Median :0.9604
Mean   :0.9974    Mean   :0.9957    Mean   :0.9526
3rd Qu.:0.9999    3rd Qu.:0.9985    3rd Qu.:0.9622
Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
  X4.....2017      X4.....2016      X5.....2017
Min.   :0.00000    Min.   :0.000000    Min.   :0.000000
1st Qu.:0.03594    1st Qu.:0.005242    1st Qu.:0.004987
Median :0.03779    Median :0.005371    Median :0.062010
Mean   :0.05076    Mean   :0.009110    Mean   :0.100950
3rd Qu.:0.04125    3rd Qu.:0.005829    3rd Qu.:0.154381
Max.   :1.00000    Max.   :1.000000    Max.   :1.000000
  X5.....2016      Output
Min.   :0.000000    Min.   :0.0
1st Qu.:0.005051    1st Qu.:0.0
Median :0.063047    Median :0.5
Mean   :0.085422    Mean   :0.5
3rd Qu.:0.135376    3rd Qu.:1.0
Max.   :1.000000    Max.   :1.0
> I1617_train<-I1617_norm[1:200,]
> show(I1617_train)
> dim(I1617_train)
[1] 200 11
> colnames(I1617_train)<- c("x117","x116","x217","x216","x317","x316","x417","x416","x517","x5
16","output")
> colnames(I1617_train)
 [1] "x117" "x116" "x217" "x216" "x317" "x316"
 [7] "x417" "x416" "x517" "x516" "output"
> I1617_test<-I1617_norm[201:400,]
> show(I1617_test)
> dim(I1617_test)
[1] 200 11
> colnames(I1617_test)<- c("x117","x116","x217","x216","x317","x316","x417","x416","x517","x51
6","output")
> colnames(I1617_test)
 [1] "x117" "x116" "x217" "x216" "x317" "x316"
 [7] "x417" "x416" "x517" "x516" "output"

```

## Appendice 2.4.2

```

> # Modello 2 con backpropagation classica
> sigmoid<- function(I1617_train) {1.0 / (1.0 + exp(-I1617_train))}
> library(neuralnet)
> I1617_model2<-neuralnet(output~x117+x116+x217+x216+x317+x316+x417+x416+x517+x516,
+ data=I1617_train,hidden=2,learningrate = 0.01,
+ learningrate.factor = list(minus = 0.1,plus = 1.0),
+ learningrate.limit=list(minus=0.1,plus=1.0),
+ threshold =0.05,stepmax=200000000,algorithm="backprop",
+ act.fct=sigmoid,linear.output=FALSE,likelihood=TRUE)
> plot(I1617_model2)
> I1617_model2
$weights
$weights[[1]]
$weights[[1]][[1]]
 [1,] 40.960953 -20.526305

```

```

[2,] 44.728029 -34.915464
[3,] 27.278825 -1.080879
[4,] -3.658360 14.719201
[5,] -15.726856 22.469486
[6,] -21.443884 13.319239
[7,] -71.122837 -3.777040
[8,] -160.005770 143.696280
[9,] 22.634289 -12.286415
[10,] -2.127908 4.992410
[11,] 15.935359 6.353479

$weights[[1]][[2]]
      [,1]
[1,] 0.3134277
[2,] -8.1241217
[3,] 36.1597578

$result.matrix
      [,1]
error 1.513589e+01
reached.threshold 4.999986e-02
steps 9.094827e+06
aic 8.027178e+01
bic 1.627297e+02
Intercept.to.l1ayhid1 4.096095e+01
x117.to.l1ayhid1 4.472803e+01
x116.to.l1ayhid1 2.727883e+01
x217.to.l1ayhid1 -3.658360e+00
x216.to.l1ayhid1 -1.572686e+01
x317.to.l1ayhid1 -2.144388e+01
x316.to.l1ayhid1 -7.112284e+01
x417.to.l1ayhid1 -1.600058e+02
x416.to.l1ayhid1 2.263429e+01
x517.to.l1ayhid1 -2.127908e+00
x516.to.l1ayhid1 1.593536e+01
Intercept.to.l1ayhid2 -2.052631e+01
x117.to.l1ayhid2 -3.491546e+01
x116.to.l1ayhid2 -1.080879e+00
x217.to.l1ayhid2 1.471920e+01
x216.to.l1ayhid2 2.246949e+01
x317.to.l1ayhid2 1.331924e+01
x316.to.l1ayhid2 -3.777040e+00
x417.to.l1ayhid2 1.436963e+02
x416.to.l1ayhid2 -1.228642e+01
x517.to.l1ayhid2 4.992410e+00
x516.to.l1ayhid2 6.353479e+00
Intercept.to.output 3.134277e-01
l1ayhid1.to.output -8.124122e+00
l1ayhid2.to.output 3.615976e+01

> #valutazione dell'attendibilità del modello 2
> I1617_risultati2<-compute(I1617_model2,I1617_test)
> I1617_risultati2
> SI2=I1617_risultati2$net.result[I1617_risultati2$net.result>0.5]
> length(SI2)
[1] 85
> prev_attend2<-I1617_risultati2$net.result
> #ora si deve misurare la correlazione tra il valore previsto
> #e e quello reale"
> cor(prev_attend2,I1617_test$output)
      [,1]
[1,] 0.4757701

```

### Appendice 2.4.3

