



Università
Ca' Foscari
Venezia

Corso di Laurea magistrale
in Economia e Finanza

Tesi di Laurea

**Reti neurali artificiali per la previsione
dell'insolvenza aziendale**

Relatore

Prof. Claudio Pizzi

Laureando

Stefano Boldrin

Matricola 857011

Anno Accademico

2019 / 2020

| Indice | Pag. |
|--|------|
| Introduzione | 1 |
| Capitolo 1: Il rischio di credito | 3 |
| 1.1 La gestione dei rischi bancari e il rischio di credito | 3 |
| 1.2 Le componenti del rischio di credito | 5 |
| 1.2.1 Probability of Default | 7 |
| 1.2.2 Loss Given Default | 8 |
| 1.2.3 Exposure at Default | 10 |
| 1.2.4 Perdita attesa e perdita inattesa | 11 |
| 1.3 Il contesto normativo | 13 |
| 1.3.1 Basilea I | 13 |
| 1.3.2 Basilea II | 15 |
| 1.3.3 Basilea III | 20 |
| 1.3.4 Verso Basilea IV | 22 |
| Capitolo 2: Tecniche statistiche tradizionali per la previsione del default | 25 |
| 2.1 Introduzione ai modelli di credit scoring | 25 |
| 2.1.1 Il ruolo dei modelli di credit scoring | 25 |
| 2.1.2 Classificatori per il credit scoring | 26 |
| 2.2 Tecniche tradizionali per lo scoring | 27 |
| 2.2.1 I primi modelli | 27 |
| 2.2.2 Modello univariato | 29 |
| 2.2.3 Analisi discriminante lineare | 29 |
| 2.2.4 Modelli di regressione | 32 |
| 2.2.5 Confronto tra LDA, Logit e Probit | 39 |
| Capitolo 3: Tecniche di machine learning | 41 |
| 3.1 Introduzione al machine learning | 41 |
| 3.1.1 Definizione | 41 |
| 3.1.2 Applicazioni | 44 |
| 3.1.3 Applicazioni in ambito finanziario | 47 |
| 3.2 Reti neurali artificiali | 49 |
| 3.2.1 Neurone biologico | 49 |
| 3.2.2 Neurone artificiale | 51 |
| 3.2.3 Multilayer Perceptron (MLP) | 57 |
| 3.3 Performance delle tecniche di machine learning | 68 |

| | |
|--------------------------------------|-----|
| 3.4 Modelli ibridi | 73 |
| Capitolo 4: Esperimento | 76 |
| 4.1 Modello Logit | 78 |
| 4.1.1 Logit18_01 | 80 |
| 4.1.2 Logit18_02 | 81 |
| 4.1.3 Logit18_03 | 82 |
| 4.1.4 Logit1718 | 83 |
| 4.1.5 Discussione dei risultati | 84 |
| 4.2 Reti neurali artificiali | 84 |
| 4.2.1 MLP 01 | 86 |
| 4.2.2 MLP 02 | 88 |
| 4.2.3 MLP 03 | 89 |
| 4.2.4 MLP 04 | 90 |
| 4.2.5 MLP 05 | 95 |
| 4.2.6 MLP 06 | 97 |
| 4.3 Sintesi dei risultati ottenuti | 98 |
| Conclusioni | 100 |
| Appendice 1 (Dataset) | 101 |
| Appendice 2 (Codice) | 101 |
| Appendice 2.1 Codice MLP (2018) | 101 |
| Appendice 2.2 Codice MLP (2017-2018) | 103 |
| Bibliografia | 106 |

Introduzione

La crisi finanziaria globale del 2007-2009 ha fatto emergere una serie di debolezze all'interno del sistema bancario, tra cui: una gestione dei rischi inadeguata, livelli di capitale e di liquidità insufficienti, una debole vigilanza e regolamentazione. Inoltre, ha reso palese l'importanza, sia ai fini di una sana e prudente gestione dell'azienda bancaria sia per la stabilità del sistema finanziario, di una corretta misurazione e gestione dei vari rischi a cui è esposto un istituto bancario. Le autorità di vigilanza hanno risposto alla crisi rivolgendo i loro sforzi nella direzione di rafforzare la capacità del sistema di resistere agli shock economici e di limitare il rischio di contagio verso l'economia reale. Ciò si è tradotto in un'ampia produzione normativa circa il contenimento dei rischi assunti dalle banche nei loro portafogli e l'aumento dei requisiti patrimoniali minimi per fronteggiare le perdite. Tra i vari rischi le autorità di vigilanza hanno posto l'accento in particolare sul rischio di credito, poiché è il fattore che impatta maggiormente sulle perdite di una banca. Nel Capitolo 1 verrà fornito un quadro generale circa la misurazione del rischio di credito, a partire dai concetti di perdita attesa e inattesa, e verrà analizzata la normativa di riferimento (Accordi di Basilea).

All'interno dei vari strumenti di misurazione del rischio di credito, il presente elaborato approfondirà i modelli di credit scoring. Essi emettono un giudizio circa il merito creditizio delle aziende a partire da una serie di indicatori economico-finanziari. Non rappresentano solo un utile strumento di supporto per le decisioni di accettazione o rifiuto delle richieste di finanziamento, ma assumono un ruolo centrale anche per la stima del patrimonio di vigilanza e del capitale di rischio adeguato. Nel Capitolo 2 verranno presentate le tecniche tradizionali di scoring più note, tra cui l'analisi discriminante lineare e la regressione logistica.

I modelli di credit scoring non sono affatto recenti, basti pensare che il noto modello Z-score di Altman è stato introdotto nel 1968. Tuttavia, negli ultimi anni sono state proposte in letteratura numerose tecniche alternative, molte delle quali basate sul machine learning, che presentano significative potenzialità. La diffusione di questo processo di analisi dei dati è divenuta al giorno d'oggi capillare: dal riconoscimento facciale ai servizi di traduzione automatica multilingue, dalla guida automatica dei veicoli alla diagnosi precoce delle malattie. Il recente sviluppo delle metodologie di machine learning è

favorito dalla crescente disponibilità di grandi dataset, da un aumento della potenza di calcolo delle macchine e dalla peculiarità di essere molto flessibili. Nel Capitolo 3 verranno introdotte alcune di queste tecniche di apprendimento automatico, in particolare il focus sarà rivolto alle Reti neurali artificiali.

Lo scopo ultimo del presente elaborato è verificare se l'applicazione di tali tecniche può portare a dei miglioramenti nelle performance rispetto ai modelli tradizionali. A tal fine, nel Capitolo 4 verrà condotto un esperimento su un campione di 370 aziende manifatturiere italiane, dove si confronteranno le capacità del modello Logit e di una Rete neurale artificiale nel prevedere l'insolvenza aziendale a partire da un set di indicatori di bilancio. Un incremento nel grado di accuratezza di tali modelli è rilevante per una banca, dal momento che favorisce una corretta allocazione del credito e consente di massimizzare il rendimento aggiustato per il rischio.

Capitolo 1: Il rischio di credito

1.1 La gestione dei rischi bancari e il rischio di credito

La misurazione e gestione dei rischi in ambito bancario è di fondamentale importanza per la profittabilità e la solvibilità degli istituti di credito, non a caso negli ultimi anni si è assistito a un crescente interesse sul tema da parte delle Autorità di Vigilanza bancaria (vedi Accordi di Basilea) e degli istituti stessi tramite numerosi investimenti nella funzione di risk management (Tedeschi, 2016).

L'assunzione di rischi è finalizzata a realizzare margini di profitto, tant'è che la funzione di risk management è coinvolta, oltre che nell'identificazione, misurazione e controllo dei vari rischi, anche nel loro pricing. Al contempo l'assunzione di rischi richiede un'adeguata dotazione di capitale bancario, che funge da garanzia imprescindibile per gli stakeholder quale presidio a tutela degli attivi bancari dal rischio di perdite future. Il capitale rappresenta una risorsa scarsa e costosa, che richiede una remunerazione in linea con la rischiosità dell'azienda bancaria. È necessario, pertanto, dotarsi di un processo che, tenendo conto anche dei vincoli imposti dal supervisore, stabilisca la quantità e la composizione ottimale del patrimonio per minimizzare il costo del capitale. A cui deve seguire una efficiente allocazione del capitale all'interno delle varie unità di business con il fine di massimizzare il rendimento aggiustato per il rischio e creare valore per gli azionisti. È possibile individuare, pertanto, una sorta di trinomio tra l'assunzione di rischi, il reperimento (e l'allocazione) di capitale e la profittabilità della banca. Ad esempio, un eccesso di capitale comporta una protezione maggiore dal rischio di perdite, ma quel capitale potrebbe essere impiegato per assumere degli ulteriori rischi (es. concessione di un finanziamento) e di conseguenza generare maggiori margini di profitto a beneficio degli azionisti. Si comprende come la funzione di risk management sia centrale nel coadiuvare la governance in questo processo e, più in generale, nella gestione dell'azienda bancaria.

La banca deve saper fronteggiare una moltitudine di rischi, a titolo esemplificativo e non esaustivo si possono citare: rischio di credito e di controparte, rischio di mercato (legato al portafoglio di negoziazione), rischio operativo, rischio di liquidità, rischio tasso di interesse e rischio tasso di cambio. Il principale fattore di rischio a livello quantitativo cui una banca è esposta è il primo; secondo i dati forniti dall'European Banking Authority

(EBA), al 31 Dicembre 2019 le attività ponderate per il rischio di credito ammontavano a circa il 74% del totale in Germania, il 75% nel Regno Unito, l'82% in Francia, l'83% in l'Italia e l'85% in Spagna (EBA, 2020).

Un'efficace gestione del rischio di credito è pertanto un elemento essenziale per il successo di lungo periodo di qualsiasi organizzazione bancaria. Tant'è che a partire dagli anni '90 le principali banche internazionali hanno investito molto in risorse umane e tecnologiche per sviluppare metodi di valutazione e gestione del rischio di credito (Tedeschi, 2016).

Il rischio di credito è definito, secondo il Comitato di Basilea (2000), semplicemente come “la possibilità che la controparte non riesca a far fronte alle proprie obbligazioni secondo i termini stabiliti”. Una definizione alternativa viene fornita da Resti e Sironi (2007), che si riferiscono al rischio di credito come alla “possibilità che una variazione inattesa del merito creditizio¹ della controparte generi una corrispondente variazione inattesa del valore di mercato dell'esposizione creditizia”.

Dalla seguente definizione emergono spunti di riflessione interessanti, in primo luogo si può notare come il concetto di rischio di credito sia più ampio rispetto al mero rischio di insolvenza della controparte (non rispetto degli obblighi contrattuali) ed includa anche il rischio di migrazione, cioè il deterioramento (downgrade) del merito creditizio. Intuitivamente un downgrade della controparte indica una maggiore rischiosità, che deve essere compensata da un tasso di interesse più elevato (dal momento che aumenta il risk premium dell'investimento), di conseguenza un maggior tasso di attualizzazione dei flussi di cassa si riflette in una riduzione del valore di mercato dell'esposizione. Si passa, pertanto, da una visione dicotomica della controparte (solvente o insolvente) a considerare un ampio numero di classi di rischio. In altre parole, le imprese classificate come solventi non costituiscono un gruppo omogeneo, presentano al loro interno una rischiosità e, di conseguenza, un pricing differente.

In secondo luogo, il concetto di rischio fa riferimento al deterioramento inatteso, ovvero non previsto nella valutazione del merito creditizio. Ad esempio, se la banca ritiene che la controparte andrà incontro a una temporanea difficoltà finanziaria l'anno seguente, ne tiene conto in sede di istruttoria ed effettua il pricing di conseguenza. Il rischio reale è connesso, invece, a eventi inattesi.

¹ Il merito di credito è la capacità di rimborsare interamente e alla scadenza pattuita il debito contratto.

Infine il rischio di credito si misura a valori di mercato, dal momento che la maggior parte delle esposizioni creditizie delle banche è rappresentato da asset illiquidi (per i quali non esiste un mercato secondario sviluppato) la banca deve dotarsi di un modello interno di pricing degli asset.

Il rischio di credito è un concetto vasto, che comprende al suo interno:

- il rischio di insolvenza della controparte, che consiste nella possibilità che il debitore non riesca a far fronte ai propri impegni finanziari nei tempi stabiliti;
- il rischio di migrazione, ovvero di deterioramento del merito creditizio;
- il rischio riguardante il processo di recupero (recovery risk), che si traduce in un tasso di recupero effettivo inferiore a quello stimato;
- il rischio esposizione, ovvero la possibilità che l'ammontare dell'esposizione verso una controparte aumenti in corrispondenza del verificarsi del default;
- il rischio spread, consiste nel possibile aumento dei tassi di interesse richiesti dal mercato (a parità di rating) causato da una maggiore avversione al rischio degli investitori;
- il rischio paese, ovvero eventi legati all'area geografica di provenienza della controparte (ad es. decisioni politiche) che impediscano il rimborso del debito;
- il rischio di concentrazione, dovuto alla detenzione di un portafoglio di crediti con un basso grado di diversificazione.

1.2 Le componenti del rischio di credito

Un modello di gestione del rischio di credito mira a stimare la distribuzione delle perdite future derivanti dall'attività creditizia associando alle possibili perdite una probabilità.

Il modello deve essere in grado di fornire una stima del livello di perdite che è ragionevole attendersi sul portafoglio e l'ammontare massimo di perdite che potrebbe verificarsi in una certa percentuale di casi e in un certo orizzonte temporale.

Il rischio di credito, infatti, può essere scisso in due componenti: una perdita attesa (expected loss, EL) e una perdita inattesa (unexpected loss, UL). La prima più che un vero e proprio rischio rappresenta un costo, in quanto è un valore atteso e perciò incorporabile nel tasso di interesse del prestito. La perdita attesa si ottiene aggregando tre elementi:

- la probabilità di insolvenza della controparte (Probability of Default, PD);

- il valore atteso dell'esposizione al momento del default (Exposure at Default, EAD);
- il tasso di perdita atteso in caso di default (Loss Given Default, LGD), pari al complementare del tasso di recupero (Recovery rate, RR).

È da sottolineare come esistono diversi gradi di insolvenza: dal semplice ritardo nel pagamento di una rata al fallimento. La scelta di quale significato attribuire a questo concetto non è di poco conto, visto che impatta sulla misurazione di PD e RR e in generale sui risultati dei modelli di misurazione del rischio di credito. Una definizione troppo restrittiva riflette tassi di default e tassi di recupero bassi, dal momento che verrebbero classificate come insolventi solamente le imprese in uno stato di estrema difficoltà finanziaria. Viceversa, una definizione più ampia includerebbe anche le imprese in difficoltà meno grave, che presentano maggiori possibilità di ripagare i loro debiti pervenendo a tassi di default e a tassi di recupero ben maggiori.

Alla luce di questa considerazione, emerge la necessità di pervenire a una definizione comune per poter confrontare le stime.

A tal proposito, le linee guida emanate dall'EBA nel 2016² stabiliscono che la banca è tenuta a classificare un'esposizione nello stato di default qualora si verificano almeno uno degli eventi sotto indicati:

- "l'ente giudica improbabile che, senza il ricorso ad azioni quale l'escussione delle garanzie, il debitore adempia integralmente alle sue obbligazioni creditizie verso l'ente stesso, la sua impresa madre o una delle sue filiazioni";
- "l'impresa è in arretrato da oltre 90 giorni su un'obbligazione creditizia rilevante. Ai fini della determinazione della rilevanza dell'esposizione devono sussistere una soglia di rilevanza assoluta pari a 500 euro (100 euro per le persone fisiche e le piccole e medie imprese con un'esposizione verso la banca inferiore a 1 milione di euro) e una relativa pari all'1% dell'importo totale delle esposizioni dell'impresa verso la banca finanziatrice" (ABI, 2019).

² Con a monte l'art. 178 del Regolamento UE n.575/2013, ossia il regolamento sui requisiti patrimoniali (CRR).

1.2.1 Probability of Default

La stima della probabilità che la controparte non sia in grado di adempiere agli obblighi contrattuali su un dato orizzonte temporale (tipicamente un anno) riveste un ruolo centrale nei modelli di gestione del rischio di credito. Il calcolo della PD si basa su una procedura a due passi. Il primo step prevede l'assegnazione di un giudizio circa il merito creditizio della controparte potendo optare tra più approcci.

I modelli analitico-soggettivi sono fondati su un'istruttoria elaborata da un analista, che tiene in considerazione informazioni sia quantitative (dati economico-finanziari) che qualitative (qualità del management, posizionamento strategico, contesto macroeconomico, ecc.). Da un lato è possibile individuare, grazie all'esperienza dell'analista, elementi che modelli di natura statistica non sono in grado di cogliere, d'altro sono giudizi dipendenti da valutazioni soggettive.

I modelli statistici restituiscono un punteggio (score), frutto della ponderazione e dell'aggregazione di un set di variabili esplicative di tipo quantitativo (in primis indicatori di bilancio e informazioni andamentali) capaci di spiegare e predire il fenomeno del default. È un metodo che fornisce giudizi oggettivi, in tempi rapidi e a basso costo. L'introduzione di modelli misti, dove l'analista interpreta e corregge i risultati dei modelli automatici di scoring, permette di superare i limiti dei due approcci e di esaltarne i vantaggi.

Infine è possibile costruire dei modelli che sfruttano i dati provenienti dal mercato azionario (come ad esempio il modello KMV di Vasicek³) oppure i credit spread dei titoli obbligazionari.

La scelta di quale approccio adottare dipende dal segmento di clientela preso in esame (retail, piccole e medie imprese, large corporate), ad esempio nel caso di credito al consumo sono più adatti modelli di scoring automatici, dove è possibile ridurre tempi e costi di formulazione del giudizio data la standardizzazione delle informazioni a disposizione. Nel caso di grandi imprese, invece, i costi di un approccio personalizzato possono essere giustificati dai profitti attesi derivanti dal prestito.

Lo scopo è comune: ordinare le controparti in base al merito di credito attraverso

³ L'idea di base consiste nel rappresentare il valore dell'equity di un'impresa come un'opzione call fittizia con sottostante il valore di mercato degli attivi aziendali e strike price pari al valore di mercato delle passività. Il default coincide con un valore dell'equity negativo, ossia quando le passività superano le attività.

l'attribuzione di un voto⁴, per poi suddividerle in una serie di classi omogenee per rischiosità. Un numero maggiore di classi consente di discriminare la clientela in modo più omogeneo e di conseguenza trarre vantaggio da una politica di pricing mirata. Successivamente si attribuisce a ogni classe una PD (fase di calibrazione) in base ai suoi tassi di insolvenza medi storici su un orizzonte temporale annuale, con l'obiettivo di esprimere la reale frequenza di default del campione studiato e pertanto dare un significato economico a questi voti.

1.2.2 Loss Given Default

La Loss Given Default (LGD) rappresenta la quota di esposizione non recuperata in caso di default, si tratta di una componente conoscibile solo al termine del processo di recupero.

La stima della LGD risulterebbe sicuramente più agevole nel caso di titoli creditizi quotati in un mercato secondario di negoziazione liquido, l'osservazione del prezzo (valore di mercato) è, infatti, un vantaggio non di poco conto (metodo "market LGD").

Dal momento che i prodotti tradizionali delle banche sono strumenti creditizi illiquidi (ad es. mutui) senza un mercato secondario di contrattazione, esse devono ripiegare sul metodo di stima "workout LGD", in cui si sfrutta un database contenente tutte le informazioni relative ai default storici. In particolare, la stima del credito recuperabile si ottiene attualizzando tutti i flussi di cassa futuri attesi provenienti dal debitore, al netto di tutti i costi legali e amministrativi legati al processo di recupero, a un tasso di sconto adeguato per tener conto del fattore temporale e dell'incertezza dei flussi. Il tasso di recupero (recovery rate, RR) è espresso come il rapporto tra questa stima e il valore dell'esposizione al momento del default (EAD), con la LGD frazione complementare del RR.

I fattori principali che influenzano la LGD attengono alle caratteristiche della garanzia, della procedura di recupero e del debitore.

Tipicamente un elemento rilevante è il rapporto tra l'importo del finanziamento e il valore della garanzia sottostante (loan to value ratio): a valori bassi corrispondono maggiori probabilità di recuperare una quota elevata del credito insoluto. Inoltre vanno tenute in

⁴ I voti possono essere espressi in lettere (rating) o in numeri (score).

considerazione la seniority della garanzia, che impatta sul realizzo o meno della stessa in via preferenziale, e il suo grado di liquidità, che incide su tempistiche e rischio di prezzo della liquidazione.

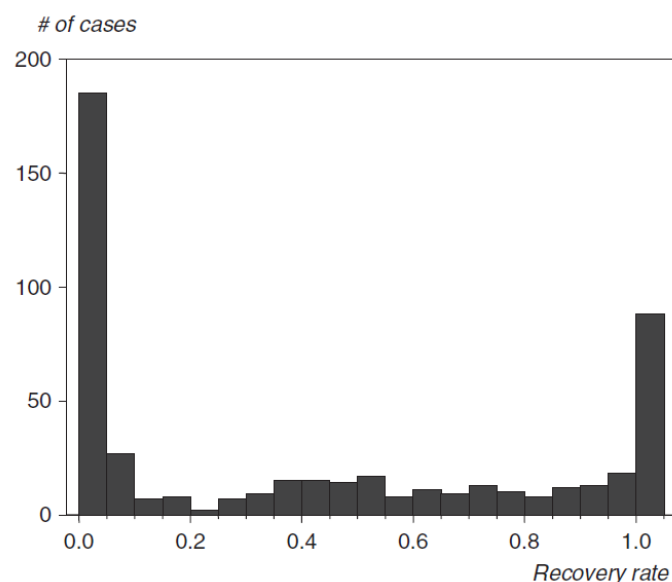
La procedura di recupero del credito dipende dall'ordinamento giuridico presente nel paese del debitore e impatta soprattutto sui tempi di realizzo, oltre che sui costi sopportati dal creditore e sull'efficace soddisfazione dello stesso.

Infine il grado di leva finanziaria del debitore e l'eventuale capacità di generare flussi di cassa in futuro sono ulteriori elementi che incidono sul livello di credito recuperabile.

È bene sottolineare che PD e LGD non sono variabili indipendenti, ma correlate positivamente. Entrambe dipendono dalle condizioni macroeconomiche, intuitivamente in caso di recessione aumenta sia la probabilità di insolvenza della controparte che il tasso di perdita derivante da tale evento.

Tornando al metodo workout LGD, sarebbe errato e poco utile calcolare semplicemente il valor medio del RR su tutte le esposizioni, dato che tipicamente la distribuzione empirica assume una forma a U, un esempio è fornito dalla Figura 1.

Figura 1. Distribuzione empirica dei tassi di recupero (RR) di una banca europea



Fonte: Resti e Sironi (2007).

Infatti, il valor medio rappresenterebbe in questo caso un dato a bassissima frequenza, è molto più probabile trovare casi con un RR vicino a 0 o 1.

L'obiettivo, pertanto, è costruire un database segmentabile in cluster con valori omogenei in base a una serie di variabili, quali: il tipo di esposizione, le caratteristiche della garanzia fornita, il loan to value ratio, ecc. Successivamente si procede al calcolo della media per ogni cluster, in modo da ottenere dati più aderenti alla realtà e soggetti a minore volatilità. In questo modo la banca associa ogni esposizione a un cluster in base alle caratteristiche della stessa. Rimane, comunque, un non trascurabile rischio che il tasso di recupero effettivo differisca da quello stimato (recovery risk).

1.2.3 Exposure at Default

L'Exposure at Default (EAD) è la componente più semplice da determinare, dipende sostanzialmente dalla forma tecnica del finanziamento.

Nei prodotti a esposizione certa, come ad esempio in presenza di un prestito con un piano di ammortamento prestabilito (mutui in primis) o di un'obbligazione, assume un valore deterministico dal momento che la banca conosce in ogni istante a quanto ammonta il credito verso la controparte.

In caso di prodotti a esposizione incerta (come le aperture di credito) il debitore beneficia di maggiore flessibilità e non è così immediato stimare il valore che l'esposizione assumerà in un istante futuro. In questa seconda ipotesi è possibile giungere a una stima dell'EAD tenendo conto dell'ammontare già erogato (il cosiddetto utilizzato) e di una quota (chiamata fattore di conversione creditizia, CCF) del credito accordato non utilizzato, che la banca prevede verrà sfruttata dalla controparte in prossimità del dissesto finanziario. Pertanto, l'unico elemento da stimare è il CCF, che è generalmente compreso tra 0 e 1, anche se in presenza di sconfinamento bancario può assumere valori superiori all'unità. La media del CCF su base storica può essere utilizzata come stima del parametro in questione.

1.2.4 Perdita attesa e perdita inattesa

Una volta stimate le tre variabili si giunge facilmente, sotto opportune assunzioni, alla determinazione della perdita attesa (EL) relativa a una singola esposizione:

$$EL = PD_{attesa} \times EAD_{attesa} \times [1 - E(RR)] \quad (1)$$

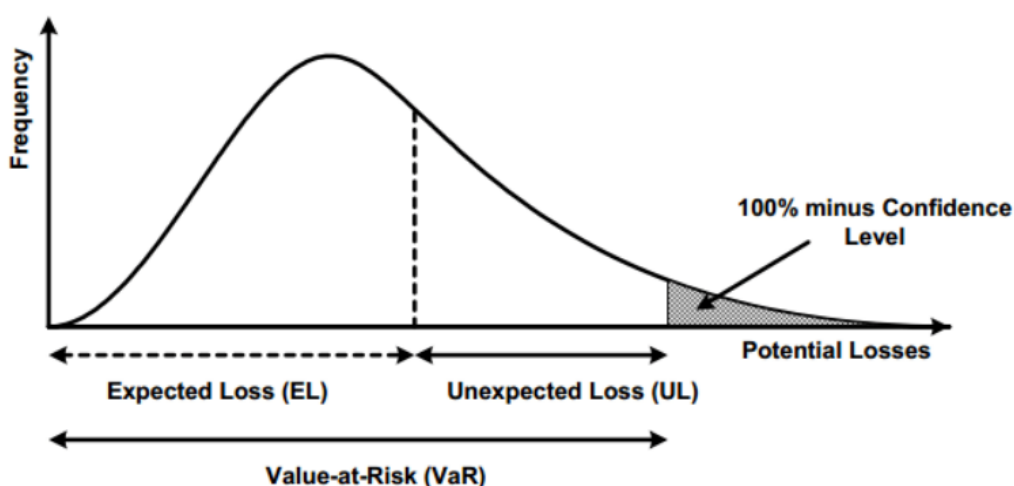
Le banche, per far fronte a questa componente, incorporano nei tassi di interesse dei finanziamenti un adeguato spread e si coprono con accantonamenti in conto economico.

Se la perdita attesa indica il valore medio della distribuzione di probabilità delle perdite future di una posizione, la perdita inattesa (UL) rappresenta la variabilità delle perdite attorno al valor medio e costituisce il vero rischio di un'esposizione. Il rischio consiste nella discrepanza tra il valore atteso e il valore effettivo ed è coperto tramite lo stanziamento di un adeguato ammontare di capitale da parte della banca. Così facendo le perdite eccedenti il valore atteso vengono assorbite dagli azionisti evitando fenomeni di sottocapitalizzazione, che possono mettere in dubbio la solvibilità della banca stessa.

Per quanto riguarda la variabilità delle perdite il calcolo non è così immediato. L'UL può essere misurata ricorrendo alla deviazione standard della distribuzione delle perdite future o, più comunemente, facendo ricorso a un percentile della distribuzione.

La distribuzione delle perdite associata a un portafoglio di crediti (Figura 2) assume tipicamente una forma asimmetrica, è concentrata su piccole perdite e presenta una coda pesante. Di norma si segue la logica del Value at Risk (VaR), che consiste nella stima della perdita massima derivante dall'esposizione al rischio di credito con un certo livello di confidenza statistica (99.9% secondo il Regolamento UE n. 575/2013 relativo ai requisiti patrimoniali) in un orizzonte temporale annuale. La perdita inattesa non è altro che la differenza tra il VaR e la perdita attesa.

Figura 2. Distribuzione teorica delle perdite e componenti del rischio di credito



Fonte: Farid J. (2014), Unexpected Loss (UL), Expected Loss & Economic Capital, FinanceTrainingCourse, 14 Dicembre 2014.

Preme sottolineare come il VaR, nonostante sia la misura di rischio più popolare, non è esente da limiti. Innanzitutto, fornisce come informazione solo la perdita associata a un determinato percentile ignorando gli scenari più sfavorevoli. Il VaR, soprattutto, non garantisce la subadditività, tale proprietà risulta essenziale per la definizione di misura di rischio coerente. La subadditività richiede che la diversificazione riduca sempre il rischio, in questo caso il VaR di un portafoglio composto da n titoli non dovrebbe essere mai superiore alla somma dei VaR calcolati sugli n singoli titoli. Una valida alternativa al VaR è l'Expected Shortfall (ES) che, tramite il calcolo del valore atteso delle perdite condizionato al superamento di un determinato percentile (VaR), cattura l'informazione lungo la coda della distribuzione e risponde ai requisiti di misura di rischio coerente.

Tra il 1994 e il 1997 vennero introdotti numerosi modelli di misurazione del rischio di credito per un portafoglio di esposizioni creditizie, tra i quali: CreditMetrics di J.P. Morgan, CreditRisk+ di Crédit Suisse e Credit Portfolio View di McKinsey. Lo scopo principale è stimare la distribuzione delle perdite associata al portafoglio per un dato orizzonte temporale, sulla quale poi poter calcolare il VaR.

L'approccio CreditMetrics sfrutta le matrici di transizione e la curva dei tassi di interesse per stimare la distribuzione del valore di mercato di un'esposizione tra un anno (logica del mark-to-market) includendo tutte le possibili variazioni del profilo di rischio della controparte. La matrice di transizione è una tabella dove sono riportate le probabilità di

variazione del rating a un anno, mentre la curva dei tassi fornisce i tassi di interesse utili per attualizzare i flussi di cassa futuri del prodotto creditizio in base alla classe di rating. Passando a una logica di portafoglio è essenziale conoscere anche le correlazioni tra le varie controparti, in questo caso vengono ricavate tramite un modello multifattoriale contenente una serie di fattori di rischio sistemico (come ad esempio il settore di riferimento e l'area geografica) e un fattore di rischio idiosincratco (specifico della singola controparte). In presenza di un numero elevato di titoli la stima della distribuzione del valore di mercato del portafoglio avviene tramite simulazioni con il metodo Monte Carlo.

