POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale

Corso di Laurea Magistrale
in Ingegneria Gestionale
percorso Finanza

Tesi di Laurea Magistrale

Sviluppo di un modello di rischio di credito mediante l'approccio di intelligenza artificiale delle reti neurali applicato al settore del noleggio e leasing di autoveicoli

Relatore
Prof. Franco Varetto

Candidato
Fabio Palmieri

Anno accademico 2019/2020

“Un giorno le macchine riusciranno a risolvere tutti i problemi,
ma mai nessuna di esse potrà porne uno”
Albert Einstein

Indice

Indice delle figure

Indice delle tabelle

Introduzione

Negli ultimi anni, l'evoluzione delle discipline informatiche ha dato notevole importanza alla progettazione di sistemi hardware e programmi software in grado di replicare i processi di ragionamento tipici dell'intelligenza umana, in due parole all'Intelligenza artificiale (AI). La possibilità che i computer imparino a svolgere lavori ad essi assegnati attraverso un meccanismo di apprendimento automatico, denominato *machine learning*, ha attirato l'attenzione di diversi ricercatori provenienti dai campi più disparati quali robotica, medicina, marketing, sicurezza, ricerca scientifica ed in particolare dalla finanza. Per quest'ultima le applicazioni sono numerose ma riconducibili in modo univoco alla costruzione di modelli in grado di sviluppare analisi predittive, che individuano i segnali contenuti nei dati aziendali per poter cogliere le informazioni di mercato e agire di conseguenza nel campo degli investimenti.

L'evoluzione tecnologica che ha coinvolto il mondo finanziario negli ultimi decenni sembra inarrestabile e ancor di più la mole di dati quotidianamente generati da queste nuove fonti. La complessità derivante dalla gestione di enormi volumi di dati si converte in possibilità, da parte degli istituti finanziari, di cogliere il maggior numero di informazioni possibili per migliorare i propri modelli di valutazione. Le moderne tecniche di *machine learning* riescono a farsi carico di questa complessità semplificandone la gestione e ottimizzando i processi decisionali di concessione del credito. L'applicazione più diffusa di tali modelli predittivi concerne il calcolo del *credit scoring*, ossia la valutazione del merito creditizio associato ad una controparte nei confronti della quale esiste un'esposizione, ma più in generale il processo di analisi del rischio di credito.

I modelli di *machine learning* comprendono al loro interno diverse tipologie di algoritmi ognuna con le proprie caratteristiche ed i propri campi di applicazione. Nel presente lavoro di tesi si focalizzerà l'attenzione sull'affascinante mondo delle reti neurali, analizzandone le diverse tipologie, le peculiarità e le principali funzioni che ne governano il funzionamento. L'obiettivo di questa tesi di laurea è quello di costruire una rete neurale in grado di riconoscere lo status di un'impresa, ossia se essa è sana o anomala, partendo dall'analisi degli indicatori economico-finanziari più rilevanti.

La tesi è articolata in sette capitoli, il primo dei quali è un breve *excursus* sull'evoluzione degli accordi di Basilea, che dal 1988 hanno giocato un ruolo fondamentale nello stabilire i requisiti minimi di adeguatezza patrimoniale volti al raggiungimento della stabilità del sistema finanziario globale. I quattro accordi regolamentari sono presentati nelle loro caratteristiche essenziali in ordine cronologico, ponendo particolare attenzione agli eventi economici che ne hanno evidenziato i limiti suggerendone le modifiche. Nel presentare le linee guida fornite negli accordi si è sottolineata l'importanza delle procedure di valutazione del rischio di credito al fine di garantire l'affidabilità degli istituti finanziari.

Tale rischio è oggetto di analisi del secondo capitolo, nel quale si sono analizzate le sue componenti specificandone la natura e le metodologie di calcolo. Si è poi proseguito con la presentazione del processo di valutazione di solvibilità effettuato dalle banche, noto come *credit scoring*, esaminando il modello statistico di regressione logistica, per poi fare una panoramica sugli attuali modelli di *machine learning*.

Dopo aver motivato l'importanza dell'analisi del rischio di credito dal punto di vista regolamentare e averne spiegato le componenti, si procede con la presentazione delle reti neurali. Il terzo capitolo inizia con lo studio dell'unità fondamentale della rete, ossia il neurone artificiale, presentando le analogie con la sua versione biologica, prosegue poi con la spiegazione delle funzioni matematiche che ne caratterizzano il funzionamento e si conclude con la presentazione delle principali tipologie di architetture adottabili ed il loro processo di addestramento.

Con il quarto capitolo si entra nel vivo del lavoro di tesi illustrando il database AIDA, messo a disposizione dal Politecnico di Torino, dal quale sono stati estratti i dati economico-finanziari delle aziende oggetto di studio. Dopo una rassegna del settore del noleggio e leasing di autoveicoli italiano per poter comprendere al meglio gli indicatori economici guida delle aziende analizzate, si descrive l'oneroso lavoro di scarico e pulizia dei dati di fondamentale importanza per la buona riuscita del modello.

Si prosegue poi con l'analisi tecnica del quinto capitolo nel quale si presenta MATLAB, ossia il software impiegato per la creazione della rete neurale riportando e spiegando nel dettaglio le righe di codice utilizzate, gli input e gli output delle funzioni del *Deep Learning Toolbox*, la personalizzazione dei parametri, i valori di default e le logiche di funzionamento che ne stanno alla base.

Dopo aver illustrato e motivato il codice implementato in MATLAB per la costruzione dei modelli, nel sesto capitolo si introducono gli strumenti impiegati nel processo di valutazione delle reti fornendo le nozioni teoriche che ne facilitano la comprensione, e si ripercorre il processo di evoluzione dell'architettura spiegando le ragioni di ogni modifica apportata.

Infine, nel settimo e ultimo capitolo si mettono in pratica gli strumenti valutativi precedentemente spiegati applicandoli alle due reti più performanti costruite passando in rassegna i loro punti di forza e di debolezza. Al fine di approfondire le performance dei modelli costruiti, si procede con l'analisi economico-finanziaria degli errori della migliore architettura ottenuta analizzando le differenze tra i valori medi degli indicatori usati come input della rete. Per concludere, si presenta una breve panoramica del contesto in cui potrebbe operare il modello definendo una procedura di utilizzo basata sulle performance ottenute.

Grazie a questo lavoro di tesi è stato possibile studiare il funzionamento delle reti neurali artificiali sia dal punto di vista teorico sia concreto attraverso la costruzione e l'analisi dei risultati di un modello capace di valutare il rischio di credito delle imprese appartenenti al settore del noleggio e leasing di autoveicoli. Le performance delle reti costruite sono illustrate nel dettaglio negli ultimi capitoli dell'elaborato e riprese nelle conclusioni finali nelle quali si riassumono le potenzialità ed i limiti di questi strumenti sempre più al centro delle attenzioni del mondo della finanza.

Capitolo I

La regolamentazione bancaria

Il Comitato di Basilea per la vigilanza bancaria è un organismo internazionale fondato nel 1974, composto dai rappresentanti delle banche centrali e dalle autorità di vigilanza del G10. Al Comitato possono partecipare non solo i dieci paesi più industrializzati al mondo ma anche quelli che adottano volontariamente le sue regole. Ad oggi, i paesi membri risultano essere: Germania, Francia, Italia, Spagna, Paesi Bassi, Belgio, Lussemburgo, Regno Unito, Svezia, Giappone, Stati Uniti, Canada e altri 14 Stati. Il Comitato fornendo standard, linee guida e raccomandazioni si propone di stabilire l'adeguatezza patrimoniale degli intermediari e di rafforzare la sicurezza e l'affidabilità del sistema finanziario. Le decisioni formulate non hanno però alcun valore giuridico ma riescono ugualmente a condizionare le legislazioni dei vari paesi che vi aderiscono su base volontaria.

In questo capitolo si illustreranno le quattro regolamentazioni emanate dal Comitato di Basilea presentandone i punti chiave e le motivazioni che hanno contribuito alla loro evoluzione nel tempo.

1.1 Basilea I: l'accordo sul capitale del 1988

L'accordo del 1988 era rivolto esclusivamente alle banche internazionali, la cui efficienza sulla gestione dei rischi e delle perdite non doveva essere compromessa in nessun modo, data la loro influenza sulla stabilità economica mondiale. Il principio guida di questa prima regolamentazione è incentrato sulla definizione di un criterio uniforme per stabilire l'adeguatezza patrimoniale dei sistemi bancari, in modo da ridurre l'aggressività di alcuni istituti di credito operanti in contesti normativi poco regolamentati. Nel 1989, infatti, ben nove delle prime dieci banche internazionali erano giapponesi e la loro supremazia era dovuta principalmente all'utilizzo di un'elevata leva finanziaria che le consentiva, dietro l'assunzione di forti rischi, di ottenere tassi di redditività maggiori.

La concezione alla base della definizione dei requisiti minimi di capitalizzazione è che ad ogni assunzione da parte della banca di un impiego corrisponde una presa in carico di un rischio, il quale, oltre a dover essere quantificato, deve essere coperto da un proporzionato ammontare di capitale proprio. In origine l'accordo faceva riferimento solamente al rischio di credito, che riguarda la possibilità che la controparte diventi inadempiente, mentre in seguito è stato inserito anche il rischio di mercato, legato alle possibili perdite inattese causate dalla variazione dei prezzi delle attività finanziarie. Inoltre, con Basilea I si divide il capitale di vigilanza in due componenti sulla base della loro qualità, la prima chiamata *Tier 1* rappresenta il patrimonio di base ed è composta dal capitale sociale e da riserve costituite da utili non distribuiti, mentre la seconda *Tier 2*, con un limite del 50% del patrimonio complessivo, costituisce il patrimonio supplementare formato dalle riserve occulte, dal debito subordinato, dai fondi rischi e dagli ibridi di capitale e di debito.

$$\frac{Capitale\ di\ vigilanza}{Risk\ Weighted\ Assets + 12.5 * Rischi\ di\ Mercato} \geq 8\% \qquad (1.1)$$

Il patrimonio di vigilanza doveva essere almeno pari al 8% delle attività ponderate per il rischio, per le quali erano previsti cinque coefficienti fissi di ponderazione in funzione del tipo di prenditore:

- 0%: cassa, crediti verso governi centrali e banche centrali dei paesi OCSE;
- 10%: crediti verso enti pubblici;
- 20%: crediti verso banche OCSE, enti bancari internazionali e banche non OCSE, con durata residua inferiore ad 1 anno;
- 50%: crediti ipotecari su immobili residenziali;
- 100%: crediti e partecipazioni in imprese private ed i restanti crediti.

1.1.1 Limiti e conseguenze di Basilea I

L'accordo del '88 ha dato il via ad un processo di regolamentazione della vigilanza bancaria volto al raggiungimento della stabilità del sistema finanziario globale. Tuttavia, essendo il precursore di questo cambiamento ha mostrato fin da subito i suoi limiti. Le cinque classi di controparte hanno dimostrato di essere troppo generiche e di comprendere attività con diversi livelli di rischiosità. Questa suddivisione grossolana ha spinto le banche a finanziare le controparti di credito più rischiose in modo da poter garantire agli azionisti rendimenti più

elevati. L'effetto di questa regolamentazione è stato opposto rispetto a quello sperato in quanto, a parità di capitale di vigilanza allocato, le banche costruivano portafogli impieghi più rischiosi rispetto ai precedenti, provocando una difficoltà nella ricerca di finanziamento da parte dei debitori più affidabili. Altri limiti della classificazione riguardano la mancanza di riferimenti alla durata e alle garanzie associate al credito. Il contributo di questi due elementi nel definire la rischiosità del credito è significante, difatti è possibile mitigarla riducendone la durata oppure associandone garanzie sicure e di valore. Inoltre, sono stati trascurati anche i possibili benefici da diversificazione ottenibili attraverso l'adozione di strumenti derivati oppure tramite un'apposita costruzione del portafoglio impieghi.

1.2 Basilea II: l'accordo dei tre pilastri

La versione definitiva del nuovo accordo sul capitale che risale a giugno 2004, si basa su tre pilastri tutti ugualmente importanti:

- *Minimum Capital Requirement*

  Il primo pilastro fissa il requisito minimo di capitale di vigilanza fornendo informazioni dettagliate sul rischio di credito e introducendo il tema del rischio operativo.
- *Supervisory Review*

  Il secondo pilastro definisce il ruolo delle autorità di vigilanza nazionali le quali, oltre a dover monitorare il rispetto dei requisiti minimi di capitale nel tempo, hanno la facoltà di aumentare i requisiti patrimoniali regolamentari.
- *Market Discipline*

  Il terzo pilastro impone dei criteri di *disclosure* ossia di divulgazione al pubblico delle informazioni riguardanti l'esposizione ai rischi e il patrimonio di vigilanza allocato. Grazie a questi obblighi di trasparenza il mercato può recepire le informazioni, elaborarle e valutare le banche fissando i prezzi.

1.2.1 Primo pilastro: Minimum Capital Requirement

Il primo pilastro stabilisce i requisiti patrimoniali minimi da rispettare aggiungendo ai precedenti rischi di credito e di mercato anche quello operativo. La banca è percepita come attività d'impresa e in quanto tale soggetta al rischio operativo, nel quale rientrano gli errori

delle risorse umane nel trattamento dei dati, nella gestione delle pratiche burocratiche legate al compimento delle operazioni finanziarie ma anche l'utilizzo errato dei sistemi informativi e nello specifico la programmazione, la cattiva gestione dei dati e i possibili guasti tecnici.

$$\frac{Capitale\ di\ vigilanza}{Rischio\ di\ Credito + Rischi\ di\ Mercato + Rischi\ Operativi} \geq 8\% \qquad (1.2)$$

L'innovazione più importate introdotta da Basilea II riguarda le ponderazioni del rischio di credito determinate attraverso l'assegnazione delle controparti a precise classi di rating. Secondo il nuovo accordo, per l'attribuzione dei debitori alle classi di rating, gli istituti di credito possono far riferimento alla valutazione effettuata dalle agenzie esterne oppure utilizzare la propria classificazione se sono provvisti di un SRI (Sistema di Rating Interno) che soddisfa i requisiti minimi.

1.2.1.1 Approccio standard

La prima metodologia per il calcolo dei requisiti patrimoniali consiste nella misurazione del rischio di credito tramite il metodo standard integrato da valutazioni esterne del merito creditizio. Nel definire l'approccio standard il Comitato ha attribuito un ruolo centrale alle ECAI (*Eligible External Credit Assessment Institution*) le quali assegnando i rating diventano un punto di riferimento nella determinazione del merito creditizio. Data la loro notevole importanza spetta alle autorità di vigilanza nazionali stabilire quali ECAI siano affidabili (ad esempio la Banca d'Italia ne ha riconosciute quattro: Fitch, Standard&Poor's, Moody's e Cerved Group), inoltre è necessario accertarsi che esse rispettino i seguenti criteri:

- Obiettività

  La metodologia del processo di assegnazione dei rating deve essere rigorosa, sistematica e sottoposta ad un continuo controllo sulla base dell'esperienza storica. Inoltre, le valutazioni devono essere soggette ad una revisione costante per tener conto dei cambiamenti delle condizioni finanziarie del momento. Per poter essere riconosciuta dalla autorità di vigilanza, ciascuna metodologia deve essere applicata almeno per dodici mesi e preferibilmente per tre anni.

- Indipendenza

  Una ECAI non deve essere influenzata da pressioni politiche, economiche o conflitti di interessi causati dalla composizione della struttura societaria dell'istituto di valutazione.

- Trasparenza

  Tutte le valutazioni devono essere disponibili a parità di condizioni a qualunque istituto nazionale o portatore di interesse.

- Pubblicità delle informazioni

  Le nozioni alle quali si fa riferimento sono: le metodologie di valutazione, la definizione di default, l'orizzonte temporale, il significato di ciascun rating, i tassi effettivi di inadempienza per ciascuna categoria di valutazioni e le matrici di migrazione, che descrivono la probabilità di un rating di cambiare classe.

- Risorse

  La disponibilità delle risorse deve essere tale da garantire un'alta qualità delle valutazioni.

- Credibilità

  È strettamente legata al rispetto dei punti precedenti e alle performance delle ECAI che sono monitorate nel tempo tramite il confronto tra i rating assegnati e la situazione reale della società valutata.

Rispetto al primo accordo del 1988 sono state effettuate notevoli modifiche alle ponderazioni delle varie esposizioni, le principali riguardano:

- Crediti verso governi

  Le ponderazioni di rischio sono attribuite sulla base del rating assegnato dalle agenzie ai governi e alle loro banche centrali.

Tabella 1.1 *Ponderazioni per crediti verso governi*

Valutazione	Da AAA ad AA-	Da A+ ad A-	Da BBB+ a BBB-	Da BB+ a B-	Inferiore a B-	Senza rating
Ponderazione	0%	20%	50%	100%	150%	100%

- Crediti verso banche

  Per questa tipologia di crediti sono proposte due alternative. La prima assegna la ponderazione immediatamente successiva rispetto ai crediti concessi al governo del paese di riferimento, fatta eccezione per i paesi non quotati o con rating da BB+ a B- per i quali è prevista una soglia massima del 100%. La seconda utilizza la

ponderazione assegnata dalle ECAI alle banche, attribuendo 50% a quelle sprovviste di valutazione e inoltre favorisce i crediti a breve termine con durata uguale o inferiore a tre mesi, per i quali assegna la ponderazione immediatamente precedente fino ad un minimo del 20%.

Tabella 1.2 *Ponderazioni per crediti verso banche: opzione 1*

Valutazione del governo	Da AAA ad AA-	Da A+ ad A-	Da BBB+ a BBB-	Da BB+ a B-	Inferiore a B-	Senza rating
Ponderazione	20%	50%	100%	100%	150%	100%

Tabella 1.3 *Ponderazioni per crediti verso banche: opzione 2*

Valutazione della banca	Da AAA ad AA-	Da A+ ad A-	Da BBB+ a BBB-	Da BB+ a B-	Inferiore a B-	Senza rating
Ponderazione	20%	50%	50%	100%	150%	50%
Ponderazione per i crediti a breve termine[17]	20%	20%	20%	50%	150%	20%

- Crediti verso le imprese

  Anche in questo caso si utilizzano le ponderazioni sulla base del rating stabilito dagli istituti di valutazione del merito creditizio. Per le aziende non quotate si adotterà una misura standard del 100% con il vincolo aggiuntivo di non poter ottenere una ponderazione migliore rispetto a quello dello Stato in cui operano.

Tabella 1.4 *Ponderazioni per crediti verso imprese*

Valutazione	Da AAA ad AA-	Da A+ ad A-	Da BBB+ a BB-	Inferiore a BB-	Senza rating
Ponderazione	20%	50%	100%	150%	100%

- Crediti garantiti da ipoteche su immobili residenziali

  La copertura di un credito con un'ipoteca su immobili residenziali consente di ridurre la ponderazione al 35%. Le autorità di vigilanza hanno il compito di verificare che questo valore di favore venga applicato solamente agli immobili residenziali.

- Categorie a più alto rischio

  In questa categoria, alla quale si applica una ponderazione maggiore o uguale al 150%, rientrano: i crediti verso i governi, ESP, banche, società di intermediazione mobiliare con rating inferiore a B-, le imprese con rating inferiore a BB-. Spetta sempre alle autorità di vigilanza nazionale decidere se aumentare la ponderazione specialmente nei casi di investimenti in *venture capital* e *private equity*.

1.2.1.2 Sistema basato sui rating interni

Il presente approccio per il calcolo dei requisiti patrimoniali minimi si basa sull'utilizzo dei sistemi di rating interni (SRI). Per sistema di rating si intende l'insieme dei metodi, dei processi organizzativi e di controllo che permette la collezione e l'elaborazione delle informazioni significative per la valutazione della rischiosità delle singole operazioni creditizie. Questa metodologia può essere utilizzata come alternativa alla precedente solamente qualora rispetti i seguenti criteri: valuta separatamente la PD e la LGD, la distribuzione dei crediti sulle varie classi avviene senza particolari concentrazioni, il rating è controllato periodicamente, il rating è utilizzato dalla banca nella gestione dei crediti e nel pricing dei prestiti, l'intermediario deve possedere un sistema di validazione dell'accuratezza e coerenza del SRI e deve rispettare i requisiti di documentazione formale sul suo funzionamento. La Banca d'Italia ha voluto specificare e integrare i precedenti requisiti nel documento "nuove disposizioni di vigilanza prudenziale per le banche" (circolare n.263 del 27 dicembre 2006), definendone altri sei:

- Documentazione del sistema di rating

  Gli intermediari hanno il dovere di documentare tutti i dettagli legati al funzionamento del proprio sistema di rating interno, spiegando le caratteristiche principali e i fondamenti teorici del modello. In particolare, devono essere documentate le definizioni di default e di perdita adottate e la loro coerenza con le indicazioni delle regolamentazioni. Inoltre, devono essere registrate e conservate tutte le informazioni che consentono di ricostruire il percorso di assegnazione del rating e le attività di controllo effettuate su di esso.
- Completezza delle informazioni

  Le banche devono adottare delle metodologie interne per la valutazione della completezza e della rilevanza dei dati e delle informazioni impiegate nel sistema di rating interno.
- Replicabilità

  Agli intermediari è richiesto non solo di definire il modello e i parametri impiegati ma anche di conservare le decisioni prese durante il percorso di assegnazione della valutazione, in modo tale da poter ripercorrere il processo di assegnazione del rating ed eventualmente ricalcolarlo per le singole posizioni.