```

> # Modello 3 con backpropagation classica
> I1617_model3<-neuralnet(output~X117+X116+X217+X216+X317+X316+X417+X416+X517+X516,
+ data=I1617_train,hidden=3,learningrate = 0.1,
+ learningrate.factor = list(minus = 0.1,plus = 1.0),
+ learningrate.limit=list(minus=0.1,plus=1.0),
+ threshold =0.05,stepmax=100000000,algorithm="backprop",
+ act.fct=sigmoid,linear.output=FALSE,likelihood=TRUE)

> plot(I1617_model3)
> I1617_model3
$weights
$weights[[1]]
$weights[[1]][[1]]
      [,1]      [,2]      [,3]
[1,] 39.086337 20.374410 -5.939750

```

```

[2,] 112.758130 -34.627767 -30.998080
[3,] 119.575886 -25.850823 -11.288536
[4,] 76.496031 -59.344658 4.966863
[5,] 74.860340 -62.329258 4.302658
[6,] -74.249238 -54.979292 15.751210
[7,] -342.996875 224.632186 -9.692697
[8,] -525.969014 -209.628622 557.985897
[9,] 5.173703 8.718858 89.184309
[10,] -33.436529 28.979406 16.977545
[11,] 48.250961 -14.288605 15.384439

$weights[[1]][[2]]
[,1]
[1,] 1.201324
[2,] -3.256830
[3,] -64.527058
[4,] 60.295678

$result.matrix
[,1]
error 1.383463e+01
reached.threshold 4.760943e-02
steps 2.978034e+06
aic 1.016693e+02
bic 2.237070e+02
Intercept.to.l1ayhid1 3.908634e+01
X117.to.l1ayhid1 1.127581e+02
X116.to.l1ayhid1 1.195759e+02
X217.to.l1ayhid1 7.649603e+01
X216.to.l1ayhid1 7.486034e+01
X317.to.l1ayhid1 -7.424924e+01
X316.to.l1ayhid1 -3.429969e+02
X417.to.l1ayhid1 -5.259690e+02
X416.to.l1ayhid1 5.173703e+00
X517.to.l1ayhid1 -3.343653e+01
X516.to.l1ayhid1 4.825096e+01
Intercept.to.l1ayhid2 2.037441e+01
X117.to.l1ayhid2 -3.462777e+01
X116.to.l1ayhid2 -2.585082e+01
X217.to.l1ayhid2 -5.934466e+01
X216.to.l1ayhid2 -6.232926e+01
X317.to.l1ayhid2 -5.497929e+01
X316.to.l1ayhid2 2.246322e+02
X417.to.l1ayhid2 -2.096286e+02
X416.to.l1ayhid2 8.718858e+00
X517.to.l1ayhid2 2.897941e+01
X516.to.l1ayhid2 -1.428861e+01
Intercept.to.l1ayhid3 -5.939750e+00
X117.to.l1ayhid3 -3.099808e+01
X116.to.l1ayhid3 -1.128854e+01
X217.to.l1ayhid3 4.966863e+00
X216.to.l1ayhid3 4.302658e+00
X317.to.l1ayhid3 1.575121e+01
X316.to.l1ayhid3 -9.692697e+00
X417.to.l1ayhid3 5.579859e+02
X416.to.l1ayhid3 8.918431e+01
X517.to.l1ayhid3 1.697755e+01
X516.to.l1ayhid3 1.538444e+01
Intercept.to.output 1.201324e+00
l1ayhid1.to.output -3.256830e+00
l1ayhid2.to.output -6.452706e+01
l1ayhid3.to.output 6.029568e+01

> #valutazione dell'attendibilità del modello 3
> I1617_risultati3<-compute(I1617_model3,I1617_test)
> I1617_risultati3
> SI3=I1617_risultati3$net.result[I1617_risultati3$net.result>0.5]
> length(SI3)
[1] 83
> prev_attend3<-I1617_risultati3$net.result
> #ora si deve misurare la correlazione tra il valore previsto
> #e e quello reale"
> cor(prev_attend3,I1617_test$output)
[,1]
[1,] 0.4471527

```



## Appendice 2.4.4

```

> # Modello 4 con backpropagation classica
> I1617_model22<-neuralnet(output~X117+X116+X217+X216+X317+X316+X417+X416+X517+X516,
+                           data=I1617_train,hidden=4,learningrate = 0.05,
+                           learningrate.factor = list(minus = 0.1,plus = 1.0),
+                           learningrate.limit=list(minus=0.1,plus=1.0),
+                           threshold =0.05,stepmax=100000000,algorithm="backprop",
+                           act.fct=sigmoid,linear.output=FALSE,likelihood=TRUE)

> plot(I1617_model22)

> I1617_model22
$weights
$weights[[1]]
$weights[[1]][[1]]
      [,1]      [,2]      [,3]
[1,] -17.559831 -45.9981011 -17.867466
[2,] -40.058417 -48.7566485 -1.972861
[3,] -36.897368 -54.8619214 -8.200246
[4,] -29.376733 -44.1455153 -4.709905
[5,] -27.874862 -42.4608458 -7.311745
[6,]  4.947747  26.3776610 13.712962
[7,] 146.515154 202.6177737  1.902714
[8,] 325.163347 292.9459890 59.365769
[9,] 31.209665  0.4274478  3.212771
[10,] -51.263374 60.9272664 49.908687
[11,] -38.621734 -17.3212532  9.833192
      [,4]
[1,]  48.04779
[2,]  10.90182
[3,]  26.34409
[4,]  14.68764
[5,]  15.63916
[6,] -17.13872
[7,] -105.42744
[8,] -262.79684
[9,] -26.44205
[10,] 52.90659
[11,] -28.77120