In questo modo si perviene a una rappresentazione del profilo di rischio della banca. I modelli di misurazione del rischio di credito, sfruttano la stima delle perdite attese e inattese per: stabilire quanti fondi propri detenere, fissare il pricing dei prestiti, misurare le performance aggiustate per il rischio, ottimizzare la composizione del portafoglio.

1.3 Il contesto normativo

1.3.1 Basilea I

Le banche, nel fissare il livello ottimale dei fondi propri, devono rispettare determinati requisiti minimi imposti dal regolatore e basati sui rischi assunti. Le norme attuali sui requisiti patrimoniali minimi per il sistema bancario si ispirano ai vari Accordi di Basilea. Fino agli anni '70 la vigilanza regolamentare di tipo prudenziale era prettamente nazionale, senza alcuna forma di collaborazione transnazionale. Nel 1974 venne costituito il Comitato di Basilea per la vigilanza bancaria, formato da banche centrali e supervisor del sistema bancario appartenenti ai paesi finanziariamente più evoluti (attualmente rappresenta 28 paesi), con lo scopo di condividere esperienze, prassi di vigilanza e informazioni. Dal punto di vista giuridico gli accordi di Basilea sono gentlemen's agreements, cioè non giuridicamente vincolanti. Tuttavia, data la notevole influenza e autorevolezza, vengono recepiti negli ordinamenti di moltissimi Stati divenendo in quel momento regole vincolanti.

Nell'Accordo di Basilea del 1988 (denominato Basilea I) vennero introdotti, per la prima volta, dal Comitato di Basilea dei requisiti di capitale uniformi per le banche su scala internazionale, recepiti in seguito dagli ordinamenti di oltre 100 paesi. Gli obiettivi del

primo accordo sono molteplici: avere banche più solide scoraggiando l'assunzione di rischi eccessivi, garantire stabilità ai mercati finanziari internazionali, prevenire la competizione tra ordinamenti nazionali sui requisiti di capitale e assicurare pertanto un minimo di regole comuni a fini concorrenziali. Le regole vincolano le banche a detenere una dotazione patrimoniale adeguata in relazione ai rischi assunti, con l'obiettivo di proteggere gli stakeholder dal rischio che gli attivi bancari subiscano delle perdite tali da minare la solvibilità della banca stessa.

Il requisito fondamentale su cui verte Basilea I si traduce in un patrimonio di vigilanza non inferiore all'8% dei Risk-Weighted Assets (RWA), ovvero delle attività ponderate per il rischio. A ogni attività viene assegnato un coefficiente di ponderazione via via crescente all'aumentare del grado di rischiosità. Pertanto la logica è che un'attività più rischiosa debba assorbire più capitale regolamentare di una meno rischiosa.

In particolare le esposizioni vengono suddivise in 4 categorie:

- valori di cassa, attività verso le amministrazioni centrali e le banche centrali (se denominate in valuta del paese debitore), altre attività verso le amministrazioni centrali e le banche centrali dell'area OCSE, attività con garanzia in denaro o in titoli di amministrazioni centrali area OCSE (0%, "no rischio");
- diritti verso banche area OCSE o garantiti dalle stesse, diritti verso banche non OCSE con scadenza residua inferiore a 1 anno, attività verso enti del settore pubblico area OCSE diversi dall'amministrazione centrale (20%, "rischio basso");
- mutui garantiti integralmente da ipoteca su immobili residenziali (50%, "rischio medio");
- attività verso il settore privato, diritti verso banche non OCSE con scadenza residua superiore a 1 anno, attività verso amministrazioni centrali non OCSE (non denominate in valuta del paese debitore), investimenti in equity e proprietà immobiliari, tutte le altre attività (100%, "rischio pieno").

Tra gli asset vengono considerati a latere anche quelli fuori bilancio (garanzie e derivati). La ponderazione è influenzata sostanzialmente dalla natura e dall'area di appartenenza (OCSE o non OCSE) del debitore, dalla presenza e dal tipo di garanzie.

Una volta giunti al calcolo dei RWA, la banca deve detenere un livello di capitale di maggiore qualità (Tier 1 Capital) non inferiore al 4% e un capitale regolamentare totale (Tier 1 e Tier 2 Capital) non inferiore all'8% dei RWA.

Basilea I, nonostante il tentativo di armonizzazione dei requisiti patrimoniali a livello internazionale per il mezzo di regole semplici da comprendere e attuare, presenta evidenti limiti. Innanzitutto, il primo Accordo è incompleto: tiene in considerazione solo il rischio di credito, il rischio di mercato verrà incluso solo nel 1996 con un emendamento. Inoltre, è incoerente, semplicistico e produce effetti distorsivi: tratta tutti i crediti verso il settore privato come aventi lo stesso rischio (in modo omogeneo) incentivando le banche a finanziare le imprese più rischiose. Analoga considerazione vale per la discriminazione paesi OCSE-non OCSE. In più l'Accordo non considera la scadenza residua dell'esposizione e trascura del tutto i benefici della diversificazione del portafoglio creditizio.

In poche parole l'Accordo, con i relativi coefficienti di ponderazione, non riesce a riflettere in modo realistico i rischi assunti nel portafoglio delle banche.

1.3.2 Basilea II

Nel giugno del 2004 si giunge a un nuovo accordo (denominato Basilea II), entrato in vigore a gennaio 2007. Il Nuovo Accordo si fonda su tre pilastri: nuove regole per calcolare i requisiti patrimoniali minimi (più sensibili al rischio), supervisione da parte delle Autorità di Vigilanza su rischi e adeguatezza patrimoniale delle banche, trasparenza degli istituti verso il mercato.

Nel primo pilastro viene incluso anche il rischio operativo (prima non presente), oltre al rischio di credito e di mercato. Nel rischio di credito la novità è rappresentata dal fatto che ora le esposizioni appartenenti a una classe (ad es. settore privato) non sono pesate allo stesso modo, ma la ponderazione dipende dal grado di rischio, calcolato attraverso dei giudizi (rating) circa il merito creditizio, e dalla presenza di garanzie.

Nell'approccio standard i giudizi sono assegnati da agenzie di rating esterne (External Credit Assessment Institutions, ECAI), autorizzate dall'Autorità di Vigilanza. I fattori di ponderazione associati ai vari rating sono stabiliti da determinate tabelle di conversione (Tabella 1). Una volta ponderati gli asset in base ai coefficienti indicati nell'Accordo, si perviene alle attività ponderate per il rischio di credito.

Tabella 1. Coefficienti di ponderazione delle esposizioni nell'approccio standard, in base alla scala di rating di Standard and Poor's e Fitch

| | AAA | AAA- | AA+ | AA | AA- | A+ | A | A- | BBB+ | BBB | BBB- | BB+ | BB | BB- | B+ | B | B- | Below B- | Unrated | Past-due |
|---|---|------|-----|-----|-----|----|------|----|------|------|------|-----|------|------|----|---|----|----------|---------|----------|
| <i>Corporates</i> | 20% | | | 50% | | | 100% | | | 150% | | | 100% | 150% | | | | | | |
| <i>Sovereign entities</i> | 0% | | | 20% | | | 50% | | | 100% | | | 150% | 100% | | | | | | |
| <i>Banks</i> | 20% | | | 50% | | | 100% | | | 150% | | | 50% | | | | | | | |
| <i>Banks, depending on the country of incorporation</i> | 20% | | | 50% | | | 100% | | | 150% | | | 100% | | | | | | | |
| <i>Retail</i> | 75% | | | | | | | | | | | | | | | | | | | |
| <i>Residential real estate mortgages</i> | 35% | | | | | | | | | | | | | | | | | | | |
| <i>Non-residential real estate mortgages</i> | From 100% to 50%, upon discretion of the national supervisory authorities | | | | | | | | | | | | | | | | | | | |
| | 150% | | | | | | | | | | | | | | | | | | | |

Fonte: Resti e Sironi (2007).

In alternativa tali giudizi possono essere emessi dalla banca stessa (approccio basato sui rating interni, IRB). Tale metodo può essere perseguito solo dalle banche con una certa dimensione, che hanno le competenze tecniche e le risorse economiche per sviluppare all'interno un sistema di rating che soddisfi determinati requisiti richiesti dall'Autorità di Vigilanza. Nell'ambito dell'IRB, per misurare il rischio internamente sono previste in realtà due metodologie:

- un approccio di base dove la banca procede alla stima della sola PD a un anno, con LGD, EAD e durata residua (maturity) prefissate dal regolatore;
- un approccio avanzato dove l'istituto è autorizzato a stimare con propri modelli tutte e quattro le componenti del rischio di credito.

Nel secondo caso il sistema di rating deve essere il centro decisionale per la concessione del credito, per la gestione dei rischi, per l'allocazione interna del capitale e per il pricing, quindi non deve essere impiegato solamente per calcolare i requisiti di capitale.

L'approccio IRB giunge al calcolo dei requisiti di capitale sfruttando un modello per il rischio di credito a partire dai quattro parametri sopra citati, oltre a due fattori relativi al portafoglio (concentrazione e correlazioni).

Nel modello unifattoriale adottato da Basilea II l'andamento futuro del valore degli asset di ogni controparte (Z_i) è influenzato da due componenti: un fattore legato a variabili macroeconomiche X e una componente specifica di rischio ε .

Ogni controparte può essere più o meno esposta al rischio sistemico e idiosincratico, a seconda del valore assunto da ω .

$$Z_i = w \cdot X + \sqrt{1 - w^2} \cdot \varepsilon_i \quad (2)$$

Con correlazione tra due esposizioni creditizie pari a:

$$\rho(Z_i, Z_j) = w^2 \quad (3)$$

Qualora il valore degli asset scendesse al di sotto di una certa soglia, identificabile con il valore dei debiti, la controparte verrebbe considerata in default. La PD è la probabilità di oltrepassare tale soglia.

Sotto l'assunzione che X , ε e Z si distribuiscano come una Normale standardizzata, è possibile giungere alla formula richiesta dal regolatore per il calcolo della perdita inattesa, che corrisponde al requisito di capitale (K) unitario:

$$\text{Capital requirement} = K = \left[LGD \cdot N \left[\frac{G(PD)}{\sqrt{1-R}} + \sqrt{\frac{R}{1-R}} \cdot G(0.999) \right] - PD \cdot LGD \right] \cdot \sigma \cdot b \quad (4)$$

dove:

- PD e LGD sono parametri stimati dal modello interno della banca;
- $N[.]$ rappresenta la funzione di ripartizione di una Normale standardizzata;
- $G[.]$ rappresenta l'inversa della funzione di ripartizione di una Normale standardizzata;
- ρ (o R) indica la correlazione ed è un valore prestabilito dal Comitato in base al segmento

del portafoglio crediti⁵;

- b è un fattore di aggiustamento utile per includere il rischio di migrazione e si ricava da una formula che è funzione della durata residua del prestito (maturity) e della PD. Quanto più la PD è bassa e la maturity elevata, tanto più la banca sarà esposta a questo tipo di rischio e pertanto dovrà affrontare maggiori requisiti regolamentari;

- σ è un fattore di scala (pari a 1.06) che mira a limitare eccessive discrepanze tra l'approccio standard e IRB.

In pratica si sottrae la perdita attesa (pari al prodotto tra PD e LGD) dal livello di possibili perdite della banca, secondo il modello fornito dal Comitato, con un intervallo di confidenza al 99,9%. Per giungere alla perdita inattesa in termini assoluti è sufficiente moltiplicare K per l'EAD (stimata dal modello interno). Il prodotto corrisponde al requisito di capitale minimo.

Il passaggio dal requisito di capitale regolamentare al RWA è immediato:

$$RWA = K \cdot 12.5 \cdot EAD \quad (5)$$

dove 12.5 corrisponde al reciproco di 0.08, ossia il coefficiente patrimoniale minimo.

Per ottenere le attività totali ponderate per il rischio, sia nell'approccio standard che IRB, è necessario aggregare anche i RWA relativi a rischio di mercato e rischio operativo.

Su tale aggregato il Comitato richiede un patrimonio di vigilanza totale (Tier 1 e Tier 2 Capital) non inferiore all'8% e un Tier 2 Capital che non superi il Tier 1 Capital.

Il secondo pilastro di Basilea II consiste nel Processo di controllo prudenziale (Supervisory Review Process, SRP), articolato in due fasi: una interna alla banca (Internal Capital Adequacy Assessment Process, ICAAP) e una esterna (Supervisory Review and Evaluation Process, SREP).

L'ICAAP si riferisce al processo interno volto a valutare l'adeguatezza patrimoniale attuale e prospettica, in relazione al proprio profilo di rischio. In particolare, nel processo ICAAP la banca è tenuta a: individuare i rischi da sottoporre a valutazione; misurare e valutare i singoli rischi e il capitale interno necessario per fronteggiarli; misurare il capitale interno

⁵ È pari a $\rho = 0,12 \times [1 + \text{EXP}(-50 \times \text{PD})]$ per le imprese (con uno "sconto" per le PMI), gli enti sovrani e le banche, invece per la clientela al dettaglio (ad es. mutui ipotecari) è prefissato a $\rho = 0.15$.

complessivo e determinare il capitale totale tenendo conto delle regole prudenziali del primo pilastro. Le regole prudenziali sono regole minime, la banca poi deve valutare se i fondi propri che sono richiesti per soddisfare i vari coefficienti prudenziali sono sufficienti nel loro insieme per proteggere la banca dai vari rischi.

Nello SREP l'Autorità di Vigilanza esamina e valuta periodicamente l'ICAAP, formula un giudizio complessivo sull'intermediario e, in caso di anomalie, richiede l'attivazione di determinate misure correttive. Le Autorità di Vigilanza, inoltre, verificano l'adeguatezza del sistema di governance e della struttura organizzativa, la validità dei processi per l'identificazione, la misurazione, il monitoraggio e la gestione dei rischi propri dell'istituto.

Il terzo pilastro si concretizza nell'obbligo per le banche di fornire al mercato report dettagliati sui rischi assunti e sul suo capitale, in questo modo gli stakeholder dovrebbero essere in grado di valutare più prontamente e correttamente la rischiosità della banca (disciplina di mercato).

Basilea II ha migliorato, in primis con l'introduzione dei rating, la sensibilità delle ponderazioni ai rischi assunti dalle banche e ha spinto gli istituti di credito a essere più consapevoli e oculati circa l'allocazione del credito. Inoltre, viene riconosciuta la possibilità di ricorrere a modelli interni per la misurazione del rischio di credito, la ratio legis è spingere le banche verso la costruzione di modelli sempre più completi e accurati. L'incentivo a investire in questo approccio viene fornito dai vantaggi, in termini di assorbimento di capitale, che ne derivano.

Dall'altro lato, questo Accordo ha creato una disparità concorrenziale tra piccole e grandi banche, dato che le prime non hanno le risorse sufficienti per adottare il metodo IRB.

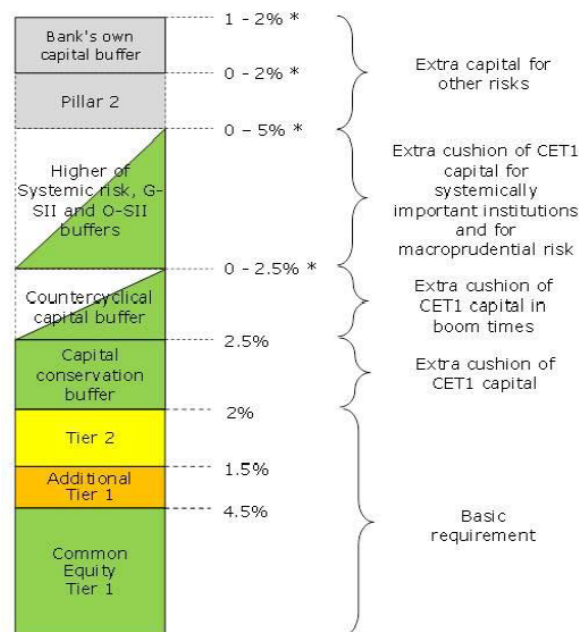
La crisi finanziaria del 2007-2009 ha fatto emergere i limiti e le criticità di questo Accordo, tra cui: il problema della pro ciclicità delle misure, la mancanza di regole per fronteggiare il rischio di crisi di liquidità e contenere la leva finanziaria, l'assenza di norme specifiche per le banche di importanza sistemica.

1.3.3 Basilea III

Alla luce di ciò, il Comitato giunge al cosiddetto Accordo di Basilea III nel Dicembre 2010. In realtà è un processo protratto nel tempo: negli anni seguenti il Comitato è intervenuto a più riprese e l'adeguamento al nuovo quadro normativo è avvenuto con gradualità. Nel 2013 l'Accordo viene recepito in Europa in parte in forma di Regolamento (CRR⁶), direttamente applicabile, in parte in forma di Direttiva (CRD IV⁷), recepita nel nostro paese nel giugno del 2015.

Le numerose misure introdotte sono volte sostanzialmente al rafforzamento e all'estensione dei tre pilastri. L'intento dell'Accordo è, infatti, migliorare la capacità delle banche di assorbire shock imprevisti derivanti da tensioni economico-finanziarie, migliorare la gestione del rischio e la governance, rafforzare la trasparenza e l'informativa delle banche al pubblico. Per quanto attiene al primo pilastro, i requisiti di capitale si fanno più stringenti dal punto di vista qualitativo e quantitativo (si veda la Figura 3).

Figura 3. Requisiti patrimoniali di Basilea III



* Assumed upper bounds (values can be higher)

Fonte: Magnus M. et al. (2017).

⁶ Regolamento UE n. 575/2013 del Parlamento europeo e del Consiglio dell'Unione Europea, o "CRR".

⁷ Direttiva 2013/36/UE del Parlamento europeo e del Consiglio dell'Unione Europea, o "CRD IV".

All'interno del Tier I Capital è stata introdotta una distinzione tra il Common Equity Tier 1 (CET1), capitale dotato di una maggiore capacità di assorbimento delle perdite, e l'Additional Tier 1 (AT1). Il coefficiente di solvibilità totale rimane all'8%, ma aumenta la qualità dei fondi propri rispetto a Basilea II, con la richiesta di un CET1 pari al 4.5% e di un AT1 al 1.5%.

Dal punto di vista quantitativo, vengono introdotte tre nuove riserve di capitale:

- un buffer (o "cuscinetto") di conservazione del capitale costituito da CET1 al 2.5% che, se violato, impone vincoli alla distribuzione di dividendi;
- un buffer anticiclico imposto dalle Autorità nelle fasi espansive del ciclo di credito, pari a un CET1 compreso tra lo 0 e il 2.5%, la ratio legis è consentire alle banche di sostenere famiglie e imprese anche nelle fasi avverse del ciclo economico;
- un buffer per le istituzioni finanziarie di rilevanza sistematica, dal momento che contribuiscono maggiormente a generare rischio sistemico, con il fine di ridurre l'azzardo morale.

A queste misure si aggiunge un ulteriore requisito patrimoniale vincolante (Pillar 2 Requirement, P2R), riferito a rischi sottostimati o non considerati nell'ambito del primo pilastro e stabilito dal supervisore nell'ambito del processo SREP.

Inoltre è previsto un indice di leva finanziaria, ovvero un rapporto tra Tier I Capital e il totale delle attività (senza alcuna ponderazione) non inferiore al 3%, atto a contenere l'effetto leva nel sistema bancario, fattore che ha amplificato la crisi finanziaria del 2007-2009. Nell'ambito del primo pilastro, il Comitato ha anche posto attenzione nel regolamentare il tema delle cartolarizzazioni e del rischio di controparte nella negoziazione di strumenti derivati.

Infine, si includono per la prima volta due requisiti per il rischio di liquidità: il Liquidity Coverage Ratio (LCR) e il Net Stable Funding Ratio (NSFR). Il primo prevede la detenzione di attività liquide di elevata qualità per coprire i fabbisogni finanziari nell'arco di 30 giorni in uno scenario di stress, ha lo scopo di fronteggiare crisi di liquidità di breve termine.

Il secondo assume un orizzonte temporale di lungo periodo e induce l'istituto a finanziarsi con una raccolta bancaria stabile.

1.3.4 Verso Basilea IV

Nel Dicembre del 2017 è stato raggiunto un accordo sul completamento di Basilea III, da alcuni ribattezzato Basilea IV per via dei più elevati requisiti di capitale che ne deriverebbero. Le misure entreranno in vigore nel 2023 ed è previsto un periodo di adeguamento fino al 2028. L'obiettivo esplicito della riforma è ridurre l'eccessiva variabilità degli attivi ponderati per il rischio (RWA), derivante (almeno in parte) dall'uso eccessivamente spinto dei modelli interni ai fini della riduzione dei requisiti di capitale, e renderli maggiormente trasparenti e comparabili tra banche diverse. Ciò anche al fine di limitare la disparità concorrenziale tra piccole e grandi banche.

La principale novità è rappresentata dal contenimento dei benefici derivanti dall'uso dei modelli interni, in primo luogo attraverso la fissazione di un output floor. In pratica, i RWA totali calcolati con l'approccio IRB non potranno essere inferiori al 72.5% dei RWA calcolati con l'approccio standardizzato⁸. Inoltre è stato fissato un livello minimo che gli input stimati internamente (PD e LGD) possono assumere.

L'uso dei modelli interni verrà limitato: non sarà più possibile impiegarli per il rischio operativo, dal momento che è previsto un unico approccio valido per tutte le banche; non si potrà ricorrere all'approccio IRB avanzato (ma solo IRB di base) per il rischio di credito di talune classi di esposizioni (verso imprese di medio-grandi dimensioni e banche).

Parimenti è stato revisionato l'approccio standardizzato per il rischio di credito migliorandone la sensibilità al rischio, vi è sostanzialmente una tabella di riferimento per le ponderazioni più dettagliata e discriminante rispetto all'Accordo precedente.

Ad esempio il coefficiente da applicare ai mutui residenziali diventa funzione del loan-to-value ratio. Infine, alle banche che adottano tale approccio è richiesta un'analisi circa l'adeguatezza del rating assegnato dall'ECAI e, se ravvisano un giudizio che sottostima il rischio associato all'esposizione, sono tenute ad applicare il coefficiente di classe superiore. Pertanto, tutte le misure vanno nella direzione di ridurre il divario esistente tra approccio IRB avanzato, utilizzato dai grandi gruppi bancari, e approccio standardizzato, impiegato da banche di dimensioni più contenute.

⁸ A seguito dell'introduzione dell'output floor, verrà rimosso il fattore di scala σ (pari a 1.06) presente nel calcolo dei RWA per il rischio di credito nell'approccio IRB. Per il resto, la formula rimane invariata rispetto a Basilea II.

Come si evince, nell'ultimo decennio legislatori e regolatori hanno risposto alle criticità emerse dalla crisi finanziaria introducendo imponenti misure di revisione della regolamentazione prudenziale per il sistema bancario a livello internazionale.

In Europa le riforme hanno avuto un impatto ancora maggiore, dal momento che si sono incanalate nel processo ambizioso dell'Unione bancaria, che ha portato a un meccanismo di vigilanza unico e una gestione armonizzata delle crisi.

Tra gli effetti positivi c'è sicuramente l'accresciuta solidità patrimoniale: tra il 2007 e il 2019 il CET1 e il Tier1 ratio delle banche italiane sono praticamente raddoppiati.

A tale risultato hanno concorso anzitutto i vari aumenti di capitale. In secondo luogo, un contributo importante è giunto dalla riduzione dei rischi nelle banche, ossia tramite il ridimensionamento del denominatore (RWA). Questi due fattori, insieme all'effetto degli utili trattenuti, hanno più che compensato la distruzione di capitale causata dalle perdite (Barbagallo, 2019). Oltre ad aver reso il sistema finanziario più stabile e resistente agli shock esterni, è maturata all'interno della banca una cultura del rischio.

Dall'altro lato la regolamentazione è diventata capillare e complessa, con il rischio di passare da una situazione pre-crisi di deregulation a un eccesso di regolamentazione.

Un eccessivo carico di compliance normativa produce oneri non indifferenti in capo alle banche, in particolare per quelle di medio-piccola dimensione. Questa situazione può essere vista come una lesione del principio generale di proporzionalità, con il rischio di indurre una forte concentrazione del mercato bancario.

Inoltre, un sensibile aumento dei requisiti di capitale e una riduzione dei rischi comporta a livello macroeconomico un rischio di stretta creditizia e un impatto negativo sulla crescita economica. Fidrmuc e Lind (2017) hanno esaminato gli studi sull'impatto macroeconomico dei più elevati requisiti patrimoniali richiesti da Basilea III. Nonostante l'incertezza delle varie stime prodotte, che variano in base alla scelta del modello e alle caratteristiche del paese preso in considerazione, in media si ottiene un effetto negativo sul prodotto interno lordo (PIL) di lungo periodo di circa 20 bps per ogni aumento di 100 bps del coefficiente patrimoniale.

Infine, il quadro normativo, con la limitazione dei rischi assunti in portafoglio e la maggiore patrimonializzazione delle banche, fa emergere il problema della loro redditività. Le banche vengono indotte a modificare, pertanto, il loro modello di business

verso la detenzione di asset meno rischiosi in portafoglio, il contenimento dei costi e l'orientamento alle commissioni.

Capitolo 2: Tecniche statistiche tradizionali per la previsione del default

2.1 Introduzione ai modelli di credit scoring

2.1.1 Il ruolo dei modelli di credit scoring

I modelli di credit scoring, come accennato in precedenza, mirano a esprimere un giudizio circa il merito creditizio della controparte attraverso l'emissione di un punteggio (o score). Lo score viene calcolato attraverso l'applicazione di un algoritmo o di un modello statistico a partire da variabili di tipo quantitativo, tipicamente indicatori economico-finanziari provenienti dai documenti contabili. L'idea di base è che la combinazione di determinati fattori chiave (indici di bilancio) possa condurre alla costruzione di un modello in grado di prevedere il fenomeno del default in modo robusto e affidabile. L'applicazione di un tale modello può essere di aiuto nella valutazione e nella prevenzione del rischio default e, più in generale, nella valutazione del rischio di credito.

Secondo Basilea IV, i modelli di credit scoring e le altre procedure meccaniche sono "ammissibili come base primaria o parziale per l'assegnazione dei rating e possono concorrere alla stima delle caratteristiche di rischio", tuttavia "si rende opportuno assegnare un ruolo al giudizio soggettivo dell'analista, che rivede e integra i risultati del modello per includere tutte le informazioni rilevanti e correggerne le debolezze".

Pertanto l'analisi quantitativa derivante da un modello di credit scoring, nonostante svolga un ruolo centrale nel processo di emissione dei rating interni, può risultare non sufficiente a catturare tutti gli aspetti inerenti l'affidabilità creditizia e lo stato di salute di un'azienda.

È bene sottolineare, prima di procedere, la distinzione tra il rating e lo score, sebbene siano due concetti correlati. Lo score è una valutazione sintetica della rischiosità creditizia del debitore, prodotta da un modello statistico automatico e basata esclusivamente su dati di tipo quantitativo, senza alcun intervento valutativo da parte dell'operatore. Lo score consente di effettuare un ordinamento in un intervallo continuo di valori.

Il rating è sempre un giudizio sintetico sul rischio di credito, ma è emesso nel contesto di un sistema di rating complesso che colloca il profilo di rischio di una controparte in una scala di rating discreta e fissata. Il sistema di rating sviluppato all'interno di una banca presenta solitamente una struttura modulare che integra, oltre al modulo finanziario-quantitativo, un'analisi andamentale (dati sul rapporto cliente-banca e cliente-sistema creditizio) e qualitativa (settore di appartenenza, qualità del management, ecc.).

Di norma queste ultime due valutazioni sono svolte avvalendosi dell'esperienza e delle competenze di un analista finanziario.

Gli Accordi di Basilea non prevedono l'adozione di uno specifico modello di rating, ma solo una validazione preventiva da parte delle autorità di vigilanza. Infatti, la scelta della struttura più idonea e di quale peso attribuire alla componente automatica rispetto al giudizio soggettivo spetta alla banca che, tra le altre cose, terrà conto anche del segmento di portafoglio considerato (ad esempio retail, piccole e medie imprese, large corporate) e delle relative informazioni a disposizione.

I modelli di credit scoring possono essere utilizzati in molti contesti all'interno di una banca:

- supporto nelle decisioni di accettazione o meno delle richieste di finanziamento. Tali modelli aumentano l'efficienza del processo decisionale, dal momento che si abbattano tempi e costi e si riducono gli errori commessi dall'uomo;
- monitoraggio della qualità degli attivi e gestione dei rischi in portafoglio;
- supporto nella classificazione dei prestiti e, di conseguenza, nella quantificazione di accantonamenti e del patrimonio di vigilanza;
- pricing del prestito grazie alla segmentazione della clientela in classi omogenee di rischio.

In sintesi, si tratta di modelli estremamente flessibili, capaci di adattarsi in diversi contesti applicativi e pertanto molto utilizzati nella valutazione del rischio di credito.

Sotto la spinta di Basilea II l'utilizzo di modelli di credit scoring è aumentato notevolmente, con finalità non solo di supporto alle decisioni di concessione del credito, ma più in generale come tecnica di analisi del rischio di credito.

2.1.2 Classificatori per il credit scoring

Il problema di fondo che si intende affrontare con le varie tecniche di scoring è la previsione dell'insolvenza di una controparte in un periodo futuro, solitamente l'orizzonte temporale è di un anno. La previsione richiede una risposta qualitativa, in questo contesto dicotomica. Il credit scoring, pertanto, può essere trattato come un problema di classificazione statistica binaria, con l'intento di assegnare correttamente le osservazioni alle due classi di appartenenza: 0 per le aziende sane e 1 per quelle

insolventi. Nel caso di modelli che restituiscono come output un valore continuo è possibile separare i risultati in due gruppi e ricondurre il problema a una classificazione binaria tramite l'utilizzo di un valore soglia (cut-off), consapevoli che ciò comporta una perdita di informazione rispetto al caso continuo.

In letteratura sono state proposte nel tempo numerose tecniche di classificazione (o classificatori) per affrontare un tale problema.

Il modello classificatore viene costruito a partire da un campione di osservazioni (training set) comprendente un numero sufficiente di imprese insolventi, ai fini di un efficace addestramento dello stesso. Per quanto riguarda le variabili predittive, la maggior parte degli studi impiega indici di struttura finanziaria-patrimoniale, redditività, liquidità, solvibilità, dimensione e crescita dell'azienda provenienti dai documenti di bilancio pubblici. Talvolta vengono integrati con informazioni macroeconomiche e settoriali.