- Integrità

  Nei casi in cui i soggetti responsabili dell'attribuzione del rating svolgano un'attività connessa al raggiungimento di obiettivi in termini di volumi o ricavi oppure abbiano il potere di decidere sull'erogazione del credito sorge un conflitto di interessi. Devono quindi essere previste cautele adeguate sul piano organizzativo e procedurale per evitare che l'assegnazione definitiva del rating possa essere condizionata da questi soggetti.
- Omogeneità

  I sistemi utilizzati dalle banche devono garantire che debitori ed attività che comportano rischi simili siano valutati allo stesso modo e quindi assegnate alla stessa classe di rating.
- Univocità

  Il carattere univoco dell'operazione di assegnazione ad una classe di rating deve riguardare sia i debitori sia le operazioni. Una volta riconosciuto che si tratta di una stessa controparte piuttosto che di una stessa attività, ogni esposizione deve essere valutata uniformemente.

1.2.2 Secondo pilastro: Supervisory Review

Il secondo pilastro avvalora l'importanza del processo di controllo prudenziale svolto dalle autorità di vigilanza nazionali, le quali non devono limitarsi a verificare che siano matematicamente rispettati i requisiti patrimoniali ma devono svolgere analisi più approfondite per valutare la coerenza dei rischi assunti dalle banche con il capitale allocato. Le attività di controllo favoriscono un dialogo tra le autorità di vigilanza e le banche incentivate a sviluppare processi interni di valutazione del capitale e della congruenza tra gli obiettivi patrimoniali e del proprio profilo di rischio. Per rimarcare l'importanza della supervisione bancaria, Basilea II presenta i quattro principi chiave che guidano il processo di controllo prudenziale.

- Le banche devono adottare delle metodologie di verifica dell'adeguatezza patrimoniale in riferimento al profilo di rischio adottato. A tal fine, dovrebbero essere effettuati *stress test* per individuare possibili scenari che potrebbero impattare negativamente sulla banca.

- Le autorità di vigilanza dovrebbero riesaminare regolarmente i processi interni di valutazione dell'adeguatezza patrimoniale, dei rischi assunti e la qualità del capitale posseduto.
- Le autorità di vigilanza devono poter richiedere alle banche di detenere un patrimonio superiore allo standard minimo obbligatorio. Il margine aggiuntivo è richiesto principalmente per tener conto delle incertezze che colpiscono i vari mercati e il settore bancario.
- Le autorità di vigilanza devono essere in grado di evitare che il patrimonio di una banca scenda sotto la soglia corrispondente al suo profilo di rischio prevenendolo tramite l'imposizione di azioni correttive quali intensificazione della vigilanza, restrizioni al pagamento dei dividendi o un piano di reintegro del patrimonio.

1.2.3 Terzo pilastro: Market Discipline

Il terzo pilastro stabilisce i requisiti informativi che devono essere resi noti affinché gli operatori del mercato finanziario possano valutare l'operatività, le esposizioni ai rischi e l'adeguatezza del patrimonio. La pubblicizzazione delle informazioni rende più sicure e solide le banche in quanto sono più propense ad operare in modo sicuro e prudente. Il mancato rispetto dei requisiti di trasparenza può portare le autorità di vigilanza a adottare provvedimenti quali il dialogo con le direzioni bancarie, le sanzioni pecuniarie o semplici richiami. Le informazioni fornite devono rispettare la frequenza semestrale e il criterio di rilevanza poiché la mancanza o errata indicazione di un'informazione rilevante è in grado di modificare il giudizio o le decisioni degli utenti che ne sono a conoscenza.

1.2.4 Limiti e conseguenze di Basilea II

L'accordo di Basilea II, nonostante sia entrato in vigore solamente nel 2008, è considerato come uno tra i colpevoli della severità della crisi finanziaria. Il Comitato stesso ha dichiarato i suoi limiti individuando sei punti chiave:

- Qualità e livello del capitale

  Gli intermediari che sono stati salvati grazie all'aiuto degli enti governativi possedevano un patrimonio superiore al minimo imposto dalla regolamentazione, ciò ha messo in dubbio che la qualità e la quantità del capitale necessarie per prevenire lo stato di insolvenza bancaria fosse in realtà di molto superiore a quanto fissato da Basilea II. In realtà quest'affermazione è vera solamente in parte poiché il patrimonio delle banche divenute insolventi aveva una quota considerevole di strumenti finanziari ibridi. Quest'ultimi sono stati introdotti a seguito di una consistente domanda da parte degli investitori desiderosi di ottenere rendimenti sicuri investendo in strumenti di debito per le banche.

- Prociclicità

  Con il termine prociclicità si fa riferimento alle dinamiche del sistema finanziario che tendono ad ampliare le fluttuazioni cicliche e quindi a migliorare le fasi espansive ed aggravare quelle recessive. La regolamentazione infatti, richiede in corrispondenza delle fasi recessive un aumento dei requisiti provocando una contrazione dell'offerta di credito ed accentuando la fase ciclica negativa. Il ragionamento appare sensato se si adotta una visione sulla singola banca ma se si guarda all'intero sistema finanziario, ad una contrazione del credito da parte di tutte le banche corrisponde un aumento della fase recessiva poiché accresce il rischio di default dei debitori causando una situazione di difficoltà per le banche.

- Incremento incontrollato della leva finanziaria

  La situazione finanziaria nel momento antecedente la crisi, vedeva le banche adottare livelli di leva finanziaria molto elevati pur sempre nel rispetto dei limiti regolamentari. L'arrivo della crisi e la conseguente contrazione della concessione del credito hanno spinto le banche ad accrescere il patrimonio riducendo i propri attivi. La velocità che ha caratterizzato questo processo di *deleveraging* ha contribuito all'instabilità dei mercati finanziari.

- Liquidità

  Nel periodo precedente la crisi il rischio di liquidità nelle banche è cresciuto a causa di una serie di fattori come ad esempio la globalizzazione dei gruppi finanziari, le cartolarizzazioni, la tecnologia e la concentrazione tra i grandi gruppi finanziari. L'aumento del rischio di controparte e la conseguente perdita di fiducia provocati dalla crisi ha causato considerevoli agitazioni ai singoli intermediari.

- Banche sistemiche

  Durante la crisi, i vari enti governativi sono dovuti intervenire per salvare diverse banche e assicurazioni, la cui interconnessione con le altre istituzioni finanziare era tale per cui un loro fallimento avrebbe potuto causare una crisi sistemica. Il problema di queste istituzioni era l'ampiezza dei loro legami con le altre entità finanziarie che ha facilitato il trasferimento degli shock economici.

- Rischi di mercato sul *trading book*

  Durante la crisi, si sono registrate importanti svalutazioni delle attività che componevano i *trading book* delle banche, uno dei motivi del crollo del valore dei portafogli di negoziazione è il fatto che tali attività erano ricondotte al *fair value*. La caduta dei mercati ha fatto emergere la fragilità dei nuovi requisiti minimi fissati dalla regolamentazione, i quali non sono stati in grado di ricoprire le perdite. Una delle cause della loro leggerezza è da ricercare nei modelli utilizzati per il calcolo dei rischi, questi sono risultati talmente reattivi da rimuovere troppo velocemente gli episodi di crisi e considerare solo quelli più recenti riferiti agli scenari ottimistici precrisi.

1.3 Basilea III: l'attuale regolamentazione

La crisi finanziaria del 2007 ha fatto emergere tutti i limiti di una regolamentazione che non è stata in grado di garantire la sopravvivenza degli intermediari rendendo necessario il salvataggio da parte degli enti governativi. Al termine di questo periodo buio per l'economia mondiale, il Comitato si è riunito ancora una volta per emanare i provvedimenti di Basilea III.

Il nuovo accordo mantiene la struttura a tre pilastri del precedente rinforzandone le basi, interviene singolarmente su ognuno di essi, fissa le direttive per le istituzioni finanziarie di rilevanza sistemica e i requisiti globali di liquidità. I cambiamenti al precedente accordo di Basilea II riguardano principalmente i requisiti patrimoniali minimi enunciati nel primo pilastro, che diventano più rigidi sia in termini di qualità sia di quantità per assicurare un maggiore assorbimento delle perdite. Le modifiche riguardano:

- Capitale

  Con Basilea III si aumentano la qualità e il livello del patrimonio di vigilanza ridefinendo la composizione del *Tier 1*, *Tier 2* ed eliminando il *Tier 3*. In particolare, il patrimonio di base deve essere composto principalmente da azioni ordinarie e riserve di utili non distribuiti (*common equity*) mentre dovranno essere esclusi gli strumenti ibridi innovativi. Inoltre, il livello minimo del *Tier 1*, utilizzato per coprire le perdite e garantire la continuità delle attività della banca (*going concern capital*) passa dal 4% al 6% delle attività ponderate per il rischio e il requisito minimo del *core Tier 1* è innalzato al 4,5% delle attività pesate per il rischio al netto degli aggiustamenti. Il *Tier 2*, utilizzato per sopperire alle perdite causate da una preventiva liquidazione della banca (*gone concern capital*) copre il 2%, che sommato al *Tier 1* consente di raggiungere l'8% regolamentare. Per di più, il nuovo accordo prevede che le banche debbano possedere un *buffer* di conservazione del capitale (*capital conservation buffer*) del 2,5% formato da *common equity*, portando la quota totale di quest'ultimo al 7%. La costruzione di tale *buffer* è garantita da limitazioni nella distribuzione degli utili da parte delle banche che non lo hanno ancora adottato. Il suo ruolo è quello di incoraggiare l'accumulo di riserve patrimoniali nei periodi economici particolarmente favorevoli e di fornire una possibile copertura aggiuntiva nei periodi di crisi. Altro cuscinetto introdotto dal nuovo accordo è il *buffer* anticiclico (*counter-cyclical buffer*) anch'esso composto da *common equity* in misura compresa tra lo 0 e il 2,5% per proteggere gli intermediari dalle possibili perdite causate da una politica eccessiva di espansione del credito. L'applicazione del *buffer* è decisa dalle autorità nazionali che valutano il surriscaldamento del ciclo creditizio sulla base della differenza tra prestiti bancari e il trend di lungo periodo del PIL.

- Copertura dei rischi

  La nuova regolamentazione stabilisce che le banche adottino un trattamento patrimoniale più cautelativo e che siano effettuate analisi più rigorose del merito creditizio nei confronti di tutti quegli strumenti finanziari innovativi responsabili della crisi del 2007. Tra questi strumenti troviamo le cartolarizzazioni provviste di rating esterno e i derivati per i quali il portafoglio di negoziazione deve essere valutato tenendo conto di un rischio incrementale derivato dai rischi di insolvenza, di migrazione di rating e di liquidità.

- Contenimento della leva finanziaria

  Per controllare la crescita dei livelli di leva finanziaria, la regolamentazione ha introdotto *leverage ratio*. Tale indice è calcolato come il rapporto tra il patrimonio di base *Tier 1* e il totale dell'attivo comprendente le esposizioni fuori bilancio. Per evitare pericolosi *deleveraging* non potrà scendere al di sotto del 3%.

$$Leverage\ ratio = \frac{Tier\ 1}{Attivo\ comprendente\ fuori\ bilancio} \geq 3\% \qquad (1.3)$$

- Liquidità

  La crisi del 2007 ha fatto emergere importanti problematiche relative alla liquidità dei mercati finanziari, per queste ragioni si è deciso di introdurre nel nuovo accordo due nuovi indicatori di liquidità minima. Il primo coefficiente chiamato *Liquidity coverage ratio* ha lo scopo di garantire la presenza di attività liquide di alta qualità (ALAQ) in grado di generare fondi per superare una fase di importanti deflussi negativi nel breve periodo. L'indice, calcolato come il rapporto tra le ALAQ e i deflussi di cassa attesi in caso di stress relativi ad un orizzonte temporale di 30 giorni, deve assumere un valore superiore ad 1.

$$Liquidity\ coverage\ ratio = \frac{ALAQ}{Deflussi\ di\ cassa_{30\ giorni}} > 1 \qquad (1.4)$$

  Le attività che rientrano nel calcolo del numeratore sono attività con basso rischio di credito, di mercato, aventi valutazione certa e con una bassa correlazione con attività rischiose. Il denominatore invece fa riferimento alla differenza tra i flussi di cassa cumulati in uscita e in entrata in un periodo di stress di 30 giorni.

  Il secondo requisito fissato dal *Net stable funding ratio*, ha il compito di assicurare l'equilibrio tra le fonti di finanziamento e il fabbisogno a medio-lungo termine determinato dalla scadenza. La stabilità è garantita da una maggiore quantità di risorse finanziarie stabili (*avaiable stable funding* o ASF) rispetto al fabbisogno di risorse stabili richiesto dalla formazione dell'attivo (*required stable funding* o RSF).

$$Net\ stable\ funding\ ratio\ = \frac{ASF}{RSF} > 1 \qquad (1.5)$$

Il numeratore è formato da: *Tier 1*, *Tier2*, azioni privilegiate e passività con scadenza maggiore o uguale ad un anno e passività con scadenza inferiore ad un anno per le quali la banca si aspetta un rinnovo oltre i dodici mesi. Tutte le componenti sono ponderate con un coefficiente che rappresenta la stabilità della fonte (da 0% per le passività con scadenza inferiore all'anno fino al 100% per le più stabili), in generale le ASF rappresentano tutte le fonti di finanziamento delle quali si potrà far uso in condizioni di stress. Il denominatore è una stima del fabbisogno di risorse stabili connesso alla struttura dell'attivo e anche in questo caso, agli elementi che lo compongono è applicata una ponderazione che riflette il livello di liquidità dell'attività in esame dando maggior peso alle attività più difficilmente liquidabili.

- Istituzioni finanziarie di importanza sistemica

  Gli intermediari che rientrano in questa tipologia devono possedere una maggiore capacità di assorbimento delle perdite sulla base della gravità di un possibile loro default. Data la loro importanza, il Comitato ha fissato i criteri quantitativi e qualitativi per poterle classificare ed assegnare, sulla base dell'importanza delle loro interconnessioni, un coefficiente patrimoniale progressivo dall'1 al 2,5% da soddisfare con *Tier 1*.

- Gradualità

  Un aspetto molto importante di Basilea III riguarda le tempistiche con le quali il Comitato ha stabilito l'entrata in vigore dei nuovi limiti per la stabilità del sistema economico. Questa gradualità è dovuta a diversi fattori, primo fra tutti è la diversità con la quale gli intermediari dei vari paesi hanno ricevuto gli aiuti da parte dello Stato. In alcuni paesi, le autorità sono intervenute con iniezioni massicce di liquidità mentre in altri, a causa delle condizioni di difficolta nelle quali riversavano le finanze pubbliche, il governo non è stato in grado di contribuire al loro salvataggio. In generale, l'applicazione progressiva di Basilea III è stata adottata in modo da evitare un possibile impatto negativo dei nuovi requisiti patrimoniali sulla crescita economica.

Tabella 1.5 *Tempistiche di entrata in vigore dei requisiti di Basilea III*

	2011	2012	2013	2014	2015	2016	2017	2018	Dal 1° gennaio 2019
Indice di leva (*leverage ratio*)	Monitoraggio regolamentare		Fase di sperimentazione 1° gennaio 2013 - 1° gennaio 2017 Informativa dal 1° gennaio 2015					Migrazione al primo pilastro	
Requisito minimo per il *common equity*			3,5%	4,0%	4,5%	4,5%	4,5%	4,5%	4,5%
Buffer di conservazione del capitale						0,625%	1,25%	1,875%	2,50%
Requisito minimo per il *common equity* più *buffer* di conservazione del capitale			3,5%	4,0%	4,5%	5,125%	5,75%	6,375%	7,0%
Introduzione delle deduzioni dal CET1 (compresi gli importi eccedenti il limite per DTA, MSR e investimenti in istituzioni finanziarie)				20%	40%	60%	80%	100%	100%
Requisito minimo per il patrimonio di base (Tier 1)			4,5%	5,5%	6,0%	6,0%	6,0%	6,0%	6,0%
Requisito minimo per il capitale totale			8,0%	8,0%	8,0%	8,0%	8,0%	8,0%	8,0%
Requisito minimo per il capitale totale più *buffer* di conservazione del capitale			8,0%	8,0%	8,0%	8,625%	9,25%	9,875%	10,5%
Strumenti di capitale non più computabili nel non-Core Tier 1 e nel Tier 2			Esclusione su un arco di 10 anni con inizio dal 2013						
Indicatore di breve termine (Liquidity Coverage Ratio)	Inizio periodo di osservazione				Introduzione requisito minimo				
Indicatore strutturale (Net Stable Funding Ratio)		Inizio periodo di osservazione						Introduzione requisito minimo	

1.3.1 Considerazioni su Basilea III

L'accordo di Basilea III, nato come risposta internazionale alla crisi finanziaria del 2007, ha inciso in maniera significativa sulla quantità aggregata di credito disponibile. Gli incrementi dei requisiti minimi di capitale e l'introduzione dei vari *buffer* hanno agito in tal senso aumentando la capacità di assorbimento di capitale e diminuendo il ricorso a forme di debito meno onerose.

Un secondo aspetto positivo è la maggiore qualità del capitale *Tier 1* che aumenta dal 2% al 7% evidenziando l'incapacità nell'assorbimento delle perdite da parte degli strumenti ibridi e innovativi che non rientrano nel patrimonio di base. Tali strumenti finanziari, durante la crisi, erano erroneamente considerati alla stregua di forme di debito e quindi si riteneva che le banche non avrebbero rinunciato a pagarne gli interessi per evitare ricadute sulla reputazione e sulle capacità di raccolta.

Altro punto di forza della regolamentazione sono le considerazioni sulle fasi del ciclo economico. Durante la crisi l'inasprimento dei requisiti di capitale ha contribuito all'instabilità del sistema finanziario quindi si è deciso di sfruttare i periodi positivi accumulando risorse destinate a coprire le perdite in quelli più bui.

1.4 Basilea IV: il futuro degli accordi sul capitale

Nel 2017 il Comitato si è riunito una quarta volta per dare vita alle riforme regolamentari che entreranno in vigore progressivamente tra il 2021 e il 2027. Lo scopo principale è quello di aumentare l'attendibilità ai calcoli delle attività ponderate per il rischio (*risk weighted assets* o RWA) e rendere più facilmente confrontabili gli indicatori patrimoniali delle banche. L'esigenza di introdurre queste nuove regole deriva dal risultato di numerose analisi che hanno evidenziato una variabilità significativa tra gli RWA di diverse banche aventi una simile rischiosità di portafoglio. Si è quindi notato che gli intermediari utilizzano i loro modelli interni, non per quantificare i rischi ma per minimizzare i requisiti patrimoniali. Per queste ragioni sono stati introdotti dei limiti all'uso dei modelli interni, nello specifico:

- È stata accresciuta la sensitività ai rischi dell'Approccio Standard tramite l'adozione di una modifica dei fattori di ponderazione degli attivi più granulare e articolata.
- Le banche devono informarsi maggiormente tramite specifiche ricerche sull'attendibilità delle valutazioni espresse dai rating esterni per sviluppare un approccio *non-rating-based*. Per gli intermediari la cui attività è in giurisdizioni che non consentono di utilizzare le valutazioni dei rating esterni possono applicare una ponderazione del 65% alle imprese *investment grade* e una del 100% alle restanti.
- Per evitare l'accumulo di rischi eccessivi nell'uso dei sistemi di rating interno si sono fissati livelli minimi (*input floor values*) di PD, LGD ed EAD riportati in Tabella 1.6.

Tabella 1.6 *Livelli minimi di PD, LGD e EAD introdotti con Basilea IV*

Minimum parameter values in the revised IRB framework[4] — Table 3

	Probability of default (PD)	Loss-given-default (LGD)		Exposure at default (EAD)
		Unsecured	Secured	
Corporate	5 bp	25%	Varying by collateral type: • 0% financial • 10% receivables • 10% commercial or residential real estate • 15% other physical	EAD subject to a floor that is the sum of (i) the on-balance sheet exposures; and (ii) 50% of the off-balance sheet exposure using the applicable Credit Conversion Factor (CCF) in the standardised approach
Retail classes:				
Mortgages	5 bp	N/A	5%	
QRRE transactors	5 bp	50%	N/A	
QRRE revolvers	10 bp	50%	N/A	
Other retail	5 bp	30%	Varying by collateral type: • 0% financial • 10% receivables • 10% commercial or residential real estate • 15% other physical	

- In aggiunta saranno introdotti degli *output floor* per stabilire che il valore complessivo degli RWA sia uguale al maggiore tra gli RWA ottenuti con gli approcci consentiti dalla regolamentazione e il 72,5% del totale degli RWA calcolati con il metodo standard. Gli *output floor* dunque fissano limite del 27,5% al risparmio in termini di minor capitale allocato da parte degli intermediari che fanno uso di un sistema di rating interno.
- Le SIFI devono adottare un ulteriore *leverage ratio buffer* per evitare un eccessivo ricorso alle fonti di debito.