$weights[[1]][[2]]
      [,1]
[1,] -26.87049
[2,]  26.00548
[3,]  28.18955
[4,]  57.19009
[5,] -49.41199

$result.matrix
      [,1]
error 1.357082e+01
reached.threshold 3.320449e-02
steps 2.618080e+06
aic 1.251416e+02
bic 2.867592e+02
Intercept.to.1layhid1 -1.755983e+01
X117.to.1layhid1 -4.005842e+01
X116.to.1layhid1 -3.689737e+01
X217.to.1layhid1 -2.937673e+01
X216.to.1layhid1 -2.787486e+01
X317.to.1layhid1 4.947747e+00
X316.to.1layhid1 1.465152e+02
X417.to.1layhid1 3.251633e+02
X416.to.1layhid1 3.120966e+01
X517.to.1layhid1 -5.126337e+01
X516.to.1layhid1 -3.862173e+01
Intercept.to.1layhid2 -4.599810e+01
X117.to.1layhid2 -4.875665e+01
X116.to.1layhid2 -5.486192e+01
X217.to.1layhid2 -4.414552e+01
X216.to.1layhid2 -4.246085e+01
X317.to.1layhid2 2.637766e+01
X316.to.1layhid2 2.026178e+02
X417.to.1layhid2 2.929460e+02
X416.to.1layhid2 4.274478e-01
X517.to.1layhid2 6.092727e+01
X516.to.1layhid2 -1.732125e+01
Intercept.to.1layhid3 -1.786747e+01
X117.to.1layhid3 -1.972861e+00
X116.to.1layhid3 -8.200246e+00
X217.to.1layhid3 -4.709905e+00
X216.to.1layhid3 -7.311745e+00
X317.to.1layhid3 1.371296e+01

```

```

X316.to.l1ayhid3      1.902714e+00
X417.to.l1ayhid3      5.936577e+01
X416.to.l1ayhid3      3.212771e+00
X517.to.l1ayhid3      4.990869e+01
X516.to.l1ayhid3      9.833192e+00
Intercept.to.l1ayhid4 4.804779e+01
X117.to.l1ayhid4      1.090182e+01
X116.to.l1ayhid4      2.634409e+01
X217.to.l1ayhid4      1.468764e+01
X216.to.l1ayhid4      1.563916e+01
X317.to.l1ayhid4      -1.713872e+01
X316.to.l1ayhid4      -1.054274e+02
X417.to.l1ayhid4      -2.627968e+02
X416.to.l1ayhid4      -2.644205e+01
X517.to.l1ayhid4      5.290659e+01
X516.to.l1ayhid4      -2.877120e+01
Intercept.to.output   -2.687049e+01
l1ayhid1.to.output     2.600548e+01
l1ayhid2.to.output     2.818955e+01
l1ayhid3.to.output     5.719009e+01
l1ayhid4.to.output    -4.941199e+01

> #Valutazione dell'attendibilità del modello 4
> I1617_risultati4<-compute(I1617_model22,I1617_test)
> SI4=I1617_risultati22$net.result[I1617_risultati4$net.result>0.5]
> length(SI4)
[1] 82
> prev_attend4<-I1617_risultati4$net.result
> cor(prev_attend4,I1617_test$output)
0.4418523

```

## Appendice 2.4.5

```

> # Modello 5 con backpropagation classica
> I1617_model5<-neuralnet(output~X117+X116+X217+X216+X317+X316+X417+X416+X517+X516,
+                          data=I1617_train,hidden=5,learningrate = 0.05,
+                          learningrate.factor = list(minus = 0.1,plus = 1.0),
+                          learningrate.limit=list(minus=0.1,plus=1.0),
+                          threshold =0.05,stepmax=10000000,algorithm="backprop",
+                          act.fct=sigmoid,linear.output=FALSE,likelihood=TRUE)

> plot(I1617_model5)

> I1617_model5
$weights
$weights[[1]]
$weights[[1]][[1]]
      [,1]      [,2]      [,3]      [,4]
[1,] -7.3711960  33.429207 -19.50773  17.155070
[2,]  1.9206560  52.246317 -16.20000  1.219090
[3,] -3.6084498  59.709365 -16.94839  6.751864
[4,] -0.4453755  37.983996 -17.56586  3.699917
[5,] -2.4475247  43.927509 -16.62481  5.974321
[6,]  39.7533606  -9.882713 -12.63654  -3.686585
[7,] -48.1961279 -216.812792  45.14320 -52.133431
[8,]  352.7156998 -183.393441  25.03138 -98.950096
[9,]  28.3900127  3.633983  4.27921  -8.549620
[10,]  19.8757546 -109.995907  102.65400  71.802495
[11,] -0.1847462  88.130158  118.96230  13.914462
      [,5]
[1,]  7.856339
[2,]  22.406237
[3,]  25.239588
[4,]  17.368305
[5,]  17.102114
[6,] -22.433980
[7,] -69.177877
[8,] -144.770657
[9,]  -2.934808
[10,] 118.100698
[11,] -72.157937

$weights[[1]][[2]]
      [,1]
[1,]  9.215927
[2,]  12.596788
[3,] -10.724823
[4,]  67.993153
[5,] -74.998502

```