Alcuni modelli includono variabili di mercato, che presentano il vantaggio di essere informazioni di alta qualità, ovvero: potenzialmente non distorte, disponibili e aggiornate in tempo reale e che incorporano tutte le informazioni sull'azienda. Nonostante da ciò discenda la possibilità di ottenere delle previsioni più accurate, nel presente elaborato non verrà preso in considerazione questo approccio, dato che il focus sarà rivolto al segmento delle piccole e medie imprese, per le quali tali dati sono raramente disponibili. Lo scopo della fase di addestramento è la ricerca dei parametri ottimali per il modello di apprendimento supervisionato, la cui efficacia sarà successivamente testata su un nuovo campione di osservazioni (test set). Le performance di un classificatore binario si valutano ricorrendo a varie metriche circa la capacità discriminatoria, la correttezza previsionale e gli errori di classificazione.

2.2 Tecniche tradizionali per lo scoring

2.2.1 I primi modelli

Il credit scoring per la valutazione del merito di credito ha mosso i primi passi nel segmento del credito al consumo, favorito dall'elevata standardizzazione delle informazioni a disposizione per questa tipologia di clientela. Il primo modello matematico per lo scoring risale al 1941 e fu proposto negli USA da David Durand, alla ricerca di pattern per discriminare i "buoni" e i "cattivi" tra coloro che facevano richiesta di una carta

di credito. Durand (1941) scoprì nel suo studio, ad esempio, che una maggiore stabilità occupazionale e residenziale dell'individuo è associata a un merito di credito maggiore, che deve tradursi pertanto in un punteggio più elevato.

Nel 1958 la Fair Isaac Corporation (FICO) costruì il primo modello di credit scoring (FICO score) dedicato a banche e società finanziarie statunitensi, tuttora il più utilizzato nell'ambito del credito al consumo. Il punteggio si basa sulla valutazione di cinque aree: la puntualità dei pagamenti, il livello di indebitamento corrente, la tipologia di finanziamenti in essere, la durata della storia creditizia e le richieste di nuovi finanziamenti.

In precedenza le decisioni sulla concessione del credito erano lasciate alle valutazioni soggettive degli analisti. L'emissione di giudizi oggettivi e uniformi in tempi molto rapidi ha consentito una riduzione dei costi e una maggiore precisione.

Un decennio dopo, l'applicazione di tali modelli iniziò a estendersi anche al segmento delle imprese, il pioniere in tale ambito di ricerca fu Edward Altman con il noto Z-score.

Da allora, grazie all'aumento della potenza di calcolo e la disponibilità di grandi quantità di dati, sono state proposte e perfezionate nel corso degli anni numerose tecniche di previsione dell'insolvenza, tutte accomunate dallo scopo di minimizzare gli errori di classificazione.

Una prima classificazione delle stesse può essere fatta distinguendo i modelli con approccio deduttivo, da quelli con approccio induttivo.

I primi si affidano al ragionamento, mediante la formulazione di assunzioni da parte del ricercatore, per tentare di spiegare le cause economiche strutturali sottostanti al default e cercano conferme delle proprie teorie nell'analisi dei dati.

I secondi fanno ricorso direttamente all'esperienza e cercano di trarre informazioni (regolarità empiriche) dai dati passati. Si tratta di modelli utili soprattutto nel caso di fenomeni complessi o impossibili da risolvere con un approccio deduttivo, il loro scopo è giungere a una buona approssimazione in tempi rapidi. Dall'altro lato presentano carenze sotto il profilo della trasparenza, infatti spesso non è chiara la struttura della relazione tra dati di input e output. Questo aspetto, in realtà, potrebbe tramutarsi in un vantaggio per i modelli induttivi se coloro che richiedono un finanziamento mirano a comprendere la logica sottostante al modello di scoring e attuare di conseguenza tutte le politiche di bilancio necessarie per influenzare il risultato a loro favore.

Nella categoria dei modelli deduttivi rientrano l'analisi discriminante lineare e i modelli di regressione; nella seconda reti neurali, algoritmi genetici e alberi decisionali.

Di seguito vengono descritti i modelli con approccio deduttivo più diffusi.

2.2.2 Modello univariato

Inizialmente venivano impiegati modelli statistici univariati, che analizzano lo stato di salute delle imprese facendo ricorso a una sola variabile.

Celebre è il lavoro di Beaver (1966), che studia la capacità di prevedere l'insolvenza da parte di una serie di indicatori contabili presi singolarmente. Un indice con un'eccellente capacità discriminativa risulta essere il rapporto Flussi di cassa/Debiti totali. La percentuale di osservazioni correttamente classificate con tale indicatore è pari all'87% con un orizzonte temporale annuale, al 79% a due anni dal fallimento. Un ulteriore indice significativo è il rapporto Utile netto/Totale attivo. Il lavoro di Beaver mostra come i dati contabili possono essere utili nel campo della previsione dell'insolvenza. Dall'altro lato le informazioni impiegate in questo approccio sono notevolmente ridotte, non sufficienti a garantire la costruzione di una tecnica previsiva affidabile, tant'è che da decenni i ricercatori si concentrano sull'analisi di più indicatori.

L'approccio multivariato considera simultaneamente più variabili e assegna a ognuna di esse un peso in relazione alla loro capacità predittiva. Tra i modelli multivariati più comuni rientrano l'analisi discriminante lineare (linear discriminant analysis, LDA) e i modelli di regressione Logit e Probit.

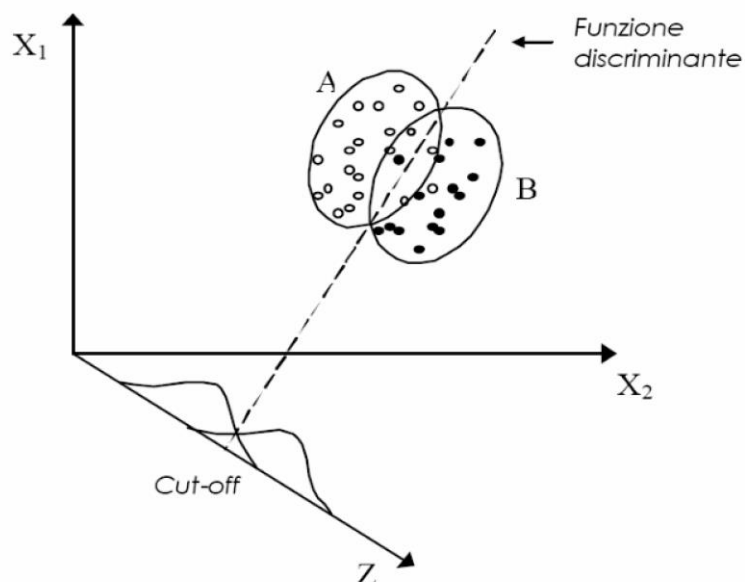
2.2.3 Analisi discriminante lineare

L'analisi discriminante lineare è stata proposta da Fisher (1936), che studia come distinguere due specie di piante di Iris attraverso quattro caratteristiche proprie dei fiori. Tale tecnica di classificazione è stata successivamente applicata per discriminare le imprese sane da quelle insolventi in base alle caratteristiche tipiche di ciascun gruppo, rintracciate nella fase di apprendimento.

L'analisi discriminante presenta il vantaggio di riuscire a ridurre la dimensione del problema attraverso il passaggio da n dimensioni (dove n è il numero di variabili) a 1

dimensione. Ciò avviene attraverso la creazione di un nuovo asse Z , sul quale vengono proiettate e sintetizzate tutte le informazioni provenienti dagli altri assi. Un esempio bidimensionale è fornito nella Figura 4.

Figura 4. Analisi discriminante lineare con due variabili X_1 e X_2



Fonte: Resti e Sironi (2007).

A ogni variabile esplicativa è associato un peso in relazione all'importanza nel prevedere il fenomeno dell'insolvenza, una semplice combinazione lineare sintetizza le informazioni in uno score. Esso idealmente rappresenta in modo conciso lo stato di salute dell'impresa nell'ottica del creditore. I coefficienti sono ottimizzati con l'intento di massimizzare la distanza tra gli score medi (centroidi) dei due gruppi e minimizzare la dispersione all'interno delle sottopopolazioni. L'ottimizzazione consiste in un procedimento oggettivo e univoco, frutto di un metodo di risoluzione analitico.

Per la classificazione delle osservazioni in uno dei due gruppi è sufficiente confrontare l'output con un valore soglia prefissato. Come punto di cut-off è possibile considerare la media dei due centroidi. In alternativa si può calibrare il valore soglia tenendo conto dell'impatto economico asimmetrico derivante dai due tipi di errori di classificazione e

della propensione al rischio del creditore⁹.

Una volta stimati i coefficienti, essi vengono mantenuti costanti ed è possibile calcolare lo score associato alle nuove osservazioni e, di conseguenza, assegnarle a una delle due classi. Nella Figura 4 si nota come le due distribuzioni non possono essere separate in modo perfetto, dal momento che esiste un'area di sovrapposizione. L'obiettivo della funzione discriminante è, pertanto, ridurre al minimo gli errori di classificazione.

Tra le critiche che i ricercatori hanno rivolto a questa metodologia vi sono le assunzioni sottostanti. I modelli LDA assumono omoschedasticità (matrici di varianza e covarianza uguali) tra le due popolazioni, si tratta di un'assunzione difficile da giustificare.

In presenza di eteroschedasticità è possibile sviluppare, ad esempio, un modello di analisi discriminante quadratica (QDA), con lo svantaggio di dover stimare un maggior numero di coefficienti. Inoltre, le variabili esplicative vengono assunte come distribuite come una Normale, tale assunzione spesso non trova riscontro nella realtà. Infatti, non è raro osservare indici contabili con valori limitati tra 0 e 1, quando la Normale assume valori compresi tra $-\infty$ e $+\infty$. L'analisi discriminante restituisce buoni risultati anche quando le assunzioni vengono violate, ma non si giunge alla soluzione ottimale e pertanto il modello perde di efficacia.

L'applicazione più famosa della LDA è quella dello Z-score di Altman. La prima versione risale al 1968 ed è costruita sulla base di un campione bilanciato di imprese industriali statunitensi quotate: 33 fallite nel periodo 1945-65 e 33 sane (Altman, 1968).

La funzione discriminante stimata è la seguente:

$$Z = 0.012 * X_1 + 0.014 * X_2 + 0.033 * X_3 + 0.006 * X_4 + 0.999 * X_5 \quad (6)$$

dove le variabili sono:

X_1 = Capitale circolante netto/Attivo totale;

X_2 = Riserve da utili/Attivo totale;

X_3 = EBIT/Attivo totale;

⁹ Un cut-off al di sotto della media conduce a un maggior numero di imprese classificate come insolventi, pertanto è frutto di una maggiore avversione al rischio da parte del soggetto finanziatore.

X_4 = Valore di Mercato del Patrimonio Netto/Valore contabile dei debiti totali;

X_5 = Ricavi/Attivo totale.

Il cut-off è posto a 2.675, un'impresa è classificata come sana se supera tale soglia.

Inoltre Altman propone due valori soglia (1.81 e 2.99) per individuare tre aree: rischio alto, zona di incertezza e rischio basso.

La capacità predittiva dello Z-score è molto buona un anno prima dell'insolvenza (circa il 95% delle imprese viene classificato correttamente), mentre l'accuratezza perde di efficacia all'aumentare dell'orizzonte temporale di previsione (72% e 48% a due e tre anni di distanza), come immaginabile. L'autore, nel corso degli anni, ha proposto varie estensioni del modello rivolgendo l'attenzione verso le imprese non industriali, le imprese non quotate o operanti in paesi emergenti.

In sostanza il modello Z-score è facile da comprendere e applicare, ma è stato costruito su un campione notevolmente ridotto e con un metodo che richiede il rispetto di assunzioni restrittive.

2.2.4 Modelli di regressione

L'analisi discriminante ha dominato per popolarità e diffusione il campo di ricerca negli anni '70 e '80, ma ormai risulta un approccio superato (Bellovary et al., 2007).

Una valida alternativa è rappresentata dalla categoria delle tecniche di scoring basate sulla regressione, che hanno visto un'ampia diffusione a partire dalla metà degli anni '80.

Il più semplice è il modello di probabilità lineare, che consiste in una regressione lineare multivariata con una variabile dipendente binaria (0 se l'evento fallimento non si verifica, 1 se si verifica). I parametri sono stimati con il metodo dei minimi quadrati ordinari (Ordinary Least Squares, OLS). Dal momento che la variabile dipendente si distribuisce come una bernoulliana, l'output della regressione può essere approssimativamente trattato come una probabilità di insolvenza. Il problema principale di questo approccio, oltre alla presenza di residui eteroschedastici che rendono le stime OLS non efficienti¹⁰, è che non garantisce valori dell'output limitati tra 0 e 1 rendendo il modello non adeguato

¹⁰ Una possibile soluzione al problema dell'eteroschedasticità è quella di utilizzare gli stimatori dei minimi quadrati ponderati (Weighted least squares, WLS).

ad affrontare problemi con risposta qualitativa. Per superare questa difficoltà è possibile ricorrere alla regressione logistica stimando un modello Logit.

Il modello Logit è un modello lineare generalizzato (GLM) che, dato il vettore delle variabili indipendenti X a disposizione, restituisce come output una stima della probabilità p che il debitore appartenga nel periodo successivo alla categoria delle imprese insolventi.

$$p = f(\alpha + \beta X) \quad (7)$$

con α costante e β parametri del modello.

Il modello si basa sull'assunzione che esistano una serie di variabili esplicative X (indicatori di bilancio) che abbiano una relazione causale con la probabilità di default dell'impresa. Si vuole sottolineare come la variabile dipendente Y sia una variabile qualitativa dicotomica, i cui esiti (sopravvivenza o default del debitore) sono osservabili, mentre p è una variabile latente. Alla relazione lineare è applicata una trasformazione non lineare f , utile a trasformare una variabile con valori reali in una probabilità.

f è una funzione sigmoide¹¹ chiamata funzione logistica standard¹², con equazione:

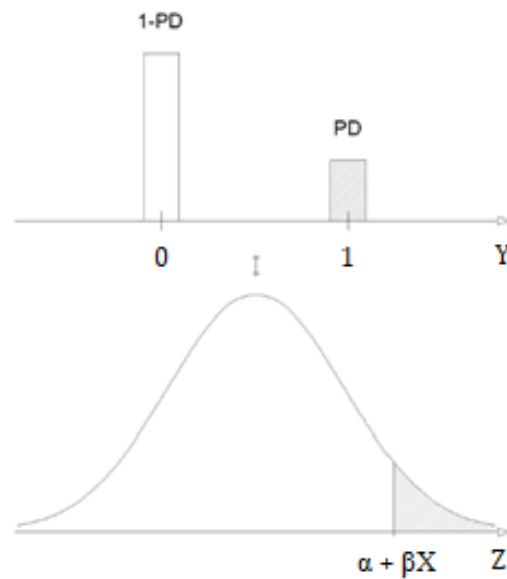
$$f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x} \quad (8)$$

Da un punto di vista teorico, il modello Logit prevede la connessione della variabile bernoulliana Y con una variabile casuale continua Z (si veda Figura 5).

¹¹ Chiamata così per via dell'andamento a S della curva.

¹² Precisamente si tratta della funzione di ripartizione della distribuzione logistica standard.

Figura 5. Relazione biunivoca tra la variabile bernoulliana Y e la variabile casuale Z



Fonte: Agnoletto M. (2017), *Il Value at Risk e la misurazione del rischio*, Tesi magistrale, Università Ca' Foscari, Venezia.

Su tale variabile di comodo si procede alla stima della soglia ($\alpha + \beta X$), con i dati a disposizione. Successivamente si calcola l'integrale dalla soglia a $+\infty$ ¹³ per giungere alla stima della probabilità di insolvenza¹⁴. Formalmente:

$$p = \int_{\alpha + \beta X}^{+\infty} f(h) dh \quad (9)$$

con $f(h)$ funzione di densità della logistica standard.

¹³ Oppure da $-\infty$ alla soglia, in base a quale significato si intende attribuire agli esiti 0 e 1 della variabile casuale Bernoulliana.

¹⁴ Tale output è, in realtà, una stima distorta della Probability of Default ed è più corretto interpretarlo come uno score, che ha lo scopo di ordinare analiticamente le controparti. Infatti, per giungere alla PD è necessario un secondo passaggio (fase di calibrazione), come accennato nel Cap. 1.

Dal momento che la probabilità p è pari a:

$$p = \frac{1}{1 + e^{-(\alpha + \beta X)}} \quad (10)$$

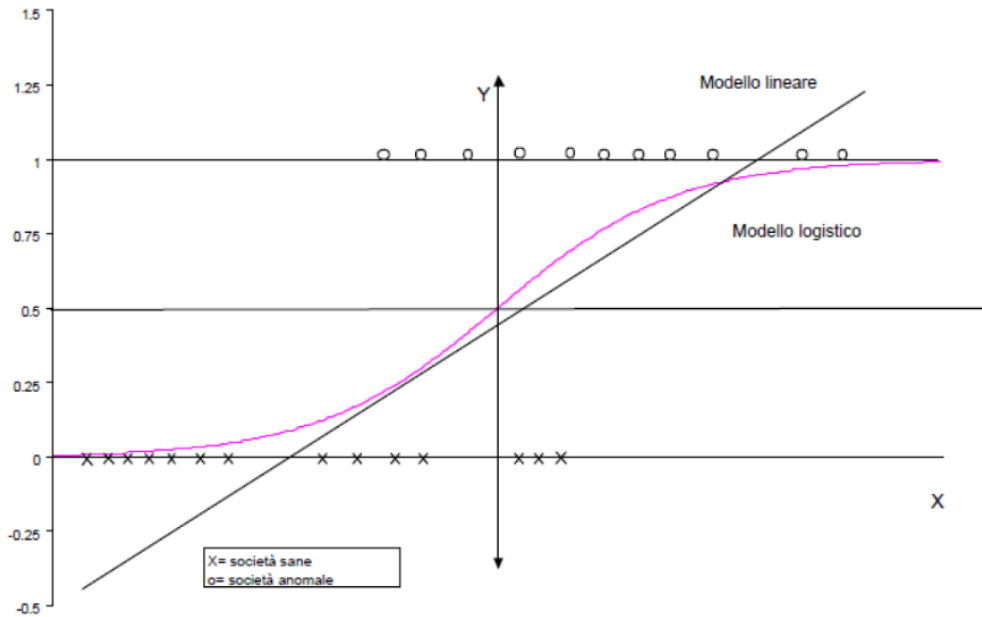
con dei semplici passaggi è possibile esplicitare la soglia $\alpha + \beta X$ come:

$$\ln \frac{p}{1-p} = \alpha + \beta X \quad (11)$$

dove il termine a sinistra è chiamato funzione logit (pari all'inversa della funzione logistica standard) e l'argomento del logaritmo è l'odd-ratio, ovvero il rapporto tra la probabilità di default p e il suo complemento a 1. In tal modo è possibile notare la differenza tra il modello di probabilità lineare, dove la probabilità p è messa in relazione diretta con $\alpha + \beta X$, e la regressione logistica, nella quale $\alpha + \beta X$ si relaziona invece con il logaritmo dell'odd-ratio.

Nella Figura 6 si comprende meglio perché una semplice regressione lineare non è pertinente in tale contesto: la pendenza della curva è costante e non varia al variare di X , pertanto viola l'assunzione di un codominio limitato tra 0 e 1. La regressione logistica, invece, impiega una funzione f con derivata prima che tende a 0. Non potendo assumere valori superiori a 1 e inferiori a 0, l'output è interpretabile come una probabilità.

Figura 6. Modello Lineare e Modello Logistico a confronto



Fonte: Szego G. e Varetto F. (1999), *Il rischio Creditizio: misura e controllo*, Utet, Torino.

Nella regressione logistica i coefficienti (α e β) vengono stimati facendo ricorso al metodo iterativo della massima verosimiglianza. Non è possibile ricorrere, invece, al metodo dei minimi quadrati ordinari (OLS) come in una regressione lineare, dal momento che il concetto di residui assume un significato differente.

Nel metodo della massima verosimiglianza si cercano quei parametri α e β tali per cui, per ogni osservazione, la probabilità di default stimata sia la più vicina possibile alla realizzazione della variabile casuale bernoulliana. In poche parole, si cerca di ottenere una probabilità vicina a 1 per le imprese fallite e vicina a 0 per quelle sopravvissute. Formalmente:

$$\ell(\alpha, \beta) = \prod_{i: y_i=1} p(x_i) \prod_{i': y_{i'}=0} (1 - p(x_{i'})) \quad (12)$$

L'obiettivo consiste nel massimizzare tale funzione di verosimiglianza¹⁵ trovando quei valori dei parametri ignoti che consentono di adattare in modo ottimale il modello ai dati. La soluzione non può essere ricavata in forma chiusa, è necessario un algoritmo iterativo per giungere alla stima dei parametri con il miglior fitting rispetto alle osservazioni.

L'analisi discriminante distingue la popolazione in due categorie e mira, attraverso dati di bilancio, a rintracciare quelle caratteristiche rilevanti per assegnare le unità osservate all'una o all'altra. Il modello logistico non intende prevedere l'appartenenza dell'impresa a una categoria, bensì il livello di difficoltà economico-finanziaria (il grado di salute) della stessa, che si traduce in una stima della probabilità di insolvenza. Se si vuole utilizzare la regressione logistica come classificatore è sufficiente fissare una soglia (ad esempio a 0.5) per suddividere le osservazioni in due classi e comparare i risultati dei due approcci.

L'approccio Logit è stato introdotto da Ohlson (1980), che utilizza un dataset composto da oltre 2000 imprese industriali quotate, fallite tra il 1970 e il 1976. Alla fine dello studio giunge alla stima dello O-score, quale alternativa allo Z-score di Altman per predire il dissesto finanziario. La formula dell'O-score¹⁶ è la seguente:

$$\begin{aligned}
 T = & - 1.32 - 0.407 \log(TA_t / GNP) + 6.03 \frac{TL_t}{TA_t} - 1.43 \frac{WC_t}{TA_t} + 0.0757 \frac{CL_t}{CA_t} \\
 & - 1.72X - 2.37 \frac{NI_t}{TA_t} - 1.83 \frac{FFO_t}{TL_t} + 0.285Y - 0.521 \frac{NI_t - NI_{t-1}}{|NI_t| + |NI_{t-1}|}
 \end{aligned}
 \tag{13}$$

dove

TA = Totale attivo e GNP = Reddito nazionale lordo;

TL = Totale debiti e WC = Capitale circolante netto;

CL = Passività correnti e CA = Attività correnti;

X è una variabile dummy che assume valore 1 se $TL > TA$, altrimenti 0;

NI = Utile netto e FFO = Fondi derivanti dalla gestione caratteristica;

¹⁵ Spesso risulta preferibile lavorare sul logaritmo della funzione di verosimiglianza, chiamata funzione di log-verosimiglianza.

¹⁶ Per convertire l'O-score in una probabilità è sufficiente utilizzarlo come input della funzione logistica standard.

Y è una variabile dummy che assume valore 1 se l'utile netto degli ultimi due anni è stato negativo, 0 altrimenti.

I tre fattori considerati più significativi dall'autore risultano: la dimensione dell'azienda, la struttura finanziaria (di cui si è tenuto conto, ad esempio, con il rapporto Totale debiti/Totale attivo) e la redditività (presa in considerazione, ad esempio, con l'introduzione dell'indice Utile netto/Totale attivo).

Nel campione usato per la stima dei parametri del modello la quota di osservazioni correttamente classificate è del 96.12% a un anno di distanza, a fronte di un valore del 91.15% ottenibile classificando tutte le aziende come sane. Ohlson ha individuato come cut-off ottimo (che minimizza la somma degli errori) il valore 0.38, che conduce nel campione di controllo a un'accuratezza previsiva dell'85%, con un'errata classificazione del 17.4% delle aziende sane e del 12.4% di quelle fallite.

In alternativa alla funzione logistica standard (modello Logit) è possibile adottare la funzione di ripartizione di una distribuzione normale standardizzata (modello Probit).

I due modelli sono assolutamente simili e producono risultati analoghi, una differenza risiede nella presenza di code più spesse nella distribuzione logistica standard rispetto alla normale, ovvero la distribuzione della prima è leptocurtica.

A introdurre l'utilizzo del modello Probit per la previsione dell'insolvenza è stato Zmijewski (1984) con lo studio di 840 aziende (di cui 40 non sane), quotate nel mercato finanziario statunitense e fallite tra il 1972 e il 1978. Il tasso di accuratezza raggiunto è del 99%.

Il modello originale di Zmijewski include solamente tre variabili esplicative:

$$Z = - 4.336 - 4.513 * X_1 + 5.679 * X_2 + 0.004 * X_3 \quad (14)$$

dove X_1 è la variabile scelta per rappresentare la redditività dell'azienda (Utile netto/Totale attivo), X_2 per la struttura del capitale (Totale debiti/Totale attivo) e X_3 per la liquidità (Attività correnti/Passività correnti).

Shumway (2001) critica il numero ridotto di variabili considerate, oltre a ritenerle statisticamente poco significative per prevedere l'insolvenza. Inoltre, sostiene che le

variabili del modello siano fortemente correlate tra loro e ciò può condurre al problema di multicollinearità.

2.2.5 Confronto tra LDA, Logit e Probit

I modelli Logit, Probit e LDA sono tra le tecniche di credit scoring più popolari in letteratura e utilizzate dagli addetti ai lavori.

È bene sottolineare come, per mettere a confronto le loro capacità predittive, non sia corretto comparare risultati provenienti da studi eterogenei per epoche temporali, tipologia di imprese incluse nel dataset e scelte metodologiche (ad es. definizione di default e selezione del campione di training e testing).

Lennox (1999) esamina un campione di 949 aziende (di cui 160 fallite) quotate nel Regno Unito, relativo al periodo 1987–1994. Egli verifica se l'adozione di uno dei tre approcci (Logit, Probit e analisi discriminante) possa comportare un vantaggio rispetto agli altri due. Il ricercatore confronta i tre modelli in termini di errori di I e II tipo tenendo conto di più punti di cut-off. I modelli Logit e Probit presentano performance previsive migliori rispetto all'analisi discriminante sia nel campione di stima che di controllo

Wu et al. (2010) raffrontano le performance di più modelli usando un ampio dataset comune, che copre un periodo compreso tra il 1980 e il 2006. I modelli considerati si riferiscono agli studi di Altman (1968), Ohlson (1980) e Zmijewski (1984) relativi rispettivamente allo Z-score, al modello Logit e al modello Probit. I risultati mostrano come il modello Logit sia notevolmente superiore rispetto al modello di Altman e ottenga performance simili, sebbene leggermente superiori, rispetto al Probit.

Kleinert (2014) mette a confronto le tre metodologie originali, ristimandone i coefficienti, in un dataset composto da imprese tedesche e belghe quotate per un totale di circa 7000 osservazioni. L'autrice conclude il suo studio accettando l'ipotesi nulla che vi sia una differenza nel tasso di accuratezza dei tre modelli considerati, con un livello di significatività del test del 5%. Il modello Logit si dimostra nettamente superiore nella previsione a 1, 2 e 3 anni rispetto all'analisi discriminante in entrambi i dataset e performa meglio del modello Probit.

Numerosi altri studi comparativi sono stati elaborati in letteratura e giungono a conclusioni analoghe. Tra le motivazioni per preferire l'adozione dell'approccio Logit vi è

la richiesta di assunzioni meno restrittive rispetto alla LDA risultando, pertanto, una tecnica più flessibile. Oltre a ciò, presenta il vantaggio di restituire risultati limitati tra 0 e 1, il che semplifica la lettura e l'interpretazione dell'output.

Dall'altro lato, quando la separazione tra le due classi è netta, i parametri stimati dalla regressione logistica soffrono di instabilità. Inoltre, in presenza di piccoli campioni e distribuzione delle variabili esplicative che si avvicina a una Normale, l'analisi discriminante presenta una maggiore stabilità. Tali condizioni, però, non si realizzano di frequente in tale contesto.

In sintesi, l'analisi discriminante, sebbene produca risultati tutto sommato accettabili, non risulta la più appropriata per affrontare il problema della previsione del default.

Il modello Logit sembra più adatto per questa tipologia di problemi, dato che è meno esigente dal punto di vista delle assunzioni ed è robusto, ciò trova riscontro nelle performance superiori rispetto ai modelli concorrenti.

In generale, i modelli statistici sono tecniche in cui la struttura è chiara e i risultati possono essere spiegati facilmente dal decisore al cliente. Il loro funzionamento è trasparente e sono semplici da implementare, oltre a ottenere buone capacità previsive. Il loro principale punto debole è l'incapacità di rilevare relazioni non lineari e complesse tra input e output, che sono rilevanti per la previsione dell'insolvenza. Inoltre, richiedono assunzioni in alcuni casi restrittive, che se non rispettate conducono a soluzioni non ottimali. A tal proposito, nel Capitolo III verranno presentate delle tecniche induttive alternative per superare tali criticità.

Capitolo 3: Tecniche di machine learning

3.1 Introduzione al machine learning

3.1.1 Definizione

L'apprendimento automatico (o machine learning, ML) è una branca dell'intelligenza artificiale¹⁷ (IA) e comprende al suo interno una serie ampia ed eterogenea di tecniche, finalizzate a prendere decisioni in modo efficiente e oggettivo. Il machine learning può essere definito come un processo di analisi dei dati, costituito da un insieme di regole formali, che presenta la peculiarità di non essere sviluppato per assolvere a un compito specifico e pertanto è applicabile in una varietà di contesti. Di norma si ha a disposizione uno specifico dataset e si vuole conoscere il processo ignoto che ha generato i dati. L'applicazione delle tecniche di machine learning mira a riconoscere gli schemi presenti all'interno dei dati e a estrarre informazioni circa la relazione esistente tra gli input e gli output. Gli algoritmi di ML apprendono il modello direttamente dagli esempi, senza assunzioni a priori sulla struttura.

L'obiettivo principale del machine learning è la capacità di generalizzare dalla propria esperienza, si tratta pertanto di un ragionamento di tipo induttivo. La generalizzazione ha successo quando, dopo aver proposto alla macchina degli esempi nella fase di apprendimento, essa è in grado di ottenere delle previsioni accurate anche su esempi mai visti prima. Una dimensione elevata del dataset a disposizione per la fase di apprendimento va a beneficio delle performance della macchina.