I possibili impatti di Basilea IV, valutati attraverso simulazioni effettuate dall'EBA (*European Banking Authority*), hanno mostrato che molto probabilmente le banche dovranno aumentare il proprio capitale. Questa previsione vale principalmente per gli istituti di importanza sistemica e per le banche che utilizzano un sistema di rating interno poiché sono proprio loro l'oggetto della regolamentazione. A risentire dell'ultimo accordo saranno anche gli enti che si occupano dell'erogazione dei mutui e dei leasing immobiliari in quanto dovranno applicare le nuove ponderazioni per il calcolo degli RWA.

In conclusione, nel sistema economico odierno in cui l'intermediazione del credito avviene attraverso gli istituti finanziari tradizionali, le nuove riforme, rendendo meno elastica l'operatività delle banche, incentiveranno l'utilizzo di mezzi propri o di strumenti alternativi di accesso al credito.

Capitolo II

Il rischio di credito

Il capitolo precedente ha mostrato l'attenzione rivolta dal Comitato di Basilea per la vigilanza bancaria nei confronti delle diverse tipologie di rischio che minano la stabilità degli intermediari finanziari. L'oggetto di questo capitolo è quello che ancora oggi rappresenta il principale fattore delle crisi bancarie ossia il rischio di credito.

Il rischio di credito è definito come la possibilità che una variazione inattesa del merito di credito di una controparte, nei confronti della quale esiste un'esposizione, generi una corrispondente variazione inattesa del valore di mercato del credito. Questa definizione sottolinea diversi aspetti, il più importante dei quali è la necessità che la variazione del merito creditizio sia inattesa, se così non fosse, sarebbe già considerata al momento della valutazione ed inclusa nel *pricing* del credito. Inoltre, il fatto che la variazione sia inattesa implica la presenza di errori di valutazione probabilmente causati da errori del modello adottato oppure da cambiamenti delle variabili economiche in gioco. Altra caratteristica evidenziata dalla definizione è che lo stato di insolvenza rappresenta solamente l'evento estremo del rischio di credito, il quale deve essere considerato in modo più ampio facendo riferimento ad una completa distribuzione di probabilità di eventi intermedi che individuano le diverse classi di rischio. Come descritto nel capitolo precedente, il rischio di credito è valutato tramite l'attribuzione di un rating, ossia un giudizio sulla capacità di un'emittente o di un'emissione di far fronte agli impegni finanziari entro certe scadenze. Tale giudizio è espresso da agenzie di rating sulla base di informazioni e valutazioni diffuse sul mercato oppure attraverso l'utilizzo di sistemi di rating interni.

Il rischio di credito è una realtà molto complessa da analizzare poiché ad esso possono essere associate diverse tipologie di rischio:

- Rischio di insolvenza
  Definito come la possibilità che il debitore non riesca ad assolvere, anche solo in parte, agli obblighi derivati dalla sua esposizione. Questo rischio è valutato attraverso l'emissione di un rating ossia un punteggio emesso da agenzie specializzate oppure dalle stesse banche.
- Rischio di migrazione
  Consiste nel rischio di un deterioramento del merito creditizio della controparte.
- Rischio di recupero
  Rappresenta l'incertezza relativa all'ammontare che sarà effettivamente recuperato in seguito alle procedure di contenzioso svolte dalla banca.
- Rischio di esposizione
  Riferito all'importo del debito che risulta al momento in cui si manifesta l'insolvenza.
- Rischio di spread
  Definito come la probabilità che a parità di merito creditizio, aumenti lo spread e quindi il premio per il rischio richiesto dal mercato di capitali.
- Rischio paese
  Consiste nel rischio relativo alla provenienza di un'esposizione in paesi caratterizzati da diverse forme di instabilità (es. politiche, legislative, economiche, sociali…).

2.1 Le componenti del rischio di credito

La variazione inattesa del valore di mercato di credito menzionata nella definizione di rischio di credito è strettamente connessa alla nozione di perdita. Tale concetto è alla base della distinzione delle due componenti del rischio di credito: la perdita attesa (*Expected Loss* o EL) e la perdita inattesa (*Unexpected Loss* o UL). Come già detto in precedenza, a costituire la vera e propria fonte di rischio è solamente la perdita inattesa in quanto non prevedibile dal modello impiegato e dunque rischiosa, mentre la prima componente è già inclusa nelle attività di *pricing* della banca in quanto nota a priori.

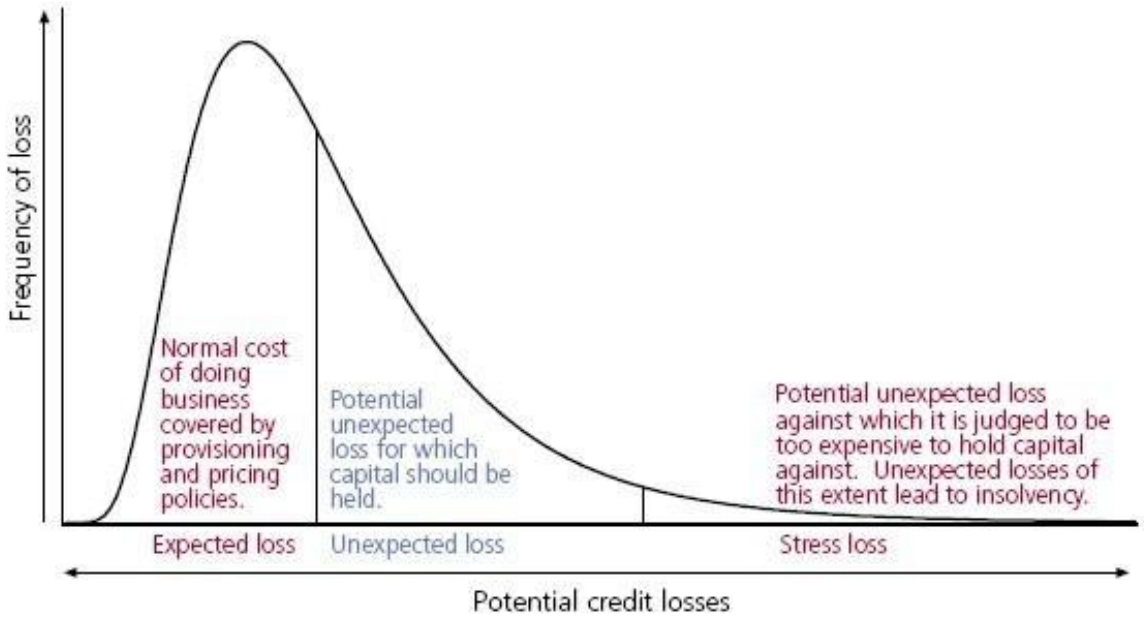

Figura 2.1 *Componenti del rischio di credito in funzione della quantità e della frequenza*

2.1.1 La perdita attesa

La perdita attesa è definita come il valore medio della distribuzione delle perdite che una banca si attende di conseguire su un singolo credito o su un portafoglio crediti. Essendo attesa rappresenta una componente stabile di costo che la banca si aspetta *ex-ante* di dover sostenere in quanto detentrice di una data esposizione creditizia.

La perdita attesa è calcolata dal prodotto di tre elementi:

- l'esposizione in caso di insolvenza (*Exposure at Default* o EAD);
- la probabilità di insolvenza (*Probability of Default* o PD);
- la percentuale di perdita in caso di insolvenza (*Loss Given Default* o LGD).

$$EL = EAD \cdot PD \cdot LGD \tag{2.1}$$

2.1.1.1 Exposure At Default

L'EAD è l'ammontare di perdita che grava sulla banca al momento del default. Dal momento che l'insolvenza si verifica in un istante futuro non sempre è possibile conoscere a priori questo valore, sulla base di questo ragionamento distinguiamo due tipologie di esposizioni:

- a valore certo: per le quali la banca conosce *ex-ante* il valore esatto del finanziamento. Es. mutuo;
- a valore incerto: per le quali il valore dell'esposizione è conosciuto con esattezza solamente *ex-post* ossia dopo il default. Es. fido di conto corrente.

Nel caso delle esposizioni a valore certo, il valore l'EAD è determinabile risalendo alla quota di debito non ancora restituita dal cliente mentre nel secondo caso la presenza di una componente aleatoria non rende il calcolo immediato. La stima dell'EAD per le esposizioni a valore incerto deve necessariamente passare dalla determinazione di due quote diverse ovvero quella di fido utilizzata (*Drawn Portion* o DP) e quella non ancora utilizzata (*Undrawn Portion* o UP). A quest'ultima è associata una terza variabile che rappresenta la percentuale di quota di fido non ancora utilizzata che si ritiene verrà utilizzata dal debitore (*Usage Given Default* o UGD).

$$EAD = DP + UP \cdot UGD \tag{2.2}$$

2.1.1.2 Probability of Default

La probabilità di default è definita come la possibilità che la controparte diventi inadempiente agli obblighi che derivano dalla sua esposizione, in tal senso la PD è strettamente connessa al merito creditizio del debitore.

A seconda dei fattori e delle caratteristiche delle quali si vuole tenere conto per il calcolo della PD, si possono distinguere tre differenti metodologie: la prima prevede che le informazioni rilevanti siano estrapolate dai dati di mercato dei capitali, la seconda utilizza modelli analitico/soggettivi in grado di valutare entrambi gli aspetti quantitativi e qualitativi ed infine la terza si avvale dei rating creditizi emanati da agenzie specializzate (Standard&Poor's, Fitch Ratings e Moody's) oppure dai sistemi di rating interni.

2.1.1.3 Loss Given Default

La LGD è definita come la perdita subita dall'impresa finanziatrice nel momento in cui la controparte diventa insolvente. Una definizione equivalente fa riferimento alla LGD come il complemento del *Recovery Rate* (RR) ossia il tasso che l'istituto di credito riesce a recuperare in seguito al default del debitore.

$$LGD = 1 - RR \tag{2.3}$$

Il tasso di recupero è influenzato da diversi fattori:

- Caratteristiche del finanziamento
  La durata, il grado di subordinazione rispetto ad altri creditori, la tipologia di contenzioso prevista per il recupero, l'eventuale presenza e il grado di liquidità e l'efficacia delle attività finanziarie o reali a garanzia del credito.
- Caratteristiche dell'impresa finanziata
  In questa categoria rientrano le informazioni relative all'area geografica e al settore produttivo di provenienza del debitore. La prima definisce la tipologia, l'efficacia e la rapidità della procedura di recupero utilizzata mentre il settore produttivo influisce sul grado di liquidità dell'impresa, ossia sull'efficienza e la velocità della trasformazione delle attività aziendali in liquidità.
- Caratteristiche dell'impresa finanziatrice
  L'efficienza dei servizi legali interni e le politiche di recupero crediti che influenzano la rapidità e l'importo recuperato.
- Fattori esterni
  La fase corrente del ciclo economico e il livello dei tassi di interesse che gravano sul valore attuale dell'importo recuperato a seguito dell'insolvenza del debitore.

Per poter stimare correttamente il valore della LGD è necessario utilizzare dei modelli in grado di considerare tutte le caratteristiche sopra elencate, a tal proposito i due approcci più utilizzati sono il *Market LGD* e il *Workout LGD*. Il primo è un modello *market based* che basa la stima della LGD sul delta di prezzo di mercato dello strumento rappresentativo dell'esposizione prima e dopo la notizia di default del suo emittente. Questa teoria si basa sul fatto che il mercato una volta recepite le nuove informazioni, le rielabora e assegna un nuovo prezzo comprensivo delle aspettative sull'entità e sui tempi di recupero dei creditori e sulla possibilità da parte dell'intermediario di poter vendere immediatamente lo strumento

finanziario relativo all'esposizione. L'approccio *Workout LGD* calcola il valore della LGD misurando i reali flussi di cassa recuperati in seguito all'evento di insolvenza considerando tutti gli elementi che determinano il *Recovery Rate*.

$$RR = \frac{\sum_{t=1}^{T} \frac{ER_t - AC_t}{(1+i)^t}}{EAD} \tag{2.4}$$

Dove:

- ER_t = *Expected Recovery*, il valore recuperato nel periodo t;
- AC_t = *Administative Cost*, l'ammontare dei costi amministrativi nel periodo t;
- EAD = *Exposure at Default;*
- i = tasso di attualizzazione dei flussi di cassa;
- T = durata del processo di recupero.

2.1.2 La perdita inattesa

La perdita inattesa è una misura del grado di dispersione del tasso di perdita attorno al proprio valore atteso, ossia attorno all'*Expected Loss*. Questa componente rappresenta la vera fonte del rischio di credito in quanto la rischiosità è legata al verificarsi di un deterioramento inatteso della qualità creditizia.

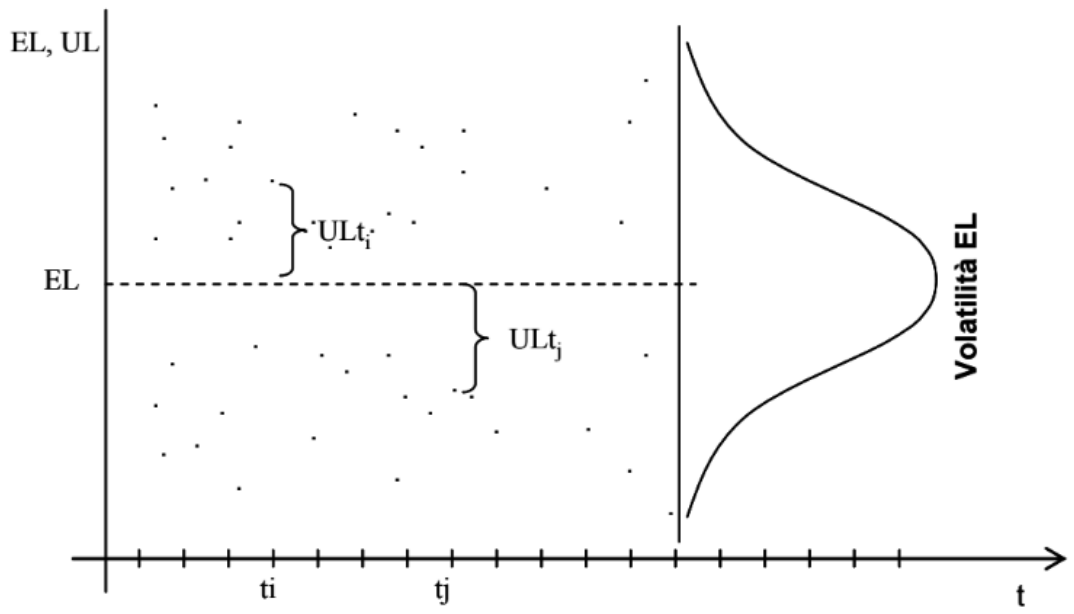

Figura 2.2 *Expected ed Unexpected Loss in funzione del tempo*

La perdita inattesa data la sua natura variabile non può essere contabilizzata in conto economico, operazione che invece avviene con la perdita attesa configurata come semplice voce di costo già inclusa nelle aspettative dell'intermediario, ma deve trovare un'adeguata copertura nel patrimonio. Altra importante differenza tra le due componenti del rischio di credito è la possibilità di attenuazione, in particolare mentre la perdita inattesa può essere ridotta tramite un'opportuna diversificazione del portafoglio, la perdita attesa non può essere mitigata ma solo stabilizzata tramite un aumento del volume delle operazioni prese in considerazione per il suo calcolo. Perciò, la perdita attesa complessiva di un portafoglio è data dalla somma delle perdite attese sulle singole esposizioni mentre la quantità da detenere in seguito all'aggregazione di più perdite inattese su più posizioni può risultare inferiore alla somma delle perdite inattese sulle singole posizioni grazie ai benefici di diversificazione.

Per valutare la perdita inattesa è sufficiente utilizzare un modello binomiale che prevede due soli eventi: default e non default.

Tabella 2.1 *Eventi con relative probabilità e perdite*

	EVENTI	
	Default	Non default
Probabilità	PD	1-PD
Perdita	LGD	0

In questo modo, le due perdite possono essere così calcolate:

$$EL = PD \cdot LGD + (1 - PD) \cdot 0 \tag{2.5}$$

$$UL = LGD \cdot \sqrt{PD(1 - PD)} \tag{2.6}$$

La (2.6) è valida solamente sotto le ipotesi di tasso di perdita LGD deterministico e di indipendenza tra la PD e LGD. Per un calcolo più accurato, si dovrebbe considerare la LGD stocastica e introdurre la covarianza tra la PD e la LGD per poter esprimere il grado di correlazione tra di esse.

Nel caso di LGD stocastica si ha:

$$UL = \sqrt{PD(1 - PD) \cdot LGD^2 + PD \cdot \sigma_{LGD}^2} \tag{2.7}$$

Rilassando entrambe le ipotesi la formula diventa:

$$\begin{aligned} UL^2 = cov(PD^2, LGD^2) + [PD(1 - PD) + PD^2] \cdot [\sigma_{LGD}^2 + LGD^2] \\ - [cov(PD, LGD) + PD \cdot LGD]^2 \end{aligned} \tag{2.8}$$

La (2.6) è la maggiormente utilizzata nonostante vi siano evidenze empiriche di una correlazione negativa tra i RR e le PD. Le basi su cui si fonda tale affermazione sono da ricercare nell'andamento del ciclo economico e dei tassi di interesse. Le fasi di depressione del ciclo infatti danno origine a momenti di difficoltà per le imprese con conseguenti aumenti delle PD e diminuzioni dei RR. Le complessità nel recupero crediti sono dovute al fatto che quest'ultimi sono per lo più costituiti da crediti verso altre imprese anch'esse in difficoltà. In questi momenti un aumento del livello dei tassi di interesse provoca una diminuzione del valore delle attività finanziarie e delle garanzie ad esse associate.

2.2 Credit scoring

Il *credit scoring* è una tecnica di valutazione statistica impiegata dagli intermediari per valutare la solvibilità dei clienti. Tale valutazione è calcolata attraverso l'uso di modelli previsionali che elaborano le informazioni disponibili dei clienti e le riducono ad un unico valore, noto come *credit score* in grado di sintetizzare il rischio creditizio. I principali vantaggi del *credit scoring* derivano dall'omogeneità di trattamento dei singoli clienti e dalla capacità di assistere in modo rapido ed efficiente al processo decisionale di concessione del prestito.

La nascita dei primi modelli di *credit scoring* è avvenuta negli anni Sessanta ad opera di Edward I. Altman, un professore di Finanza della Stem Shool of Business presso la New York University, che diede vita al modello *Z-Score*. Altman analizzò i dati di bilancio di 33 aziende in stato di default e 33 aziende solide e sviluppò un modello in grado di prevedere, con un grado di accuratezza del 95%, la probabilità di fallimento di un'impresa negli anni successivi. L'aspetto più interessante dello *Z-Score* è che, nonostante i limiti, possedeva tutti gli elementi che ancora oggi caratterizzano un modello di *credit scoring*: la selezione delle

variabili, la scelta del modello, la selezione del campione sul quale costruire il modello, la fase di verifica con dati diversi da quelli usati per la costruzione e l'impiego di test per la validazione dei risultati.

Nel corso degli anni, l'evoluzione delle tecniche statistiche e delle capacità computazionali a disposizione ha consentito lo sviluppo di nuovi modelli di diversa natura, partendo dai *logit* e *probit* di Martin ed Ohlson fino ad arrivare alle più recenti tecniche di *machine learning*. Nonostante le numerose differenze, tutti i modelli perseguono la minimizzazione della funzione di errore, calcolata sui valori delle probabilità di default assegnate ad ogni società del campione in esame.

2.2.1 Modelli statistici: la regressione logistica

Procedendo con ordine, i primi modelli previsionali sviluppati sono stati di tipo statistico. Il caso più semplice è rappresentato dalla regressione lineare a variabile singola che consente di studiare la relazione tra due variabili attraverso l'equazione di una retta generica:

$$Y = \alpha + \beta X + \varepsilon \tag{2.9}$$

La variabile Y, detta variabile dipendente o di risposta, deve essere prevista attraverso l'analisi dei valori della variabile X, chiamata variabile indipendente o predittore. Gli altri tre parametri della (2.9) sono: l'intercetta α che rappresenta il valore di Y quando la variabile indipendente X assume valore nullo, il coefficiente di regressione β che descrive la pendenza della retta, ossia il cambiamento della variabile dipendente Y corrispondente ad una variazione unitaria del predittore X, e la componente di errore ε di natura casuale in quanto non può essere spiegata dalla variabile indipendente X.