```
[6,] -10.881821
```

```
$result.matrix
```

```

                                [,1]
error                          1.174344e+01
reached.threshold              4.824468e-02
steps                          1.765091e+06
aic                             1.454869e+02
bic                             3.466842e+02
Intercept.to.l1ayhid1         -7.371196e+00
X117.to.l1ayhid1              1.920656e+00
X116.to.l1ayhid1              -3.608450e+00
X217.to.l1ayhid1              -4.453755e-01
X216.to.l1ayhid1              -2.447525e+00
X317.to.l1ayhid1              3.975336e+01
X316.to.l1ayhid1              -4.819613e+01
X417.to.l1ayhid1              3.527157e+02
X416.to.l1ayhid1              2.839001e+01
X517.to.l1ayhid1              1.987575e+01
X516.to.l1ayhid1              -1.847462e-01
Intercept.to.l1ayhid2         3.342921e+01
X117.to.l1ayhid2              5.224632e+01
X116.to.l1ayhid2              5.970936e+01
X217.to.l1ayhid2              3.798400e+01
X216.to.l1ayhid2              4.392751e+01
X317.to.l1ayhid2              -9.882713e+00
X316.to.l1ayhid2              -2.168128e+02
X417.to.l1ayhid2              -1.833934e+02
X416.to.l1ayhid2              3.633983e+00
X517.to.l1ayhid2              -1.099959e+02
X516.to.l1ayhid2              8.813016e+01
Intercept.to.l1ayhid3        -1.950773e+01
X117.to.l1ayhid3              -1.620000e+01
X116.to.l1ayhid3              -1.694839e+01
X217.to.l1ayhid3              -1.756586e+01
X216.to.l1ayhid3              -1.662481e+01
X317.to.l1ayhid3              -1.263654e+01
X316.to.l1ayhid3              4.514320e+01
X417.to.l1ayhid3              2.503138e+01
X416.to.l1ayhid3              4.279210e+00
X517.to.l1ayhid3              1.026540e+02
X516.to.l1ayhid3              1.189623e+02
Intercept.to.l1ayhid4        1.715507e+01
X117.to.l1ayhid4              1.219090e+00
X116.to.l1ayhid4              6.751864e+00
X217.to.l1ayhid4              3.699917e+00
X216.to.l1ayhid4              5.974321e+00
X317.to.l1ayhid4              -3.686585e+00
X316.to.l1ayhid4              -5.213343e+01
X417.to.l1ayhid4              -9.895010e+01
X416.to.l1ayhid4              -8.549620e+00
X517.to.l1ayhid4              7.180249e+01
X516.to.l1ayhid4              1.391446e+01
Intercept.to.l1ayhid5        7.856339e+00
X117.to.l1ayhid5              2.240624e+01
X116.to.l1ayhid5              2.523959e+01
X217.to.l1ayhid5              1.736831e+01
X216.to.l1ayhid5              1.710211e+01
X317.to.l1ayhid5              -2.243398e+01
X316.to.l1ayhid5              -6.917788e+01
X417.to.l1ayhid5              -1.447707e+02
X416.to.l1ayhid5              -2.934808e+00
X517.to.l1ayhid5              1.181007e+02
X516.to.l1ayhid5              -7.215794e+01
Intercept.to.output           9.215927e+00
l1ayhid1.to.output            1.259679e+01
l1ayhid2.to.output            -1.072482e+01
l1ayhid3.to.output            6.799315e+01
l1ayhid4.to.output            -7.499850e+01
l1ayhid5.to.output            -1.088182e+01

```

```

> #Valutazione dell'attendibilità del modello 5
> I1617_risultati5
> I1617_risultati5<-compute(I1617_model5,I1617_test)
> SI5=I1617_risultati5$net.result[I1617_risultati5$net.result>0.5]
> length(SI5)
[1] 76
> prev_attend5<-I1617_risultati5$net.result
> #ora si deve misurare la correlazione tra il valore previsto
> #e e quello reale"
> cor(prev_attend5,I1617_test$output)
                                [,1]
[1,] 0.4369027

```

## Appendice 2.4.6

```

> #Imprese solventi
> IS1617<-I1617_test[I1617_test$output>0,c("X117","X116","X217","X216","X317","X316","X417","X
416","X517","X516","output")]
> dim(IS1617)
[1] 100 11
> #Imprese non solventi
> IF1617<-I1617_test[I1617_test$output<1,c("X117","X116","X217","X216","X317","X316","X417","X
416","X517","X516","output")]
> dim(IF1617)
[1] 100 11

> #Modello 2
> #Imprese solventi classificate correttamente
> I1617_risultatiS2<-compute(I1617_model2,IS1617)
> SIS32=I1617_risultatiS2$net.result[I1617_risultatiS2$net.result>0.5]
> length(SIS32)
[1] 63
> #Imprese non solventi classificate correttamente
> I1617_risultatiF2<-compute(I1617_model2,IF1617)
> SIF32=I1617_risultatiF2$net.result[I1617_risultatiF2$net.result<0.5]
> length(SIF32)
[1] 78

> #Modello 3
> #Imprese solventi classificate correttamente
> I1617_risultatiS3<-compute(I1617_model3,IS1617)
> SIS33=I1617_risultatiS3$net.result[I1617_risultatiS3$net.result>0.5]
> length(SIS33)
[1] 60
> #Imprese non solventi classificate correttamente
> I1617_risultatiF3<-compute(I1617_model3,IF1617)
> SIF33=I1617_risultatiF3$net.result[I1617_risultatiF3$net.result<0.5]
> length(SIF33)
[1] 77