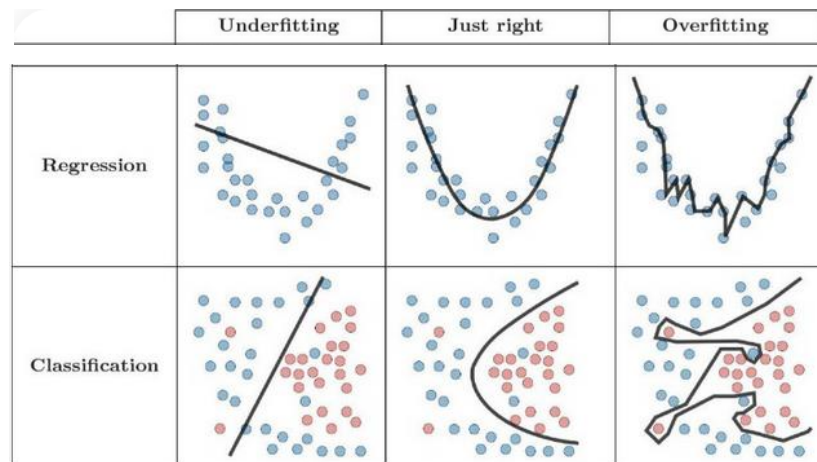
Il trade-off tra bias e varianza è uno dei modi per quantificare l'errore di generalizzazione. Il bias (o distorsione) rappresenta la differenza tra il valore previsto e il valore osservato. Un modello con bias elevato (e bassa varianza) non apprende abbastanza dai dati nella fase di addestramento e ipotizza una struttura più semplificata della funzione effettiva. In tal caso il modello manifesta uno scarso adattamento ai dati osservati (o underfitting), come evidenziato nella Figura 7. Un simile modello otterrà scarse performance sia nel dataset di addestramento sia in quello di test.

La varianza attiene alla variabilità delle previsioni del modello. Un modello con alta varianza (e poca distorsione) si concentra eccessivamente sugli esempi dimostrando un sovra-adattamento (o overfitting) rispetto ai dati (si veda a tal proposito sempre la Figura

¹⁷ L'intelligenza artificiale è definibile come la scienza che studia e sviluppa macchine capaci di svolgere compiti che normalmente richiedono l'impiego dell'intelligenza umana.

7), con una scarsa capacità di generalizzare. Si ipotizza una struttura eccessivamente complessa. Di conseguenza le performance previsive saranno ottime nel dataset di addestramento, ma pessime nei nuovi dati proposti alla macchina.

Figura 7. Esempi di underfitting e overfitting nel caso di regressione e classificazione

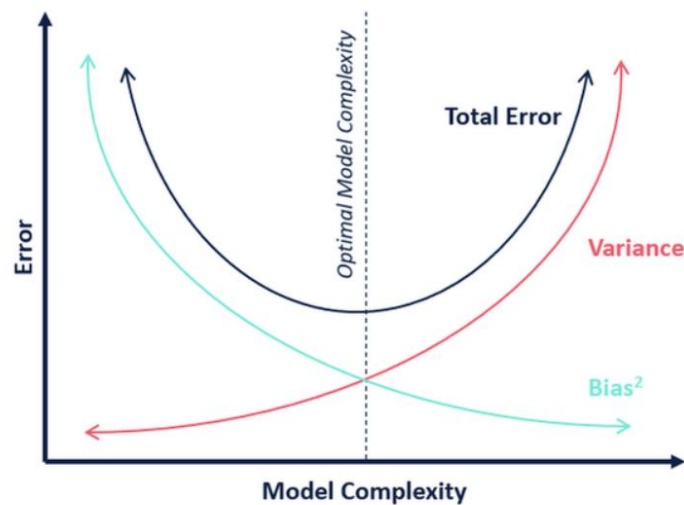


Fonte: Inouye David, *Introduction to Machine Learning (and Notation)*, Slide del corso “Introduction to Machine Learning”, Purdue University, 28/08/2019¹⁸.

Per ottenere buone prestazioni e giungere a una generalizzazione soddisfacente, la macchina deve essere in grado di stimare un modello con una complessità in linea con quella della funzione sottostante i dati. La relazione tra complessità del modello ed errore generato è illustrata nella Figura 8. È indispensabile trovare il giusto equilibrio tra bias e varianza per minimizzare l’errore totale.

¹⁸ Disponibile a <https://www.davidinouye.com/course/ece57000-fall-2019/lectures/lec3-2-intro-to-ml.pdf>

Figura 8. Trade-off tra bias e varianza



Fonte: AI Pool (2019), *Bias-Variance Tradeoff in Machine Learning*¹⁹.

Esistono tre tipologie di apprendimento automatico:

- apprendimento supervisionato (supervised learning), nel quale i dati vengono forniti alla macchina sotto forma di coppie di input e di output. Lo scopo dell'addestramento è rintracciare in modo approssimato la relazione sconosciuta che lega i dati in ingresso (input) ai risultati (output) desiderati. Si tratta probabilmente della metodologia ML più diffusa;
- apprendimento non supervisionato (unsupervised learning), in cui le classi non sono note a priori. L'algoritmo dovrà imparare la distribuzione dei dati di input nello spazio e raggrupparli sulla base di caratteristiche comuni. In poche parole, questi metodi consentono la creazione di classi omogenee (cluster) di dati;
- apprendimento per rinforzo (reinforcement learning), nel quale un agente compie una serie di azioni interagendo con un ambiente circostante per raggiungere un determinato obiettivo. Tale metodologia di apprendimento replica il modo in cui gli esseri umani e altri esseri viventi imparano: a ogni azione eseguita l'ambiente restituisce una ricompensa che può essere positiva (indica che l'apprendimento sta avvenendo nel modo corretto) o negativa (è necessario modificare il comportamento). Dopo un certo numero di iterazioni l'agente apprende in modo approssimato la funzione di rinforzo ottimale, ovvero la sequenza di azioni che massimizza le ricompense attuali e future.

¹⁹ Disponibile a <https://ai-pool.com/a/s/bias-variance-tradeoff-in-machine-learning>

Le tecniche di ML sono impiegate per la risoluzione di tre categorie di problemi:

- classificazione, dove l'output è rappresentato da valori qualitativi (discreti) e l'obiettivo dell'apprendimento consiste nell'assegnare le singole unità osservate alla giusta classe di appartenenza;
- regressione, in cui l'algoritmo deve essere in grado di stimare una risposta di tipo quantitativo con dominio continuo;
- clustering, dove le unità osservate vengono suddivise in gruppi non noti a priori.

Di norma, per compiti di classificazione e regressione si utilizzano algoritmi di apprendimento supervisionato, mentre per il clustering si adottano tecniche di apprendimento non supervisionato.

3.1.2 Applicazioni

Le prime ricerche sull'apprendimento automatico risalgono agli anni '50 e sono strettamente connesse al campo dell'intelligenza artificiale. Negli anni seguenti emergono alcune limitazioni di queste tecniche, che conducono a un calo di interesse e di investimenti in questo settore di ricerca. Dagli anni '90 si riscopre il machine learning e lo si differenzia dall'IA classica. Da allora due fattori hanno permesso di applicare questa tecnologia a vaste aree di ricerca e di diffondersi: l'aumento della capacità computazionale delle macchine e la disponibilità di enormi quantità di dati, favorita dall'avvento di Internet che ha facilitato la raccolta e la trasmissione delle informazioni.

Il machine learning è attualmente utilizzato per affrontare molti problemi di natura pratica, di seguito verranno forniti alcuni esempi in merito.

Programmi di riconoscimento ottico dei caratteri vengono solitamente costruiti tramite tecniche di apprendimento profondo (o deep learning). Tali applicazioni sono di grande utilità, ad esempio, per la gestione della corrispondenza da parte delle aziende che effettuano servizi postali. Lo scopo è estrapolare in modo automatico i caratteri presenti in un documento e trasformarli in un formato leggibile dalla macchina. L'addestramento consiste nel fornire una serie di esempi di immagini con la relativa soluzione. Di particolare complessità è il riconoscimento della scrittura a mano libera, dove gli stili di scrittura sono molto eterogenei e la classificazione dei caratteri non è immediata.

Le reti neurali artificiali, in particolar modo le reti neurali ricorrenti, offrono un'accurata performance anche nel campo del riconoscimento vocale. Negli ultimi anni Amazon, Apple, Google e Microsoft si sono cimentati con successo nello sviluppo di assistenti virtuali vocali (rispettivamente con Alexa, Siri, Google Assistant e Cortana). Gli input, ossia i comandi vocali, sono rappresentati da onde sonore, che vengono convertite prima in singole lettere e poi in frasi intere. Il testo viene successivamente associato a un'istruzione, che viene eseguita dal dispositivo.

Sempre all'interno delle tecniche di elaborazione e comprensione del linguaggio naturale trova spazio il tema della traduzione automatica di testi. Il compito di tali algoritmi è estremamente complesso per via dell'ambiguità intrinseca di questo linguaggio. Uno dei problemi principali è la comprensione del contesto da parte della macchina. Il modo migliore per affrontare problemi di questo tipo è imparare dall'esperienza incrementando con il tempo l'accuratezza dei risultati. Un noto esempio è il servizio Google Translate, che restituisce in tempo reale traduzioni in più di cento lingue diverse. Sebbene i risultati siano imperfetti, si tratta di tecnologie relativamente recenti con ampi margini di miglioramento.

Un altro compito in cui le tecniche di ML sono efficienti e pertanto largamente utilizzate è il riconoscimento facciale in tempo reale, il quale funge da metodo di identificazione biometrico delle persone. Principalmente vengono impiegate tecniche di deep learning, quali le reti neurali convoluzionali profonde. L'input è costituito da una o più immagini del volto, che vengono confrontate con i volti memorizzati nel database e usate per stabilire l'identità della persona (oltre a fornire una stima circa l'età, il genere, l'etnia, il riconoscimento di emozioni, ecc.). Il funzionamento di tali sistemi si basa sull'identificazione di alcuni punti chiave del volto umano, che vengono collegati tra loro per effettuare delle misurazioni (ad es. la distanza tra gli occhi). Si tratta di una tecnica che negli ultimi anni ha iniziato a diffondersi nel campo della sicurezza (ad es. negli aeroporti) e per l'accesso alle app dello smartphone.

Per quanto riguarda la guida dei veicoli senza l'intervento umano (guida automatica), si applicano modelli di apprendimento per rinforzo. L'ambiente con cui essi devono interagire è costituito da strade urbane e autostradali, dalle quali raccolgono informazioni circa il posizionamento e gli ostacoli presenti. Queste auto, infatti, sono dotate di un gran numero di sensori, telecamere e sistemi di localizzazione al fine di comprendere il

contesto in cui si trovano e agire di conseguenza. L'obiettivo è ambizioso e coinvolge la gestione di scenari di traffico molto complessi, che richiedono decisioni estremamente rapide e accurate.

Inoltre, gli algoritmi di machine learning possono essere di grande aiuto in campo sanitario, principalmente nella diagnosi precoce delle malattie. Gli input sono rappresentati da una molteplicità di informazioni personali sullo stato di salute del paziente (ad es. età, sesso, stile di vita, storia clinica, sintomi, struttura dei geni, radiografie, ecc.) e lo scopo è identificare in anticipo la presenza di anomalie e schemi sottostanti.

A titolo esemplificativo Microsoft ha sviluppato il progetto di ricerca InnerEye, che sfrutta un algoritmo derivante dall'unione di reti neurali e foreste casuali per lo studio di radiografie tridimensionali. Il progetto mira a identificare con precisione dove è localizzato il tumore. Sostanzialmente l'algoritmo analizza ogni voxel dell'immagine per estrarre informazioni (ad es. sul colore, la luminosità e l'intensità) e ottenere una classificazione delle cellule in sane e non sane. Ciò si dimostra molto utile ai fini della progettazione dell'operazione chirurgica e consente di ottenere una maggiore accuratezza e velocità rispetto allo stesso compito svolto da un medico.

A marzo il gruppo Alibaba ha sviluppato un sistema di diagnosi per il Covid-19, successivamente introdotto in più di cento ospedali della provincia dell'Hubei. Lo strumento è in grado di discriminare la polmonite derivante da Covid-19 rispetto alle altre polmoniti comuni tramite la comparazione di immagini TAC. Il sistema fornisce una risposta in 20 secondi, rispetto agli usuali 5-15 minuti impiegati da un medico, con un'elevata accuratezza. Nonostante il business di Alibaba non sia presente in campo sanitario, l'azienda ha addestrato le tecniche di apprendimento automatico a loro disposizione con migliaia di osservazioni dimostrando l'ampia versatilità di queste tecniche.

Infine, possono essere di aiuto nel complesso processo di scoperta di nuovi farmaci.

Un esempio è la prima molecola (DSP-1181) creata a scopo farmaceutico dall'intelligenza artificiale. La società Exscientia ha costruito un software per riuscire a selezionarla all'interno di decine di milioni di combinazioni possibili utilizzando una serie di informazioni sulla patologia in questione (malattia di Wilson). La cosa sorprendente è la rapidità con cui è stata scoperta la molecola: sono stati impiegati soltanto dodici mesi, a

fronte di un processo che può prolungarsi anche per cinque anni. L'efficienza di una tale tecnica può contribuire a rendere i farmaci più accessibili, dal momento che si risparmiano tempi e costi nel processo di ricerca. Infine, si sottolinea come la flessibilità propria di queste tecniche consente potenzialmente di estendere l'applicazione a ogni patologia.

3.1.3 Applicazioni in ambito finanziario

Machine learning e intelligenza artificiale sono in grado di offrire importanti applicazioni anche in ambito finanziario. L'industria dei servizi finanziari si coniuga bene con tali tecnologie, poiché è possibile sfruttare la disponibilità di una grande quantità di dati strutturati con un'ampiezza storica non indifferente.

Tramite l'elaborazione del linguaggio naturale è possibile sviluppare strumenti per la previsione dell'andamento del mercato azionario. Uno dei quali prevede l'estrazione automatica di informazioni da differenti fonti testuali (news, social network, report finanziari, ecc.) e la conseguente analisi del sentiment del mercato, ovvero la classificazione del testo come positivo, neutro o negativo. Per condurre un tale compito possono essere adottate reti neurali del tipo Long short-term memory (LSTM).

Sempre in tema di investimenti finanziari, molteplici sono le tecniche di machine learning utilizzabili, a titolo esemplificativo:

- Algoritmi genetici (GA) e Particle Swarm Optimization (PSO) per ottimizzare l'allocazione dei titoli in un portafoglio (Corazza et al., 2013);
- Group method of data handling (GMDH) per la previsione di serie storiche finanziarie (Corazza, 2000);
- Algoritmi di Reinforcement learning per la creazione di sistemi di trading automatici applicati al mercato azionario (Corazza et al., 2019);
- PSO per ottimizzare i parametri di indicatori e oscillatori di analisi tecnica impiegati in sistemi di trading automatici (Corazza et al., 2019).

Per quanto concerne il settore creditizio, si citano due casi (Mybank e Modefinance) per far comprendere le potenzialità di queste tecniche in tale ambito.

Mybank è una banca interamente online, legata al gruppo cinese Alibaba, che nei primi 5 anni di attività ha concesso credito per oltre 300 miliardi \$ a quasi 21 milioni di piccole e micro imprese cinesi, segmento spesso considerato troppo rischioso dalle banche tradizionali. Mybank raccoglie tutta una serie di informazioni, di tipo bancario e non²⁰, che vengono elaborate da tecniche di apprendimento automatico al fine di accettare o meno la richiesta di finanziamento. Tra le variabili considerate rientrano i dati sui pagamenti²¹, ma anche dati provenienti dall'attività della controparte, in qualità di venditore, su piattaforme di e-commerce, ad esempio: quante volte non ha consegnato la merce e a quanto ammontano i suoi flussi di cassa giornalieri medi.

La banca presenta, nel segmento delle piccole imprese, un tasso di approvazione dei prestiti di gran lunga maggiore rispetto alle banche tradizionali, con costi e tempi drasticamente ridotti. La concessione del credito richiede, infatti, un tempo sensibilmente ridotto (3 minuti) rispetto alle banche tradizionali (circa 30 giorni) e con costi operativi unitari nettamente inferiore (in media 3¥ per prestito contro 2000¥ dei concorrenti). Inoltre, il rapporto tra i crediti deteriorati e il totale dei crediti erogati (NPL ratio) si attesta a fine 2019 all'1.3%, contro una media del settore del 3.22%.

Modefinance è un'agenzia di rating italiana autorizzata dall'Autorità europea degli strumenti finanziari e dei mercati (ESMA) e specializzata nella valutazione del rischio di credito. Modefinance, tramite la metodologia automatica di credit scoring MORE (acronimo di Multi Objective Rating Evaluation), dichiara di essere in grado di valutare il merito di credito di tutte le aziende e di tutti gli istituti bancari del mondo.

MORE sfrutta al suo interno tutta una serie di tecniche di machine learning, tra le quali le reti neurali e gli algoritmi genetici. L'analisi si focalizza principalmente su redditività, liquidità, solvibilità ed efficienza della controparte, ma tiene conto anche di elementi quali l'area geografica, il settore economico di appartenenza, la struttura societaria e i diversi principi contabili adottati. Lo score viene poi integrato con il giudizio soggettivo degli analisti di Modefinance per giungere all'attribuzione di un rating. L'assegnazione di un rating è un elemento di particolare interesse per le piccole e medie imprese (PMI), dal

²⁰ Va sottolineato come l'utilizzo di big data nel settore creditizio cinese è favorito da una legislazione praticamente inesistente sul diritto alla privacy.

²¹ L'azionista di riferimento della banca è la Ant Financial, una delle principali società fornitrici di servizi di pagamento online in Cina.

momento che consente di ridurre l'asimmetria informativa tra le parti e di accedere, pertanto, al mercato del credito a condizioni più favorevoli.

Infine, il machine learning presenta grandi potenzialità anche nel settore assicurativo, dato che dall'analisi di un enorme quantità di dati sul comportamento dei clienti è possibile proporre dei prodotti in linea con le loro specifiche esigenze e pervenire a una stima accurata dei premi, coerente con il profilo di rischio della controparte.

3.2 Reti neurali artificiali

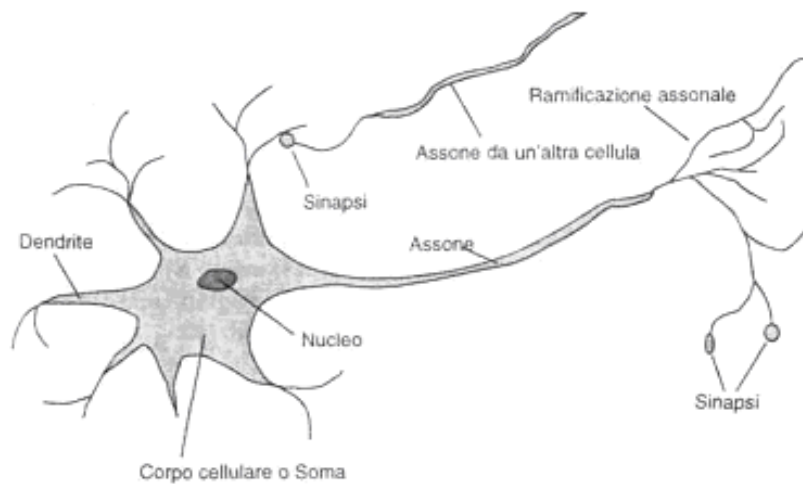
3.2.1 Neurone biologico

Le Reti neurali artificiali (Artificial Neural Networks, ANN) sono una categoria di tecniche di apprendimento supervisionato, che possono rappresentare una valida alternativa ai metodi di credit scoring tradizionali. Di seguito viene descritto il loro funzionamento e le performance ottenute in letteratura.

Le ANN rientrano nel campo della biomimesi, ossia “la disciplina che studia e imita le caratteristiche degli esseri viventi come modello cui ispirarsi per il miglioramento di attività e tecnologie umane” (Treccani, 2010). In questo caso, l'idea di fondo consiste nell'implementare degli utili algoritmi prendendo come riferimento la struttura e il funzionamento del cervello umano.

Il cervello è composto da circa cento miliardi di cellule nervose (o neuroni) altamente specializzate che raccolgono, elaborano e trasmettono informazioni sotto forma di impulsi nervosi. Il neurone costituisce l'unità funzionale del sistema nervoso, la cui struttura semplificata è presente in Figura 9.

Figura 9. Struttura di un neurone biologico



Fonte: Immagine tratta da Pensieri diretti (blog).

I dendriti di un neurone sono delle piccole ramificazioni, che ricevono le informazioni in entrata dai recettori sensoriali o da altri neuroni e le trasmettono al corpo cellulare.

Il corpo cellulare (o soma) contiene il nucleo ed è l'unità che si occupa di elaborare le informazioni giunte dai vari dendriti.

Infine l'assone è un prolungamento, che trasmette il segnale in uscita agli altri neuroni.

La connessione tra l'assone del neurone trasmittente e il dendrite del neurone ricevente avviene tramite la sinapsi.

La cellula nervosa è racchiusa da una membrana attraversata da canali ionici, i quali mettono in collegamento l'interno e l'esterno della cellula. Nel neurone sono, infatti, presenti degli ioni con carica positiva (ad es. il sodio, il potassio e il calcio) e ioni con carica negativa (ad es. il cloruro) e la membrana è di tipo semipermeabile, ossia permette il passaggio solo di certi ioni bloccando gli altri.

Il neurone a riposo mantiene una differenza di potenziale elettrico tra l'esterno e l'interno di circa -70 millivolt (mV). L'ambiente all'interno della cellula è pertanto più elettronegativo di quello esterno, questa differenza è data principalmente da una maggiore concentrazione di ioni sodio all'esterno della cellula. Quando la cellula viene eccitata da uno stimolo esterno si aprono i canali del sodio, che fanno affluire ioni sodio dall'esterno e, come conseguenza, il potenziale di membrana si muove verso 0 mV (fase

di depolarizzazione). Se il neurone raggiunge il livello di soglia critico (circa -55 mV) allora si genera un potenziale d'azione di ampiezza costante, che dà origine all'impulso nervoso. Altrimenti nessun impulso verrà generato seguendo il principio del tutto-o-nulla.

Un neurone viene considerato più o meno attivo in funzione della quantità di potenziali d'azione generati in un arco di tempo (frequenza di scarica) e non dell'ampiezza degli stessi, dal momento che il voltaggio è lo stesso per tutti.

Per quanto riguarda il potenziale di membrana, questo scende nuovamente verso valori negativi (fase di ripolarizzazione) grazie al lavoro della pompa sodio-potassio e il potenziale ritorna gradualmente ai valori di partenza.

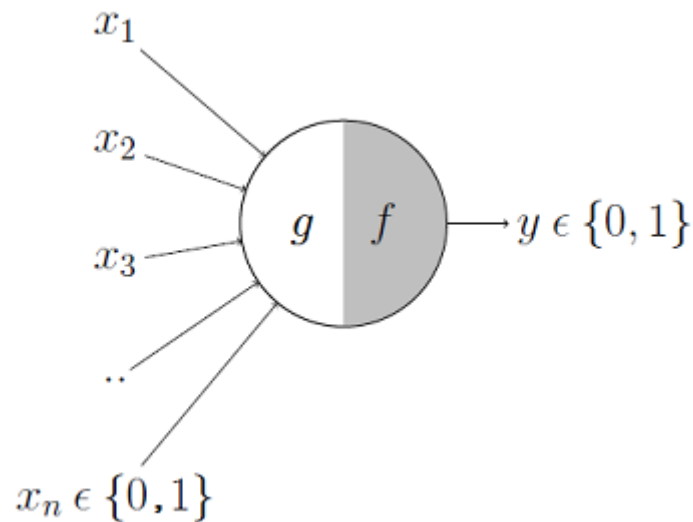
I dendriti di un neurone ricevono dalle sinapsi sia segnali elettrici eccitatori (o potenziali postsinaptici eccitatori, PPSE) che inibitori (o potenziali postsinaptici inibitori, PPSI).

Di norma un singolo segnale eccitatore non è in grado di attivare un potenziale d'azione. Pertanto un aspetto cruciale è l'aggregazione di questi stimoli, ovvero: gli stimoli generati da più sinapsi su un unico dendrite (sommazione spaziale) e gli stimoli generati dalla stessa sinapsi in rapida successione (sommazione temporale) vengono sommati nel corpo cellulare per aumentare il potenziale di membrana. In questo modo all'aumentare dei segnali eccitatori aumentano le probabilità di superare la soglia e generare l'impulso nervoso.

3.2.2 Neurone artificiale

Il neurone artificiale costituisce il tentativo di descrivere matematicamente il funzionamento di un neurone biologico. McCulloch e Pitts (1943) proposero il primo modello matematico del neurone (si veda Figura 10).

Figura 10. Rappresentazione grafica del neurone di McCulloch e Pitts



Fonte: Chandra A. L., *McCulloch-Pitts Neuron — Mankind's First Mathematical Model Of A Biological Neuron*, in Towards data science (blog), 24/07/2018²².

Il modello originale presenta una struttura elementare e un funzionamento molto semplice. I dendriti ricevono le informazioni in ingresso, ossia gli input x . Il corpo cellulare è responsabile dell'aggregazione prima (funzione g) e della trasformazione poi (funzione f) di questi input. L'assone infine trasmette la risposta (output) y del neurone.

Nel modello originale gli input e l'output sono costituiti da variabili booleane, ossia possono assumere come valori solo 0 (paragonabili ai segnali inibitori) e 1 (paragonabili ai segnali eccitatori).

L'aggregazione (funzione g) consiste in una semplice sommatoria, del tipo:

$$g(x_1, x_2, x_3, \dots, x_n) = g(\mathbf{x}) = \sum_{i=1}^n x_i \quad (15)$$

Il risultato viene utilizzato come input della funzione di trasformazione (funzione f) a gradino:

²² Disponibile a <https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1>

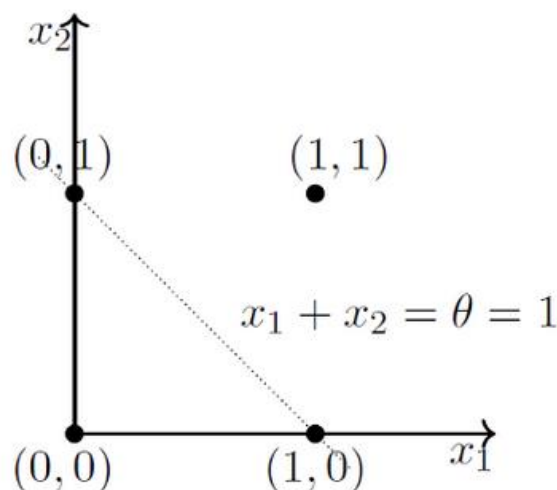
$$y = f(g(\mathbf{x})) = \begin{cases} 1 & \text{if } g(\mathbf{x}) \geq \theta \\ 0 & \text{if } g(\mathbf{x}) < \theta \end{cases} \quad (16)$$

Dove θ è il valore soglia, il neurone si attiva solo se il risultato della sommatoria è maggiore o uguale a tale livello.

Tramite il neurone di McCulloch-Pitts è possibile rappresentare alcune funzioni booleane, come ad esempio le funzioni AND, OR, NOR e NOT. Ogni funzione ha un proprio valore soglia, computato manualmente, che suddivide i dati nelle due classi. A titolo esemplificativo la funzione AND richiede che tutti gli input siano eccitatori (valore pari a 1), pertanto la soglia è fissata a n , ossia pari al numero di input presenti nel neurone. Nel caso della funzione OR la soglia è uguale a 1 indipendentemente dal numero di input, infatti è sufficiente la presenza di un solo input eccitatore.

Analizzando il problema dal punto di vista geometrico (come in Figura 11), risulta evidente come il modello di McCulloch-Pitts svolga un compito di classificazione binaria di tipo lineare.

Figura 11. Interpretazione geometrica della funzione logica OR con due input booleani



Fonte: Chandra A. L., *McCulloch-Pitts Neuron - Mankind's First Mathematical Model Of A Biological Neuron*, in *Towards data science (blog)*, 24/07/2018²³.

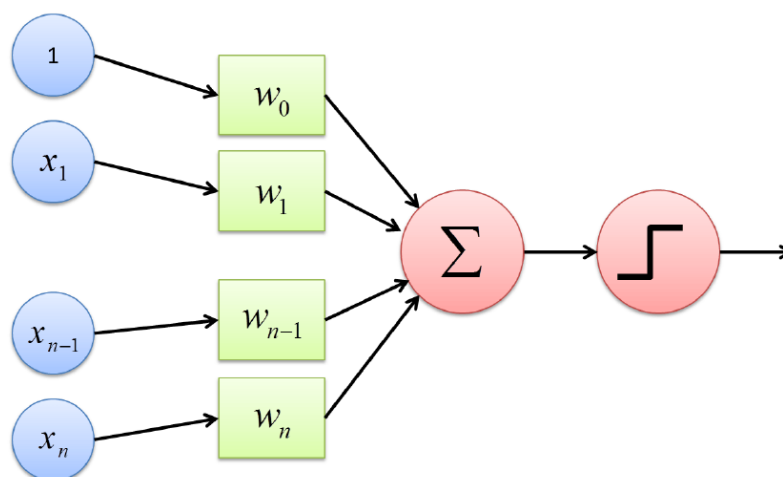
²³ Disponibile a <https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1>

In questo caso una semplice funzione lineare è sufficiente per separare correttamente le due classi: le combinazioni sotto soglia con output 0 e quelle sopra soglia con output 1. Ciò non vale per funzioni più complesse, come le funzioni non lineari, per risolvere le quali è necessario ricorrere a modelli più evoluti.

Il modello di McCulloch e Pitts è un modello semplificato, che presenta alcune evidenti limitazioni: non considera input di tipo non booleano, richiede il computo manuale della soglia, considera gli input egualmente importanti (non vengono pesati) e non è in grado di affrontare funzioni non separabili linearmente.

Rosenblatt (1958) riesce a superare alcune di queste criticità con il noto perceptrone (Figura 12), ossia un modello di neurone artificiale più generale rispetto a quello di McCulloch-Pitts.

Figura 12. Struttura del perceptrone di Rosenblatt



Fonte: Corazza Marco, *Supervised learning methods*, Slide del corso Fintech, Università Ca' Foscari di Venezia.

Gli input x del perceptrone appartengono ora all'insieme dei numeri reali²⁴ e viene introdotta una costante (o bias) $x_0 = 1$. Dal punto di vista biologico gli input rappresentano i segnali elettrici in entrata, che il neurone riceve da altri neuroni attraverso i dendriti. A ogni input è associato un peso sinaptico w , che indica quanto è forte la connessione tra due neuroni consecutivi. Questi coefficienti sono ignoti inizialmente e sono oggetto di

²⁴ Si supera, quindi, il vincolo di variabili booleane del modello di McCulloch-Pitts.