La logica della regressione lineare semplice può essere complicata a piacere, ad esempio è possibile includere un maggior numero di predittori in modo da spiegare meglio il comportamento della variabile di risposta Y. Questi modelli statistici rientrano nella famiglia degli algoritmi di regressione lineare multipla che, date k variabili indipendenti, possono essere generalizzati con la seguente formula:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \tag{2.10}$$

La (2.10) presenta la stessa struttura della (2.9) l'unica differenza riguarda la presenza di molteplici coefficienti di regressione $\beta_1, \beta_2, \dots, \beta_k$ definiti parziali in quanto esprimono l'entità della variazione della variabile Y rispetto al cambiamento del singolo predittore $X_1, X_2, \dots, X_k$ ad essi associato. Questi modelli risultano particolarmente versatili in quanto, apportando opportune trasformazioni, è possibile crearne delle versioni generalizzate che linearizzano le relazioni tra le variabili esplicative anche laddove non lo fossero in origine, adattando in questo modo il modello al caso di studio.

Un particolare modello lineare generalizzato è la regressione logistica. Essa è utilizzata quando la variabile Y, della quale si vuole spiegare il comportamento, è dicotomica, definita in 0 e 1, e avente distribuzione binomiale. Le variabili di tipo dicotomico sono utilizzate principalmente per problemi di classificazione, ovvero la determinazione della probabilità di appartenenza ad un gruppo piuttosto che ad un altro. Il modello di regressione logistica non è di tipo lineare in quanto la variabile Y non può ammettere valori tra $-\infty$ e $+\infty$ ma deve avere un comportamento asintotico in corrispondenza dei valori 0 e 1. La Figura 2.3 ne mostra l'andamento al variare della X:

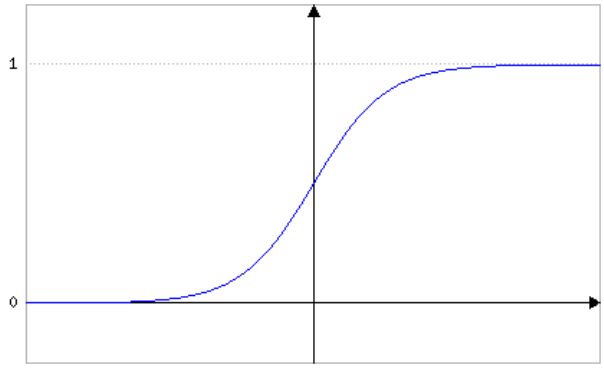

Figura 2.3 *Funzione logistica*

Nelle regressioni lineari la variabile di risposta è calcolata come il valore medio della variabile Y condizionata all'assunzione di un dato valore di X. In termini matematici si ha[1]:

$$E[Y|X] = \mu_Y \tag{2.11}$$

[1]Formule tratte da: http://psiclab.altervista.org/MetTecPsicClinica2017/2.1.RegressioneMultipla_Logistica2016.pdf.

Diversamente, nella regressione logistica il valor medio è uguale alla probabilità che la variabile Y sia pari ad 1 condizionata all'assunzione del valore x della variabile indipendente X, ossia:

$$P(Y = 1|X = x) = \pi(x) = \alpha + \beta X \tag{2.12}$$

La probabilità così espressa possiede ancora un campo di esistenza tra $-\infty$ e $+\infty$, tuttavia è possibile restringerlo al range (0;1) applicando prima la trasformazione esponenziale e poi quella logistica:

$$P(Y = 1|X = x) = \pi(x) = \frac{e^{\alpha+\beta X}}{1 + e^{\alpha+\beta X}} \tag{2.13}$$

A questo punto, è possibile esprimere le probabilità mediante il rapporto tra le due categorie 0 e 1, considerando la complementarietà degli eventi ossia $P(Y = 0) = P(Y = 1|X = x)$. Tale rapporto prende il nome di *odd* ed in genere si calcola dividendo le frequenze osservate in una classe con quelle osservate nell'altra:

$$odds(Y = 1|X = x) = \frac{\pi(x)}{1 - \pi(x)} \tag{2.14}$$

$$odds(Y = 1|X = x) = \frac{\frac{e^{\alpha+\beta X}}{1 + e^{\alpha+\beta X}}}{1 - \frac{e^{\alpha+\beta X}}{1 + e^{\alpha+\beta X}}} = \frac{\frac{e^{\alpha+\beta X}}{1 + e^{\alpha+\beta X}}}{\frac{1}{1 + e^{\alpha+\beta X}}} = e^{\alpha+\beta X} \tag{2.15}$$

Calcolando il *logit*, ossia il logaritmo dell'*odd*, e applicando le proprietà dei logaritmi si ottiene:

$$\ln\big(odds(Y = 1|X = x)\big) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta X \tag{2.16}$$

Nel caso di regressione logistica multipla avente k variabili indipendenti è possibile ripercorrere lo stesso ragionamento arrivando alla seguente conclusione:

$$\pi(x) = \frac{e^{\alpha+\beta_1 X_1+\beta_2 X_2+ \ldots+\beta_k X_k}}{1 + e^{\alpha+\beta_1 X_1+\beta_2 X_2+ \ldots+\beta_k X_k}} \tag{2.17}$$

Riassumendo, è possibile affermare che i concetti di probabilità *odd* e *logit* trattati, forniscono sfumature diverse di uno stesso concetto. La probabilità è riferita all'appartenenza ad un gruppo, l'*odd* definisce quanto è più o meno probabile un evento rispetto ad un altro e il *logit*, calcolato come il logaritmo dell'*odd*, rende possibili le trasformazioni algebriche.

2.2.2 Modelli di machine learning

Negli ultimi anni gli algoritmi di *machine learning* hanno destato l'interesse di numerosi ricercatori soprattutto in campo economico-finanziario. La loro peculiarità è la possibilità di imparare da dati appartenenti a campioni opportunamente creati e quindi di poter elaborare una previsione su nuovi dati. Rispetto ai modelli più classici per i quali bisogna definire le operazioni che l'algoritmo dovrà svolgere, i modelli di *machine learning* consentono ai computer di prendere decisioni ed imparare autonomamente a svolgere il compito. Fondamentalmente, questi modelli valutano, sulla base dei dati usati per imparare, un possibile default di un'impresa quando certi indicatori sono più alti o più bassi del dovuto.

Le tecniche di *machine learning* sono usate principalmente in presenza di sistemi il cui livello di complessità risulta troppo elevato a causa del numero delle variabili in gioco per capire l'equazione che li governa. L'output di una tecnica è quindi un modello in grado di svolgere analisi predittive su nuovi dati basandosi su proprietà non note a priori e imparate attraverso l'utilizzo di un *dataset* di training. Il concetto alla base di questo ragionamento è la generalizzazione, in quanto una volta che l'algoritmo è stato alimentato con i dati di training, questo deve poter poi mappare una funzione sulla quale saranno passati i dati di test per verificare la correttezza della generalizzazione.

La precisione delle analisi svolte con le tecniche di *machine learning* è strettamente connessa alla qualità delle informazioni contenute nel database impiegato nella fase di addestramento del modello, proprio questa forte dipendenza dai campioni utilizzati rappresenta il punto debole dei modelli di *machine learning*. Per poter migliorare sensibilmente l'accuratezza di questi modelli, il database impiegato dovrebbe tra le altre cose racchiudere un numero di osservazioni tale da poter ritenere statisticamente significative le inferenze svolte su di esso.

Altri due attributi che deve possedere un database per poter essere ritenuto di qualità, sono la rappresentatività territoriale di tutte le imprese nazionali e l'estensione della serie storica esaminata. La prima caratteristica è necessaria per poter somministrare al modello, imprese con strutture finanziarie simili e quindi ricevere come risposta uno score attendibile mentre la serie storica di riferimento dovrebbe essere tale da poter analizzare un periodo maggiore ad una congiuntura economica.

In generale, è possibile classificare i diversi algoritmi di *machine learning* sulla base della tipologia di apprendimento e della categoria di appartenenza. Queste due distinzioni consentono di comprendere la teoria e la logica di funzionamento dei vari algoritmi. È possibile distinguere tre categorie di apprendimento:

- Apprendimento supervisionato (*Supervised Learining*)

  Con il termine supervisione si intende che le informazioni fornite all'elaboratore sono associazioni di output agli input. Queste coppie di valori sono da intendersi come esempi ideali i cui output desiderati sono già noti a priori. Le categorie di algoritmi adottati in questo campo sono le tecniche di regressione e di classificazione. La prima è utilizzata quando è necessario trovare una relazione tra una serie di variabili predittive e una variabile di output che può assumere valori continui, mentre si ricorre alla classificazione quando le informazioni sono di tipo categorico e consentono di stabilire se un oggetto ha le caratteristiche più simili per appartenere ad una classe piuttosto che ad un'altra.

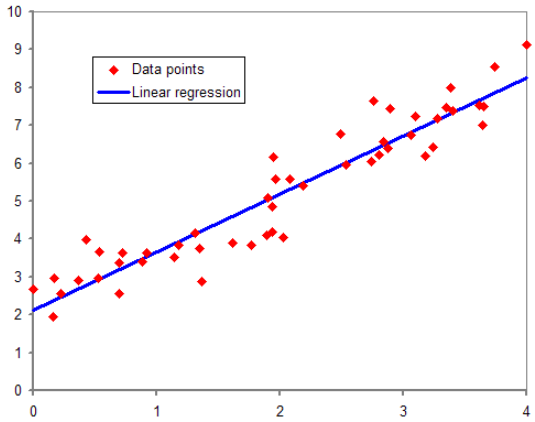

Figura 2.4 *Regressione*

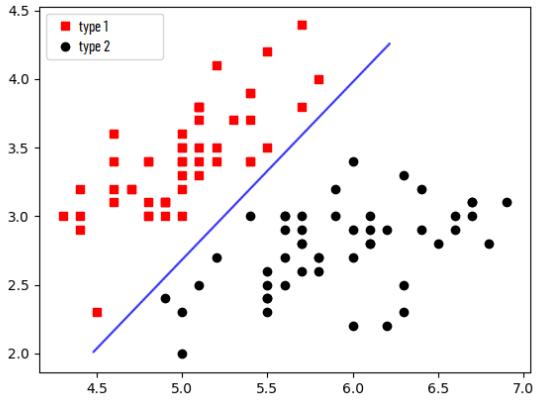

Figura 2.5 *Classificazione*

- Apprendimento non supervisionato (*Unsupervised Learning*)

  In questa tipologia gli input consegnati al sistema informatico sono riclassificati sulla base di caratteristiche che li rendono coerenti tra di loro. Le classi non sono note a priori ma osservando la natura dei dati si estrapolano le informazioni necessarie per identificarle. Per affrontare i problemi di apprendimento non supervisionato si utilizzano tecniche di *clustering* che consentono di selezionare e raggruppare i dati in insiemi omogenei.

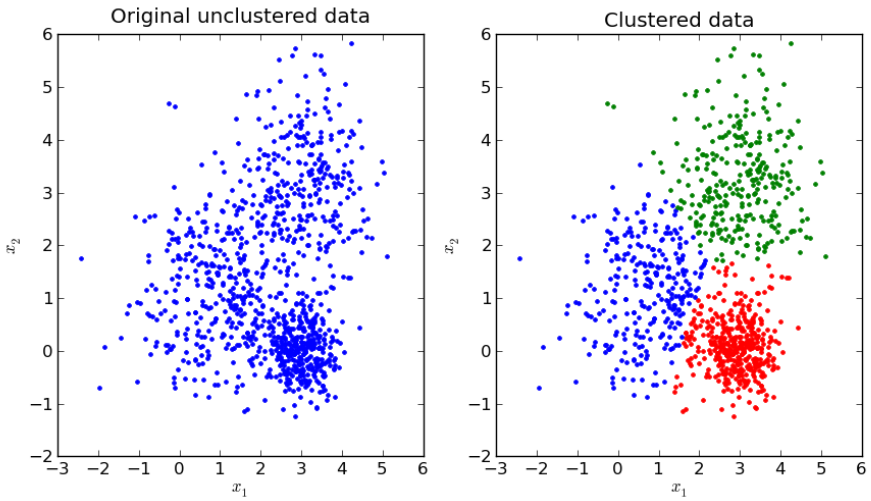

Figura 2.6 *Clustering*

- Apprendimento per rinforzo (*Reinforcement Learning*)

  L'ultima metodologia di apprendimento avviene attraverso l'interazione del sistema informatico con l'ambiente che lo circonda. L'elaboratore inizia operando in modo casuale per poi ricevere dei segnali di ricompensa detti "rinforzi" che, valutando la qualità delle azioni intraprese, permettono al sistema di migliorare la propria performance.

Un particolare algoritmo di classificazione che sta destando notevole interesse in campo economico-finanziario, sono le reti neurali. La loro architettura, volta a riprodurre artificialmente l'organizzazione e il funzionamento del cervello umano, può essere impiegata in diversi campi in virtù delle sue capacità di apprendimento, classificazione e generalizzazione. Nel prossimo capitolo, dopo un breve *excursus* sulle analogie con le strutture celebrali umane, si analizzeranno le principali caratteristiche e funzioni algebriche che governano tali modelli.

Capitolo III

Le reti neurali

Il cervello, il ragionamento e la memoria hanno da sempre suscitato interesse dell'essere umano. Notevoli sforzi sono stati adoperati in tale direzione ma solamente i più recenti progressi nella biologia molecolare e delle neuroscienze computazionali, hanno consentito di capire a fondo il funzionamento del sistema nervoso e del suo organo principale: il cervello.

Il cervello costituisce il centro del sistema nervoso e, come un computer, ci consente di memorizzare le informazioni. Sia il cervello sia i computer possono immagazzinare informazioni ma i loro meccanismi sono molto differenti. La memoria centrale dei computer è un elemento fisico integrato sulla scheda madre interfacciato direttamente con il processore CPU per consentire un flusso continuo di memorizzazione e scambio di dati. Nel cervello non sono presenti parti fisiche destinate allo *storage*, infatti il neurone in sé non ha alcuna capacità di accumulo delle informazioni. Dunque, il cervello può essere considerato come una gigantesca rete di neuroni la cui associazione richiama informazioni specifiche.

In questo capitolo, si focalizzerà l'attenzione sulle reti neurali in quanto imitazione del funzionamento del cervello. Quest'ultimo è composto da associazioni di neuroni mentre le prime sono formate da connessioni pesate di nodi.

3.1 Il neurone biologico

I neuroni rappresentano le unità cellulari che costituiscono il sistema nervoso e per mezzo delle connessioni sinaptiche consentono il passaggio delle informazioni. Il cervello umano contiene circa 100 miliardi di neuroni e ciascuno di essi è collegato con altri 1.000. La struttura tipica di un neurone, come si evince dalla Figura 3.1, è costituita da tre elementi, ciascuno con un compito predefinito:

- Pirenoforo (o soma)

  È il corpo cellulare ed il centro metabolico del neurone stesso, contiene il nucleo e l'apparato biosintetico per la produzione dei costituenti di membrana, degli enzimi e di altre sostanze chimiche di cui la cellula ha bisogno per svolgere le proprie funzioni.

- Assone

  È un prolungamento di lunghezza variabile che consente il trasporto, sotto forma di potenziale d'azione, degli impulsi effettori dal centro integrativo del neurone verso le cellule bersaglio.

- Dendriti

  Sono estensioni sottili di forma tubulare che tendono a suddividersi più volte conferendo al neurone la sua struttura caratteristica simile alla chioma di un albero. Il loro compito consiste nella ricezione dei segnali da parte di altre cellule nervose che comunicano in specifiche zone chiamate sinapsi o spine dendritiche.

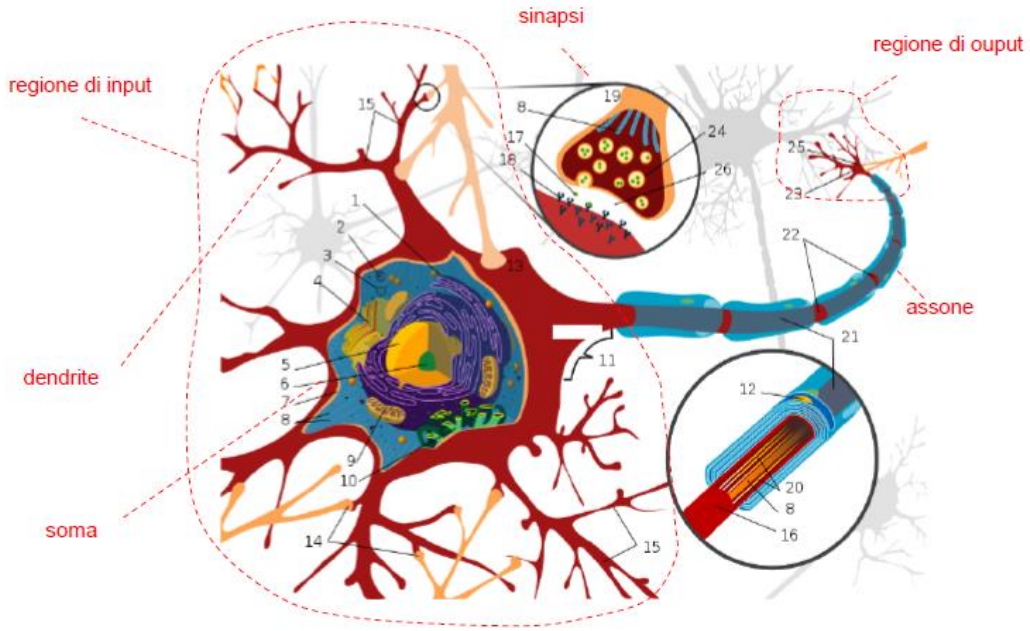

Figura 3.1 *Neurone biologico*

Il trasferimento delle informazioni avviene per mezzo di processi elettro-chimici. L'ingresso di ioni attraverso le sinapsi dei dendriti provoca una differenza di potenziale tra il soma e l'esterno. Non appena il potenziale oltrepassa un valore soglia il neurone emette un impulso (*spike*) che si propaga nell'assone rilasciando ioni. I segnali ricevuti in ingresso sono sommati e l'informazione trasmessa può essere di natura eccitatoria o inibitoria. Nel primo caso il neurone si attiva e genera a sua volta informazioni attraverso le sinapsi in uscita mentre nel secondo caso blocca l'impulso.

Un aspetto strettamente collegato ai processi di apprendimento e di memorizzazione è la modifica dei pesi delle sinapsi nota come *reweighting*. La regola di apprendimento attualmente più diffusa è la legge di Hebb secondo la quale "*se un neurone A è abbastanza vicino ad un neurone B da contribuire ripetutamente e in maniera duratura alla sua eccitazione, allora ha luogo in entrambi i neuroni un processo di crescita o di cambiamento metabolico tale per cui l'efficacia di A nell'eccitare B viene accresciuta*"[2].

3.2 Il neurone artificiale: modello di McCulloch e Pitts

Il primo modello computazionale del neurone biologico fu proposto da McCulloch e Pitts nel trattato "*A logial calculus of the ideas immanent in nervous activity*" del 1943. Secondo tale modello gli elementi che costituiscono il neurone biologico possono essere schematizzati nel seguente modo:

- le connessioni tra le sinapsi e i dendriti sono considerate come linee di input;
- le differenze di potenziale sono i segnali di input provenienti dagli altri neuroni;
- il pirenoforo o soma rappresenta il centro del sistema e quindi l'unità di computazione elementare;
- l'assone è il canale di output che trasporta il segnale oppure lo blocca.

Per poter tradurre a livello logico il funzionamento del neurone McCulloch e Pitts hanno dovuto apportare ulteriori semplificazioni al modello biologico. In particolare:

- l'efficacia della connessione tra due neuroni, spiegata dalla legge di Hebb, è rappresentata dal valore w_i associato ad ogni linea di input;
- i valori di input sono rappresentati da variabili binarie x_i che assumono valore pari ad 1 per indicare la presenza di un potenziale d'azione in ingresso mentre ad un valore nullo è associata una mancanza del segnale;
- un'altra variabile binaria y_i è utilizzata per segnalare la presenza di un potenziale d'azione sull'assone.

La Figura 3.2 mostra la semplicità del primo modello di McCulloch e Pitts evidenziandone la natura binaria delle variabili in ingresso e in uscita.