> #Modello 4
> I1617_risultatiS4<-compute(I1617_model22,IS1617)
> SIS34=I1617_risultatiS4$net.result[I1617_risultatiS4$net.result>0.5]
> length(SIS34)
[1] 59
> #Imprese non solventi classificate correttamente
> I1617_risultatiF4<-compute(I1617_model22,IF1617)
> SIF34=I1617_risultatiF4$net.result[I1617_risultatiF4$net.result<0.5]
> length(SIF34)
[1] 77

> #Modello 5
> I1617_risultatiS5<-compute(I1617_model5,IS1617)
> SIS35=I1617_risultatiS5$net.result[I1617_risultatiS5$net.result>0.5]
> length(SIS35)
[1] 57
> #Imprese non solventi classificate correttamente
> I1617_risultatiF5<-compute(I1617_model5,IF1617)
> SIF35=I1617_risultatiF5$net.result[I1617_risultatiF5$net.result<0.5]
> length(SIF35)
[1] 81

```

## Appendice 2.5.1

```

> library(readxl)
> I7525 <- read_excel("Dati/Campione TS75 VS25/TS75 VS25 definitivo.xlsx")
> str(I7525)
Classes 'tbl_df', 'tbl' and 'data.frame':    400 obs. of  6 variables:
 $ x1
 %
2017: num  98 -2025 48 8 12 ...
 $ x2
 %
2017: num  -10 -1510 15 32 10 0 0 0 -30 0 ...
 $ x3
 %
2017: num  4 -284 4 9 60 3 14 3 -2 38 ...
 $ x4
 %

```

```

2017: num  7896 -95 68 70 10450 ...
$ x5
%
2017: num  76 225 34 69 100 268 136 78 1 106 ...
$ Output : num  1 0 1 1 1 1 1 0 0 1 ...
> dim(I7525)
[1] 400 6
> #Normalizzazione delle variabili
> normalize2<-function(I7525){return((I7525-min(I7525))/(max(I7525)-min(I7525)))}
> I7525_norm<-as.data.frame(lapply(I7525, normalize2))
> summary(I7525_norm)
  x1.....2017      x2.....2017      x3.....2017
Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.9994   1st Qu.:0.9998   1st Qu.:0.9983
Median :0.9995   Median :0.9998   Median :0.9984
Mean   :0.9961   Mean   :0.9972   Mean   :0.9957
3rd Qu.:0.9997   3rd Qu.:0.9998   3rd Qu.:0.9985
Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
  x4.....2017      x5.....2017      Output
Min.   :0.000000   Min.   :0.000000   Min.   :0.0
1st Qu.:0.03592    1st Qu.:0.00513    1st Qu.:0.0
Median :0.03779    Median :0.06224    Median :0.5
Mean   :0.05075    Mean   :0.10094    Mean   :0.5
3rd Qu.:0.04129    3rd Qu.:0.15458    3rd Qu.:1.0
Max.   :1.00000    Max.   :1.00000    Max.   :1.0
> #TS=75% VS=25%
> I7525_train<-I7525_norm[1:300,]
> dim(I7525_train)
[1] 300 6
> colnames(I7525_train)<- c("x1","x2","x3","x4","x5","output")
> colnames(I7525_train)
[1] "x1" "x2" "x3" "x4" "x5" "output"
> show(I7525_train)

> I7525_test<-I7525_norm[301:400,]
> show(I7525_test)
> dim(I7525_test)
[1] 100 6
> colnames(I7525_test)<- c("x1","x2","x3","x4","x5","output")
> colnames(I7525_test)
[1] "x1" "x2" "x3" "x4" "x5" "output"

> library(neuralnet)
> # Modello 2 con backpropagation classica
> I7525_model<-neuralnet(output~x1+x2+x3+x4+x5,
+                          hidden=2,data=I7525_train,learningrate = 0.05,
+                          learningrate.factor = list(minus = 0.1,plus = 1.0),
+                          learningrate.limit=list(minus=0.1,plus=1.0),
+                          threshold =0.05,stepmax=20000000,algorithm="backprop",
+                          act.fct=sigmoid,linear.output=FALSE,likelihood=TRUE)
> plot(I7525_model)

> I7525_model
$weights
$weights[[1]]
$weights[[1]][[1]]
      [,1] [,2]
[1,] -28.016399  47.50471
[2,]  5.192952 -22.48023
[3,] -9.332819  19.54461
[4,] 15.664648 -29.15968
[5,] 312.606077 -456.69797
[6,] 18.398831  8.09344

$weights[[1]][[2]]
      [,1]
[1,]  0.6802075
[2,]  4.7432955
[3,] -4.6407260

$result.matrix
      [,1]
error 2.540028e+01
reached.threshold 4.998894e-02
steps 2.739275e+06
aic 8.080056e+01
bic 1.363573e+02
Intercept.to.1layhid1 -2.801640e+01
x1.to.1layhid1 5.192952e+00
x2.to.1layhid1 -9.332819e+00
x3.to.1layhid1 1.566465e+01
x4.to.1layhid1 3.126061e+02
x5.to.1layhid1 1.839883e+01
Intercept.to.1layhid2 4.750471e+01
x1.to.1layhid2 -2.248023e+01
x2.to.1layhid2 1.954461e+01

```

```

X3.to.1layhid2      -2.915968e+01
X4.to.1layhid2      -4.566980e+02
X5.to.1layhid2       8.093440e+00
Intercept.to.output 6.802075e-01
1layhid1.to.output  4.743296e+00
1layhid2.to.output  -4.640726e+00

> #Valutazione dell'attendibilità del modello 2
> I7525_risultati<-compute(I7525_model,I7525_test)
> I7525_risultati