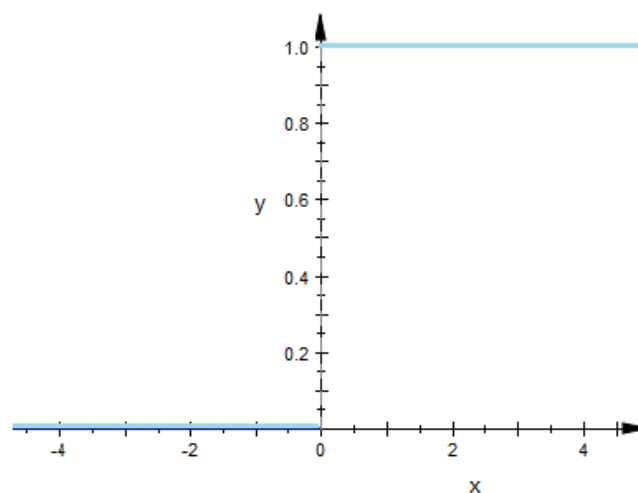
apprendimento da parte della macchina.

Il corpo cellulare processa le informazioni ricevute effettuando due operazioni: aggrega questi segnali elettrici e li trasforma in un nuovo segnale elettrico, che corre lungo l'assone e viene rilasciato ad altri neuroni tramite le sinapsi. L'aggregazione consiste in una semplice somma pesata degli input²⁵:

$$\sum_{i=0}^n w_i x_i = w_0 x_0 + \dots + w_n x_n \quad (17)$$

Il passaggio da tale aggregazione all'output finale avviene tramite l'impiego di una funzione di trasformazione (o funzione di attivazione) a gradino, chiamata funzione gradino di Heaviside²⁶. È evidente come tale funzione (Figura 13), non sia né continua né differenziabile.

Figura 13. Funzione gradino di Heaviside



Fonte: Calculus How To (2019), *Heaviside Function (Unit Step Function)*²⁷.

²⁵ In modo analogo, l'aggregazione può essere vista come il prodotto tra il vettore dei pesi w e il vettore degli input x .

²⁶ In alternativa è possibile utilizzare una delle varie funzioni di attivazione proposte in letteratura, a cui si fa cenno in seguito.

²⁷ Disponibile a <https://calculushowto.com/heaviside-function/>

La trasformazione costringe l'output ad assumere solo due valori (0 e 1), in questo modo il perceptrone può fungere da classificatore binario.

Dal punto di vista geometrico, lo scopo del perceptrone è costruire un iperpiano che separi nel miglior modo possibile l'input space, ossia lo spazio n-dimensionale contenente tutti i possibili input, in due classi. A tal fine è necessario addestrare il modello affinché impari i parametri ottimali, che minimizzano una determinata funzione di perdita.

Nella ricerca dei pesi sinaptici ottimali, gli esempi vengono forniti al perceptrone uno alla volta (online learning). In questo modo i parametri vengono aggiornati di continuo e in modo graduale. Di norma il vettore dei pesi viene inizializzato in maniera completamente casuale e seguendo un approccio iterativo si aggiustano leggermente i parametri a ogni passo.

In particolare sono tre le fasi che caratterizzano il processo di apprendimento, per ognuna delle M osservazioni (coppie di input-output): il perceptrone restituisce un output, si computa l'errore di previsione (ovvero la differenza tra il valore stimato e quello effettivo) e si modificano i pesi tramite la regola delta (o delta rule).

Nel caso di funzione di attivazione lineare, la delta rule assume la forma (semplificata):

$$w_i^{t+1} = w_i^t + \Delta w_i^t = w_i^t + \alpha * (y_j - o_j) * x_{i,j} \quad (18)$$

dove:

- w_i sono i pesi sinaptici con $i = 0, \dots, n$;
- t conta il numero delle epoche²⁸, con $t = 1, \dots, k$;
- α è una costante chiamata tasso di apprendimento. Di norma si fissa un valore prossimo allo zero (ad esempio 0.01), in modo da ottenere un apprendimento più lento e preciso;
- $(y-o)$ è la differenza tra l'output desiderato e quello stimato dal perceptrone, ossia l'errore di previsione;
- x_{ij} è l' i -esimo input della j -esima osservazione.

²⁸ Gli esempi, infatti, vengono riproposti alla macchina più volte.

In pratica, se la previsione è corretta non vi è necessità di modificare i pesi. Altrimenti i pesi vengono aggiustati in modo da far convergere l'output del perceptrone a quello desiderato. In un'epoca tale procedura viene ripetuta per tutti gli n pesi sinaptici e per tutte le M coppie di input-output presenti nel dataset di addestramento.

La fase di addestramento (o training) è conclusa quando, dopo numero adeguato di epoche k , si ottiene una funzione di perdita contenuta.

Minsky e Papert (1969) mostrano come il modello di Rosenblatt non riesca ad apprendere la funzione XOR booleana. L'operatore logico XOR restituisce l'output "Vero" se gli input sono diversi tra loro, se gli input sono uguali il risultato è "Falso".

Il principale limite di questo modello è di non essere in grado di risolvere problemi di classificazione con dati non linearmente separabili, come appunto la funzione XOR.

Il lavoro di Minsky e Papert generò un pessimismo circa le potenzialità delle reti neurali artificiali, che condusse a un periodo (il cosiddetto inverno dell'intelligenza artificiale) di calo degli investimenti e di interesse in questo campo di ricerca.

3.2.3 Multilayer Perceptron (MLP)

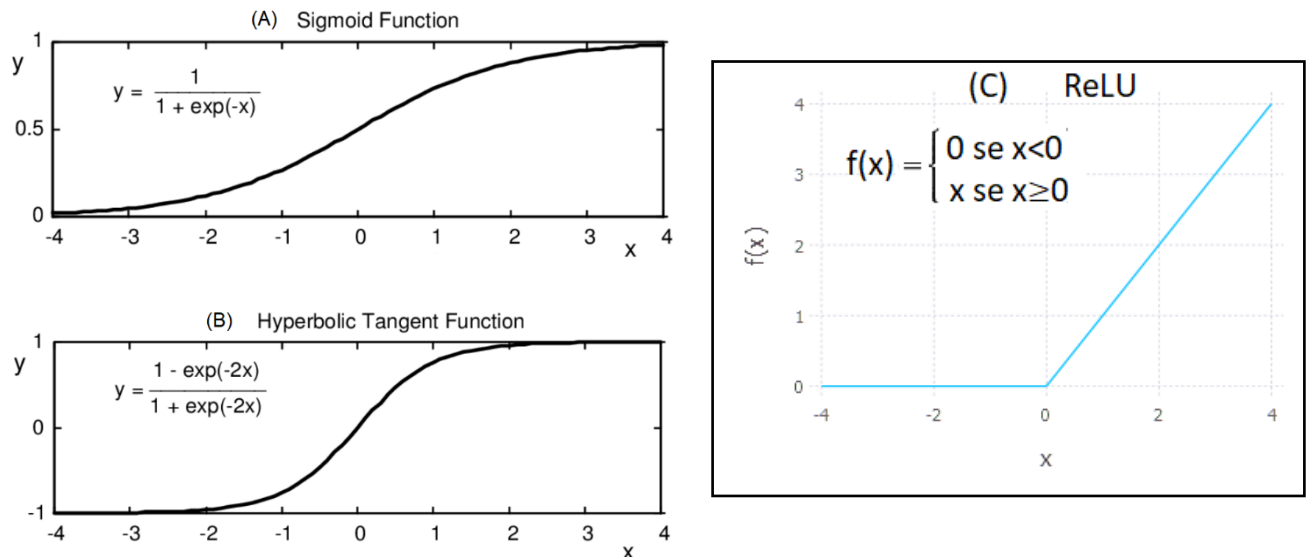
In realtà negli anni seguenti si comprese come il perceptrone, nonostante lo svolgimento di compiti semplici, potesse costituire il modello di partenza per la progettazione di reti più complesse. Le limitazioni del perceptrone di Rosenblatt, infatti, possono essere superate modificando la funzione di attivazione e, soprattutto, la struttura del modello.

Per quanto riguarda il primo aspetto, è utile che la funzione di attivazione abbia alcune proprietà:

- continuità. In questo modo l'informazione che viene persa con una funzione a gradino può essere ora sfruttata. L'apprendimento migliora in presenza di una funzione continua in tutto l'intervallo;
- differenziabilità. In modo da poter applicare per l'ottimizzazione dei pesi il metodo di discesa del gradiente;
- non linearità. Questa proprietà è necessaria per introdurre la non linearità nel modello, dal momento che la combinazione lineare di più funzioni di attivazione lineari restituisce sempre una funzione lineare vanificando l'introduzione di una struttura più complessa della rete.

Tra le funzioni di attivazione più comuni, rappresentate graficamente nella Figura 14, rientrano: la funzione sigmoide logistica, tangente iperbolica e rettificatore (o ReLU).

Figura 14. Funzioni di attivazione a confronto:
funzione logistica (a), tangente iperbolica (b), ReLU (c)



Fonte: Jain Pawan, *Complete Guide of Activation Functions*, in Towards data science (blog), 12/06/2019²⁹.

La funzione di trasformazione tangente iperbolica è simile alla logistica, ma è centrata sullo zero e restituisce valori compresi tra (-1;1), invece che tra (0;1). Inoltre, per input con valori intorno allo zero, presenta una pendenza più marcata rendendo la modifica dei pesi più repentina in questo tratto rispetto alla logistica. Entrambe presentano il rischio di scomparsa del gradiente, che è un problema riguardante l'addestramento delle reti con il metodo di discesa del gradiente. Infatti, l'aggiornamento dei pesi è proporzionale al gradiente della funzione di errore (rispetto ai pesi stessi), che per valori in ingresso estremi tende a zero. In questo modo l'aggiornamento dei coefficienti sostanzialmente si interrompe e sono bassissime le possibilità di cambiare stato. Una funzione che non soffre di questo problema è la ReLU³⁰. Si tratta di una funzione semplice ed efficiente nella fase di apprendimento, tuttavia trasforma tutti i valori negativi in zero rendendo impossibile

²⁹ Disponibile a <https://towardsdatascience.com/complete-guide-of-activation-functions-34076e95d044>

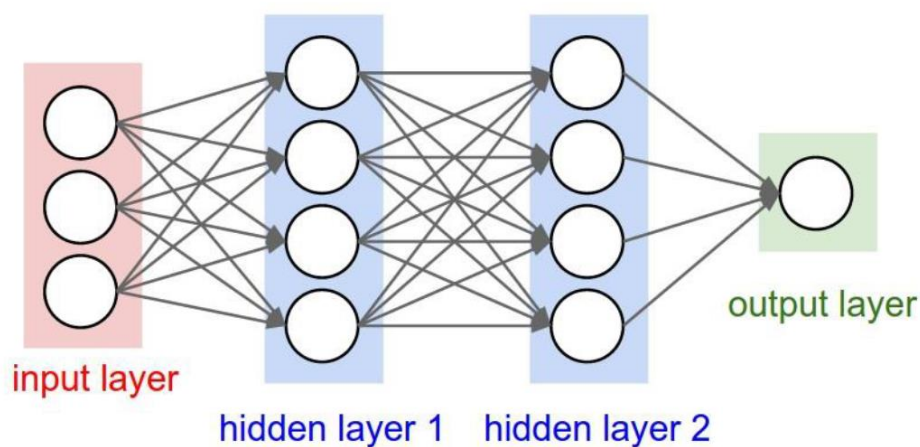
³⁰ Sebbene non sia differenziabile in tutto il dominio, come le due precedenti, è comunque possibile sfruttarla nell'apprendimento con discesa del gradiente.

un'attivazione futura del neurone.

Va sottolineato che la scelta della funzione di attivazione dipende dal contesto di applicazione, dato che non ne esiste una superiore alle altre.

Per quanto concerne la struttura, tra lo strato di input e quello di output è possibile introdurre uno o più strati di neuroni nascosti (hidden layer). Il modello che ne deriva, si veda la Figura 15, è chiamato Percettrore multistrato (Multilayer perceptron, MLP).

Figura 15. Multilayer perceptron (con due hidden layer) feedforward



Fonte: McGonagle John et al., *Feedforward Neural Networks*, in Brilliant (blog)³¹.

Il Multilayer perceptron rappresenta la struttura più popolare di rete neurale artificiale e, in particolare, rientra nella classe delle reti neurali di tipo feedforward, ossia "con flusso in avanti". Ogni layer, infatti, è connesso solo con il layer successivo e le informazioni procedono in un'unica direzione (in avanti) rispetto agli input.

Il MLP è una rete neurale completamente connessa, ovvero tutte le unità di uno strato sono connesse alle unità dello strato successivo. A ciò si aggiunge l'unità di bias, presente negli strati nascosti e d'ingresso. Ogni unità (o nodo) della rete funge da neurone artificiale, l'unica eccezione è costituita dai nodi dello strato di ingresso, che hanno il solo compito di ricevere gli input dall'ambiente esterno e trasmetterli al livello successivo.

Le unità degli hidden layer impiegano una funzione di attivazione di tipo non lineare,

³¹ Disponibile a <https://brilliant.org/wiki/feedforward-neural-networks/>

mentre per l'unità (o le unità) di output è possibile scegliere una funzione di tipo lineare o non lineare.

Il risultato è una rete di neuroni strettamente connessi tra loro, che è in grado di risolvere compiti più complessi rispetto a un semplice perceptrone. La superiorità di questa struttura rispetto al singolo perceptrone è confermata dagli studi di Cybenko (1989) e DARPA (1989).

Cybenko dimostrò che, con una rete neurale feedforward con almeno un hidden layer, è sempre possibile approssimare una funzione continua con la precisione desiderata.

In particolare, data una qualsiasi funzione sigmoide continua σ , limitata nell'intervallo (0;1), la sommatoria finita $G(x)$ della forma:

$$G(x) = \sum_{j=1}^N \alpha_j \cdot \sigma(y_j^T x + \theta_j) \quad (19)$$

consente di approssimare qualsiasi funzione continua $f(x)$ in modo che

$$|G(x) - f(x)| < \epsilon \quad (20)$$

dove:

- ϵ assume un valore piccolo a piacere;
- α_j sono i coefficienti della combinazione lineare;
- y_j corrisponde al vettore dei pesi sinaptici, x al vettore di input e θ_j all'unità di bias.


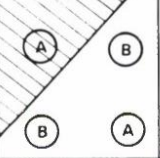



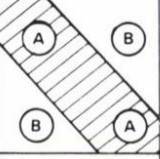
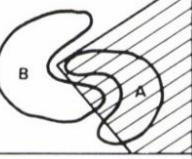
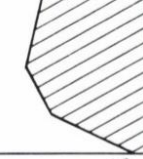

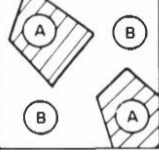
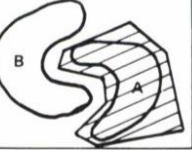

Si noti come $\sigma(y_j^T x + \theta_j)$ equivale al valore restituito dai neuroni dello strato nascosto.

È bene precisare che, per giungere a un livello di accuratezza ϵ , la struttura e i parametri della rete devono essere aggiustati di conseguenza per tenere conto della complessità della funzione oggetto di studio.

L'agenzia governativa statunitense DARPA (Defense Advanced Research Projects Agency), in un report del 1989, evidenzia le capacità classificatorie delle reti neurali.

Nella Figura 16 vengono presi in considerazione tre modelli: un perceptrone e due Multilayer perceptron (con rispettivamente 1 e 2 livelli nascosti). Per ognuno di essi viene valutata la capacità di tracciare confini decisionali (decision boundary), che separano le osservazioni in due regioni (A e B). La zona ombreggiata rappresenta la regione classificata dal modello come A.

Figura 16. Tipologia di confini decisionali prodotti da una rete neurale al variare del numero di hidden layer

| STRUCTURE | TYPES OF DECISION REGIONS | EXCLUSIVE OR PROBLEM | CLASSES WITH MESHERD REGIONS | MOST GENERAL REGION SHAPES |
|--|---------------------------|---|--|---|
|  SINGLE-LAYER | HALF PLANE |  |  |  |
|  TWO-LAYER | TYPICALLY CONVEX |  |  |  |
|  THREE-LAYER | ARBITRARY |  |  |  |

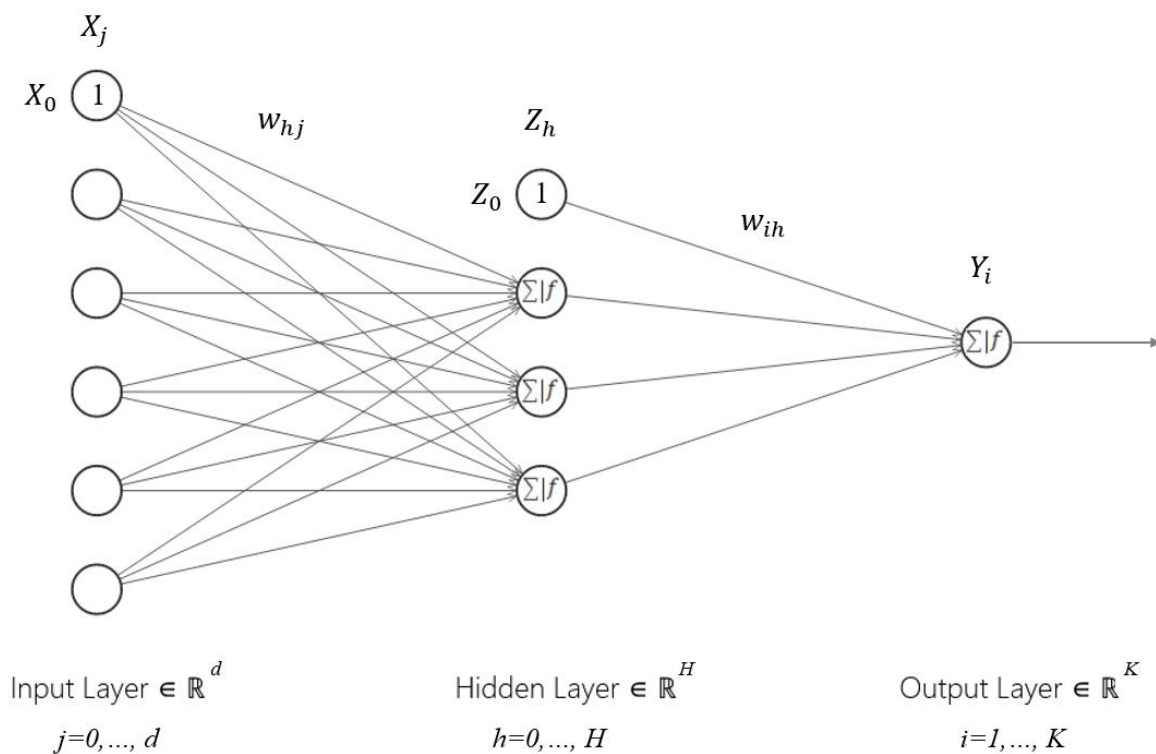
Fonte: DARPA (1989).

Nel primo caso il perceptrone è in grado di rilevare solo relazioni lineari attraverso la creazione di un semipiano che separa le regioni A e B. L'introduzione di uno strato di neuroni nascosti consente di risolvere problemi non lineari (come il problema XOR) con la creazione di una regione convessa. Infine, un secondo hidden layer aumenta la capacità di classificazione della rete, che è ora in grado di rilevare una regione in modo arbitrario.

Pertanto, dai due studi emerge come la rete neurale artificiale MLP feedforward si comporti come un ottimo approssimatore universale di funzioni continue e classificatore non lineare.

Di seguito viene descritto il meccanismo di apprendimento dei parametri della rete. I parametri ignoti della rete neurale MLP (si veda Figura 17) non sono altro che i pesi w_{hj} e w_{ih} associati alle connessioni tra neuroni, che devono essere appresi dalla macchina nella fase di training.

Figura 17. Struttura di un Multilayer perceptron con 1 livello nascosto



Fonte: elaborazione dell'autore.

Di norma viene impiegato un algoritmo di retropropagazione dell'errore (o error backpropagation algorithm), che costituisce una generalizzazione della regola Delta vista in precedenza (applicabile soltanto al singolo perceptrone).

Inizialmente ai parametri ignoti vengono assegnati dei valori casuali, di norma intorno allo zero per non saturare le funzioni di attivazione sigmoidee³². Per ogni osservazione il

³² Come accennato in precedenza la logistica, così come altre funzioni sigmoidee, in presenza di valori in ingresso elevati (positivi o negativi) restituisce un valore in uscita vicino a 0 o 1. Queste due regioni sono problematiche per l'ottimizzazione basata sulla discesa del gradiente, poiché il gradiente tende a 0 e l'apprendimento di conseguenza si "interrompe".

prodotto tra il vettore degli input X e la matrice dei pesi W_{hj} viene computato e il vettore risultante viene sottoposto a una trasformazione non lineare (sigmoide solitamente).

I valori ottenuti (vettore Z) fungono da input per lo strato di neuroni successivo. L'aggregazione risultante dalla ponderazione di questi nuovi "input" Z con i pesi W_{ih} viene sottoposta alla funzione di trasformazione dell'unità di output, che restituisce così la risposta della rete. Pertanto, la rete neurale può essere vista come una serie di funzioni composte, che procedono dal livello di input verso quello di output.

Nel passo successivo viene comparato l'output stimato dal modello y con quello desiderato t . Il confronto avviene tramite il calcolo di una funzione di costo (o funzione di perdita) E , del tipo:

$$E = \frac{1}{2} \sum_{i=1}^K (t_i - y_i)^2 \quad (21)$$

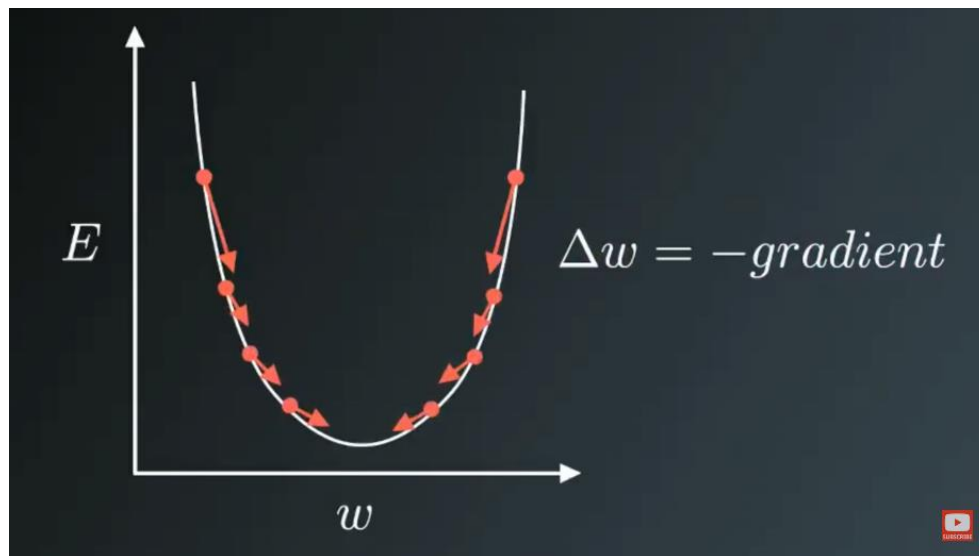
dove K è il numero di nodi presenti nello strato di output (per la Figura 17 $K=1$).

In caso di divergenza tra t e y , è necessario correggere l'errore di stima commesso dalla rete modificandone i parametri. I pesi vengono quindi aggiornati, solitamente con il metodo di ottimizzazione di discesa del gradiente³³, in modo da far convergere gradualmente questi due valori e minimizzare, di conseguenza, la funzione di perdita E .

Il gradiente di una funzione, formato da un vettore di derivate parziali, fornisce la direzione di massima pendenza per scendere da un punto e avvicinarsi al minimo della funzione (si veda la Figura 18).

³³ Nel caso in cui il dataset di training contenga un numero elevato di osservazioni, un metodo di ottimizzazione alternativo è la discesa stocastica del gradiente.

Figura 18. Rappresentazione grafica del concetto di discesa del gradiente



Fonte: Siraj Raval, *Backpropagation in 5 Minutes (tutorial)*, Video YouTube, 02/04/2017³⁴.

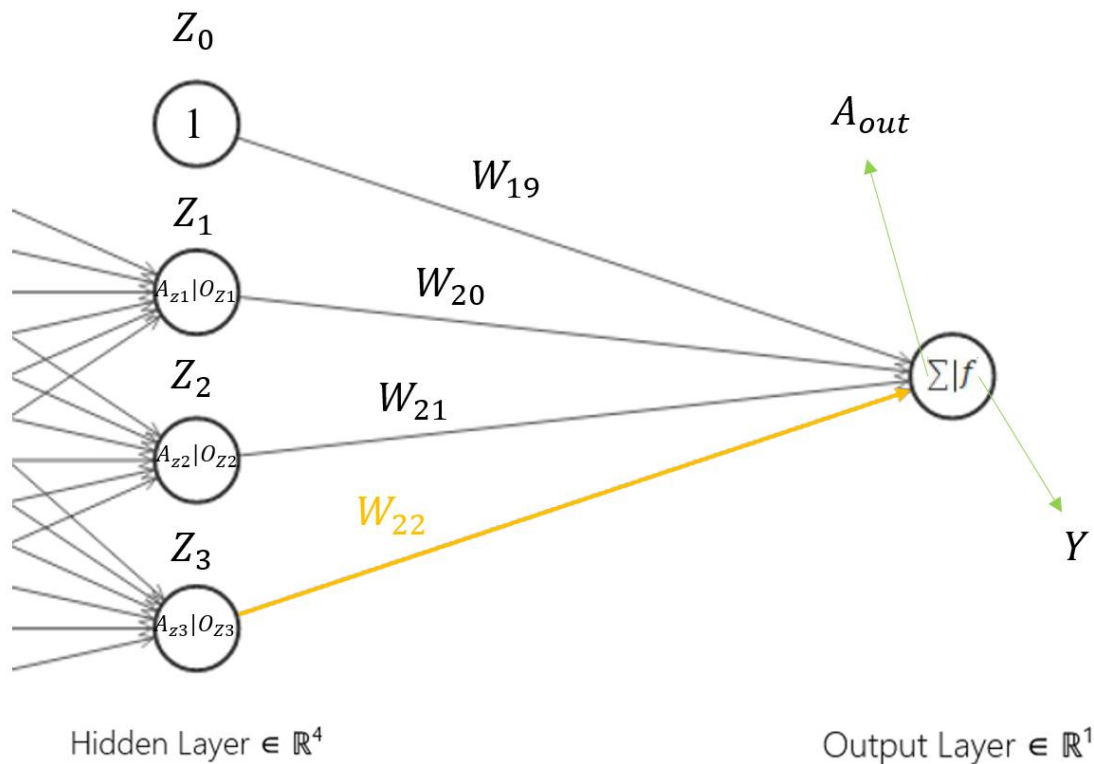
In questo contesto siamo interessati al gradiente della funzione di costo E rispetto a ogni singolo parametro w , che fornisce l'impatto di una variazione del parametro sull'errore totale. Il computo di tali gradienti non è immediato, dal momento che nel Multilayer Perceptron entrano in gioco funzioni composte complesse, per le quali è necessario ricorrere alla regola della catena (o chain rule).

Di seguito viene fornito un esempio circa il funzionamento di questo meccanismo: l'obiettivo è aggiornare il peso w_{22} della Figura 19³⁵.

³⁴ Disponibile a <https://www.youtube.com/watch?v=q555kfFUCM>

³⁵ La Figura 19 si riferisce alla parte finale del modello presente in Figura 17.

Figura 19. Strato nascosto e strato di output di un Multilayer perceptron



Fonte: elaborazione dell'autore.

L'idea alla base dell'algoritmo di backpropagation consiste nell'assegnare a ciascun neurone una porzione dell'errore complessivo. Dal momento che non è possibile attribuire direttamente un errore alle unità non appartenenti al livello di output, l'algoritmo propaga all'indietro (lungo la rete neurale) l'errore complessivo invertendo la direzione del flusso di informazioni.

L'applicazione della chain rule consente di scrivere la derivata parziale di E rispetto al peso w_{22} come:

$$\frac{\partial E}{\partial w_{22}} = \frac{\partial E}{\partial Y} \times \frac{\partial Y}{\partial A_{out}} \times \frac{\partial A_{out}}{\partial w_{22}} \quad (22)$$

Si procede ora con il calcolo delle tre derivate parziali. La prima misura di quanto varia la funzione di costo al variare dell'output y ed è pari a:

$$\frac{\partial E}{\partial Y} = y - t \quad (23)$$

La seconda derivata esprime l'impatto su y di una variazione del valore assunto da A_{out} , pari al prodotto tra il vettore degli "input" forniti dall'hidden layer e il vettore di pesi intercorrenti tra questo livello e quello di output. Supponendo una funzione di attivazione logistica

$$y = \frac{1}{1 + e^{-A_{out}}} \quad (24)$$

si avrà:

$$\frac{\partial Y}{\partial A_{out}} = y \times (1 - y) \quad (25)$$

Infine, l'ultima componente risulta pari a:

$$\frac{\partial A_{out}}{\partial w_{22}} = \frac{\partial(1 \times w_{19} + O_{Z1} \times w_{20} + O_{Z2} \times w_{21} + O_{Z3} \times w_{22})}{\partial w_{22}} = O_{Z3} \quad (26)$$

Pertanto, aggregando le tre derivate parziali si ottiene:

$$\frac{\partial E}{\partial w_{22}} = \frac{\partial E}{\partial Y} \times \frac{\partial Y}{\partial A_{out}} \times \frac{\partial A_{out}}{\partial w_{22}} = (y - t) \times y \times (1 - y) \times O_{Z3} \quad (27)$$

Avendo a disposizione tutti i dati, è possibile procedere all'aggiornamento di w_{22} :

$$w_{22}^{t+1} = w_{22}^t + \Delta w_{22}^{t+1} = w_{22}^t - \alpha \times \frac{\partial E}{\partial w_{22}} \quad (28)$$

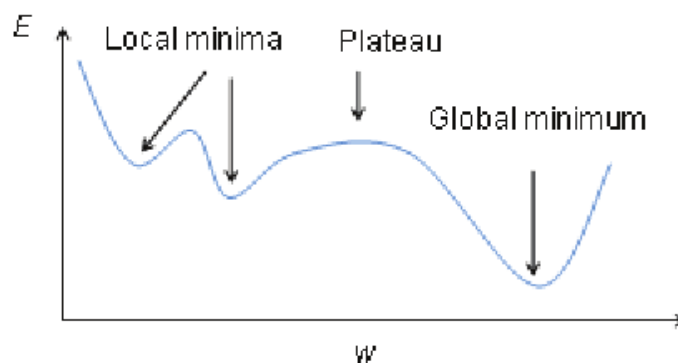
dove α è il tasso di apprendimento (o learning rate).