[2] Legge di Hebb tratta da:
https://it.wikipedia.org/wiki/Donald_Olding_Hebb

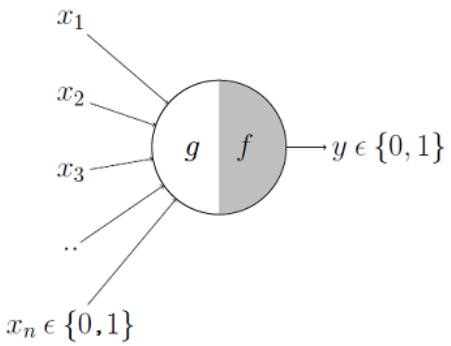

Figura 3.2 *Modello di McCulloch e Pitts*

Il funzionamento del neurone artificiale comincia con la ricezione degli input x_i, i quali sono moltiplicati per i rispettivi pesi w_i per tenere conto dell'efficacia della connessione. Se tale somma pesata supera il valore soglia T allora il neurone si attiverà è l'output y varrà 1 altrimenti 0. È possibile formalizzare il ragionamento precedente in termini matematici nel seguente modo:

$$y = \theta(h - T) \quad con \quad h = \sum_{i=1}^{n} w_i \, x_i \tag{3.1}$$

$$\theta(x) = \begin{cases} 1 & se\ x > 0 \\ 0 & altrimenti \end{cases} \tag{3.2}$$

Pertanto, il neurone artificiale avrà un output binario con valore pari ad 1 quando la somma pesata dei suoi input supera la soglia T, in caso contrario il segnale in uscita sarà pari a 0.

3.3 Il neurone moderno

La Figura 3.3 schematizza il funzionamento di un neurone artificiale nella sua versione più moderna. Esso si presenta come un'evoluzione del modello precedente introducendo alcune importanti novità.

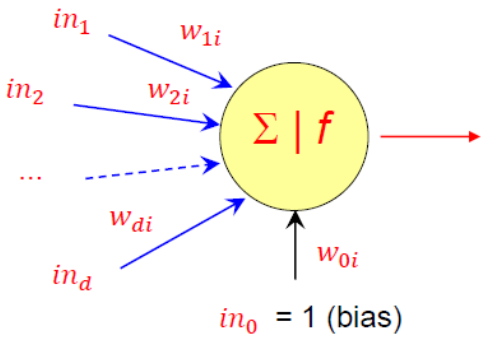

Figura 3.3 *Neurone artificiale moderno*

$$neuron_i = \sum_{j=1}^{d} w_{ji} \cdot in_j + w_{0i} \tag{3.3}$$

$$out_i = f(neuron_i) \tag{3.4}$$

Dove:

- $in_1, in_2, \dots in_d$ sono i d segnali di input che il neurone *i-esimo* riceve dagli altri neuroni;
- $w_{1i}, w_{2i}, \dots w_{di}$ rappresentano i pesi (*weight*) che definiscono l'efficacia delle diverse connessioni;
- w_{0i} è il *bias* un peso aggiuntivo al quale è sempre associato un input in_0 pari a 1;
- $neuron_i$ è il suo livello di eccitazione ossia il suo potenziale interno;
- out_i è il valore in uscita del neurone *i-esimo*;
- $f(\cdot)$ è la funzione di attivazione che sulla base del livello di eccitazione determina l'output del neurone.

In altri termini, se si considera un generico neurone della rete, esso riceverà in ingresso d valori di input ciascuno dei quali moltiplicato per un peso w_{di} più il valore del *bias* w_{0i}. L'output è calcolato come somma pesata e la (3.3) mostra che maggiori sono i pesi maggiore è l'importanza associata quel preciso valore in ingresso. Ad esempio, se si annulla il valore di un peso w_{ji}, il segnale e l'informazione ad esso associata non avrà alcun effetto sulla rete. Dunque, in una rete neurale, le informazioni sono immagazzinate in termini di pesi e di *bias* i quali determinano il valore dell'output.

Il neurone moderno migliora il modello di McCulloch e Pitts conferendo la possibilità di operare anche con valori non binari, ora i neuroni possono sia ricevere in ingresso sia restituire valori continui. Altre novità importanti riguardano l'introduzione del parametro *bias* e la funzione di attivazione. Il primo, come già precedentemente accennato, è utilizzato per regolare il funzionamento del neurone in quanto si aggiunge alla somma pesata degli input contribuendo alla definizione del suo potenziale interno. Tale potenziale è poi processato da una funzione di attivazione che restituisce l'output finale ossia la risposta del neurone ai valori in ingresso.

3.4 Le funzioni di attivazione

Le funzioni di attivazione sono equazioni matematiche che determinano l'output di una rete neurale. La funzione è applicata ad ogni singolo neurone della rete e stabilisce se deve essere attivato o meno, ossia se l'informazione può passare allo strato successivo. Nei neuroni biologici sono di tipo binario in quanto l'impulso si propaga solamente al superamento di un certo livello di eccitazione interno. Generalmente, le reti neurali artificiali replicano il comportamento di quelle biologiche utilizzando funzioni continue, non-lineari e differenziabili. La non-linearità è necessaria se si vuole creare un modello in grado di processare informazioni di input complesse mentre la continuità e la differenziabilità derivano da esigenze matematiche di calcolo.

A seconda delle necessità richieste dalla costruzione del modello previsionale, è possibile ricorrere a differenti tipologie di funzioni di attivazione. Le più comuni sono: *step function*, *linear function*, *standard logistic function*, *hyperbolic tangent function*, *Rectified Linear Units* (*ReLU*) e *Leaky ReLU*.

3.4.1 Step function

La funzione di attivazione che replica più fedelmente il comportamento dei neuroni biologici è la "funzione a gradino". Essa si basa sull'esistenza di un valore soglia di attivazione, se la somma pesata dei valori in ingresso è superiore a tale valore allora il neurone è attivato e trasmette allo strato successivo esattamente lo stesso segnale.

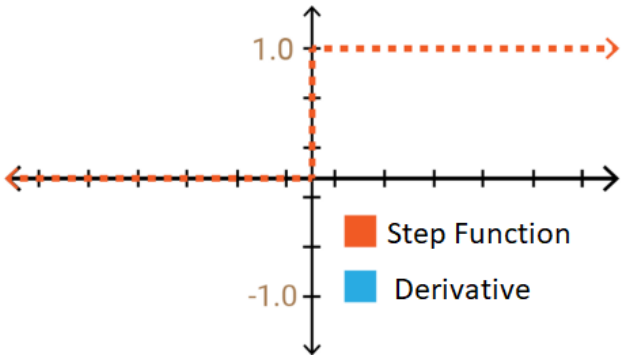

Figura 3.4 *Step function e derivata*

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{3.5}$$

$$f'(x) = 0 \tag{3.6}$$

Il principale problema derivante dall'utilizzo di questa funzione è che la *step function* non è derivabile in $x = 0$ e ha derivata nulla in qualsiasi altro punto. Una tra le procedure più efficienti per allenare una rete multistrato, come sarà spiegato meglio in seguito, è usare il metodo del *gradient descent* con la *backpropagation* ed uno dei requisiti richiesti da quest'algoritmo è che la funzione di attivazione sia differenziabile. Questo implica che il metodo *gradient descent*, che prevede piccole modifiche ai valori dei pesi e ai *bias* per ottenere modesti cambiamenti nell'output della rete, non permette di migliorare progressivamente l'aggiornamento dei pesi.

3.4.2 Linear function

Le funzioni di attivazione di tipo lineare anche note con il nome di funzioni d'identità, generano valori di output proporzionali agli input. La formula e la derivata di una generica funzione sono:

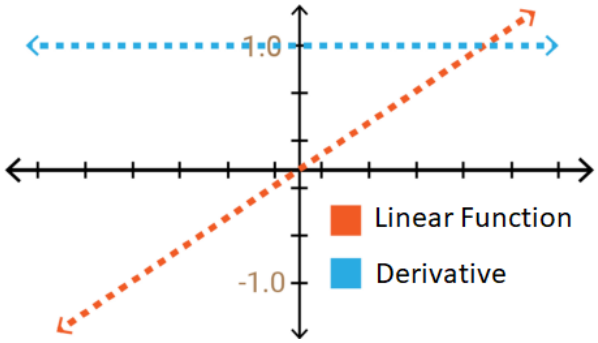

Figura 3.5 *Linear function e derivata*

$$f(x) = k \cdot x \tag{3.7}$$

$$f'(x) = k \tag{3.8}$$

Questo tipo di funzione nonostante generi valori di facile interpretazione non è molto utilizzato a causa del valore costante della derivata. Infatti, se la derivata è sempre pari a k significa che non c'è nessun legame con la variabile di input x e quindi ogni volta che si

attua la *backpropagation* il gradiente rimane lo stesso e il processo di apprendimento fallisce. Inoltre, siccome la combinazione di funzioni lineari è anch'essa una funzione lineare non importa quanti strati si impieghino in quanto l'output finale sarà semplicemente una trasformazione lineare dell'input.

3.4.3 Standard logistic function e hyperbolic tangent function

La *standard logistic function* conosciuta anche con il nome di sigmoide e l'*hyperbolic tangent function* sono entrambe curve ad "S" aventi equazioni e derivate pari a:

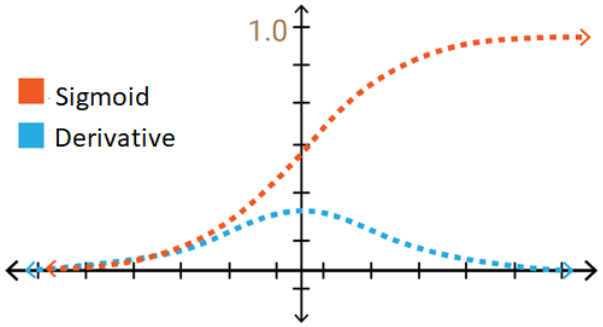

Figura 3.6 *Sigmoid function e derivata*

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.9}$$

$$f'(x) = \frac{\partial}{\partial x}\left(\frac{1}{1 + e^{-x}}\right) = \frac{e^{-x}}{(1 + e^{-x})^2} = f(x)(1 - f(x)) \tag{3.10}$$

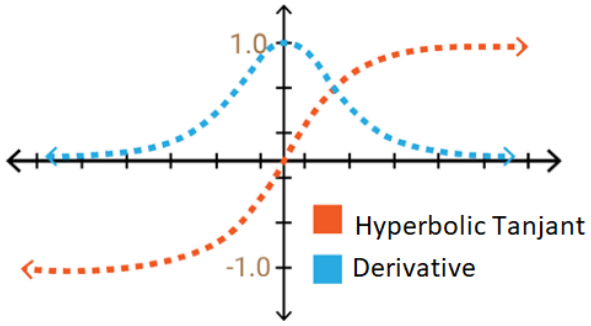

Figura 3.7 *Hyperbolic tangent function e derivata*

$$f(x) = \tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \tag{3.11}$$

$$f'(x) = 1 - f(x)^2 \tag{3.12}$$

La più grande differenza è il campo di esistenza, per la prima si estende da 0 a 1 mentre la tangente iperbolica va da -1 a +1. Queste differenze di maggior ampiezza e centratura sullo 0 del campo di esistenza, provocano una maggiore difficoltà nel processo di ricerca del gradiente ottimale, che risulta leggermente più lento per la sigmoide. Ad ogni modo si tratta di ottime funzioni di attivazione in quanto non binarie, derivabili e non lineari. Queste qualità unite alla facilità di comprensione dei risultati le rendono tra le funzioni di attivazione più usate, anche se più recentemente sono state soggette a critiche. La più importante riguarda il *vanishing gradient problem*, ossia il fatto che per valori molto grandi o molto piccoli di x la variabile y non cambia e quindi il modello avrà difficoltà nell'addestramento e nell'ottenere una buona performance per questi punti.

3.4.4 Rectified Linear Units function e Leaky ReLU function

La *Rectified Linear units* (*ReLU*) e la *Leaky ReLU* sono state introdotte per risolvere il v*anishing gradient problem* nei casi in cui i dati di input causavano una particolare lentezza nel raggiungimento della soluzione ottima del modello. Le funzioni e il loro andamento sono riportate di seguito:

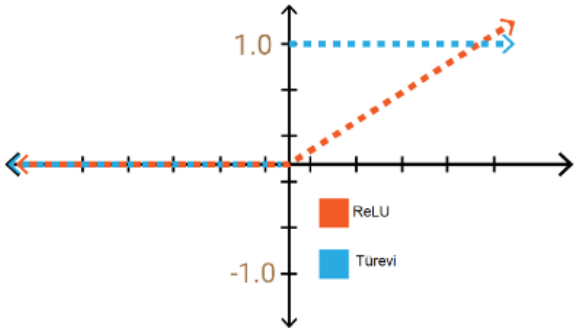

Figura 3.8 *Rectified Linear Unit e derivata*

$$f(x) = \max(0, x) \tag{3.13}$$

$$f'(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases} \tag{3.14}$$

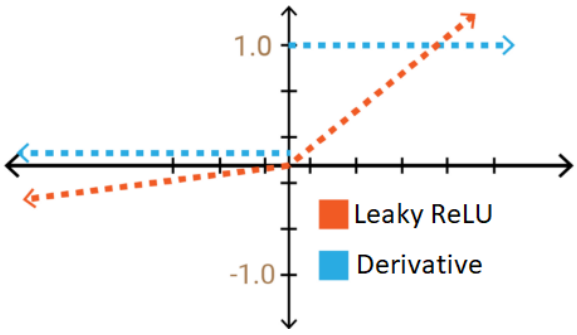

Figura 3.9 *Leaky ReLU e derivata*

$$f(x) = \begin{cases} kx, & x \leq 0 \\ x, & x > 0 \end{cases} \tag{3.15}$$

$$f'(x) = \begin{cases} k, & x \leq 0 \\ 1, & x > 0 \end{cases} \tag{3.16}$$

Il principale vantaggio rispetto alle due funzioni precedenti è che siccome l'output di certi neuroni è 0, solamente alcuni saranno attivi. Il fatto che non tutti i neuroni siano attivati allo stesso tempo rende la rete più efficiente a livello computazionale rispetto alla sigmoide e alla tanh.

Dai due grafici si può facilmente cogliere la differenza tra le due funzioni ovvero che la *Leaky ReLU* presenta una leggera inclinazione per i valori negativi di x. Il motivo di questa lieve pendenza è da ricercare nuovamente nell'algoritmo di *backpropagation*, in questa regione infatti i pesi degli input non sono aggiornati e ciò può causare la presenza di *dead neurons* ossia neuroni che non saranno mai attivati. Le due nuove funzioni tendono ad assumere un valore medio superiore allo 0 provocando uno slittamento positivo della media nei vari strati della rete e quindi un rallentamento del processo di apprendimento.

È possibile affermare che non esista un'unica funzione di attivazione ottimale a priori poiché per ognuna di esse esistono vantaggi e svantaggi. La sigmoide, per esempio, può essere usata se si ritiene che la tanh o le altre funzioni abbiamo un campo di esistenza eccessivamente ampio per il modello utilizzato. La *ReLU* potrebbe essere la soluzione migliore se il problema principale della rete è la complessità computazionale del modello. La *Leaky ReLU* invece, potrebbe risultare un'ottima opzione per risolvere il *vanishing gradient problem*. Pertanto, la scelta della funzione di attivazione si riduce ad un problema di ottimizzazione da risolvere sulla base delle informazioni disponibili e dei requisiti richiesti dal proprio modello di intelligenza artificiale.

3.5 L'architettura delle reti

Una rete neurale è una connessione di neuroni sistemati in architetture più o meno complesse. Con il termine architettura si fa riferimento alla disposizione dei neuroni nella rete e agli schemi di connessione tra di essi. Il modello e la struttura della rete determinano il suo funzionamento e in particolare come l'informazione si trasmette dagli input agli output. La scelta dell'architettura della rete influenza la performance del modello, diventa quindi necessario conoscere le principali strutture di rete per scegliere quella che consente di ottenere la soluzione ottima.

L'architettura delle reti neurali si è evoluta nel corso degli anni per soddisfare le più moderne esigenze applicative. Le prime reti costruite note come *single layer neural network* erano formate unicamente da strati di input e di output, si è proceduto con l'aggiunta degli strati nascosti, per i quali non vi è alcun vincolo sul numero di neuroni che possono contenere, prima reti con strati singoli *shallow network o vanilla network*, poi con strati multipli *deep neural network*. Tutte queste reti rientrano nella categoria delle reti a strati, per le quali il segnale entra negli strati di input e si propaga nei successivi dove avviene il processamento dei dati.

3.5.1 Feed-Forward Neural Networks

Le architetture più comunemente usate sono le reti *feed-forward*, così chiamate per sottolineare che l'unico verso di propagazione dell'informazione è in avanti. Lo scopo principale di una rete di questo tipo è l'approssimazione di una generica funzione f in grado di mappare i valori di x con gli output y. In sostanza, una rete *feed-forward* osservando i valori in uscita $y = f(x;\theta)$ apprende i valori dei parametri θ che meglio approssimano f. In accordo con quanto espresso dal "*Universal approximation theorem*", le reti *feed-forward* aventi almeno uno strato intermedio di *hidden neurons* sono in grado di approssimare qualsiasi funzione continua in R^n. In altre parole, il teorema afferma che non importa quale sia la funzione che si sta tentando di imparare, ci sarà sempre una rete *feed-forward* in grado di approssimare la funzione che si sta cercando.

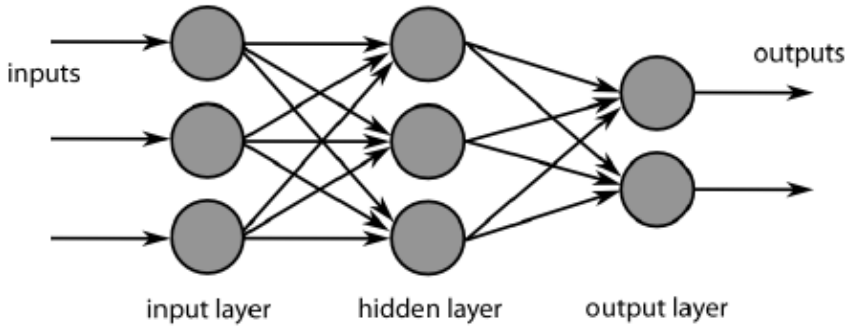

Figura 3.10 *Feed-Forward Neural Network a tre strati: input, hidden e output*

Con riferimento alla Figura 3.10, il primo stato è chiamato *input layer* e i neuroni che lo compongono *input neurons*, essi hanno il compito di trasmettere i segnali in ingresso così come sono senza calcolare alcuna somma pesata o far uso di funzioni di attivazione. I segnali di input sono quindi duplicati e inviati a tutti i nodi intermedi chiamati *hidden neurons* andando a formare una struttura completamente interconnessa. Al contrario dello strato precedente, questi e tutti gli altri neuroni della rete, sono considerati attivi in quanto elaborano i dati che ricevono in input effettuando le opportune trasformazioni. I neuroni nascosti appartenenti agli *hidden layers* sono così chiamati in quanto collocati tra input e output e quindi non accessibili dall'esterno della rete. Infine, nello strato sull'estrema destra del grafo, o *output layer*, si dispongono gli *output neurons* dai quali si ricava la risposta finale della rete.

3.5.2 Convolutional Neural Networks

Le reti convoluzionali sono state sviluppate da Yann LeCun ed i suoi collaboratori nel 1998 per la costruzione di un modello di riconoscimento di cifre scritte a mano. Il principale campo applicativo di tali reti riguarda l'identificazione da parte di un computer di immagini e video, infatti solitamente ci si serve di queste reti per classificare un'immagine e identificarne il contenuto.

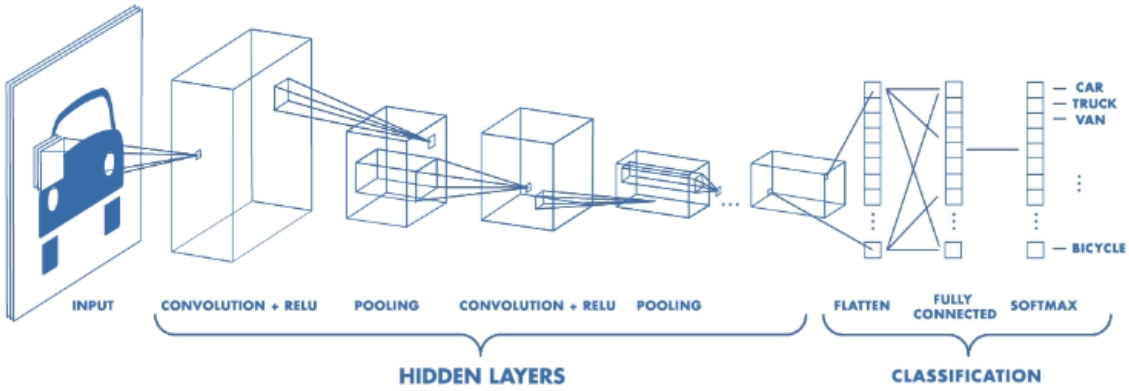

Figura 3.11 *Convolutional Neural Network*

Le *convolutional neural networks* risultano molto diverse dalle altre reti anche in termini di architettura. La Figura 3.11 mostra i diversi livelli in cui è strutturata la rete. Il primo livello ha la funzione di recepire gli input esterni solitamente rappresentati da un insieme di pixel convertiti in cifre. I dati passano ai livelli convoluzionali, rappresentanti il cuore del programma, che individuano la presenza nelle immagini di elementi schematici quali curve ed angoli. L'output dei livelli convoluzionali diventa l'input dei *ReLU* che utilizzando la stessa funzione di attivazione annullano i valori negativi degli strati precedenti aumentando la non-linearità del modello. L'immagine è poi semplificata dal livello *pool* che evidenzia le caratteristiche emerse dall'analisi dei livelli convoluzionali. Infine, la rete elabora la propria previsione nel livello *fully connected* generando come output un vettore di dimensione pari ad N, dove N è il numero delle classi tra le quali il programma deve scegliere, contenente la probabilità che l'immagine appartenga a quella determinata classe.