> SI7525=I7525_risultati$net.result[I7525_risultati$net.result>0.5]
> length(SI7525)
[1] 44

> prev_attend7525<-I7525_risultati$net.result
> #ora si deve misurare la correlazione tra il valore previsto
> #e e quello reale"
> cor(prev_attend7525,I7525_test$output)
      [,1]
[1,] 0.5541945

> #Imprese solventi
> I7525_test
> IS7525<-I7525_test[I7525_test$output>0,c("X1","X2","X3","X4","X5","output")]
> dim(IS7525)
[1] 50 6

> #Imprese non solventi
> IF7525<-I7525_test[I7525_test$output<1,c("X1","X2","X3","X4","X5","output")]
> dim(IF7525)
[1] 50 6

> #Modello 2
> #Imprese solventi classificate correttamente
> I7525_risultatis<-compute(I7525_model,IS7525)
> SIS7525=I7525_risultatis$net.result[I7525_risultatis$net.result>0.5]
> length(SIS7525)
[1] 36
> #Imprese non solventi classificate correttamente
> I7525_risultatiF<-compute(I7525_model,IF7525)
> SIF7525=I7525_risultatiF$net.result[I7525_risultatiF$net.result<0.5]
> length(SIF7525)
[1] 42

```

## Appendice 2.5.2

```

> library(readxl)
> I2575 <-read_excel("Dati/Campione TS25 vs75/TS25 vs75 definitivo.xlsx")
> str(I2575)
Classes 'tbl_df', 'tbl' and 'data.frame':      400 obs. of  6 variables:
 $ X1
%
2017: num  22 -2 45 40 -82 18 -186 5 57 1 ...
 $ X2
%
2017: num  11 0 83 0 0 6 -61 0 0 -5 ...
 $ X3
%
2017: num  4 3 4 -1 -31 5 -88 3 14 1 ...
 $ X4
%
2017: num  40 4 1119 1 3 ...
 $ X5
%
2017: num  55 222 111 42 3 86 144 29 380 134 ...
 $ Output
      : num  1 1 1 1 1 0 0 1 1 0 ...
> dim(I2575)
[1] 400 6
> #Normalizzazione delle variabili
> normalize3<-function(I2575){return((I2575-min(I2575))/(max(I2575)-min(I2575)))}
> I2575_norm<-as.data.frame(lapply(I2575, normalize3))
> summary(I2575_norm)
  x1.....2017      x2.....2017      x3.....2017
Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.9994   1st Qu.:0.9998   1st Qu.:0.9983
Median :0.9995   Median :0.9998   Median :0.9984
Mean   :0.9961   Mean   :0.9972   Mean   :0.9957
3rd Qu.:0.9997   3rd Qu.:0.9998   3rd Qu.:0.9985
Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
  x4.....2017      x5.....2017      Output
Min.   :0.000000   Min.   :0.000000   Min.   :0.0
1st Qu.:0.03592   1st Qu.:0.00513   1st Qu.:0.0

```

```

Median :0.03779   Median :0.06224   Median :0.5
Mean    :0.05075   Mean    :0.10094   Mean    :0.5
3rd Qu.:0.04129   3rd Qu.:0.15458   3rd Qu.:1.0
Max.    :1.00000   Max.    :1.00000   Max.    :1.0

> #TS=25% VS=75%
> I2575_train<-I2575_norm[1:100,]
> dim(I2575_train)
[1] 100 6
> show(I2575_train)
> colnames(I2575_train)<- c("x1","x2","x3","x4","x5","output")
> colnames(I2575_train)
[1] "x1" "x2" "x3" "x4" "x5" "output"
> I2575_test<-I2575_norm[101:400,]
> show(I2575_test)
> dim(I2575_test)
[1] 300 6
> colnames(I2575_test)<- c("x1","x2","x3","x4","x5","output")
> colnames(I2575_test)
[1] "x1" "x2" "x3" "x4" "x5" "output"

> # Modello 2 con backpropagation classica
> I2575_model<-neuralnet(output~x1+x2+x3+x4+x5,
+                         hidden=2,data=I2575_train,learningrate = 0.001,
+                         learningrate.factor = list(minus = 0.1,plus = 1.0),
+                         learningrate.limit=list(minus=0.1,plus=1.0),
+                         threshold =0.01,stepmax=100000000,algorithm="backprop",
+                         act.fct=sigmoid,linear.output=FALSE,likelihood=TRUE)
> plot(I2575_model)

> I2575_model
$weights
$weights[[1]]
$weights[[1]][[1]]
      [,1] [,2]
[1,] -0.213089 1.8231754
[2,] -2.089058 3.1730243
[3,] -1.425052 -0.4241892
[4,] 1.035230 -4.4176792
[5,] 26.953672 -29.5872850
[6,] 7.658565 4.1131425

$weights[[1]][[2]]
      [,1]
[1,] 5.963497
[2,] 23.192487
[3,] -35.260267

$result.matrix
      [,1]
error 8.016558e+00
reached.threshold 9.999998e-03
steps 1.485489e+06
aic 4.603312e+01
bic 8.511067e+01
Intercept.to.l1ayhid1 -2.130890e-01
X1.to.l1ayhid1 -2.089058e+00
X2.to.l1ayhid1 -1.425052e+00
X3.to.l1ayhid1 1.035230e+00
X4.to.l1ayhid1 2.695367e+01
X5.to.l1ayhid1 7.658565e+00
Intercept.to.l1ayhid2 1.823175e+00
X1.to.l1ayhid2 3.173024e+00
X2.to.l1ayhid2 -4.241892e-01
X3.to.l1ayhid2 -4.417679e+00
X4.to.l1ayhid2 -2.958729e+01
X5.to.l1ayhid2 4.113143e+00
Intercept.to.output 5.963497e+00
l1ayhid1.to.output 2.319249e+01
l1ayhid2.to.output -3.526027e+01