In modo analogo è possibile procedere alla modifica dei coefficienti w_{19} , w_{20} e w_{21} .

Per quanto riguarda i parametri posti tra il livello di input e il livello nascosto, il meccanismo della chain rule rimane lo stesso, solo che la presenza di funzioni composte con un numero elevato di funzioni concatenate aumenta la complessità dei calcoli.

L'algoritmo di retropropagazione dell'errore generalmente si comporta male nelle regioni piane della superficie dell'errore (Figura 20). In questi casi il gradiente tende a zero e aumenta, pertanto, il rischio di rimanere bloccati in minimi locali o in plateau.

Figura 20. Rappresentazione dell'errore E rispetto ai parametri w



Fonte: Temel T. (2010), *Biologically-Inspired Learning: An Overview and Application to Odor Recognition*, System and Circuit Design for Biologically-Inspired Intelligent Learning, pp. 59-92.

Un'alternativa per rendere più efficiente l'apprendimento consiste nell'inserimento di una componente (detta momentum):

$$w_{22}^{t+1} = w_{22}^t + \Delta w_{22}^{t+1} = w_{22}^t - \alpha \times \frac{\partial E}{\partial w_{22}} + \beta \times \Delta w_{22}^t \quad (29)$$

Questo termine consente di accelerare la modifica dei parametri nelle regioni piatte, in modo da cercare di uscire dai minimi locali e raggiungere il minimo globale.

L'aggiornamento dei pesi avviene dopo ogni osservazione fornita alla rete (online learning) e si ripete fino alla completa presentazione di tutte le coppie input-output presenti nel dataset. In questo modo la rete ha compiuto un'epoca. Tale processo viene reiterato per un certo numero di epoche, fino a quando la funzione di costo non raggiunge un livello predeterminato oppure si innesca un meccanismo di interruzione dell'apprendimento. Conclusa la fase di addestramento, la rete può passare alla fase di prova (o testing), nella quale riceve nuovi input. In questo modo è possibile valutare la bontà dell'apprendimento della rete neurale in base alle sue performance previsionali.

3.3 Performance delle tecniche di machine learning

Il primo studio che applica le reti neurali artificiali nel campo della previsione dell'insolvenza è quello di Odom e Sharda (1990). Gli autori impiegano una rete a 3 livelli, con rispettivamente 5 nodi di input (le 5 note variabili di Altman (1968)), 5 nodi nello strato nascosto e 1 nodo nello strato di output. Il modello restituisce un valore compreso tra 0 e 1 e le società vengono classificate come sane o non sane in base al superamento o meno della soglia, fissata a 0.5. L'apprendimento della rete è basato sulla regola della retropropagazione dell'errore, una critica che viene mossa a tale algoritmo è il numero eccessivo di iterazioni (quasi 200.000) necessarie in questo studio.

Il dataset è composto da un campione bilanciato (50% sane e 50% fallite) di 129 imprese statunitensi e il benchmark scelto per confrontare le performance è l'analisi discriminante. Nel campione di controllo la rete neurale artificiale prevede correttamente

l'81.48% delle imprese fallite, contro il 59.26% dell'analisi discriminante. Se si aumenta la numerosità delle imprese sane nel campione di addestramento, rispettivamente all'80% e al 90% del totale (situazione più realistica), la rete continua a classificare correttamente un numero superiore di imprese fallite rispetto al benchmark. Un discorso analogo vale per il grado di accuratezza, ovvero la percentuale di imprese classificate correttamente sul totale. Per quanto riguarda l'errore di I tipo, ossia la quota di controparti fallite classificate come sane, il benchmark batte la rete neurale nel campione bilanciato (10.71% contro il 17.86%) e nel campione 80%-20% (14.29% contro il 21.43%), ma non in quello 90%-10% (21.43% contro il 14.29%). Nonostante il campione a disposizione sia ridotto, i risultati mostrano come le reti neurali artificiali possano costituire una valida alternativa alle tecniche tradizionali, quale l'analisi discriminante.

Analogamente Altman et al. (1994) esplorano le potenzialità di questi modelli usando come benchmark l'analisi discriminante. L'analisi si basa su un campione di oltre 1000 imprese italiane relative al periodo 1982-1992. Dallo studio si evince come le reti neurali siano in grado di classificare con accuratezza le imprese in base al loro stato di salute, con risultati che si avvicinano e, in alcuni casi superano, quelli prodotti dall'analisi discriminante. Dall'altro lato vengono evidenziate alcune loro criticità, quali: la lunga durata della fase di addestramento, la necessità di ottimizzare gli iperparametri³⁶ del modello, il problema dell'overfitting e la scarsa trasparenza dei risultati.

Zhang et al. (1997) conducono un confronto tra un modello di rete neurale artificiale a tre livelli e la regressione logistica. Il dataset, bilanciato tra le due classi, contiene 220 osservazioni di imprese manifatturiere statunitensi, relative al periodo 1980-1991.

Le variabili utilizzate sono le 5 di Altman (1968), con l'aggiunta di un sesto indicatore (Attivo corrente/Passivo corrente).

I ricercatori adottano la tecnica della convalida incrociata (o cross-validation), che consiste nel partizionare il dataset in k sotto-campioni di uguale numerosità (in questo caso $k = 5$). Uno dei quali viene utilizzato per la fase di testing e i restanti per quella di training. Il processo si ripete k volte, ovvero fino a quando tutti i sotto-campioni non vengono impiegati almeno una volta nella fase di testing. La convalida incrociata aiuta a

³⁶ Gli iperparametri sono valori che definiscono la struttura della rete e dell'algoritmo. Esempi sono: il numero di livelli nascosti nella rete, il numero di nodi presente in ogni strato e il tasso di apprendimento. La loro ottimizzazione ha lo scopo di migliorare le prestazioni del modello.

rilevare meglio la presenza di overfitting e riduce il problema del campionamento asimmetrico, che può derivare dalla suddivisione del dataset in due sole parti (training e testing). In tutti e 5 i sotto-campioni l'accuratezza della rete neurale artificiale è superiore a quella della regressione logistica (in media 80.46% contro 78.18%) e la differenza è statisticamente significativa al livello del 5%. In 4 casi su 5 la rete riesce a classificare correttamente un numero maggiore di imprese non sane rispetto al benchmark (in media 81.82% contro 78.18%); dall'altro lato performa peggio in 2 casi su 5 per quanto riguarda la corretta classificazione delle imprese sane (in media 76.09% contro 78.18%).

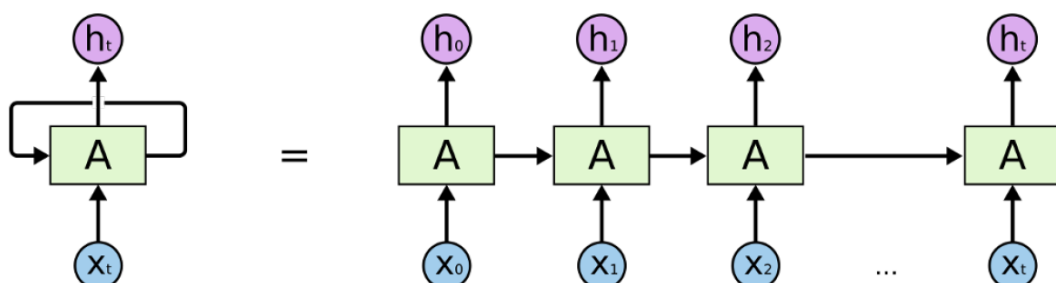
L'ANN ha complessivamente sovraperformato rispetto al benchmark probabilmente per le sue peculiarità; è infatti una tecnica flessibile, non parametrica, che non richiede assunzioni restrittive e che è capace di affrontare problemi di classificazione complessi con dati non linearmente separabili. Nonostante la superiorità dimostrata da questo tipo di modelli nello studio in questione, il loro utilizzo nel mondo reale è ostacolato principalmente dal fatto di possedere un meccanismo di funzionamento opaco e con risultati difficilmente interpretabili. A tal proposito, vengono spesso appellati come scatole nere (black-box).

Nadali (2017) mette a confronto un Multilayer perceptron e una rete neurale ricorrente (Recurrent Neural Network, RNN) per prevedere il fallimento di piccole e medie imprese italiane. Il campione, non bilanciato, è formato da oltre 700.000 osservazioni.

Il modello MLP, scelto come benchmark, include come input solo i dati relativi all'ultimo bilancio disponibile prima del fallimento, quindi offre una visione statica del problema³⁷. Tale modello viene comparato con una RNN con un livello nascosto di neuroni A (si veda Figura 21), che può essere considerata come un'evoluzione del MLP.

³⁷ Il MLP non è in grado di distinguere l'ordinamento temporale degli input.

Figura 21. Funzionamento di una rete neurale ricorrente (RNN)



A Recurrent Neural Network in its looped and “unrolled” representation. X_t are the inputs at time t , that are transferred to A (a neuron or, more commonly, a layer of neurons). For every t , A produces an output h_t and also sends it as an input to the network in $t+1$.

Fonte: Nadali (2017).

La RNN possiede una memoria interna, infatti l'output h_t dipende sia dai dati correnti X_t forniti alla rete dall'esterno sia dagli output prodotti dalla rete stessa nelle epoche precedenti. Tale struttura si adatta bene a problemi di natura sequenziale e di previsione di serie storiche. Il modello RNN sfrutta, in questo esperimento, 5 anni di dati di bilancio assumendo un comportamento dinamico.

Il benchmark raggiunge in media un'accuratezza dell'85% nel campione di testing, il risultato migliore (85.5%) è stato ottenuto con rispettivamente 100 e 25 nodi nei due livelli nascosti. L'inserimento di ulteriori neuroni non migliora necessariamente le prestazioni della rete. Il MLP in questo esperimento tende a classificare la quasi totalità delle imprese come sane, ciò è causato dalla presenza di un campione di addestramento fortemente sbilanciato verso questa classe: l'81% delle osservazioni è costituito da imprese sane. Di conseguenza si giunge a un errore di I tipo particolarmente elevato (circa il 65%) e a un errore di II tipo trascurabile (tra il 2-3%).

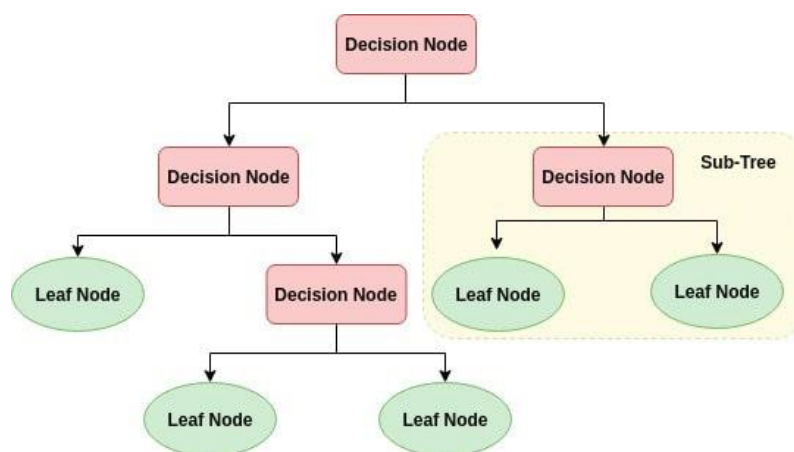
La RNN restituisce risultati migliori: un'accuratezza di circa l'88% e un errore di I tipo sensibilmente ridotto, ma ancora elevato (attorno al 42%). Tali prestazioni possono essere ascritte alla maggior complessità della rete e all'impiego di un numero maggiore di input. Considerando un campione di addestramento con classi bilanciate, l'errore di I tipo scende al di sotto del 19%, così come quello di II tipo. Pertanto, le due tipologie di errori

sono sensibili alla composizione del campione. Ne consegue che è possibile inserire una quota più o meno elevata di imprese fallite per giungere a un rapporto tra errore di I e II tipo in linea con l'avversione al rischio del soggetto decisore.

Nonostante l'accuratezza si attesti circa all'82% con questo nuovo campione, complessivamente si può considerare il risultato come positivo, dal momento che la minimizzazione dell'errore di I tipo è un fattore decisivo per ridurre le perdite delle società che erogano credito.

Al di fuori delle reti neurali artificiali, in letteratura sono state proposte numerose altre tecniche di machine learning. Un albero decisionale (decision tree) costituisce un tipico modello di apprendimento supervisionato, capace di svolgere compiti di classificazione e regressione. È una tecnica molto semplice e di facile interpretazione, formata sostanzialmente da una lista ordinata di regole del tipo "Se vale la condizione X, allora esegui l'istruzione Y". Le condizioni permettono di discriminare le osservazioni in base ai valori assunti dalle variabili di input. Le istruzioni conducono a una decisione di classificazione oppure a un'altra regola. Graficamente (Figura 22) le regole sono rappresentate da nodi decisionali (decision node) e le classi da foglie (leaf node).

Figura 22. Rappresentazione grafica di un albero decisionale



Fonte: Navlani Avinash, *Decision Tree Classification in Python*, in Datacamp (blog), 28/12/2018³⁸.

³⁸ Disponibile a <https://www.datacamp.com/community/tutorials/decision-tree-classification-python>

Le regole e il loro ordine di esecuzione devono essere apprese dalla macchina a partire da un dataset di coppie di input-output.

Le applicazioni più interessanti si possono ottenere combinando (in vario modo) più modelli ad albero tra loro, in questo modo si ottengono risultati più accurati e stabili. Moscatelli et al. (2019), ad esempio, sviluppano una foresta casuale (random forest, RDF) e un gradient boosted tree (GBT), che risultano essere superiori in precisione e capacità discriminatoria rispetto alla regressione logistica e all'analisi discriminante. Si tratta di modelli che si dimostrano essere particolarmente validi in presenza di dataset con un numero elevato di osservazioni e di variabili.

L'esperimento si basa su un dataset contenente circa 300.000 aziende italiane per il periodo 2011-2017. Come input sono state selezionate 26 variabili di carattere finanziario-creditizio, oltre al settore economico e all'area geografica.

Dallo studio emerge come i modelli di machine learning tendano a concedere credito a un numero minore di soggetti, ma meno rischioso e per un ammontare maggiore. I tassi di default (e le perdite) sono inferiori rispetto ai modelli benchmark.

3.4 Modelli ibridi

Di recente in letteratura è stato proposto l'utilizzo di modelli ibridi, ossia frutto dell'unione di metodologie di analisi differenti. L'obiettivo è riuscire a combinare i vantaggi di più tecniche e, allo stesso tempo, superare le debolezze insite in ognuna di esse.

Tsai e Chen (2010) mirano a rintracciare la miglior combinazione tra tecniche di classificazione e di clustering per la costruzione di modelli ibridi.

Nella prima categoria vengono utilizzate quattro tecniche di apprendimento supervisionato: l'algoritmo C4.5 ossia un albero decisionale, un classificatore bayesiano, la regressione logistica (LR) e il MLP. Nella seconda rientrano due tecniche di apprendimento non supervisionato: l'algoritmo K-means e algoritmo Expectation Maximization (EM). Una componente svolge il compito di pre-processing dei dati (rilevazione dei valori anomali e riduzione dei dati), l'altra procede a classificare o a clusterizzare.

Il dataset, relativo al periodo 2004-2006, è composto da quasi 13.000 debitori di una banca commerciale taiwanese. I modelli ibridi con le performance migliori, in termini di accuratezza delle previsioni e massimizzazione dei profitti della banca, si sono rivelati essere quelli che sfruttano una combinazione di due tecniche di classificazione.

Con l'uso di due regressioni logistiche si ottiene una corretta classificazione dell'82.94% delle osservazioni, mentre con l'uso di una regressione logistica e di un MLP il valore è pari al 83.44%. Quando si considerano le tecniche di clustering le prestazioni sono deludenti, a eccezione del modello EM-LR (accuratezza dell'80.16%).

Pertanto la scelta degli autori propende verso una combinazione di due classificatori, in particolar modo il modello ibrido che si è dimostrato superiore agli altri deriva dalla combinazione tra la regressione logistica e il MLP.

Nel loro studio, Yim e Mitchell (2002) ambiscono a sviluppare dei modelli ibridi in grado di sovraperformare sia le tecniche statistiche tradizionali, sia le reti neurali artificiali.

I modelli statistici tradizionali (analisi discriminante e regressione logistica) supportano il lavoro di una rete MLP in due modi. Nel primo caso preselezionano le variabili, in modo da ridurre il rischio di overfitting. Nel secondo, allo scopo di condensare le informazioni a disposizione, stimano come output delle probabilità di fallimento che costituiscono gli input del MLP. Per quanto riguarda le performance, il modello ibrido migliore è costruito selezionando le variabili tramite la regressione logistica e fornendo come input al MLP le probabilità di fallimento stimate con l'analisi discriminante. In generale le reti neurali ibride si sono dimostrate superiori rispetto alle tecniche statistiche tradizionali e, in alcuni casi, anche al MLP.

Un altro metodo per costruire modelli ibridi consiste nel combinare una serie di classificatori in parallelo. Successivamente i loro output vengono aggregati in base a una determinata regola per restituire la decisione finale. Ad esempio, con la maggioranza semplice l'output che riceve il maggior numero di voti viene selezionato come output finale. In alternativa è possibile assegnare un peso diverso ai vari output, in funzione delle capacità previsive dimostrate in precedenza dai classificatori.

In letteratura l'unione di più classificatori sembra avere successo solo se i singoli modelli sono diversificati tra loro. Ciò può avvenire ad esempio impostando iperparametri diversi o utilizzando un dataset di addestramento differente.

Tsai e Wu (2008) adottano tale metodologia a partire da una serie di MLP. Le singole reti

neurali artificiali vengono diversificate prima in base al numero di nodi presenti nel livello nascosto e al numero di epoche impiegate per l'addestramento, successivamente addestrando i classificatori con dati differenti. L'aggregazione degli output avviene seguendo la regola della maggioranza semplice.

L'esperimento, effettuato su tre dataset diversi, non consente di esprimere un giudizio definitivo. Tuttavia, il modello ibrido sovraperforma il classificatore singolo solo nel secondo dataset e ottiene risultati molto simili nel terzo. Tenendo in considerazione anche l'errore di I tipo, il modello ibrido non presenta vantaggi significativi rispetto al singolo MLP.

Nonostante non sia presente un'ampia letteratura in merito, lo studio di nuovi modelli ibridi è sicuramente un tema di ricerca da approfondire, dal momento che racchiude in sé grandi potenzialità.

Capitolo 4: Esperimento

Nel seguente capitolo viene testata l'affidabilità dei modelli Logit e Multilayer Perceptron nel prevedere l'insolvenza aziendale. Il primo assume il ruolo di benchmark ed è stato individuato durante l'analisi della letteratura di riferimento (si veda il paragrafo 2.2).

Il compito richiesto a tali tecniche consiste essenzialmente nel valutare lo stato di salute di un campione di aziende attraverso la restituzione di un punteggio. La fissazione di un valore soglia consente poi di classificare le aziende in due gruppi.

Tali modelli possono rappresentare un punto di partenza per lo sviluppo di sistemi più complessi e completi, in grado di coadiuvare le banche, le società finanziarie e varie tipologie di creditori nella gestione delle posizioni creditizie.

Il dataset³⁹, estratto dalla banca dati AIDA, comprende un campione di 185 imprese fallite e 185 imprese attive⁴⁰. Le osservazioni si riferiscono a società di capitali non quotate che svolgono un'attività manifatturiera (in base al codice ATECO). Tali criteri di selezione sono stati imposti al fine di ottenere una popolazione sufficientemente omogenea.

Come accennato in precedenza, la variabile dipendente è di tipo binario. La prima classe, identificabile con il valore 1, include le imprese del dataset dichiarate fallite nell'arco del 2019 (anno di riferimento). La seconda classe, etichettata con 0, raggruppa le imprese attive a fine 2019. Situazioni giuridiche intermedie, come le fusioni, le scissioni e altre procedure concorsuali, non sono state prese in considerazione, dal momento che riducono le capacità del modello di discriminare le due classi.

Il dataset completo è stato suddiviso in due campioni: uno necessario per stimare il modello (training set) e uno sul quale effettuare la previsione (testing set). La quota destinata ai due sotto campioni è fissata inizialmente al 50% per il training e al 50% per il testing. L'estrazione dei sotto campioni, avvenuta in modo casuale, è stata condotta in modo da ottenere un bilanciamento tra le due classi: il primo set contiene 93 imprese attive e 93 imprese fallite, il secondo 92 osservazioni per classe.

Per quanto riguarda le variabili esplicative, sono stati individuati undici indicatori di redditività, crescita, struttura finanziaria, liquidità, solvibilità ed efficienza:

³⁹ Per ottenere il dataset si rimanda all'Appendice 1.

⁴⁰ Lavorare con campioni bilanciati tra le due classi consente di ottenere delle stime più precise incrementando il potere predittivo del modello.

- “ROA”, indice di redditività dato dal rapporto tra il Risultato operativo e il Totale dell’attivo, concettualmente esprime il rendimento generato dagli asset aziendali;
- “Margine sulle vendite”, dove al numeratore è presente l’Utile prima degli ammortamenti, degli interessi e delle imposte (o EBITDA) e al denominatore i Ricavi derivanti dalle vendite. Indica la marginalità che riesce a conseguire l’azienda vendendo i propri beni e/o servizi;
- “Utili trattenuti/Attivo”, dove il numeratore rappresenta la somma degli utili reinvestiti e/o delle perdite maturate durante l’intera vita della società. Tale variabile evidenzia il ricorso a fondi generati internamente quale fonte di finanziamento dell’Attivo;
- “Crescita dei Ricavi”, misura la variazione percentuale del fatturato negli ultimi tre anni⁴¹, ovvero la capacità di espansione dell’azienda;
- “Crescita degli Utili”, indice che rileva l’andamento degli utili prodotti dalla società negli ultimi tre anni;
- “Grado di indipendenza da terzi”, si tratta di un indice di leva finanziaria che rapporta il Patrimonio netto ai Debiti totali. L’analisi si concentra sulla composizione delle fonti di finanziamento dell’attivo tra fonti proprie e fonti di terzi. Un eccessivo ricorso alla leva finanziaria aumenta la rischiosità aziendale;
- “Indice di indipendenza finanziaria”, indice di leva finanziaria dato dal rapporto tra il Patrimonio netto e il Totale Attivo;
- “Capitale Circolante Netto/Attivo”, dove la grandezza al numeratore misura la capacità dell’azienda di coprire i Debiti a breve con le componenti dell’Attivo più liquide;
- “Indice di liquidità” pari a $(\text{Attivo circolante} - \text{Rimanenze}) / \text{Debiti a breve}$.
Mostra la capacità dell’azienda di ripagare i Debiti a breve termine con la parte dell’Attivo più facilmente convertibile in moneta. Indica l’esistenza o meno di uno squilibrio finanziario tra fonti e fabbisogni di breve termine;

⁴¹ L’arco temporale triennale consente di ottenere valori più stabili, dal momento che i dati annuali sono soggetti a maggiore variabilità.

- “Grado di copertura degli interessi passivi”, indice di solvibilità che rapporta l’EBITDA al Totale degli oneri finanziari. Evidenzia se l’azienda, con i margini conseguiti (ossia l’EBITDA), è in grado di coprire almeno i costi per interessi;

- “Rotazione del capitale investito”, pari al rapporto tra i Ricavi da vendite e il Totale Attivo. È un indicatore di efficienza, che mostra l’abilità di generare fatturato dal complesso delle attività aziendali a disposizione.

Per misurare l’accuratezza della classificazione si fa riferimento alla matrice di confusione. Tale matrice consiste in una semplice tabella dove nelle righe vengono indicate le osservazioni classificate in una determinata classe e nelle colonne la loro effettiva classe di appartenenza. Nel caso di classificazione binaria sono 4 i valori presenti nella tabella: il numero di osservazioni correttamente classificate rispettivamente nella classe 0 e 1, le osservazioni attribuite per errore alla classe 0 (errore di I tipo) e quelle etichettate erroneamente come appartenenti alla classe 1 (errore di II tipo).

4.1 Modello Logit

Il modello Logit è stato stimato con il software econometrico gretl.

In particolare sono state sviluppate 4 versioni dello stesso.

Le prime tre utilizzano come variabili esplicative gli indici riferiti all’anno 2018 e fanno riferimento a tre differenti tentativi di generazione del training e testing set. L’obiettivo è verificare se fornendo al modello osservazioni diverse si giunge comunque a risultati simili, in termini di variabili significative e potere predittivo.

Nel quarto tentativo sono state incluse anche le variabili ritardate di un periodo (anno 2017), per verificare se ciò contribuisce a migliorare la capacità predittiva.

Il dataset contenente le variabili del 2018 è stato sottoposto a un’analisi preliminare di collinearità, tramite lo strumento “Fattore di inflazione della varianza” (Variance Inflation Factor, VIF), i cui risultati sono riportati in Figura 23.

Figura 23. Risultati del Variance Inflation Factor (VIF)

```

Fattori di Inflazione della Varianza (VIF)
Valore minimo possibile = 1.0
Valori oltre 10.0 indicano un problema di collinearità

CrescitaRicaviin3anniEUR      1.010
  IndicediliquiditA2018      2.452
  Rotazcapinvestitovolte2    1.230
    EBITDAVendite2018      1.401
      ROA2018              7.027
UtiliportatianoovoAttivo     77.566
  Gradodiindipdaterzi201    2.861
CrescitaUtili3anniEUR201     1.013
  CCNAttivo2018             4.221
  PNAttivo2018              99.680
  Coperturainteressi2018    1.075

```

La multicollinearità, ovvero l'alta correlazione tra variabili esplicative, può incidere sulla precisione e sulla stabilità dei coefficienti stimati.

In questo caso, l'eliminazione della variabile "Patrimonio Netto/Attivo" riporta il VIF abbondantemente al di sotto del valore 10 (Figura 24).

Figura 24. Risultati del Variance Inflation Factor (VIF)

```

Fattori di Inflazione della Varianza (VIF)
Valore minimo possibile = 1.0
Valori oltre 10.0 indicano un problema di collinearità

CrescitaRicaviin3anniEUR      1.010
  IndicediliquiditA2018      2.375
  Rotazcapinvestitovolte2    1.206
    EBITDAVendite2018      1.399
      ROA2018              2.437
UtiliportatianoovoAttivo     2.384
  Gradodiindipdaterzi201    2.424
CrescitaUtili3anniEUR201     1.013
  CCNAttivo2018             3.708
  Coperturainteressi2018    1.075

```

I modelli finali sono stati costruiti tramite una strategia di specificazione dal generale al particolare, ossia viene esclusa una variabile alla volta in base al p-value. La soglia del p-value è stata impostata al 0.05, come è solito considerare in letteratura.

Di seguito sono riportate le 4 versioni del modello Logit con la relativa matrice di confusione.

4.1.1 Logit18_01

Tabella 2. Modello Logit18_01

| | <i>Coefficiente</i> | <i>Errore Std.</i> | <i>z</i> | <i>p-value</i> | |
|---------------------------------------|---------------------|--------------------|----------|----------------|-----|
| costante | 1.70397 | 0.462065 | 3.688 | 0.0002 | *** |
| Rotazione del Capitale investito 2018 | -1.68610 | 0.451724 | -3.733 | 0.0002 | *** |
| Grado di indipendenza da terzi 2018 | -2.08491 | 0.631693 | -3.301 | 0.0010 | *** |
| CCN/Attivo 2018 | -0.0242931 | 0.00730829 | -3.324 | 0.0009 | *** |

| | | | | |
|-----------------------|-----------|--|---------------------|----------|
| Media var. dipendente | 0.500000 | | SQM var. dipendente | 0.501350 |
| R-quadro di McFadden | 0.488079 | | R-quadro corretto | 0.457053 |
| Log-verosimiglianza | -65.99964 | | Criterio di Akaike | 139.9993 |
| Criterio di Schwarz | 152.9023 | | Hannan-Quinn | 145.2281 |

Tabella 3. Matrice di confusione del modello Logit18_01

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 84 (45.65%) | 12 (6.52%) | 96 52.17% |
| | 1 | 8 (4.35%) | 80 (43.48%) | 88 47.83% |
| Tot. | | 92 50% | 92 50% | 184 100% |

4.1.2 Logit18_02

Tabella 4. Modello Logit18_02

| | <i>Coefficiente</i> | <i>Errore Std.</i> | <i>z</i> | <i>p-value</i> | |
|---------------------------------------|---------------------|--------------------|----------|----------------|-----|
| costante | 0.726122 | 0.516865 | 1.405 | 0.1601 | |
| Rotazione del Capitale investito 2018 | -1.24275 | 0.530330 | -2.343 | 0.0191 | ** |
| ROA 2018 | -0.110821 | 0.0302379 | -3.665 | 0.0002 | *** |
| Grado di indipendenza da terzi 2018 | -1.52246 | 0.611278 | -2.491 | 0.0128 | ** |
| CCN/Attivo 2018 | -0.0237667 | 0.00959480 | -2.477 | 0.0132 | ** |

| | | | | |
|-----------------------|-----------|--|---------------------|----------|
| Media var. dipendente | 0.500000 | | SQM var. dipendente | 0.501350 |
| R-quadro di McFadden | 0.611879 | | R-quadro corretto | 0.573097 |
| Log-verosimiglianza | -50.03865 | | Criterio di Akaike | 110.0773 |
| Criterio di Schwarz | 126.2060 | | Hannan-Quinn | 116.6133 |

Tabella 5. Matrice di confusione del modello Logit18_02

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 86 (46.74%) | 11 (5.98%) | 97 52.72% |
| | 1 | 6 (3.26%) | 81 (44.02%) | 87 47.28% |
| Tot. | | 92 50% | 92 50% | 184 100% |

4.1.3 Logit18_03

Tabella 6. Modello Logit18_03

| | <i>Coefficiente</i> | <i>Errore Std.</i> | <i>z</i> | <i>p-value</i> | |
|---------------------------------------|---------------------|--------------------|----------|----------------|-----|
| costante | 1.33735 | 0.492560 | 2.715 | 0.0066 | *** |
| Rotazione del Capitale investito 2018 | -0.907226 | 0.427490 | -2.122 | 0.0338 | ** |
| Grado di indipendenza da terzi 2018 | -7.39888 | 1.22178 | -6.056 | <0.0001 | *** |

| | | | | |
|-----------------------|-----------|--|---------------------|----------|
| Media var. dipendente | 0.500000 | | SQM var. dipendente | 0.501350 |
| R-quadro di McFadden | 0.599565 | | R-quadro corretto | 0.576296 |
| Log-verosimiglianza | -51.62621 | | Criterio di Akaike | 109.2524 |
| Criterio di Schwarz | 118.9297 | | Hannan-Quinn | 113.1740 |

Tabella 7. Matrice di confusione del modello Logit18_03

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 82 (44.57%) | 16 (8.70%) | 98 53.27% |
| | 1 | 10 (5.43%) | 76 (41.30%) | 86 46.73% |
| Tot. | | 92 50% | 92 50% | 184 100% |

4.1.4 Logit1718

Tabella 8. Modello Logit1718

| | <i>Coefficiente</i> | <i>Errore Std.</i> | <i>Z</i> | <i>p-value</i> | |
|---------------------------------------|---------------------|--------------------|----------|----------------|-----|
| costante | 1.40628 | 0.461480 | 3.047 | 0.0023 | *** |
| Crescita dei Ricavi 2018 | -0.0111439 | 0.00415583 | -2.682 | 0.0073 | *** |
| Rotazione del Capitale investito 2018 | -0.958197 | 0.415528 | -2.306 | 0.0211 | ** |
| EBITDA/Vendite 2017 | -0.0405729 | 0.00910198 | -4.458 | <0.0001 | *** |
| Grado di indipendenza da terzi 2018 | -4.83743 | 0.826318 | -5.854 | <0.0001 | *** |
| Crescita degli Utili 2017 | 0.000198089 | 7.76327e-05 | 2.552 | 0.0107 | ** |

| | | | | |
|-----------------------|-----------|--|---------------------|----------|
| Media var. dipendente | 0.500000 | | SQM var. dipendente | 0.501350 |
| R-quadro di McFadden | 0.547182 | | R-quadro corretto | 0.500643 |
| Log-verosimiglianza | -58.37973 | | Criterio di Akaike | 128.7595 |
| Criterio di Schwarz | 148.1139 | | Hannan-Quinn | 136.6026 |

Tabella 9. Matrice di confusione del modello Logit1718

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 83 (45.11%) | 14 (7.61%) | 97 52.72% |
| | 1 | 9 (4.89%) | 78 (42.39%) | 87 47.28% |
| Tot. | | 92 50% | 92 50% | 184 100% |

4.1.5 Discussione dei risultati

In tutti i casi risultano significative le variabili “Rotazione del Capitale investito” e “Grado di indipendenza da terzi”, segnale della loro importanza nello spiegare l’insolvenza aziendale. La variabile “Capitale Circolante Netto/Attivo” è presente nei primi due modelli, ma non nel terzo e quarto. Queste tre variabili esplicative si riferiscono, rispettivamente, a un indice di efficienza dell’azienda, di leva finanziaria e di liquidità. Come prevedibile, esiste una relazione negativa tra l’andamento delle suddette variabili e la probabilità di default.