3.5.3 Recurrent Neural Networks

Le *Recurrent Neural Networks* sono state presentate da Jeffrey Elman nel suo studio "*Finding structure in time*" del 1990. I modelli costruiti con queste reti sono molto performanti in quanto prevedono sia collegamenti con gli strati precedenti che verso lo stesso livello. Le associazioni ricorsive conferiscono alla rete la capacità mnemonica di prendere le decisioni sulla storia passata ossia in base all'esperienza maturata precedentemente.

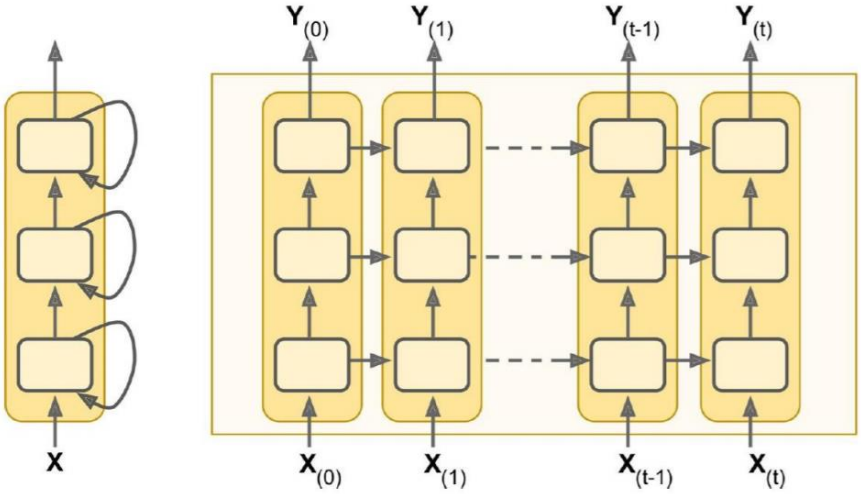

Figura 3.12 *Recurrent Neural Network*

La Figura 3.12 mostra l'architettura tipica delle reti neurali ricorrenti. L'unità elementare che le compone è la cella, essa è caratterizzata da uno strato interno $h_{(t)}$ composto da un numero fisso di neuroni per ogni istante temporale. Inoltre, è possibile notare come ad ogni istante di tempo t, una determinata cella riceva in ingresso, non solo l'output della cella antecedente $x_{(t)}$ ma anche la propria risposta emessa nell'istante precedente $h_{(t-1)}$. Il valore assunto dallo stato nell'istante t e quindi l'output $y_{(t)}$ della cella ad esso corrispondente vale:

$$y_{(t)} = h_{(t)} = f(h_{(t-1)}, x_{(t)}) \tag{3.17}$$

Le *recurrent neural networks* nella loro forma base non riescono a sfruttare la capacità mnemonica degli strati più lontani, in quanto gli effetti dei primi input tendono a svanire nella rete. Nei campi in cui la memoria a lungo termine gioca un ruolo importante, si utilizzano strutture formate da celle più complesse come le *Long Short-Term Memory* (LSTM) e le *Gated Recurrent Unit* (GRU).

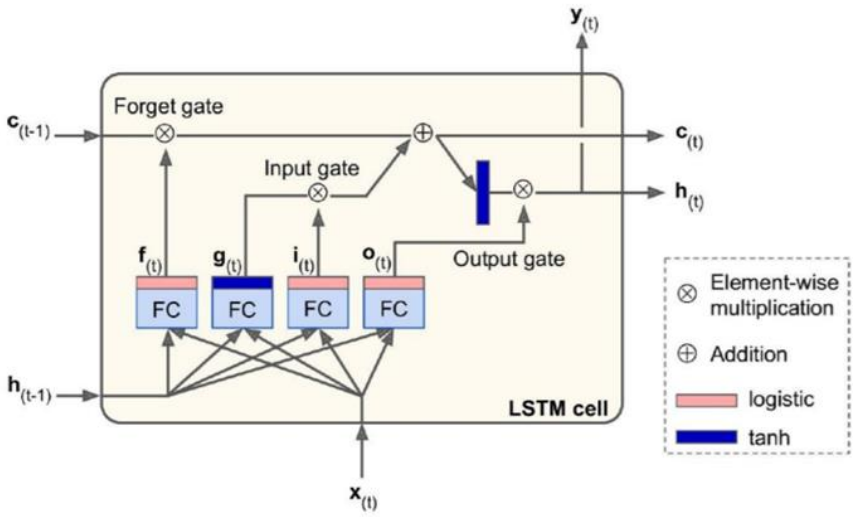

Figura 3.13 *Cella LSTM*

Nelle celle LSTM due fattori contribuiscono alla determinazione del valore finale del suo stato. Il primo $h_{(t)}$ è uguale all'output della rete $y_{(t)}$ e, come nel caso precedente, fa riferimento alla memoria a breve termine mentre il secondo $c_{(t)}$ conferisce alla rete la capacità di applicare ragionamenti a lungo termine. Nella fase di apprendimento, la cella impara sia quali informazioni aggiungere al segnale in ingresso $x_{(t)}$ (*input gate*), sia quali sono i dati meno importanti e quindi da dimenticare $c_{(t-1)}$ (*forget gate*). Infine, la cella elabora il proprio output combinando l'input $x_{(t)}$ con le informazioni della memoria a lungo termine.

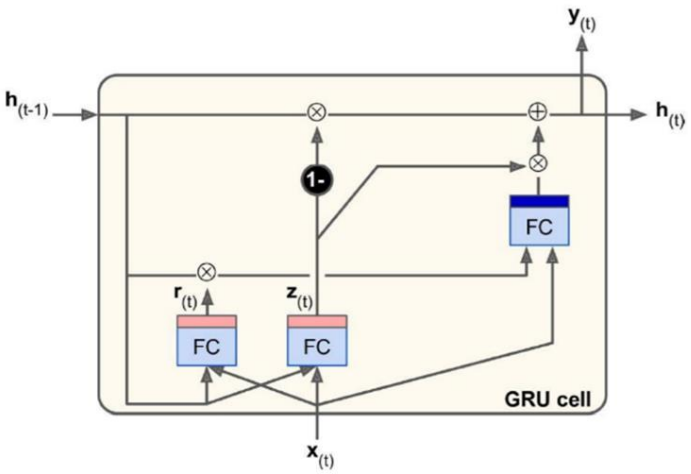

Figura 3.14 *Cella GRU*

Un'applicazione più semplicistica ma con proprietà analoghe della versione precedente è la cella GRU. Esse posseggono un solo stato di memoria $h_{(t)}$ ed un solo gate (*update gate*) per stabilire se e quanto dimenticare per poter aggiungere informazioni al nuovo output.

3.6 L'addestramento delle reti MLP

Dopo aver passato in rassegna le principali architetture utilizzate per la costruzione dei modelli artificiali, è necessario analizzare il loro funzionamento ed in particolare l'aspetto che più affascina e rende speciali questi modelli ossia l'apprendimento. Come già anticipato, le reti *feed-forward* sono le strutture più diffuse tra i modelli previsionali e solitamente sono costituite da almeno tre strati (input, *hidden* e output) prendendo il nome di *Multilayer Precepton* (MLP).

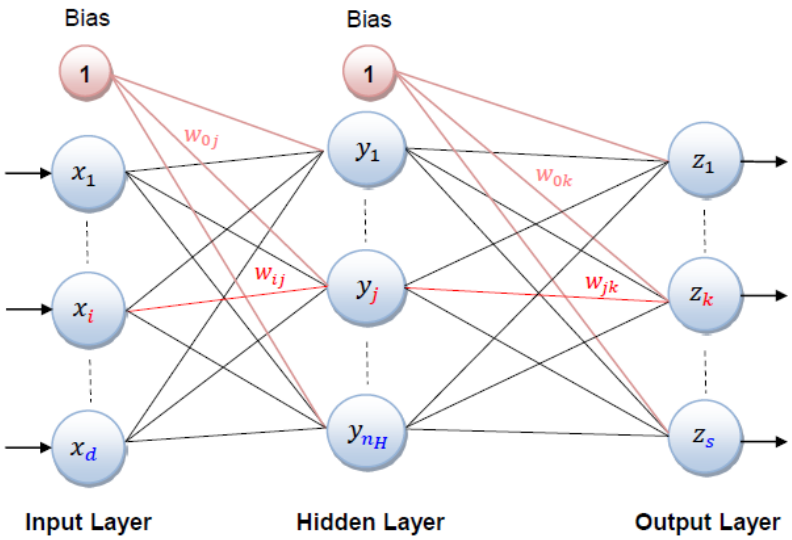

Figura 3.15 *Rete Multilayer Precepton a tre strati: input, hidden e output*

La caratteristica di queste reti è il verso della propagazione dell'informazione, essa avviene solamente in avanti dallo strato di input a quello di output rendendo l'interpretazione del modello più semplice. Riprendendo la nomenclatura della Figura 3.15 è possibile esprimere l'output di un *k-esimo* nodo di una rete.

$$z_k = f\left(\sum_{i=1}^{n_h} w_{ij} \cdot y_j + w_{0k}\right) = f\left(\sum_{i=1}^{n_h} w_{ij} \cdot f\left(\sum_{i=1}^{d} w_{ij} \cdot x_i + w_{0j}\right) + w_{0k}\right) \quad (3.18)$$

La formula evidenzia come l'output z_k sia influenzato dai valori dei pesi w_{ij} e dai *bias* w_{0j}, quindi è possibile affermare che le reti neurali immagazzinano le informazioni in termini di pesi e *bias*. Questo significa che il processo di addestramento di una rete, deve necessariamente passare attraverso la modifica di questi valori.

Nonostante il modello di McCulloch e Pitts risalga al 1943 per aggiungere uno strato *hidden* al fine di costruire la prima rete MLP si è dovuto attendere fino al 1986 quando Rumelhart, Hinton & Williams hanno presentato l'algoritmo di *error backpropagation*. Lo sviluppo di questo algoritmo ha richiesto ben 40 anni di ricerca durante i quali i ricercatori non sono stati in grado di escogitare alcun metodo efficiente per l'allenamento degli strati nascosti. Dall'aneddoto storico si coglie la centralità dell'algoritmo di apprendimento per il corretto funzionamento di una rete neurale.

3.6.1 Error backpropagation

Gli algoritmi di apprendimento hanno come scopo quello di determinare i valori dei pesi e dei *bias* in grado di stimare i parametri del modello. L'algoritmo di *backpropagation* è solitamente impiegato nei modelli che fanno uso di apprendimento supervisionato in quanto l'output z generato dalla rete in risposta all'input x è ripetutamente confrontato con l'output desiderato t. Lo scostamento tra questi due valori prende il nome di errore relativo e la somma quadratica di questi errori è chiamata *loss function:*

$$J(w,x) = \frac{1}{2}\sum_{c=1}^{s}(t_c - z_c)^2 \tag{3.19}$$

Dove s è il numero di classi di risposta che in un modello binomiale sono 0 e 1.

La funzione di perdita è quindi una funzione di errore che misura di quanto la previsione del modello si discosta dall'output desiderato. Successivamente si procede alla minimizzazione di $J(w,x)$ modificando i pesi in direzione opposta al gradiente di J, ossia contrariamente alla direzione di massima crescita della funzione e quindi dell'errore, per questo motivo la procedura è anche nota con il nome di *gradient descent*. Invece, il termine *backpropagation* (retropropagazione) fa riferimento alle regole di derivazione utilizzate per la minimizzazione della *loss function*.

La minimizzazione della funzione di perdita avviene attraverso il calcolo di due derivate distinte che consentono di ottenere rispettivamente le espressioni di aggiornamento dei pesi di *hidden-output* w_{jk} e di *input-hidden* w_{ij}.

Nel seguito si riportano i passaggi relativi al calcolo delle due derivate[3]:

- Derivata *hidden-output* w_{jk}:

$$\frac{\partial J}{\partial w_{jk}} = \frac{\partial}{\partial w_{jk}}\left(\frac{1}{2}\sum_{c=1}^{s}(t_c - z_c)^2\right) = (t_k - z_k)\frac{\partial(-z_k)}{\partial w_{jk}} =$$

$$= (t_k - z_k)\frac{\partial(-f(neuron_k))}{\partial w_{jk}} = -(t_k - z_k)\cdot\frac{f(neuron_k)}{\partial neuron_k}\cdot\frac{\partial neuron_k}{\partial w_{jk}} =$$

$$= -(t_k - z_k)\cdot f'(neuron_k)\cdot\frac{\partial\sum_{s=1}^{n_h} w_{sk}\cdot y_s}{\partial w_{jk}} = -(t_k - z_k)\cdot f'(neuron_k)\cdot y_j$$

ponendo $$\delta_k = (t_k - z_k)\cdot f'(neuron_k)$$

allora $$\frac{\partial J}{\partial w_{jk}} = -\delta_k \cdot y_j$$

quindi l'*hidden-output* w_{jk} sarà così aggiornato:

$$w_{jk} = w_{jk} + \eta\cdot\delta_k\cdot y_j = w_{jk} + \Delta w_{jk} \tag{3.20}$$

- Derivata *input-hidden* w_{ij}:

$$\frac{\partial J}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}}\left(\frac{1}{2}\sum_{c=1}^{s}(t_c - z_c)^2\right) = -\sum_{c=1}^{s}(t_c - z_c)\cdot\frac{\partial z_c}{\partial w_{ij}} =$$

$$= -\sum_{c=1}^{s}(t_c - z_c)\cdot\frac{\partial z_c}{\partial neuron_c}\cdot\frac{\partial neuron_c}{\partial w_{ij}} =$$

$$= -\sum_{c=1}^{s}(t_c - z_c)\cdot f'(neuron_c)\cdot\frac{\partial neuron_c}{\partial w_{ij}}$$

$$\frac{\partial neuron_c}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}}\sum_{r=1}^{n_h} w_{rc}\cdot y_r = \frac{\partial}{\partial w_{ij}}\sum_{r=1}^{n_h} w_{rc}\cdot f(neuron_r) =$$

$$= \frac{\partial}{\partial w_{ij}}\left(w_{jc}\cdot f\left(neuron_j\right)\right) = w_{jc}\frac{\partial f\left(neuron_j\right)}{\partial neuron_j}\cdot\frac{\partial neuron_j}{\partial w_{ij}} =$$

[3]Calcoli tratti da:
http://bias.csr.unibo.it/maltoni/ml/DispensePDF/8_ML_RetiNeurali.pdf.

$$= w_{jc} \cdot f'(neuron_j) \cdot \frac{\partial}{\partial w_{ij}} \sum_{q=1}^{d} w_{qj} \cdot x_q = w_{jc} \cdot f'(neuron_j) \cdot x_i$$

$$\frac{\partial J}{\partial w_{ij}} = -\sum_{c=1}^{s} \delta_c \cdot w_{jc} \cdot f'(neuron_j) \cdot x_i = -x_i \cdot f'(neuron_j) \sum_{c=1}^{s} w_{jc} \cdot \delta_c$$

ponendo $$\delta_j = f'(neuron_j) \cdot \sum_{c=1}^{s} w_{jc} \cdot \delta_c$$

allora $$\frac{\partial J}{\partial w_{ij}} = -\delta_j \cdot x_i$$

quindi l'*hidden-output* w_{jk} sarà così aggiornato:

$$w_{ij} = w_{ij} + \eta \cdot \delta_j \cdot x_i = w_{ij} + \Delta w_{ij} \tag{3.21}$$

Dove η è il *learning rate* o tasso di apprendimento, che determina l'entità dell'aggiornamento dei pesi ad ogni iterazione e quindi la velocità della convergenza alla soluzione attesa. Per valori troppo piccoli di tale parametro la convergenza sarà lenta, mentre per valori troppo elevati si rischia di ottenere oscillazioni o divergenze.

In generale, è possibile riassumere il processo di addestramento di una rete neurale in 6 steps:

- Step 1: inizializzazione dei pesi
  La fase di inizializzazione dei pesi è necessaria affinché il modello possa iniziare ad elaborare le prime risposte. Successivamente, si passano al modello le coppie di input, formate da dato in ingresso e output desiderato, il quale elabora la propria risposta.
- Step 2: calcolo dell'errore
  Una volta che il modello ha calcolato il proprio output, si calcola l'errore della previsione come la differenza tra output della rete e l'output desiderato.
- Step 3: calcolo dell'aggiornamento dei pesi
  Il calcolo dell'aggiornamento dei pesi consiste nella definizione del parametro Δw_{ij}. Tale valore può essere quantificato utilizzando un qualsiasi algoritmo di apprendimento, come ad esempio l'*error backpropagation.*

- Step 4: modifica dei pesi per la riduzione dell'errore previsionale

  Dopo aver calcolato l'entità dei vari aggiornamenti, i pesi sono finalmente aggiornati utilizzando, a seconda delle necessità computazionali, uno dei tre metodi seguenti: SGD*, batch* o *mini-batch*.

  La sigla SGD sta per *stochastic gradient descent,* in questo metodo l'errore è calcolato per ogni dato di training e i pesi sono aggiornati immediatamente, ciò significa che il metodo SGD modificherà i pesi tante volte quanti sono i training data. Solitamente, il comportamento del processo di addestramento del metodo SGD è randomico e l'iterazione di ciascun elemento del *dataset* provoca forti irregolarità nell'andamento della *loss function*.

  Nel metodo *batch*, invece, gli errori sono calcolati per tutti i dati di training e successivamente si calcolano gli aggiornamenti dei pesi. Una volta completato tale calcolo, si utilizza la media degli aggiornamenti per modificare i pesi. Data la natura del procedimento per il calcolo del valore Δw_{ij} , l'addestramento della rete impiega molto tempo.

  Il valor medio degli aggiornamenti Δw_{ij} è così espresso:

$$\Delta w_{ij} = \frac{1}{N} \sum_{k=1}^{N} \Delta w_{ij}(k) \tag{3.22}$$

  Dove $\Delta w_{ij}(k)$ è l'aggiornamento del *k-esimo* dato di training ed N è il numero totale dei training data.

  Il metodo *mini-batch* nasce dalla combinazione dei due metodi precedenti in modo da ottenere la velocità di calcolo del SGD e la stabilità del *batch*. Nel *mini-batch* i dati di training sono divisi in più gruppi e la rete è addestrata utilizzando per ogni gruppo selezionato il metodo *batch*.
- Step 5: ripetizione dello step 2 e dello step 3 per ogni dato utilizzato per l'addestramento della rete.
- Step 6: ripetizione dallo step 2 allo step 5 fino a quando l'errore raggiunge un livello accettabile.

Il processamento di tutti i dati di training dallo step 2 allo step 5 è chiamato *epoch*. In altre parole, si definisce *epoch* quanto un intero *dataset* è passato attraverso la rete neurale una sola volta e quindi quante volte si ripete lo step 6. Come si è visto, a seconda del metodo utilizzato è possibile suddividere i dati di input in *batches* più piccoli, il numero dei *batches* necessari al completamento di una *epoch* è chiamato iterazione.

3.7 Loss functions

Le funzioni di perdita sono impiegate nel processo di ottimizzazione dei parametri della rete neurale e il loro obiettivo è la minimizzazione della perdita del modello tramite l'aggiornamento dei pesi. Per calcolare l'errore del modello durante il processo di ottimizzazione si fa ricorso alle *loss functions* che esprimono una misura della distanza tra la risposta della rete e l'output desiderato.

In generale è possibile affermare che le funzioni di perdita giochino un ruolo importante nella definizione dei parametri del modello e, in un certo senso, riducono tutti gli aspetti positivi e negativi di un sistema complesso in un unico valore scalare, che consente di classificare e comparare la soluzione considerata.

3.7.1 Mean Square Error

Nella spiegazione dell'algoritmo di *backpropagation* è stata introdotta la funzione di perdita considerata come la funzione di default per i problemi di regressione, ossia la *Mean Square Error*. Come suggeriscono il nome e la (3.23), essa è calcolata come la media delle differenze quadratiche tra il valore predetto e l'output desiderato.

$$J(w,x) = \frac{1}{2}\sum_{c=1}^{s}(t_c - z_c)^2 \tag{3.23}$$

Per costruzione, il risultato è sempre positivo, indipendentemente da segno della risposta della rete e il valore limite 0 rappresenta la situazione in cui il modello è in grado di spiegare perfettamente la realtà. L'elevazione al quadrato dà maggior peso agli errori più grandi in valore assoluto rispetto a quelli più piccoli penalizzando le previsioni che si discostano maggiormente dall'output desiderato.