> #Valutazione dell'attendibilità del modello 2
> I2575_risultati<-compute(I2575_model,I2575_test)
> I2575_risultati

> SI2575=I2575_risultati$net.result[I2575_risultati$net.result>0.5]
> length(SI2575)
[1] 144
> prev_attend2575<-I2575_risultati$net.result
> #ora si deve misurare la correlazione tra il valore previsto
> #e e quello reale"
> cor(prev_attend2575,I2575_test$output)
      [,1]
[1,] 0.5387822

```

```
> #Imprese solventi
> I2575_test
> IS2575<-I2575_test[I2575_test$output>0,c("x1","x2","x3","x4","x5","output")]
> dim(IS2575)
[1] 150 6
> #Imprese non solventi
> IF2575<-I2575_test[I2575_test$output<1,c("x1","x2","x3","x4","x5","output")]
> dim(IF2575)
[1] 150 6
> #Modello 2
> #Imprese solventi classificate correttamente
> I2575_risultatiS<-compute(I2575_model,IS2575)
> SIS2575=I2575_risultatiS$net.result[I2575_risultatiS$net.result>0.5]
> length(SIS2575)
[1] 105
> #Imprese non solventi classificate correttamente
> I2575_risultatiF<-compute(I2575_model,IF2575)
> SIF2575=I2575_risultatiF$net.result[I2575_risultatiF$net.result<0.5]
> length(SIF2575)
[1] 111
```



## Bibliografia

### Z-Score e analisi discriminante:

- Altman Edward I., *Corporate Financial Distress and Bankruptcy*, Third Edition Edith Hotchkiss), 2006.
- Altman Edward I., *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, The Journal of Finance, Vol. 23, No. 4, Sep., 1968.
- Altman Edward I., *The Link Between Default and Recovery Rates: Theory, Empirical Evidence and Implications*, Andrea Sironi e Andrea Resti, The Journal of Business, 2005.
- Altman Edward I., *Z-Score Models' Application to Italian Companies Subject to Extraordinary Administration*, Alessandro Danovi, e Alberto Falini, Journal of Applied Finance – No. 1, 2013.
- Beaver William H., *Financial Ratios As Predictors of Failure*, Journal of Accounting Research, Vol. 4, 1966
- Beaver William H., *The Information Content of Annual Earnings Announcements*, Journal of Accounting Research, Vol. 6, 1968.
- Bluhm Christian, Overbeck Ludger, Wagner Christoph, *An Introducing of Credit Risk Modelling*, Chapman & Hall/CRC, 2003.
- Bottani Pietro, Cipriani Letizia e Serao Francescomaria, *Analisi del rischio di insolvenza di una PMI tramite l'utilizzo del modello dello Z-Score*, Articolo pubblicato sul n°1 /2004 di Amministrazione e Finanza, edito da Ipsoa.
- Carmichael Douglas R., *Z Score: A Guide to Failure Prediction*, PhD, CPA Baruch College, February 1995.
- Frye Jon, *Depressing Recoveries*, Emerging Issues Series Supervision and Regulation Department Federal Reserve Bank of Chicago ottobre 2000 (S&R-2000-8)
- Iannucci Jacopo, *Analisi discriminante, regressione logistica e reti per la stima delle probabilità di fallimento. Applicazioni ed estensioni del metodo Z-score*, Università degli Studi di Roma Tre, Facoltà di Scienze M.F.N., Tesi di Laurea in Matematica, 2004.
- Mossman Charles E., Bell Geoffrey G., Swartz L. Mick e Turtle Harry, *An empirical comparison of bankruptcy models*, The Financial Review 33, 1998
- Ohlson James A., *Financial Ratios and the Probabilistic Prediction of Bankruptcy*, Journal of Accounting Research, Vol. 18, No. 1. (Spring, 1980), pp. 109-131.
- Scott James, *The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models*, Journal of Banking and Finance 5 317-344. North-Holland Publishing Company, 1981.

- Senese Vincenzo Paolo, *Regressione Multipla e Regressione Logistica: concetti introduttivi ed esempi*, I Edizione, ottobre 2014.
- Shumway Tyler, *Forecasting Bankruptcy More Accurately: A Simple Hazard Model*, 2002, *The Journal of Business*, Vol. 74, No. 1 (January 2001), pp. 101-124
- Stanghellini E., *Introduzione ai metodi statistici per il credit scoring*, Springer-Verlag Italia, Milano, 2009
- Yen-Ting Hu e Perraudin William, *The Dependence of Recovery Rates and Defaults*, Birkbeck College, 2002
- Zmijewski Mark E., *Methodological Issues Related to the Estimation of Financial Distress Prediction Models*, *Journal of Accounting Research*, Vol. 22, Studies on Current Econometric Issues in Accounting Research, pp. 59-82, 1984.

### **Reti Neurali:**

- Altman E.I., Marco G., Varetto F. *Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the italian experience)*, *Journal of Banking and finance*, 18, pp. 505-529, 1994.
- Atiya A.F., *Bankruptcy prediction for credit risk using neural networks: a survey and new result*, *IEEE Transactions on neural networks*, vol. 12, n. 4, 2001.
- Back B., Sere K., G. Oosterom e Wezel M.C., *A comparative study of neural networks in bankruptcy prediction*, *Turku Centre for Computer Science, Technical Report*, 1995.