Il secondo modello risulta superiore agli altri, con un’accuratezza pari al 90.76% e un errore del I tipo del 5.98%. Risultati soddisfacenti sono stati raggiunti anche dal primo modello (rispettivamente con valori dell’89.13% e 6.52%), mentre l’introduzione delle variabili ritardate di un periodo non ha apportato miglioramenti (87.50% e 7.61%).

Infine il terzo modello offre dei risultati deludenti (85.87% e 8.70%), ma ciò è conseguenza della presenza di sole due variabili esplicative significative.

4.2 Reti neurali artificiali

I Multilayer Perceptron (MLP) sono stati realizzati mediante il software R⁴², in particolare facendo ricorso al pacchetto “neuralnet”. L’intero dataset è stato sottoposto a una trasformazione che limita le variabili esplicative nell’intervallo [0;1]. In questo modo gli input non sono dipendenti dal fattore di scala, ciò si traduce in una fase di addestramento più rapida e in una minore probabilità di rimanere bloccati in minimi locali.

Per quanto riguarda la struttura della rete neurale artificiale:

- i nodi di input sono pari alle variabili esplicative incluse nel modello;
- si è ritenuto sufficiente l’inserimento di un solo strato nascosto⁴³, come è prassi in letteratura;
- nell’ultimo livello è presente un unico nodo. Le imprese con un output superiore alla

⁴² Gli script di R, contenenti i comandi utilizzati, sono presenti in Appendice 2.

⁴³ Solo in un caso (modello “MLP06”) si è rivelato utile “allungare” la rete con l’introduzione di un secondo livello, data la presenza di un elevato numero di input.

soglia (0.50) vengono classificate come fallite, quelle al di sotto come sane.
 Le unità della rete utilizzano una funzione di attivazione sigmoidea logistica.

La flessibilità che contraddistingue questo tipo di reti ha permesso la creazione di una pluralità di modelli, con lo scopo di rintracciare gli iperparametri ottimali.

La diversificazione ha avuto per oggetto: il numero di nodi presenti nel livello nascosto, l'algoritmo di ottimizzazione, la funzione di costo, il tasso di apprendimento, la frazione del dataset destinata al training e al testing set e gli indici impiegati come input della rete. La Tabella 10 illustra in dettaglio la configurazione scelta per ogni versione del MLP.

Tabella 10. Parametri scelti per la costruzione dei vari modelli Multilayer Perceptron

| | MLP 01 | MLP 02 | MLP 03 | MLP 04 | MLP 05 | MLP 06 |
|-----------------------------|---------------|---------------|---------------|---------------------------|--------------------|---------------|
| Nodi nell'hidden layer | 4 | 4 | 4 | 2/4 | 4 | 5 + 3 |
| Algoritmo di ottimizzazione | Backprop. | Backprop. | Rprop- | Backprop. / Rprop- | Backprop. / Rprop- | Backprop. |
| Funzione di costo | SSE | Cross-entropy | SSE | SSE | SSE | SSE |
| Tasso di apprendimento | 0.05 | 0.05 | - | 0.05/- | 0.05/- | 0.05 |
| Training e testing | 50-50% | 50-50% | 50-50% | 50-50% | 70-30% | 50-50% |
| Input | 2018 | 2018 | 2018 | 2018 (solo significative) | 2018 | 2017 e 2018 |

Di seguito vengono forniti ulteriori dettagli circa le impostazioni della rete neurale. In particolare, per quanto riguarda il numero delle unità presenti nel livello nascosto, il trade-off è tra una maggiore complessità della rete e un maggior rischio di overfitting. È bene trovare un equilibrio tra le due esigenze.

Un algoritmo di ottimizzazione alternativo al classico backpropagation (illustrato nel paragrafo 3.2.3) è rappresentato dal Resilient Backpropagation (o Rprop-), proposto da Riedmiller e Braun (1993). Il Rprop- aggiorna i parametri solo in ragione del segno della derivata parziale della funzione di costo rispetto al singolo peso, senza considerare la sua

dimensione. Se la derivata parziale cambia di segno, ossia il precedente aggiornamento è stato più ampio del necessario, allora il successivo aggiornamento del peso viene ridimensionato, altrimenti viene amplificato per accelerare la convergenza al valore ottimale. Coerentemente con il processo di minimizzazione della funzione di costo, se la derivata parziale è positiva tale aggiornamento viene sottratto dal valore del parametro al passo precedente, viceversa viene aggiunto.

Per la funzione di costo sono state prese in considerazione sia la somma dei quadrati dei residui (sum of squared errors, SSE) sia la funzione di entropia incrociata (cross-entropy, CE). Dal punto di vista teorico, quest'ultima è la più adatta per affrontare problemi di classificazione binaria, ma potrebbe creare problemi nella convergenza.

Il parametro "threshold" della rete neurale consente di interrompere l'ottimizzazione dei parametri. Ad esempio, se è impostato a 0.01, è necessario un cambiamento nella funzione di costo tra un'iterazione e l'altra inferiore all'1% per arrestare la fase di addestramento. Se tale criterio non viene soddisfatto, l'addestramento si interrompe al raggiungimento del numero massimo di iterazioni.

La scelta del tasso di apprendimento influenza la velocità con cui la rete impara i parametri. Un valore troppo basso comporta un maggior numero di iterazioni necessarie per giungere a una soluzione accettabile e il rischio di rimanere bloccati in un minimo locale; dall'altro lato un valore elevato rende difficile la convergenza.

Di seguito si riportano le versioni del MLP che hanno ottenuto i risultati più significativi.

4.2.1 MLP 01

Il primo modello impiega un classico algoritmo di retropropagazione dell'errore e come funzione di costo la somma dei quadrati dei residui (SSE). Tra le varie configurazioni i risultati migliori si sono registrati con l'utilizzo di 4 nodi nel livello nascosto e un tasso di apprendimento pari al 5%⁴⁴. Utilizzando come training set le stesse identiche osservazioni del primo modello Logit ("Logit18_01") e come input tutti gli indici riferiti all'anno 2018 a disposizione, si ottiene la rete rappresentata in Figura 25.

⁴⁴ Risultati soddisfacenti sono stati ottenuti anche con un valore del 3%.

Figura 25. Modello MLP 01

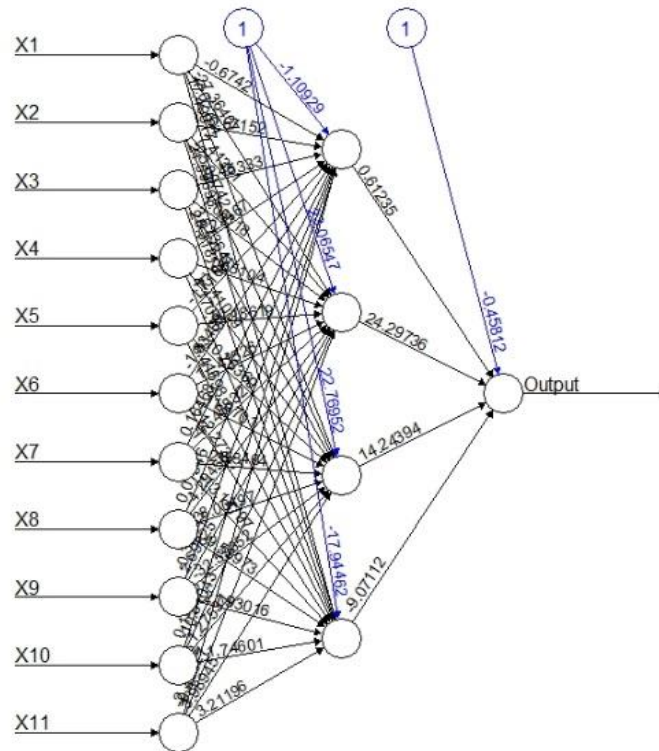


Tabella 11. Matrice di confusione del modello MLP 01

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 81 (44.02%) | 14 (7.61%) | 95 51.63% |
| | 1 | 11 (5.98%) | 78 (42.39%) | 89 48.37% |
| Tot. | | 92 50% | 92 50% | 184 100% |

Nonostante un'accuratezza dell'86.41% e un errore del I tipo inferiore all'8%, la rete non riesce a raggiungere le performance ottenute dall'analogo modello Logit.

Se invece si selezionano in modo casuale le osservazioni da includere nel training set, si riesce a migliorare leggermente l'errore totale portandolo dal 13.59% al 13.04%, ma l'errore di I tipo sale all'8.70%.

Tabella 12. Matrice di confusione del modello MLP 01 (training set alternativo)

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|---------------|
| | | 0 | 1 | |
| Prevista | 0 | 84 (45.65%) | 16 (8.70%) | 100 54.35% |
| | 1 | 8 (4.35%) | 76 (41.30%) | 84 45.65% |
| Tot. | | 92 50% | 92 50% | 184 100% |

4.2.2 MLP 02

Nella seconda versione l'unica differenza è rappresentata dalla scelta della funzione di entropia incrociata come funzione di costo.

Figura 26. Modello MLP 02

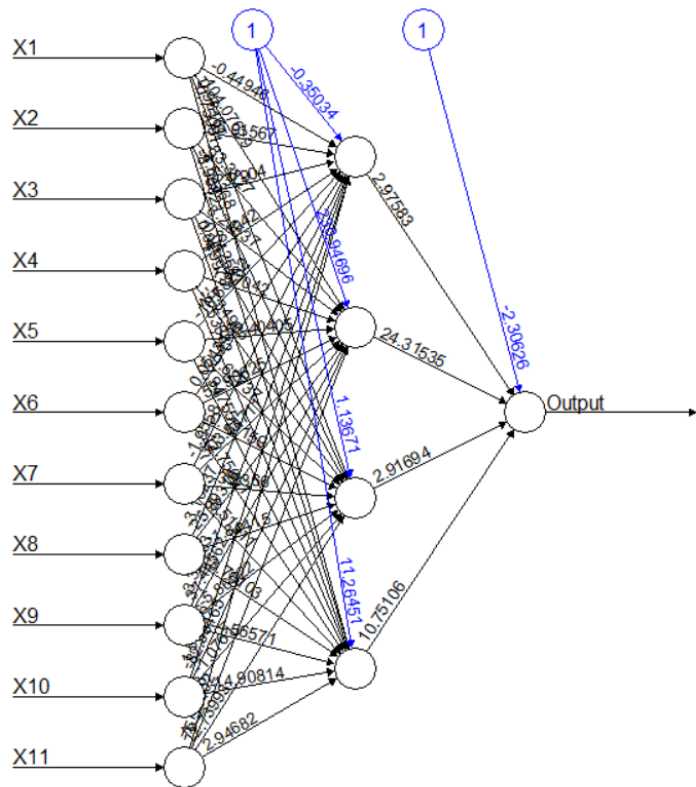


Tabella 13. Matrice di confusione del modello MLP 02

| Classe | | Osservata | | Tot. |
|----------|---|-----------------------|-----------------------|---------------|
| | | 0 | 1 | |
| Prevista | 0 | 89 (48.37%) | 14 (7.61%) | 103 55.98% |
| | 1 | 3 (1.63%) | 78 (42.39%) | 81 44.02% |
| Tot. | | 92 50% | 92 50% | 184 100% |

In questo caso si giunge a un tasso di accuratezza identico al miglior modello Logit (90.76%), nonostante un errore del I tipo superiore (7.61% contro 5.98%).

4.2.3 MLP 03

Nel terzo modello è stato utilizzato l'algoritmo di ottimizzazione Resilient Backpropagation (Rprop-) in combinazione con la funzione di costo SSE.

Figura 27. Modello MLP 03

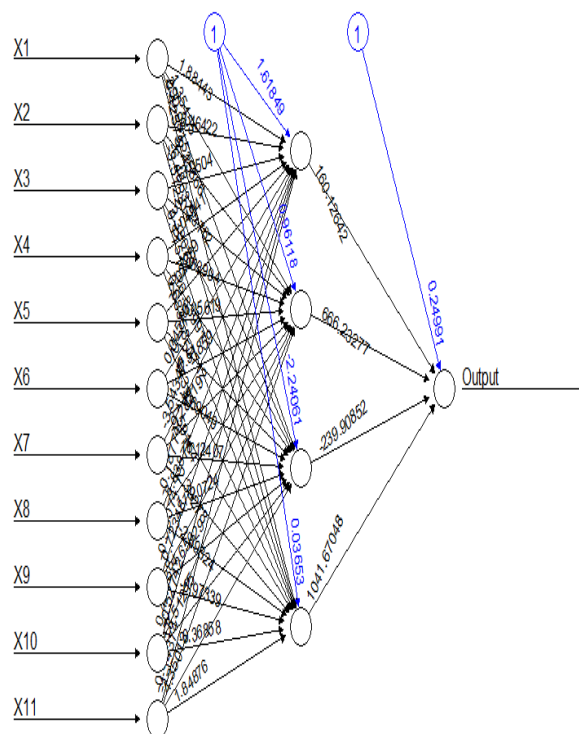


Tabella 14. Matrice di confusione del modello MLP 03

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 84 (45.65%) | 11 (5.98%) | 95 51.63% |
| | 1 | 8 (4.35%) | 81 (44.02%) | 89 48.37% |
| Tot. | | 92 50% | 92 50% | 184 100% |

L'errore di I tipo è molto basso, allo stesso livello del modello benchmark, mentre l'accuratezza è leggermente inferiore (89.67%).

In generale sia il secondo che il terzo modello forniscono una buona classificazione, pertanto nelle versioni successive verranno condotti tentativi sia con l'algoritmo Backpropagation sia con il Resilient Backpropagation (Rprop-).

4.2.4 MLP 04

Successivamente sono state incluse solo le variabili esplicative risultate significative nella migliore versione della regressione logistica ("Logit18_02"), ovvero:

- "Rotazione del Capitale investito (2018)";
- "ROA (2018)";
- "Grado di indipendenza da terzi (2018)" e
- "CCN/Attivo (2018)".

Lo scopo è verificare se la riduzione degli input e, di conseguenza, dei parametri da stimare rende la rete più efficiente ed efficace.

Con l'algoritmo di backpropagation i migliori risultati sono stati ottenuti rispettivamente con 2 e 4 nodi nel livello nascosto.

Figura 28. Modello MLP 04 con l' algoritmo di backpropagation e due nodi nell'hidden layer

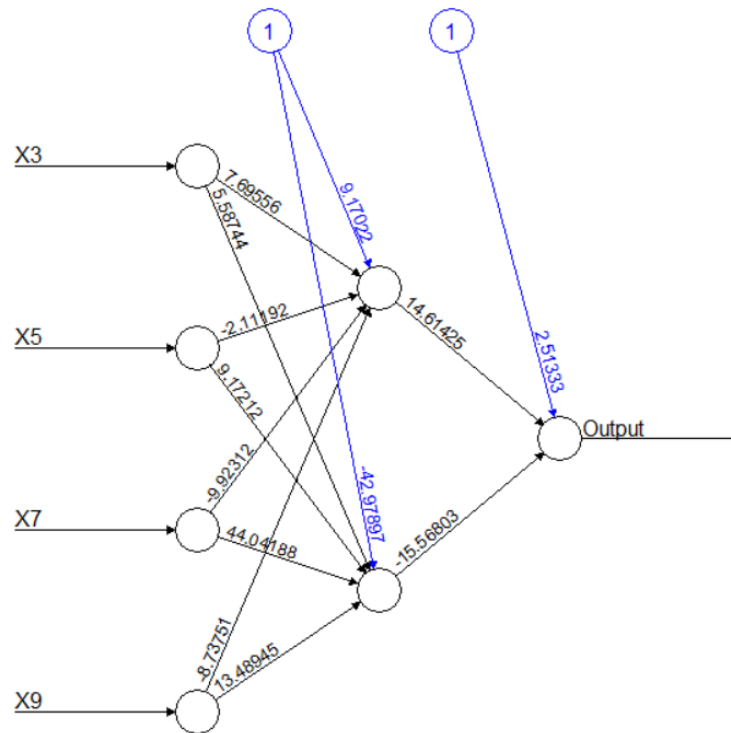


Tabella 15. Matrice di confusione del modello MLP 04

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 80 (43.48%) | 10 (5.43%) | 90 48.91% |
| | 1 | 12 (6.52%) | 82 (44.57%) | 94 51.09% |
| Tot. | | 92 50% | 92 50% | 184 100% |

Figura 29. Modello MLP 04 con l'algoritmo di backpropagation e quattro nodi nell'hidden layer

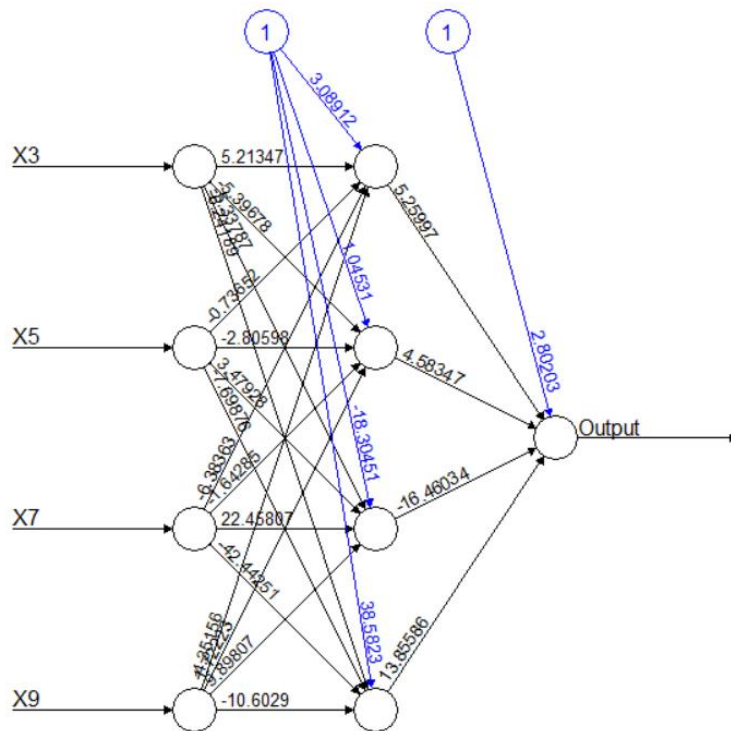


Tabella 16. Matrice di confusione del modello MLP 04

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 79 (42.93%) | 10 (5.43%) | 89 48.37% |
| | 1 | 13 (7.07%) | 82 (44.57%) | 95 51.63% |
| Tot. | | 92 50% | 92 50% | 184 100% |

Queste reti neurali riescono a ottenere un errore del I tipo (5.43%) inferiore a tutti i modelli visti finora, migliorando anche il risultato del benchmark (5.98%). Dall'altro lato aumenta leggermente l'errore di II tipo portando l'accuratezza (pari all'88.05% nella prima rete e all'87.50% nella seconda) a valori leggermente inferiori a quelli ottimali.

Considerando l'algoritmo di ottimizzazione "Rprop-" migliorano sia l'accuratezza (rispettivamente 89.13% e 89.67%) e sia l'errore di II tipo, ma cresce l'errore di I specie (6.52% e 7.07%).

Figura 30. Modello MLP 04 con algoritmo di ottimizzazione Rprop-

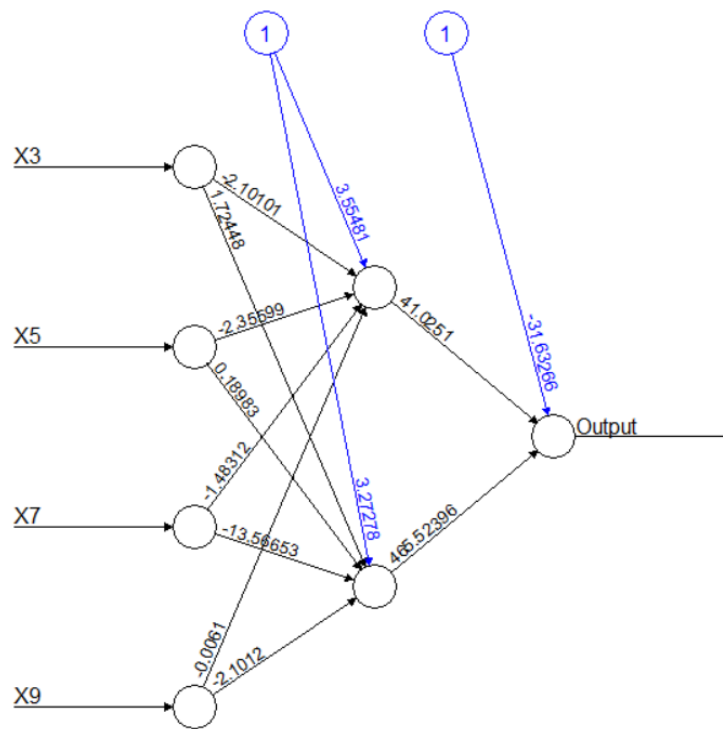


Tabella 17. Matrice di confusione del modello MLP 04

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 84 (45.65%) | 12 (6.52%) | 96 52.17% |
| | 1 | 8 (4.35%) | 80 (43.48%) | 88 47.83% |
| Tot. | | 92 50% | 92 50% | 184 100% |

Figura 31. Modello MLP 04 con algoritmo di ottimizzazione Rprop- (seconda versione)

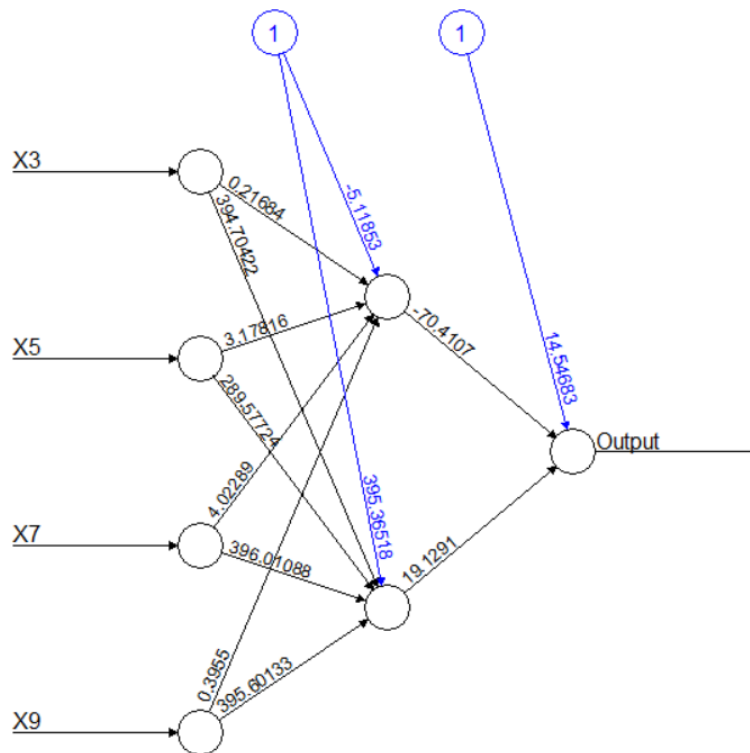


Tabella 18. Matrice di confusione del modello MLP 04

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 86 (46.74%) | 13 (7.07%) | 99 53.80% |
| | 1 | 6 (3.26%) | 79 (42.93%) | 85 46.20% |
| Tot. | | 92 50% | 92 50% | 184 100% |

4.2.5 MLP 05

Nella quinta versione del modello è stata ampliata la quota di dataset destinata alla fase di addestramento (70% contro il precedente 50%).

La rete avendo ora a disposizione più osservazioni dovrebbe essere in grado di apprendere meglio la struttura esistente tra input e output. Dall'altro lato, la riduzione del numero di osservazioni nel campione di test aumenta la sensibilità delle metriche di valutazione delle prestazioni.

Figura 32. Modello MLP 05 con algoritmo di ottimizzazione backpropagation

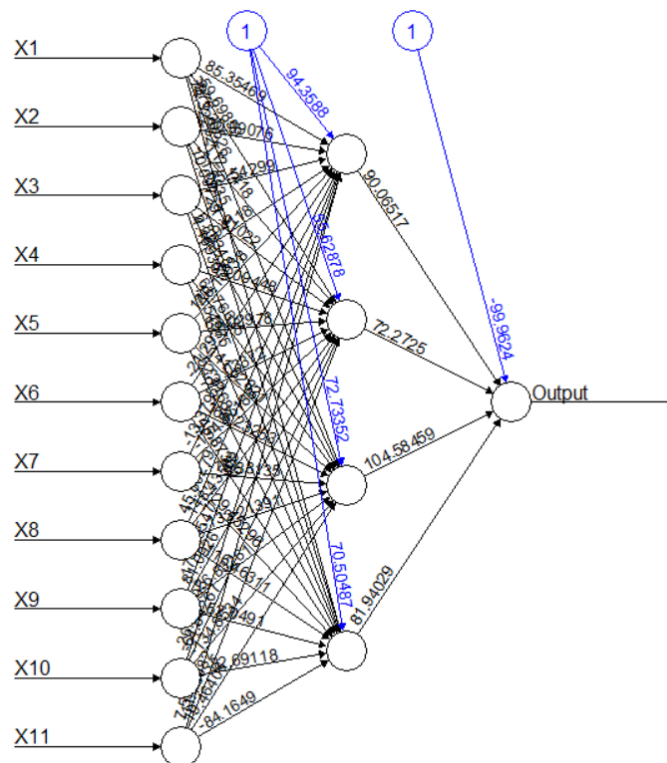


Tabella 19. Matrice di confusione del modello MLP 05

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 50 (45.45%) | 4 (3.64%) | 54 49.09% |
| | 1 | 5 (4.55%) | 51 (46.36%) | 56 50.91% |
| Tot. | | 55 50% | 55 50% | 110 100% |

Figura 33. Modello MLP 05 con algoritmo di ottimizzazione Rprop-

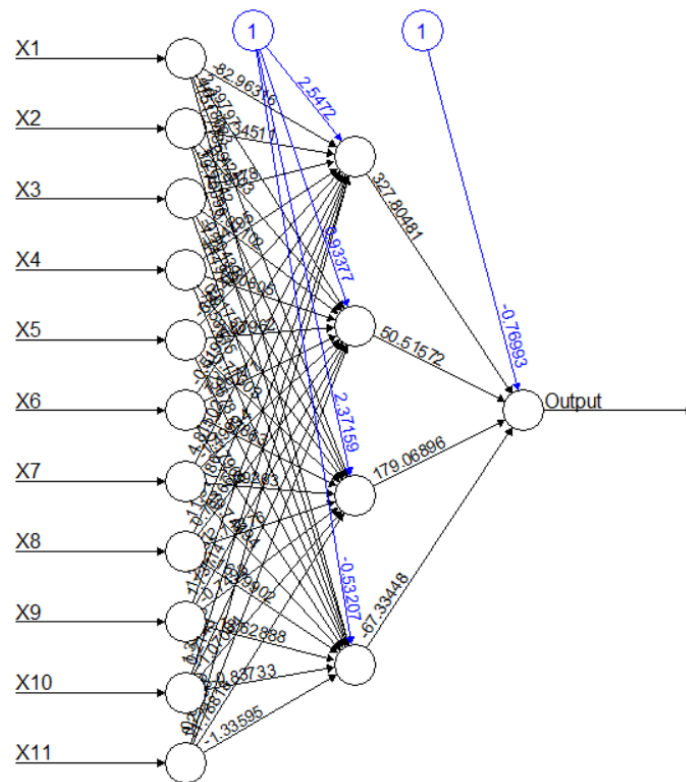


Tabella 20. Matrice di confusione del modello MLP 05

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|-------------|
| | | 0 | 1 | |
| Prevista | 0 | 50 (45.45%) | 5 (4.55%) | 55 50% |
| | 1 | 5 (4.55%) | 50 (45.45%) | 55 50% |
| Tot. | | 55 50% | 55 50% | 110 100% |

La prima rete risulta superiore a tutti gli altri modelli, con un grado di accuratezza del 91.81% e un errore di I tipo del 3.64%. Anche la seconda fornisce ottime performance: un'accuratezza al 90.90% e un errore di I tipo al 4.55%.