3.7.2 Cross-entropy e one-hot targets

Le funzioni *cross-entropy* sono solitamente utilizzate nei problemi di classificazione multi-classe per i quali ad ogni classe è assegnato un unico numero intero. Il valore di *cross-entropy* tra la distribuzione discreta p e la distribuzione q che determina la distanza di q da p è così calcolata:

$$H(p,q) = -\sum_{v} p(v) \cdot \log\left(q(v)\right) \tag{3.24}$$

Anche in questo caso il risultato è sempre positivo e il valore limite pari a 0 si raggiunge quando la risposta della rete coincide con l'output desiderato. Inoltre, è importante notare che la quantità $q(v)$ essendo l'output della rete, sarà stata processata da una funzione di attivazione e, nel caso della sigmoide, sarà necessariamente diversa da 0 anche se vi potrà tendere asintoticamente, evitando in questo modo questioni relative alla non esistenza della funzione logaritmica.

Una casistica particolare della (3.24), che ha riscontrato enorme successo nella struttura delle reti è la corrispondente forma binaria anche nota come *one-hot target*. Se i valori target assumono la forma {0,1} allora la *loss function* diventa:

$$J(w,x) = H(t,z) = -\log\left(z_g\right) \tag{3.25}$$

Dove z_g è il valore della risposta della rete ottenuta in seguito all'applicazione di una generica funzione di attivazione.

Per sopperire a particolari esigenze computazionali di alcune tipologie di rete, recentemente sono state introdotte altre funzioni di perdita, tuttavia le più comunemente usate rimangono la *Mean Square Error* e la *cross-entropy* in tutte le sue forme.

In questo capitolo si sono descritte le principali strutture e funzioni che definiscono il funzionamento dei vari modelli. Come illustrato nei vari paragrafi, le reti neurali sono completamente personalizzabili sulla base delle esigenze richieste dal modello. La scelta di ogni parametro spetta al creatore della rete che, conoscendo a fondo i fini del modello, decide quali strutture e funzioni sono più adatte a rappresentare la situazione che si vuole descrivere. L'importanza di queste scelte non deve essere sottovalutata in quanto ogni caratteristica della rete influenza in modo diretto la performance del modello. Pertanto, per costruire un modello performante è necessario conoscere tutte le possibili architetture e funzioni in modo da scegliere quelle che più si addicono al modello previsionale.

Capitolo IV

Il campione in analisi

La scelta del settore di riferimento per la costruzione del modello di *credit scoring* è ricaduta sul noleggio e leasing di autoveicoli, individuato dal codice ATECO 771. Le principali attività svolte dalle aziende analizzate riguardano la locazione dei veicoli, la gestione delle flotte aziendali e il car sharing. In questo capitolo, si effettuerà una breve panoramica del settore riguardante i principali andamenti che hanno caratterizzato il mondo del noleggio nel periodo di riferimento, ossia il quinquennio 2014 – 2018. Infine, si illustreranno nel dettaglio le operazioni di pulizia dei dati effettuate per ottenere il campione di analisi, sulla base del quale si è costruito il modello.

4.1 L'evoluzione del noleggio in Italia

Il settore, secondo quanto riportato nel XVIII rapporto sul noleggio veicoli dall' Associazione Nazionale Industria dell'Autonoleggio e Servizi Automobilistici (ANIASA), segmenta le attività in base alla durata, distinguendo tra quelle di breve periodo, che coprono un periodo massimo di 1 mese, da quelle di medio e lungo periodo, che raggiungono i 60 mesi.

Nel 2018, il mercato dell'auto, ha subito un'interruzione di una crescita durata 4 anni. Lo stop è stato causato da differenti motivi riguardanti i nuovi sistemi di omologazione, la riduzione delle pratiche dei chilometri zero, l'interruzione dei benefici derivanti dalle politiche di superammortamento e principalmente la frenata dell'economia italiana. La crescita del PIL, che aveva caratterizzato il 2016 e il 2017 è rallentata fino a far registrare negli ultimi due trimestri del 2018 una leggera recessione. Questa frenata, accompagnata da un periodo di forte indecisione politica ha causato una diminuzione delle attività produttive e di conseguenza del mercato dell'auto.

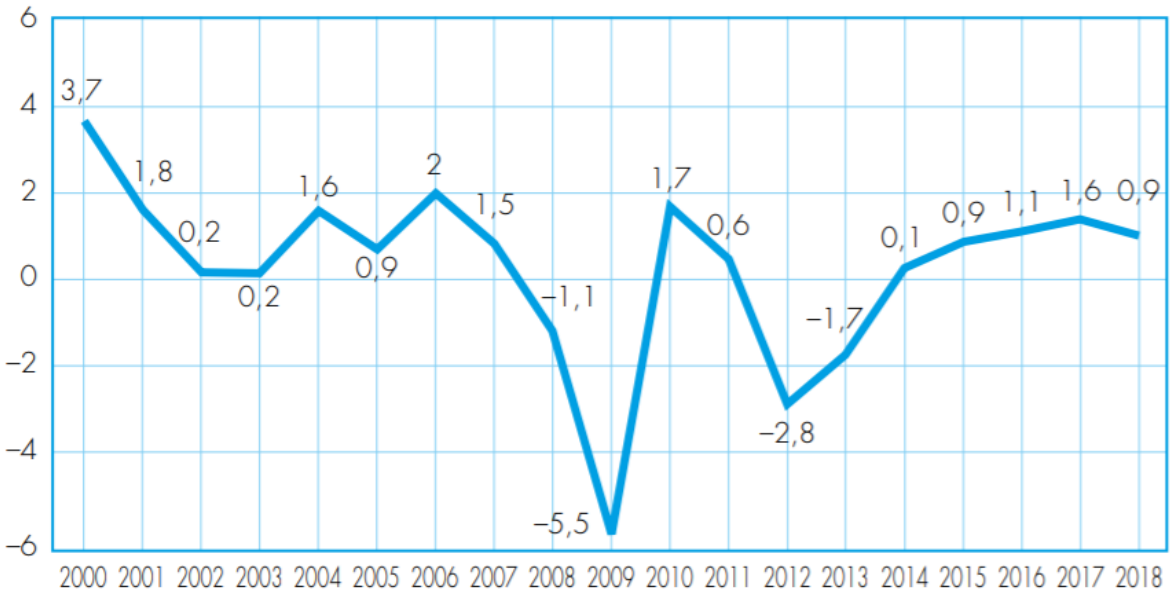

Figura 4.1 *Andamento del PIL in Italia dal 2000 al 2018 in termini di variazione percentuale*

La fragilità dell'economia italiana si è riversata nella produzione delle auto, il numero delle immatricolazioni è infatti calato rispetto al 2017, ciò è dovuto in parte all'andamento irregolare degli indici di fiducia delle imprese e delle famiglie ma soprattutto alla riduzione dello stock di auto da parte delle case automobilistiche.

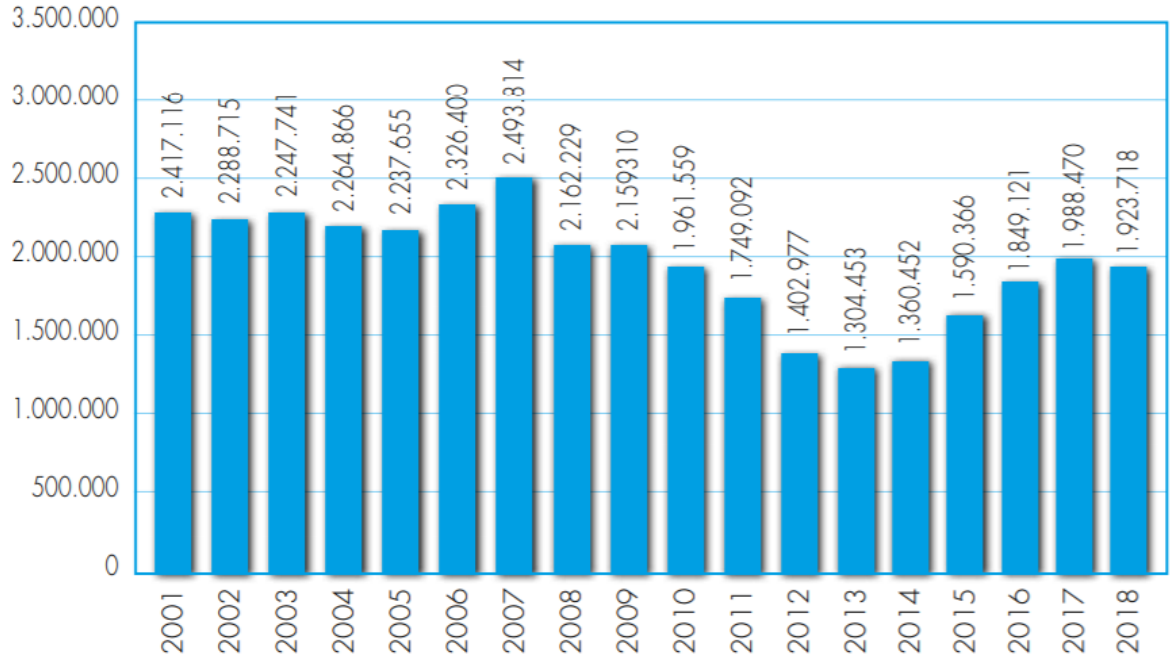

Figura 4.2 *Trend immatricolato vetture in Italia dal 2001 al 2018*

Anche il valore complessivo del mercato delle auto in crescita dal 2013 ha subito una leggera diminuzione rimanendo però al di sopra dei 38 miliardi di euro. Tale riduzione, in parte dovuta ai maggiori sconti applicati sugli acquisti delle auto, si è riscontrata anche nei volumi che hanno fatto segnare un -3% rispetto al 2017.

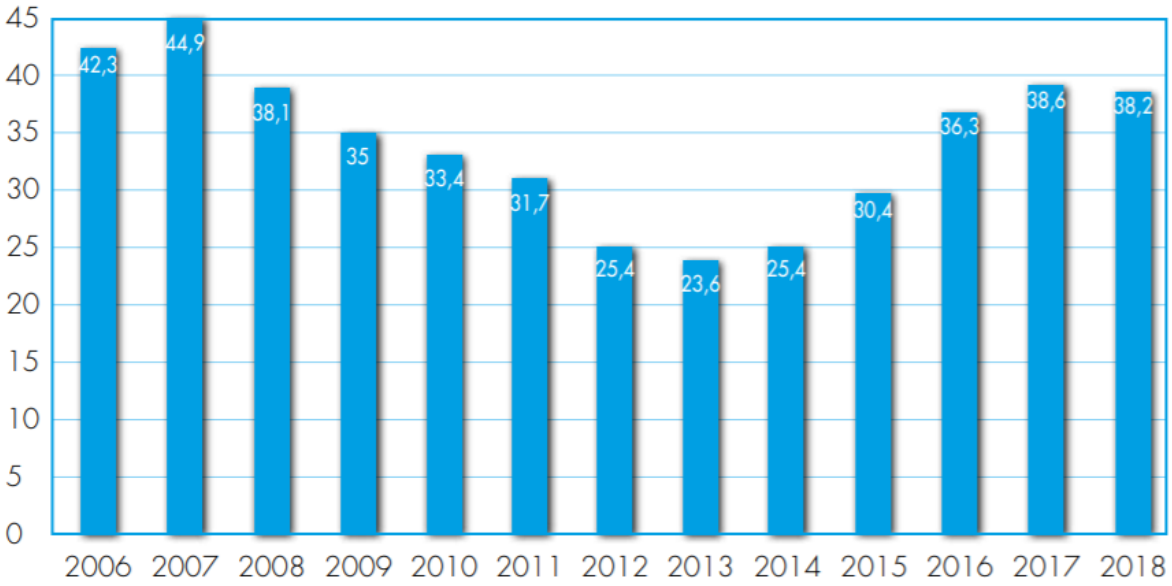

Figura 4.3 *Mercato auto nuove dal 2006 al 2018 in Italia. Valore al netto di sconti e incentivi. Dati in miliardi di euro*

Nel 2018 il settore delle auto è stato trainato dall'aumento del *rent-a-car*, cresciuto principalmente attraverso la leva del prezzo, infatti la riduzione dei prezzi è stata di incentivo per aumentare i volumi di vendita. In generale, con il trascorrere degli anni la riduzione del prezzo giornaliero del noleggio è stata accompagnata da un aumento dei volumi, della durata media e del chilometraggio.

La prima tipologia di aziende, caratterizzate da un noleggio a breve termine, ha confermato il moderno trend del *low-cost*. Questo segmento ha fatto registrare un primato storico nella stagione estiva con 179.000 unità impiegate con un aumento del 4,7% e del 3,2% rispettivamente dei giorni e dei numeri di noleggi. Al contrario, si sono verificati una riduzione del 3,3% dell'utilizzo medio della flotta e delle diminuzioni in termini di fatturato per noleggio (-1,4%) e per giorno di noleggio (-2,8%).

Il calo del fatturato ha messo in luce la questione della riduzione del valore dei margini. Per risolvere questo problema, negli ultimi anni, si è assistito ad una maggiore attenzione da parte delle aziende alla proposta di servizi aggiuntivi innovativi da affiancare a quelli essenziali. Le varie aziende, tramite la ricerca e l'innovazione, perseguono una diversificazione atta ad aumentare la qualità e ridurre la concorrenza ormai sempre più dura in questo settore.

Altro trend per le aziende di noleggio a breve termine è l'*e-commerce*, la sua rapida crescita ha attirato l'attenzione di numerosi *broker online* che hanno investito sull'intermediazione del noleggio. Questo trend ha però inasprito la concorrenza e reso ancor più difficile la differenziazione di prodotto, spostata sempre di più verso attività *aftermarket* quali la gestione del credito.

Le necessità di diversificazione nascono dalla particolarità della domanda che caratterizza il settore. Essa risulta difficilmente generalizzabile, per poter soddisfare le esigenze individuali dei singoli consumatori è necessario proporre offerte flessibili ed il più personalizzabili possibile. Ultimamente le aziende si stanno cimentando nella sfida del *pay-per-use*, ossia a soddisfare una domanda caratterizzata da un elevato numero di clienti occasionali, servizi flessibili e di alta qualità e sempre più tecnologica e vicina al cliente tramite app, call center e siti web.

Nel XVIII rapporto ANIASA, delinea, grazie agli studi derivati dalla collaborazione con la società di consulenza Bain & Company, il profilo del tipico cliente del noleggio come un uomo, sempre più giovane, più informato e maggiormente disposto a effettuare scelte più ecologiche, che decide di noleggiare l'auto soprattutto per ragioni lavorative.

La seconda tipologia di imprese presenti nel settore, si occupa del noleggio a medio e lungo termine. Tali aziende hanno confermato la loro crescita, ormai continua da otto anni, chiudendo il 2018 con una flotta di oltre 900.000 unità ed un fatturato di circa 5,5 miliardi di euro. Per questa classe di aziende, i clienti rivelazione degli ultimi anni si confermano i privati, quest'ultimi hanno siglato 25.000 contratti nel 2017, per poi raggiungere i 40.000 nel 2018, facendo presagire il superamento di 50.000 contratti nel 2019.

4.1.1 Analisi del fatturato

Per poter comprendere al meglio quali siano le variabili che guidano il settore è necessario presentare una breve analisi delle componenti del fatturato che hanno consentito al settore di raggiungere nel 2018 un valore pari a 1,229 miliardi di euro. Questo risultato, che rappresenta un record per il settore, deriva per il 67% dagli elementi essenziali quali la durata del noleggio, la distanza percorsa e gli oneri automobilistici, aeroportuali e ferroviari. L'altra grossa fetta (28%) proviene dai servizi aggiuntivi richiesti dai clienti. In questa categoria rientrano oltre ai classici optional del veicolo come il navigatore, il parcheggio assistito, il controllo automatico della velocità e il sistema anticollisione, anche le coperture assicurative per furto o incidenti. È importante notare come quest'ultima componente sia in constante aumento negli anni, segnalando un +3% rispetto al 2017 ma più in generale un trend di maggior focus sulla proposizione di servizi. Il 5% rimanente è classificato nella voce "altro fatturato" e solitamente include delle tasse legate alla gestione di eventi straordinari quali le multe, incidenti, furti, il mancato ritiro del veicolo già prenotato e il servizio di rifornimento carburante.

Tabella 4.1 *Distribuzione del fatturato per tipologia di servizio*

	2018	2017	Var. %
Fatturato totale	1.228.618.834	1.207.347.196	1,8%
- fatturato di base	823.662.852	844.585.389	−2,5%
- fatturato su servizi accessori	339.132.963	295.930.416	14,6%
- altro fatturato	65.823.020	66.831.391	−1,5%

Dividendo il fatturato rispettivamente per i giorni di noleggio, il numero di noleggi e per la flotta media si ottengono quelli che possono essere considerati come i tre indicatori guida di questo settore ovvero il fatturato per giorno, il fatturato per noleggio e il fatturato per veicolo. L'andamento del fatturato per giorno è calato del 2,8% rispetto al 2017 passando da 34,8 a 33,8 milioni, confermando il trend precedentemente anticipato del ricorso alla leva del prezzo per catturare maggiore clientela ed aumentare i volumi di vendita. Gli effetti della diminuzione delle tariffe si sono riscontrati anche nell'allungamento della durata media da 6,7 a 6,8 giorni (+1,5%) e dalla conseguente riduzione del fatturato per noleggio che dai 232 euro del 2017 ha raggiunto i 229. Infine, il 2018 ha visto decrescere anche l'ultimo indicatore con una flotta media capace di generare solamente 9.500 euro, 600 in meno rispetto al 2018. Nonostante un peggioramento generale dei valori assunti dai tre indicatori, il fatturato totale

è risultato in crescita. Tuttavia, gli indicatori suggeriscono che questo miglioramento non è corrisposto ad un aumento della capacità del settore di generare maggior valore quanto piuttosto al ricorso della leva del prezzo e quindi ad una maggiore accessibilità alle offerte.

4.1.2 Prospettive future

Come si è potuto notare, i numeri del 2018, ma più in generale l'andamento del quinquennio 2014 – 2018, hanno mostrato l'interruzione di una crescita che sembrava destinata a protrarsi per almeno altri tre anni. All'inizio brillante del 2018, il noleggio a lungo termine è entrato in difficoltà con un settembre che ha fatto registrare un -35,5% rispetto all'anno precedente, concludendo l'anno con un +1,3% contrariamente alle crescite a doppia cifra che avevano caratterizzato gli ultimi anni. Il rallentamento dell'ultimo quadrimestre, secondo quanto affermano gli analisti, sarebbe dovuto ad un clima di generale incertezza sul futuro causato anche dei decreti dalla Legge di Bilancio del 2019 che lasciano intuire una possibile penalizzazione dei clienti tramite un nuovo sistema di assegnazione bonus-malus.

Il 2019 è iniziato sulla stessa linea di come si era concluso il 2018, con il noleggio a lungo termine che nel febbraio ha segnalato 4.073 richieste in meno, ossia una diminuzione del 14,69% rispetto all'anno precedente, mentre per il breve termine un passivo di 4.673 veicoli, vale a dire il 17,8% in meno. La contrazione della domanda nei primi due mesi dell'anno ha provocato una riduzione del numero di immatricolazioni e quindi di vetture nuove consegnate, -7,5% e -2,3% rispetto a gennaio e febbraio 2018.

Dai dati del 2018 si presagisce che i prossimi anni potrebbero rappresentare la svolta per questo settore. Le forti incertezze non aiutano a comprendere quali possano essere i possibili risvolti per le aziende interessate. Per queste ragioni risulta interessante sviluppare un modello di rischio di credito in grado di valutare se un'azienda è sana o meno sulla base della sua situazione patrimoniale, in modo da prevedere quali imprese saranno in grado di affrontare questo periodo di cambiamento.

4.2 La banca dati AIDA

La ricerca delle informazioni riguardanti le aziende del settore del noleggio e leasing di autoveicoli è avvenuta in AIDA, la banca dati di Bureau Van Dijk, azienda di Moody's Analytics. La Bureau Van Dijk è un'azienda leader del settore IT che supporta i propri clienti offrendo prodotti e servizi di qualità grazie alle collaborazioni con i più rinomati Information Provider di tutto il mondo, tra i quali: Fitch, Standard & Poor's, Moody's, Reuters, Economist Intelligence Unit, Capital Intelligence, Jordans, Creditreform…

La banca dati AIDA, acronimo che sta per Analisi Informatizzata Delle Aziende, contiene le informazioni di tutte le società di capitali attive o fallite operanti nel territorio italiano. Le informazioni in costante aggiornamento riguardano più di un milione di imprese in serie storica fino ad un massimo di 10 anni. Il database contiene informazioni anagrafiche e finanziarie, il settore economico e il codice di attività nazionale (ATECO) ed internazionale (NACE, NAICS, SIC), la serie storica dei bilanci e i dati sull'azionariato, le partecipazioni e gli esponenti delle società.