- Chilanti, *Analisi e previsione delle insolvenze: un approccio neurale*, *Finanza Imprese e Mercati*, 1993.
- Coats P., Fant L., *Recognizing financial distress patterns using a neural network tool*, *Financial Management*, vol. 22, pp. 142–155, 1993.
- D'Annunzio Nadia e Falavigna Greta, *Modelli di analisi e previsione del rischio di insolvenza. Una prospettiva delle metodologie applicate*, *Working Paper Ceris-CNR*, 2004.
- De Almeida, Dumontier, *Neural networks, accounting numbers and bankruptcy prediction*, *Association Française de Comptabilité, Comptabilité et Nouvelles Technologies*, Mai, pp 269-286, 1993.
- Fausett Laurene, Englewood Cliffs, *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*, NJ, Prentice-Hall, 1994.
- Fenga L., *L'utilizzo di reti neurali per la stima anticipata e la previsione di serie storiche stagionali. Un'analisi comparativa con modelli della classe ARMA su serie ISTAT delle presenze turistiche in Italia*.

- Floreano Dario e Mattiussi Claudio, *Manuale sulle Reti Neurali*, Il Mulino, 2002
- Foster W. R., Collopy F., Ungar L. H., *Neural network forecasting of short, noisy time series*, Computers in Chemical Engineering, 16, 4, pp. 293-297, 1992.
- Hebb D. O., *The Organization of Behavior: A Neuropsychological Theory*, McGill University, New York John Wiley & Sons, Inc. London Chapman & Hall, Limited, 1949.
- Hecht-Nielsen Robert, *Neurocomputing*, Reading, MA, Addison-Wesley, 1990.
- Hertz J., Krogh A. e Palmer R., *Introduction to the theory of Neural Computation*, Reedwood City, CA, Addison-Wesley, 1991.
- Hill T., Marquez L., O'Connor M., Remus W., *Artificial neural network models for forecasting and decision making*, NN4CAST6, wp 5, 1993.
- Hopfield J. J., *Neural networks and physical systems with emergent collective computational abilities*, Proc. NatL Acad. Sci. USA Vol. 79, pp. 2554-2558 Biophysics, April 1982.
- Hornick, Stinchcombe e White, *Multilayer feedforward networks are universal approximators*, 1989, Neural Networks, vol. 2, pp 359-366.
- Kiviluoto K, *Predicting bankruptcies with the self-organizing map*, Neurocomputing, vol. 21, pp. 191–201, 1998.
- Lee K. C., Han I., Kwon Y., *Hybrid neural network models for bankruptcy predictions*, Decision Support Systems, 18, 1996.
- McCulloch Warren S. e Pitts Walter, *A logical calculus of the ideas immanent in nervous activity*, Bulletin of Mathematical Biophysics, Volume 5, 1943.
- Minsky e Papert, *Perceptrons. Expanded Edition*, Cambridge, MA, MIT Press, 1988
- Odom M., Sharda R., *A neural network model for bankruptcy prediction*, in Proc. Int. Joint Conf. Neural Networks, San Diego, CA, 1990.
- Perez Muriel, *Artificial neural networks and bankruptcy forecasting: a state of the art*, Springer-Verlag London Limited, 2006.
- Raghupathi, Schkade e Raju, *A neural network approach to bankruptcy prediction*, NN in Finance and Investing: Using AI to improve real-world performance, TRIPPI/TURBAN Irwin Professional Publishing, revised 1996, pp. 227-241, 1991.
- Rehkugler H. and Zimmermann H. G., *Klassifikation von Unternehmen mittels KNN*, Kerling M., Poddig T., in Neuronale Netze in der Ökonomie, Eds. München, Germany, 1994.
- Rosenblatt Frank, *Principles of Neurodynamics*, New York, Spartan Books, 1962.
- Sexton R.S., Dorsey R.E., Johndon J., *Toward global optimization of neural networks: a comparison of the genetic algorithm and backpropagation*.
- Sironi A., Marsella M., *La misurazione e la gestione del rischio di credito. Modelli, strumenti e politiche*, Bancaria editrice, 1998.

- Stent Gunther S., A Physiological Mechanism for Hebb's Postulate of Learning, Proc. Nat. Acad. Sci. USA Vol. 70, No. 4, pp. 997-1001, April 1973.
- Tam K. Y., Kiang M. Y., *Managerial applications of neural networks: the case of bank failure predictions*, Management Science, 38, 7, pp. 926-947, 1992.
- Tam K. Y., Kiang M. Y., *Predicting bank failures: a neural network approach*, Applied Artificial Intelligence, 4, pp. 265-282, 1990.
- Turban E., Trippi R., *Neural Networks in Finance and Investing. Using artificial neural intelligence to improve real-world performance*, Probus Publishing Company, Chicago Illinois, Cambridge, England.
- Willshaw David e Dayan Peter, *Optimal Plasticity from Matrix Memories: What Goes Up Must Come Down*, Centre for Cognitive Science and Department of Physics, University of Edinburgh, Edinburgh, Scotland, 1990.
- Windrow, Winter e Baxter, *Learning Phenomena in Layered Neural Networks*, IEEE Press, 1987
- Yim J., Mitchell H., *A comparison of corporate failure models in Australia: Hybrid neural networks, logit models and discriminant analysis*, School of Economics and Finance, WP No. 10, 2002.