4.2.6 MLP 06

Nel sesto e ultimo modello sono state introdotte le variabili esplicative ritardate di un periodo (anno 2017). Ciò ha reso necessaria l'introduzione di un secondo livello nascosto. La configurazione ottimale della rete è stata raggiunta con 5 e 3 unità nascoste.

Figura 34. Modello MLP 06

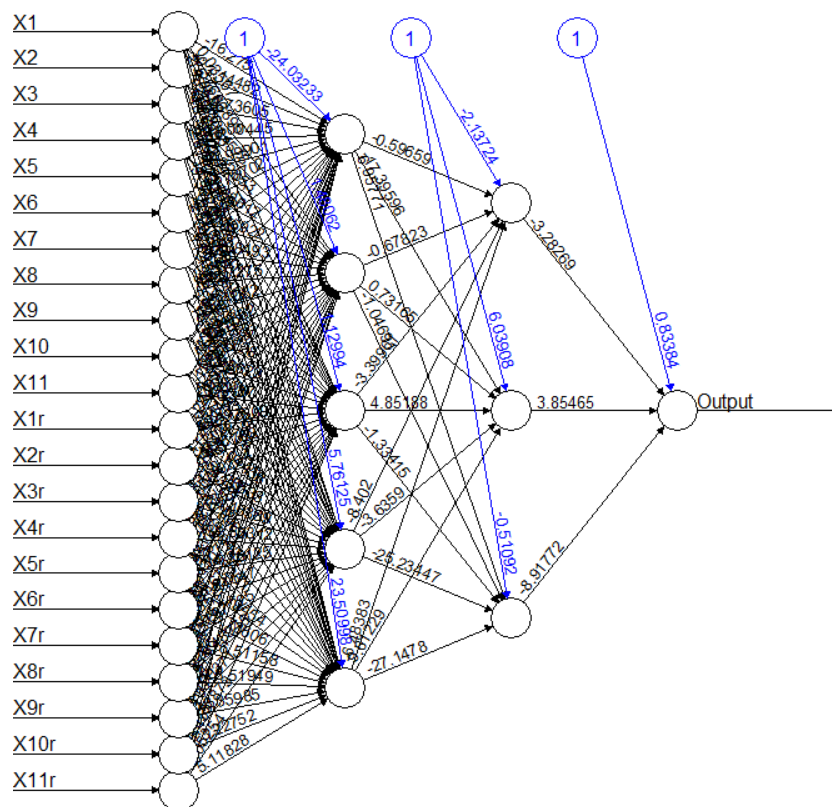


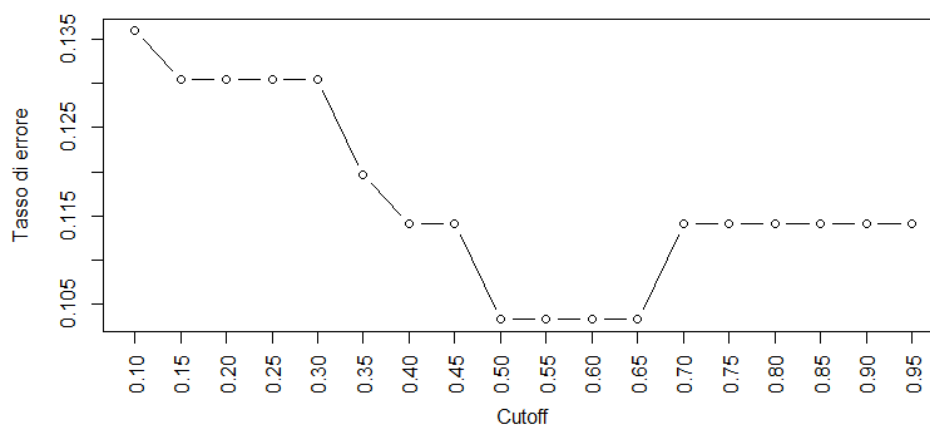
Tabella 21. Matrice di confusione del modello MLP 06

| Classe | | Osservata | | Tot. |
|----------|---|----------------|----------------|--------------|
| | | 0 | 1 | |
| Prevista | 0 | 80 (43.48%) | 7 (3.80%) | 87 47.28% |
| | 1 | 12 (6.52%) | 85 (46.20%) | 97 52.72% |
| Tot. | | 92 50% | 92 50% | 184 100% |

Della relativa matrice di confusione si vuole evidenziare la bassa percentuale di aziende fallite erroneamente classificate come sane (3.80%), mentre l'accuratezza presenta un valore sostanzialmente in linea con gli altri modelli (89.68%).

Infine, mettendo in relazione il tasso di errore complessivo di questo modello con il valore soglia, il valore 0.50 si conferma la scelta ottimale per suddividere le osservazioni nelle due classi.

Figura 35. Relazione tra il tasso di errore complessivo e il punto di cut-off



4.3 Sintesi dei risultati ottenuti

Lo scopo di un classificatore binario è massimizzare il numero di osservazioni correttamente classificate (ossia il grado di accuratezza). In questo contesto molta attenzione è rivolta anche alla minimizzazione del numero di aziende classificate come sane, ma risultate fallite (errore di I tipo). Ciò perché, dal punto di vista di chi finanzia le aziende, vi è un'evidente asimmetria tra le due tipologie di errore: la perdita derivante dalla concessione del credito a un soggetto non sano è di gran lunga superiore al mancato guadagno per non aver finanziato un soggetto sano erroneamente classificato come non meritevole di credito.

La Tabella 22 riepiloga i risultati ottenuti focalizzandosi su queste due metriche.

Tabella 22. Riassunto delle prestazioni dei vari modelli

| Modello | Corretta classificazione azienda sana | Corretta classificazione azienda fallita | Accuratezza | Errore I tipo | Errore II tipo | Algoritmo | Nodi nascosti |
|-------------------|---|--|-------------|------------------|-------------------|------------------|------------------|
| MLP 05 | 45.45% | 46.36% | 91.81% | 3.64% | 4.55% | backprop. | 4 |
| MLP 05 | 45.45% | 45.45% | 90.90% | 4.55% | 4.55% | rprop- | 4 |
| Logit18_02 | 46.74% | 44.02% | 90.76% | 5.98% | 3.26% | - | - |
| MLP 02 | 48.37% | 42.39% | 90.76% | 7.61% | 1.63% | backprop. | 4 |
| MLP 06 | 43.48% | 46.20% | 89.68% | 3.80% | 6.52% | backprop. | 5+3 |
| MLP 03 | 45.65% | 44.02% | 89.67% | 5.98% | 4.35% | rprop- | 4 |
| MLP 04 | 46.74% | 42.93% | 89.67% | 7.07% | 3.26% | rprop- | 2 |
| Logit18_01 | 45.65% | 43.48% | 89.13% | 6.52% | 4.35% | - | - |
| MLP 04 | 45.65% | 43.48% | 89.13% | 6.52% | 4.35% | rprop- | 2 |
| MLP 04 | 43.48% | 44.57% | 88.05% | 5.43% | 6.52% | backprop. | 2 |
| MLP 04 | 42.93% | 44.57% | 87.50% | 5.43% | 7.07% | backprop. | 4 |
| Logit1718 | 45.11% | 42.39% | 87.50% | 7.61% | 4.89% | - | - |
| MLP 01 | 45.65% | 41.30% | 86.95% | 8.70% | 4.35% | backprop. | 4 |
| MLP 01 | 44.02% | 42.39% | 86.41% | 7.61% | 5.98% | backprop. | 4 |
| Logit18_03 | 44.57% | 41.30% | 85.87% | 8.70% | 5.43% | - | - |

Il modello che riesce a classificare più correttamente lo stato di salute delle aziende sembra essere la rete neurale denominata MLP05. Il risultato può essere spiegato dal maggior numero di osservazioni a disposizione della rete nella fase di addestramento, che ha consentito una migliore stima dei parametri.

Per quanto riguarda l'errore di I tipo le performance migliori sono attribuibili alle reti neurali MLP05 e MLP06. In particolare, quest'ultima beneficia dell'introduzione di un maggior numero di variabili esplicative. Spesso le reti neurali hanno ottenuto un errore di I tipo inferiore rispetto ai modelli di regressione logistica.

Conclusioni

Lo studio dei modelli di credit scoring ha fatto emergere alcuni loro limiti, tra cui:

- la mancanza di un'analisi qualitativa e andamentale della controparte;
- un'eccessiva semplificazione delle informazioni a disposizione. Un solo score deve sintetizzare un intero set di dati;
- la mancanza di informazioni di alta qualità. I dati contabili sono, infatti, informazioni ritardate e potenzialmente distorte.

Nonostante ciò, presentano degli indubbi vantaggi, che hanno favorito la loro diffusione: riescono a produrre una valutazione oggettiva sul merito di credito della controparte in tempi rapidi e a basso costo.

Dall'esperimento si evince come, in generale, le Reti neurali artificiali siano in grado di fornire prestazioni in linea con i modelli benchmark e talvolta superiori. Pertanto, possono rappresentare un'ottima alternativa al modello Logit e alle altre tecniche statistiche tradizionali, soprattutto in presenza di una grande quantità di dati e informazioni da analizzare. Un ulteriore aumento nel grado di accuratezza può giungere dal ricorso a strutture di reti neurali più evolute o a tecniche di machine learning alternative (ad esempio foreste casuali e macchine a vettori di supporto). Inoltre, sarebbe interessante sperimentare l'introduzione di modelli ibridi, frutto dell'unione di metodologie di analisi differenti, e l'inclusione di informazioni alternative, come ad esempio i dati sulle transazioni finanziarie e sulle attività compiute online dalla controparte.

Il principale limite alla diffusione di queste tecniche di apprendimento automatico è rappresentato dalla mancanza di trasparenza. Il modello Logit consente di individuare facilmente le cause che hanno prodotto un determinato score e comprendere le aree su cui l'azienda deve intervenire. Le nuove tecniche di credit scoring, invece, sono più sofisticate, opache e difficili da interpretare e spiegare rispetto alle tecniche tradizionali. Lo sforzo dei ricercatori attualmente è rivolto allo sviluppo di metodologie per rendere i modelli di machine learning più interpretabili, senza sacrificare la loro accuratezza.

Appendice 1 (Dataset)

Il dataset utilizzato nell'esperimento (Cap. 4) è accessibile al seguente link:

<https://drive.google.com/drive/folders/1fNv0hC2DE-V35Hx-Gri48-blXAiCjfbk?usp=sharing>

Appendice 2 (Codice)

Per quanto riguarda i modelli Multilayer Perceptron impiegati nell'esperimento (Cap. 4), si riportano i due script di R contenenti il codice necessario per la loro costruzione e per la valutazione delle loro performance. Il primo si riferisce ai modelli MLP che utilizzano come variabili esplicative solo i dati riferiti al 2018, il secondo ai modelli MLP che includono come input anche le informazioni del 2017.

Appendice 2.1 Codice MLP (2018)

```
library(readxl)
library(writexl)
library(neuralnet)
library(caret)
#importazione dei dati
Dataset <- read_excel("C:/Users/boldr/Desktop/R/ANN/Dataset.xlsx", sheet = "2018")[, -1]
View(Dataset)
str(Dataset)
dim(Dataset)
#normalizzazione dei dati (tra 0 e 1)
normalize=function(x){return((x-min(x))/(max(x)-min(x)))}
Dataset_norm=as.data.frame(lapply(Dataset,normalize))
summary(Dataset_norm)
#suddivisione del dataset in training e testing
Indice=createDataPartition(Dataset_norm$Output,p=0.5,list=FALSE)
Train=Dataset_norm[Indice,]
Test=Dataset_norm[-Indice,]
dim(Train)
dim(Test)
```

```

#ridenominazione delle variabili
colnames(Train)=c("X1","X2","X3","X4","X5","X6","X7","X8","X9",
,"X10","X11","Output")
head(Train)
colnames(Test)=c("X1","X2","X3","X4","X5","X6","X7","X8","X9",
,"X10","X11","Output")
head(Test)
#funzione logistica
sigmoid=function(x) {1.0 / (1.0 + exp(-x))}
#MLP01
MLP=neuralnet(Output~X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11,data=T
rain,hidden=4,learningrate=0.05,threshold
=0.02,
stepmax=1e7,algorithm="backprop",act.fct=sigmoid,
linear.output=FALSE, err.fct = "sse", likelihood=TRUE)
#MLP02
MLP=neuralnet(Output~X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11,data=T
rain,hidden=4,learningrate=0.05,threshold
=0.02,
stepmax=1e7,algorithm="backprop",act.fct=sigmoid,
linear.output=FALSE, err.fct = "ce", likelihood=TRUE)
#MLP03
MLP=neuralnet(Output~X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11,data=T
rain,hidden=4,threshold =0.02, stepmax=1e7,algorithm="rprop-
",act.fct=sigmoid, linear.output=FALSE, err.fct = "sse",
likelihood=TRUE, learningrate.limit = NULL, learningrate.factor
= list(minus = 0.5,plus = 1.2), learningrate = NULL)
#MLP04 (con backprop.)
MLP=neuralnet(Output~X3+X5+X7+X9,data=Train,hidden=2,learningr
ate=0.05,threshold =0.02, stepmax=1e7,algorithm=
"backprop",act.fct=sigmoid, linear.output=FALSE, err.fct =
"sse", likelihood=TRUE)
#MLP04 (con rprop-)
MLP=neuralnet(Output~X3+X5+X7+X9,data=Train,hidden=2,threshold
=0.02, stepmax=1e7,algorithm="rprop-",act.fct=sigmoid,
linear.output=FALSE, err.fct = "sse", likelihood=TRUE,
learningrate.limit = NULL, learningrate.factor = list(minus =
0.5,plus = 1.2), learningrate = NULL)
plot(MLP)
print(MLP)
#Previsione nel testing set
Previsione=compute(MLP,Test)
Previsione$net.result
#post-processing output
Previsione_MLP=ifelse(Previsione$net.result>0.5,1,0)
#Performance del modello
diff=cbind(Previsione_MLP,Test$Output,Previsione$net.result)
diff
#matrice di confusione
table(factor(Previsione_MLP),factor(Test$Output))
#Calcolo del tasso di errore al variare del punto di cut-off
Tasso_errore=function(trues, predicted_prb, cutoff){
preds=ifelse(predicted_prb<cutoff,0,1)

```

```

tab=table(preds, trues)
round((tab[1,2]+tab[2,1])/sum(tab), 4)
}
Tasso_errore(Test$Output, Previsione$net.result, 0.5)
cutoff_value=c(0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5,0.55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95)
errRate=capture.output(for (cutoff_value in cutoff_value) {
  cat(Tasso_errore(Test$Output, Previsione$net.result,
cutoff_value), "\n")
})
errRate=as.numeric(errRate)
errRate
cutoff_value=c(0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5,0.55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95)
plot(cutoff_value,errRate,type = "b",xlab = "Cutoff", ylab =
"Tasso di errore",main = "Andamento del tasso di errore al
variare del punto di cutoff", xaxt="n")
axis(1,at=seq(0.05,0.95,by=0.05),las=2)

```

Appendice 2.2 Codice MLP (2017-2018)

```

library(readxl)
library(writexl)
library(neuralnet)
library(caret)
#importazione dei dati
Dataset_1718 <-
read_excel("C:/Users/boldr/Desktop/R/ANN/Dataset.xlsx", sheet =
"2017+2018")[, -1]
View(Dataset_1718)
str(Dataset_1718)
dim(Dataset_1718)
#normalizzazione dei dati (tra 0 e 1)
normalize=function(x){return((x-min(x))/(max(x)-min(x)))}
Dataset_norm_1718=as.data.frame(lapply(Dataset_1718,normalize)
)
summary(Dataset_norm_1718)
#suddivisione del dataset in training e testing
Indice_1718=createDataPartition(Dataset_norm_1718$Output,p=0.5
,list=FALSE)
Train_1718=Dataset_norm_1718[Indice_1718,]
Test_1718=Dataset_norm_1718[-Indice_1718,]
dim(Train_1718)
dim(Test_1718)
#ridenominazione delle variabili
colnames(Train_1718)=c("X1","X1r","X2","X2r","X3","X3r","X4","
X4r","X5","X5r","X6","X6r","X7","X7r","X8","X8r","X9","X9r","X
10","X10r","X11","X11r","Output")

```

```

head(Train_1718)
colnames(Test_1718)=c("X1", "X1r", "X2", "X2r", "X3", "X3r", "X4", "X
4r", "X5", "X5r", "X6", "X6r", "X7", "X7r", "X8", "X8r", "X9", "X9r", "X1
0", "X10r", "X11", "X11r", "Output")
head(Test_1718)
#funzione logistica
sigmoid=function(x) {1.0 / (1.0 + exp(-x))}
#MLP06 (con backprop.)
MLP_1718=neuralnet(Output~X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11+X
1r+X2r+X3r+X4r+X5r+X6r+X7r+X8r+X9r+X10r+X11r,data=Train_1718,h
idden=c(5,3),learningrate=0.05,threshold
=0.02,
stepmax=1e7,algorithm="backprop",act.fct=sigmoid,
linear.output=FALSE, err.fct = "sse", likelihood=TRUE)
plot(MLP_1718)
#MLP06 (con rprop-)
MLP_1718=neuralnet(Output~X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11+X
1r+X2r+X3r+X4r+X5r+X6r+X7r+X8r+X9r+X10r+X11r,data=Train_1718,h
idden=c(4,2),threshold =0.02, stepmax=1e7,algorithm="rprop-
",act.fct=sigmoid, linear.output=FALSE, err.fct = "sse",
likelihood=TRUE, learningrate.limit = NULL, learningrate.factor
= list(minus = 0.5,plus = 1.2), learningrate = NULL)
plot(MLP_1718)
#Previsione nel testing set
Previsione_1718=compute(MLP_1718,Test_1718)
head(Previsione_1718$net.result)
Previsione_MLP_1718=ifelse(Previsione_1718$net.result>0.5,1,0)
#post-processing output
#Performance del modello
diff=cbind(Previsione_MLP_1718,Test_1718$Output)
head(diff)
table(factor(Previsione_MLP_1718),factor(Test_1718$Output))
#matrice di confusione
#Calcolo del tasso di errore al variare della soglia
Tasso_errore=function(trues, predicted_prb, cutoff){
  preds=ifelse(predicted_prb<cutoff,0,1)
  tab=table(preds, trues)
  round((tab[1,2]+tab[2,1])/sum(tab), 4)
}
Tasso_errore(Test_1718$Output, Previsione_1718$net.result, 0.5)
cutoff_value=c(0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5,0.
55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95)
errRate=capture.output(for (cutoff_value in cutoff_value) {
  cat(Tasso_errore(Test_1718$Output, Previsione_1718$net.result,
cutoff_value),"\n")
})
errRate=as.numeric(errRate)
errRate
cutoff_value=c(0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4,0.45,0.5,0.
55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95)

```

```
plot(cutoff_value,errRate,type = "b",xlab = "Cutoff", ylab =  
"Tasso di errore",main = "Andamento del tasso di errore al  
variare del punto di cutoff", xaxt="n")  
axis(1,at=seq(0.05,0.95,by=0.05),las=2)
```

Bibliografia

- ABI (2019), *Le nuove regole europee in materia di default*, disponibile a https://www.abi.it/DOC_Mercati/Crediti/Credito-alle-imprese/Tavolo%20CIRI%20-%20Guida%20semplice%20default/Guida%20Default_imprese.pdf ;
- Alpaydin E., *Introduction to Machine Learning*, The MIT Press, 2014;
- Altman E. I. (1968), *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, *The Journal of Finance*, Vol. 23, No. 4., pp. 589-609;
- Altman E. I., Marco G., Varetto F. (1994), *Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)*, *Journal of Banking and Finance*, Vol. 18, pp. 505-529;
- Atiya A. (2001), *Bankruptcy prediction for credit risk using neural networks: a survey and new results*, *IEEE Transactions on neural networks*, Vol. 12, No. 4, pp. 929-935;
- Aziz M. A., Humayon A. D. (2006), *Predicting corporate bankruptcy: where we stand?*, *Corporate Governance*, Vol. 6, pp. 18-33;
- Banca d'Italia, Circolare n. 285 del 17 dicembre 2013, *Disposizioni di vigilanza per le banche*;
- Banca d'Italia (2020), *Relazione annuale 2019*, disponibile a https://www.bancaditalia.it/pubblicazioni/relazione-annuale/2019/rel_2019.pdf ;
- Barbagallo C. (2019), *Crisi e regolamentazione finanziaria: cambiamenti e prospettive*, lavoro presentato al Convegno "Un decennio di regole finanziarie anticrisi, come la gestione delle banche italiane ha risposto alle domande dell'economia", Modena, 1 marzo 2019, disponibile a <https://www.bancaditalia.it/pubblicazioni/interventi-vari/int-var-2019/Barbagallo-01.03.2019.pdf> ;
- Basel Committee on Banking Supervision (1988), *International Convergence of Capital Measurement and Capital Standards*, disponibile a <https://www.bis.org/publ/bcbs04a.pdf> ;
- Basel Committee on Banking Supervision (2000), *Principles for the Management of Credit Risk*, disponibile a <https://www.bis.org/publ/bcbs75.pdf> ;

- Basel Committee on Banking Supervision (2004), *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*, disponibile a <https://www.bis.org/publ/bcbs107.pdf> ;
- Basel Committee on Banking Supervision (2005), *An Explanatory Note on the Basel II IRB Risk Weight Functions*, disponibile a <https://www.bis.org/bcbs/irbriskweight.pdf> ;
- Basel Committee on Banking Supervision (2010), *Basel III: A global regulatory framework for more resilient banks and banking systems*, disponibile a https://www.bis.org/publ/bcbs189_dec2010.pdf ;
- Basel Committee on Banking Supervision (2017), *Basel III: Finalising post-crisis reforms*, disponibile a <https://www.bis.org/bcbs/publ/d424.pdf> ;
- Beaver W. H. (1966), *Financial ratios predictors of failure*, Journal of Accounting Research, Vol. 4, pp. 71-111;
- Becker W. E., Kennedy P. E. (1992), *A Graphical Exposition of the Ordered Probit*, Econometric Theory, Vol. 8, pp. 127-131;
- Bellovary J. L., Don E. Giacomino, Akers M. D. (2007), *A Review of Bankruptcy Prediction Studies: 1930 to Present*, Journal of Financial Education, Vol. 33, pp. 1-42;
- Chandra A. L., *McCulloch-Pitts Neuron — Mankind's First Mathematical Model Of A Biological Neuron*, in Towards data science (blog), 24/07/2018, disponibile a <https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1> ;
- Coats P. K., Franklin L. (1993), *Recognizing Financial Distress Patterns Using a Neural Network Tool*, Financial Management, Vol. 22, No. 3, pp. 142-155;
- Collins R. A., Green R. D. (1982), *Statistical Methods for Bankruptcy Forecasting*, Journal of Economics and Business, Vol. 34, pp. 349-354;
- Corazza M. (2000), *Un approccio "Group Method of Data Handling" alla soft-computation: i polinomi approssimanti di Ivakhnenko*, Università Ca' Foscari di Venezia;
- Corazza M., Fasano G., Gusso R. (2013), *Particle Swarm Optimization with non-smooth penalty reformulation, for a complex portfolio selection problem*, Applied Mathematics and Computation, Vol. 224, pp. 611-624;

- Corazza M., Fasano G., Gusso R., Pesenti R. (2019), *A comparison among Reinforcement Learning algorithms in financial trading systems*, Working paper, No. 33, Dipartimento di Economia, Università Ca' Foscari di Venezia;
- Corazza M., Parpinel F., Pizzi C. (2019), *Can PSO improve TA-based trading systems?*, Neural Advances in Processing Nonlinear Dynamic Signals, Springer International Publishing, Vol. 102, pp. 277-288;
- Cybenko G. (1989), *Approximation by superpositions of a sigmoidal function*, *Mathematics of Control, Signals, and Systems*, Vol. 2, pp. 303–314;
- DARPA (1989), *DARPA neural network study final report*, Technical Report No. 840;
- Durand D. (1941), *Risk Elements in Consumer Instalment Financing*, Working Paper, NBER;
- EBA (2020), *EU-wide transparency exercise*, disponibile a <https://eba.europa.eu/risk-analysis-and-data/eu-wide-transparency-exercise> ;
- Fidrmuc J., Lind R. (2017), *Macroeconomic impact of Basel III: Evidence from a meta-analysis*, National Bank of Slovakia, Working Paper WP 5/2017;
- Fischer R. A. (1936), *The Use Of Multiple Measurement In Taxonomic Problems*, *Annals of Eugenics*, Vol. 7, pp. 179-188;
- Gambacorta L., Huang Y., Qiu H., Wang J. (2019), *How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm*, Bank for International Settlements, Working Paper, No. 834;
- Izzi L., Oricchio G., Vitale L., *Basel III Credit Rating Systems. An Applied Guide to Quantitative and Qualitative Models*, Palgrave Macmillan, 2012;
- James G., Witten D., Hastie T., Tibshirani R., *An Introduction to Statistical Learning: with Applications in R*, New York, Springer, 2013;
- Kleinert M. K. (2014), *Comparison of accounting-based bankruptcy prediction models of Altman (1968), Ohlson (1980), and Zmijewski (1984) to German and Belgian listed companies during 2008 -2013*, Tesi magistrale, University of Twente;

- Lennox C. (1999), *Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches*, Journal of Economics and Business, Vol. 51, pp. 347–364;
- Louzada F., Ara A., Fernandes G. B. (2016), *Classification methods applied to credit scoring: A systematic review and overall comparison*, Surveys in Operations Research and Management Science, Vol. 21, pp. 117-134;
- Luo J., Zhang D. (2019), *Jack Ma's \$290 Billion Loan Machine Is Changing Chinese Banking*, Bloomberg, 28/07/2019;
- Magnus M., Margerit A., Mesnard B., Korpas A. (2017), *Upgrading the Basel standards: from Basel III to Basel IV?*, European Parliament, Briefing paper, disponibile a [https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_BRI\(2016\)587361](https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_BRI(2016)587361) ;
- McCulloch W.S., Pitts W. (1943), *A logical calculus of the ideas immanent in nervous activity*, Bulletin of Mathematical Biophysics, Vol. 5, pp. 115-133;
- Minsky M., Papert S. (1969), *Perceptrons: An Introduction to Computational Geometry*, The MIT Press;
- Modefinance (2017), *The Multi Objective Rating Evaluation(MORE)*, disponibile a <https://cra.modefinance.com/pdf/modefinanceCorporatesScoringMethodology.pdf> ;
- Modefinance (2019), *Credit rating e credit score: quali differenze?*, disponibile a <https://www.modefinance.com/it/blog/2019-01-24-credit-rating-e-credit-score-quali-differenze-> ;
- Moscatelli M., Narizzano S., Parlapiano F., Viggiano G. (2019), *Corporate default forecasting with machine learning*, Banca d'Italia, Working Paper No. 1256;
- Nadali L. (2017), *Can an Artificial Neural Network automate the credit rating process of small and medium-sized enterprises?*, Tesi di Laurea magistrale, Università Ca' Foscari di Venezia;
- Odom M. D., Sharda R. (1990), *A neural network model for bankruptcy prediction*; Conference Paper, International Joint Conference on Neural Networks, San Diego, CA;
- Ohlson J. A. (1980), *Financial Ratios and the Probabilistic Prediction of Bankruptcy*, Journal of Accounting Research, Vol. 18, pp. 109-131;

- Parlamento europeo e Consiglio dell'UE, Regolamento (UE) n. 575/2013, *Requisiti prudenziali per gli enti creditizi e le imprese di investimento*;
- Resti A., Sironi A., *Risk Management and Shareholders' Value in Banking*, Wiley, 2007;
- Riedmiller M., Braun H. (1993), *A direct adaptive method for faster backpropagation learning: The RPROP algorithm*, lavoro presentato alla IEEE International Conference on Neural Networks (ICNN), San Francisco, 28 Marzo-1 Aprile 1993, pp. 586-591;
- Rosenblatt F. (1958), *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, Psychological Review, Vol. 65, pp. 386-408;
- Shen L., Wang B. (2020), *MYbank Served Over 20 Million SMEs as of 2019, Further Spurring the Growth of China's Small and Micro Businesses*, Businesswire, 27/04/2020;
- Shumway T. (2001), *Forecasting Bankruptcy More Accurately: A Simple Hazard Model*, The Journal of Business, Vol. 74, pp. 101-24;
- Singh S., *Understanding the Bias-Variance Tradeoff*, in Towards data science (blog), 21/05/2018, disponibile a <https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229>;
- Sironi A. (2000), *I rating interni e i modelli per la gestione del rischio di credito*, lavoro presentato alla giornata di studio "Modelli per la gestione del rischio di credito. I ratings interni", Perugia, 11 ottobre 1999;
- Tedeschi R. (2016), *Breve storia del risk management nelle banche*, L'Atlante, 4 ottobre 2016;
- Treccani (2010), *Etimologia del termine "biomimesi"*, disponibile a http://www.treccani.it/magazine/lingua_italiana/domande_e_risposte/lessico/lessico_194.html# ;
- Tsai C., Wu J. (2008), *Using neural network ensembles for bankruptcy prediction and credit scoring*, Expert Systems with Applications, Vol. 34, pp. 2639-2649;
- Tsai C., Chen M. (2010), *Credit rating by hybrid machine learning techniques*, Applied Soft Computing, Vol. 10, pp. 374-380;

- Vinci, A. (2020), *Il primo farmaco creato dall'intelligenza artificiale: al via i test sugli esseri umani*, Corriere della Sera, 03/02/2020;
- Vinci, A. (2020), *Coronavirus, diagnosi da record con Alibaba: responso in 20 secondi (e accuratezza del 96%)*, Corriere della Sera, 05/03/2020;
- Wu Y., Gaunt C., Gray S. (2010), *A comparison of alternative bankruptcy prediction models*, Journal of Contemporary Accounting & Economics, Vol. 6, pp. 34–45;
- Yim J., Mitchell H. (2002), *A comparison of corporate failure models in Australia: Hybrid neural networks, Logit models and Discriminant analysis*, MIT, Working Paper No. 10;
- Zhang G., Hu M. Y., Patuwo B. E., Indro D. C. (1997), *Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis*, European Journal of Operational Research, Vol. 116, pp. 16-32;
- Zmijewski M. E. (1984), *Methodological Issues Related to the Estimation of Financial Distress Prediction Models*, Journal of Accounting Research, Vol. 22, pp. 59-82.