4.3 Scarico dei dati

L'operazione di scarico dati è riuscita grazie all'abbonamento messo a disposizione dal Politecnico di Torino. Tramite la rete Wi-Fi dell'Università, sono stati scaricati i dati di 2.241 aziende contrassegnate dal codice ATECO 771 che individua il settore "noleggio e leasing di autoveicoli". Per poter effettuare una buona analisi si è deciso di consultare le informazioni relative al periodo 2014 – 2018 in modo da avere per ogni azienda un quinquennio di riferimento, salvo le società che hanno interrotto le loro attività in questo periodo per le quali naturalmente si sono considerati solo i dati disponibili.

Le informazioni estratte riguardano due macrocategorie: i dati anagrafici e i dati economico-finanziari. Nel primo gruppo assumono particolare rilevanza per il lavoro di tesi lo stato giuridico ("ditta attiva", "ditta in liquidazione", "ditta in fallimento", "ditta sospesa", "ditta inattiva", "ditta cessata", "ditta cessata per trasferimento") e l'eventuale procedura subita. Sulla base delle informazioni contenute in quest'ultimo campo si sono assegnati cinque diversi flag ciascuno corrispondente ad una condizione societaria:

- 0 = società senza particolari segnalazioni
  Procedura subita: nessuna.

- 1 = società anomale

  Procedura subita: concordato preventivo, fallimento, amministrazione giudiziaria, accordo di ristrutturazione dei debiti, chiusura del fallimento, altre cause, liquidazione giudiziaria, motivo non precisato, stato di insolvenza, sequestro giudiziario, concordato fallimentare, amministrazione controllata, cancellazione per comunicazione piano di riparto, amministrazione straordinaria, chiusura per fallimento o liquidazione, decreto cancellazione tribunale, liquidazione coatta amministrativa, scioglimento per atto dell'autorità, sequestro conservativo di quote, bancarotta.
- 2 = società sane in condizioni particolari

  Procedura subita: liquidazione volontaria, scioglimento e liquidazione, scioglimento, chiusura della liquidazione, chiusura dell'unità locale, cessazione di ogni attività, cancellata d'ufficio ai sensi art. 2.490 c.c. (bilancio di liquidazione), liquidazione, scioglimento e messa in liquidazione, chiusura per liquidazione, scioglimento senza messa in liquidazione, cessazione delle attività nella provincia, cessazione d'ufficio.
- 3 = società sane in condizioni particolari

  Procedura subita: fusione mediante incorporazione in altra società, scissione, trasferimento sede all' estero, fusione mediante costituzione di nuova società, cessione azienda.
- 4 = società sana in condizioni particolari

  Procedura subita: cessata, cancellata dal registro impresa, trasferimento in altra provincia, cancellata d'ufficio a seguito istituzione cciaa di fermo, di monza...

Sulla base della precedente classificazione si sono individuati tre diversi flag capaci di sintetizzare lo status della società senza dover ricorrere ad analisi più approfondite. I flag in questione sono:

- FLAG DI STATUS S/A – SOCIETA': indica se la società è sana o anomala
  = 0 se la società è sana la procedura rientra nei casi 0, 3 o 4
  = 1 se la società è anomala e la procedura rientra nei casi 1, 2
- FLAG DI STATUS S/A – ANNO: individua l'anno in cui è avvenuta procedura
  = 0 se la società non ha subito la procedura nell'anno in considerazione
  = 1 se la società ha subito la procedura nell'anno in considerazione

- FLAG soc sana in liquidazione – SOC: precisa l'anomalia della società
  = 0 se la società è sana e la procedura rientra nei casi 0, 3 o 4
  = 1 se la società è anomala e la procedura rientra nel caso 1
  = 2 se la società è anomala e la procedura rientra nel caso 2

Nel medesimo file excel si sono aggregate le voci di stato patrimoniale e conto economico per ottenere la versione di bilancio abbreviato. Partendo da queste voci, tramite apposite formule, si sono calcolati 42 indicatori suddivisi in tre macro-categorie:

- Redditività
  Valutano la capacità di un'impresa di produrre reddito e generare risorse nel corso degli anni.
- Produttività e struttura operativa
  Stabiliscono se i fattori produttivi sono impiegati in modo più o meno efficiente dall'azienda.
- Liquidità e struttura finanziaria
  Esprimono la capacità dell'impresa di onorare in modo tempestivo gli impegni assunti analizzando i legami temporali tra le attività e le fonti di finanziamento.

4.4 Pulizia dei dati e colonne di controllo

Dopo aver riordinato i dati, assegnato i flag e calcolato i principali indicatori si è proceduto con un'operazione di pulizia dei dati o *data cleaning*. In generale, quando si opera con database contenenti una grande mole di dati, come nel caso di AIDA, è indispensabile assicurarsi che siano corretti ed il più completi possibile. Dunque, prima di procedere con le analisi statistiche, è stato necessario svolgere un accurato controllo della qualità dei dati correggendo e completando le informazioni laddove risultavano parziali o errate.

Per prima cosa si è deciso di non considerare le righe contenenti informazioni relative ad aziende con valore di attivo netto pari a zero, in quanto indice di bilanci inesistenti o non caricati in AIDA. Si sono poi convertiti tutti i valori *n.d.* in 0 e si sono corretti gli eventuali errori #DIV/0! e #NUM! in modo da sostituire gli elementi *non-machine-readable* con valori compatibili per le operazioni compiute dai software utilizzati.

Per poter svolgere i controlli sulla correttezza dei dati esportati da AIDA si è fatto uso di tre colonne di controllo che, tramite semplici operazioni di addizione e sottrazione, hanno

verificato la corrispondenza tra voci di attivo e passivo del bilancio. Infatti, è noto che nello stato patrimoniale, la somma delle attività debba necessariamente corrispondere alla somma delle passività. Andando più nel dettaglio, le tre colonne verificano le seguenti condizioni:

- Colonna X: composizione dell'attivo
  Verifica che sottraendo all'attivo netto, calcolato come somma di tutte le voci che lo compongono, il valore totale delle attività e aggiungendo i crediti verso soci e le azioni proprie il risultato sia nullo.
- Colonna AI: uguaglianza tra attivo netto e passivo netto
  Accerta l'uguaglianza tra attivo e passivo scomponendo i due valori nelle singole voci da cui sono composte.
- Colonna BS: controllo del conto economico
  Ripercorre il calcolo del risultato netto partendo dai ricavi e lo confronta con l'utile/perdita d'esercizio.

Per ogni bilancio disponibile si è controllato che tutte le tre colonne assumessero valori nulli e in caso contrario si sono ricercate le cause degli errori. È interessante notare come la maggior parte delle incongruenze fosse causata da omissioni di alcune voci parziali, la cui assenza non consentiva la ricostruzione delle voci aggregate e quindi il corretto bilanciamento dei valori. Come è possibile intuire, le operazioni di pulizia dei dati costituiscono un grande onere in termini di tempo impiegato e richiedono un grande impegno ma, ad ogni modo, rappresentano il punto di partenza per la costruzione di un modello statistico robusto.

4.5 Preparazione dei dati

Una volta conclusa la pulizia dei dati, è necessario predisporli affinché possano essere recepiti come input dal modello. Queste operazioni, così come le precedenti, risultano particolarmente delicate in quanto ogni mancanza o errore si ripercuote sul corretto funzionamento e sulle performance del modello predittivo. L'organizzazione e la predisposizione dei dati, soprattutto quando si opera con database di grandi dimensioni, consentono di ottenere il massimo beneficio dall'analisi statistica, infatti se si lavora con informazioni di pessima qualità difficilmente si riusciranno ad ottenere i risultati pianificati.

Per prima cosa, si sono calcolati i valori del 5° e del 95° percentile di tutti gli indicatori in modo da poter allineare i valori degli *outliers* a tali soglie. Con il termine *outliers*, si fa riferimento ad un insieme di osservazioni che assumono valori estremi rispetto agli altri dati disponibili, per questa ragione sono da considerarsi anomale e devono essere allineate con i valori assunti dalle altre osservazioni. Per tutti i casi in cui gli indicatori presentavano valori nulli o negativi a denominatore, l'allineamento è avvenuto sulla base del loro significato economico calcolando la regressione dell'indicatore con il FLAG DI STATUS S/A – ANNO, in modo da individuarne il verso, ossia se all'aumentare del valore assunto dall'indicatore corrispondesse una migliore o peggiore performance aziendale.

La preparazione dei dati è proseguita con il calcolo delle matrici di correlazione. Nella prima matrice si sono riportate le correlazioni tra gli indicatori in modo da valutare il grado di sovrapposizione dei segnali. Nella seconda invece, si sono calcolate le correlazioni tra i singoli indicatori ed i FLAG DI STATUS S/A – SOCIETA'. La predisposizione di tali matrici rappresenta il primo di una serie di step atti ad individuare le variabili da includere nel modello. Questo procedimento prende il nome di *feature selection* e consiste nell'individuare le caratteristiche rilevanti che influenzano i dati e di conseguenza la performance del modello. Lo scopo finale è la scelta ottima delle variabili e quindi la selezione del minor numero di indicatori che meglio descrivono il problema in considerazione. La *feature selection* prevede dunque che gli indicatori fortemente correlati tra loro, e quindi in grado di spiegare gli stessi comportamenti, siano scartati per dar spazio ad altri valori di input che apportano informazioni aggiuntive migliorando l'accuratezza del modello predittivo. La scelta delle variabili rilevanti rientra nella procedura di preparazione dei dati ma rimane comunque un processo ripetuto nel tempo sulla base di un confronto continuo con le performance del modello. L'algoritmo di *feature selection* inizia con l'individuazione nella matrice di correlazione tra indicatori e flag, degli indicatori maggiormente correlati con il flag. In seguito, si prosegue con l'eliminazione progressiva degli indicatori che risultano molto correlati tra loro per evitare problemi di collinearità imperfetta. Il processo si conclude con la scelta degli indicatori per poi ripetersi al termine della fase di calcolo per cercare di migliorare la qualità dei risultati ottenuti.

Infine, in vista dell'alimentazione al modello finale si sono riportati i dati anagrafici, i flag e gli indicatori di ogni osservazione. In particolare, si è ridotto il campo di esistenza degli indicatori ai valori compresi tra 0 ed 1 estremi inclusi. Detto x_i il dato in considerazione, dove il pedice i rappresenta il numero dell'osservazione, la normalizzazione si ottiene attraverso l'applicazione della seguente formula:

$$Norm(x_i) = \frac{x_i - \min_i\{x_i\}}{\max_i\{x_i\} - \min_i\{x_i\}} \tag{4.1}$$

Con il quarto capitolo si è voluto fornire tramite una breve analisi del settore di riferimento, un quadro generale dell'andamento economico e dei fattori chiave che guidano il mondo del leasing e noleggio di autoveicoli. In seguito, si sono illustrati il database AIDA impiegato per l'estrazione delle imprese oggetto di studio e le relative operazioni di pulizia e predisposizione dei dati necessarie per la buona riuscita del modello. In sostanza, questo capitolo costituisce il punto di partenza per la costruzione del modello che proseguirà nel prossimo con un'analisi più tecnica del codice implementato.

Capitolo V

La costruzione della rete neurale

La scelta del software per la costruzione della rete neurale è ricaduta su MATLAB. Il Politecnico di Torino ha reso possibile non solo il download gratuito della sua versione più recente ma anche di alcuni *tools* aggiuntivi che hanno facilitato la scrittura del programma. Il *Matrix Laboratory*, più comunemente noto come MATLAB, è un ambiente scritto in C utilizzato principalmente per il calcolo numerico e l'analisi statistica. La particolarità che ne rende il suo uso così diffuso è la possibilità di utilizzare molteplici strumenti, chiamati *tools* o applicazioni, che eseguono *task* di calcolo applicabili ad un'ampia varietà di studi.

L'applicazione impiegata per questo lavoro è il *Deep Learning Toolbox*, precedentemente conosciuto come *Neural Network Toolbox*. Tale strumento fornisce un *framework* che assiste gli utenti nella costruzione di reti neurali con parametri standard ed è proprio la personalizzazione dei parametri della rete a costituire il vero limite di questo *tool*. Per tali ragioni, nel presente lavoro di tesi, non si è fatto uso diretto dell'applicazione ma, attraverso la scrittura di righe di codice, si sono richiamate le sue principali funzioni in modo da personalizzare i parametri e le logiche di funzionamento laddove possibile. In particolare, in questo capitolo si illustreranno passo dopo passo le righe di codice che definiscono il flusso di esecuzione del programma, spiegando le ragioni del loro impiego e il significato che sta alla base.

5.1 Importazione dei dati

```
%% Import data
%Flag
Flag = readtable('Flag.xlsx');
Flag = Flag{:,:};
%Indicatori
Indicatori = readtable('Indicatori.xlsx');
Indicatori = Indicatori{:,:};
```

Il punto di partenza per la costruzione di un modello è l'importazione dei dati. Logicamente, nessuna attività può proseguire se prima non si sono definiti i dati sui quali si opererà. L'importazione dei dati consiste nel loro caricamento da fonti esterne, nel caso in esame tali fonti sono costituite da fogli excel nei quali, come illustrato in precedenza, si sono effettuate le opportune operazioni di pulizia.

I dati in questione sono suddivisi in tre tipologie:

- Dati anagrafici
  Contengono le informazioni anagrafiche di tutte le aziende in esame incluse le eventuali procedure subite e le date ad esse associate.
- Flag
  Si tratta del FLAG DI STATUS S/A – SOCIETA' ossia la variabile di risposta o variabile dipendente del modello.
- Indicatori
  Sono i valori che sintetizzano le informazioni di bilancio e conto economico normalizzati tra 0 ed 1, ovvero le variabili indipendenti attraverso le quali si vuole spiegare il comportamento della variabile FLAG DI STATUS S/A – SOCIETA'.

Procedendo con ordine, i primi dati ad essere importati sono i dati anagrafici. Per questi, si è deciso di utilizzare uno *script* a parte per gestirne la complessità. I dati anagrafici comprendono diversi formati di celle (stringhe, numeri, date…) e per preservare ognuno di essi è stato necessario importare il file manualmente e generare lo *script* ad esso associato, in modo da richiamare automaticamente il file di origine ad ogni generazione della rete.

```
%% Import data from spreadsheet
% Script for importing data from the following spreadsheet:
%
%    Workbook: C:\Users\Fabio Palmieri\Desktop\Dati
aziende\Dati anagrafici.xlsx
%    Worksheet: Foglio1

% Setup the Import Options
opts = spreadsheetImportOptions("NumVariables", 22);
```

```
% Specify sheet and range
opts.Sheet = "Foglio1";
opts.DataRange = "A2:V6939";

% Specify column names and types
opts.VariableNames = ["Anno", "nosservazione", "Num",
"RagioneSociale", "AnnoCostituzione", "Comune", "Provincia",
"Regione", "CodiceISTATRegione", "CodiceISTATProvincia",
"CodiceISTATComune", "FormaGiuridica", "CodiceFiscale",
"DescrizioneConsolidata", "IFRS",
"Modellodicontabilitbilancio", "ATECO2007Codice",
"StatoGiuridico", "Mercato", "ProceduraCessazione",
"Datadiinizioproceduracessazione",
"Datadichiusuradellaprocedura"];
opts.SelectedVariableNames = ["Anno", "nosservazione", "Num",
"RagioneSociale", "AnnoCostituzione", "Comune", "Provincia",
"Regione", "CodiceISTATRegione", "CodiceISTATProvincia",
"CodiceISTATComune", "FormaGiuridica", "CodiceFiscale",
"DescrizioneConsolidata", "IFRS",
"Modellodicontabilitbilancio", "ATECO2007Codice",
"StatoGiuridico", "Mercato", "ProceduraCessazione",
"Datadiinizioproceduracessazione",
"Datadichiusuradellaprocedura"];
opts.VariableTypes = ["string", "string", "string", "string",
"string", "string", "string", "string", "string", "string",
"string", "string", "string", "string", "string", "string",
"string", "string", "string", "string", "string", "string"];
opts = setvaropts(opts, [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,
12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22],
"WhitespaceRule", "preserve");
opts = setvaropts(opts, [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,
12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22],
"EmptyFieldRule", "auto");
```

```
% Import the data
Datianagrafici = readtable("C:\Users\Fabio
Palmieri\Desktop\Dati aziende\Dati anagrafici.xlsx", opts,
"UseExcel", false);

% Clear temporary variables
clear opts
```

Per quanto riguarda il flag e gli indicatori, essendo costituiti unicamente da celle numeriche, è stato sufficiente utilizzare la funzione `readtable` fornendole in input il nome del file da richiamare. Successivamente, si sono assegnati i valori alle rispettive variabili creando delle matrici contenenti i dati di tutte le righe e le colonne dei file excel richiamati.

Importando i flag e gli indicatori, si sono definiti indirettamente il numero dei neuroni degli strati di input e di output. Come precedentemente spiegato, i neuroni di input hanno il compito di trasmettere il segnale senza modificarlo, per questo motivo ad ogni indicatore corrisponderà un neurone di input che conserverà i valori di quel particolare indicatore per tutte le aziende osservate. Lo stesso ragionamento vale per il neurone di output, unico in quanto ad esso è associato un solo flag.

5.2 Struttura della rete e suddivisione dei dati nei set

```
%% Define the architecture
net = patternnet([12 13 11]);
```

Fissati il primo e l'ultimo strato non resta che definire la struttura intermedia della rete ovvero gli *hidden layers*. La funzione `patternnet` riceve in ingresso un vettore riga avente dimensione pari al numero di strati nascosti e valori che indicano la quantità di *hidden neurons* per ogni strato.

```
%% Set up division of data
net.divideFcn = 'divideind';
net.divideParam.trainInd = 1:4626;
net.divideParam.testInd = 4627:5782;
net.divideParam.valInd = 5783:6938;

% Max number of failures
net.trainParam.max_fail = 15;
```

Arrivati a questo punto è necessario definire il metodo con il quale si divideranno i dati nei tre set di studio. Il criterio applicato di default da MATLAB è l'assegnazione casuale, come si può facilmente intuire i risultati ottenuti con questo metodo non sono costanti, quindi la performance risultante è del tutto casuale rendendo difficile qualsiasi tipo di confronto tra risultati di strutture diverse. Per rimuovere questo elemento di casualità e per un confronto omogeneo tra i risultati si è utilizzata la divisione `'divideind'`, grazie alla quale è stato possibile fissare la suddivisione del campione in analisi in tre set distinti sulla base degli indici, ossia la loro posizione all'interno della matrice `Indicatori`. I tre set sono:

- *Training set*
  Contiene le osservazioni utilizzate nel processo di apprendimento della rete. Sulla base di questi dati si calcolano i pesi delle connessioni tra i neuroni.
- *Validation set*
  È utilizzato per comparare le performance dell'algoritmo di previsione creato sulla base del *training set* e per interrompere l'allenamento della rete attraverso il meccanismo chiamato *validation stop*. Questo criterio valuta l'andamento del valore della funzione di perdita, nel caso in esame la *cross-entropy*, nelle varie iterazioni della fase di apprendimento. Ad una diminuzione del valore della *cross-entropy* corrisponde un miglioramento della qualità del modello, al contrario, l'aumento della funzione di perdita è interpretato come un segnale di allontanamento dalla soluzione ottima. Il *validation stop* interrompe il processo di addestramento non appena il valore della *cross-entropy* cresce per un numero di *epoch* consecutive stabilito dalla funzione `net.trainParam.max_fail`. Il modello considera poi come soluzione migliore, o *best epoch*, l'iterazione in cui la funzione di perdita ha assunto valore minimo.

- *Test set*

  Racchiude i dati impiegati nella valutazione finale del modello.

Per la costruzione di questa tipologia di modelli, la teoria prevede l'utilizzo di un campione per la stima dei parametri pari ai 2/3 delle osservazioni ed il restante 1/3 per il controllo. Dovendo suddividere il campione di controllo nei set di *validation* e *test*, si è deciso di dividere equamente le osservazioni assegnando 1/6 dei dati ad entrambi i set. Dunque, la totalità delle osservazioni, pari a 6.938, è stata ripartita cercando di assegnare al set di *training* e di *test* una percentuale simile di società anomale, in modo da rendere più omogenea l'applicazione di quanto imparato nella fase di apprendimento sul *test set*. In particolare, come si evince dal codice soprariportato, la suddivisione è così avvenuta:

- *Training set:* dalla 1 alla 4.626 per un totale di 4.626 osservazioni (2/3), di cui 634 con FLAG DI STATUS S/A – SOCIETA' pari a 1 (13,71%);
- *Test set*: dalla 4.627 alla 5.782 per un totale di 1.156 osservazioni (1/6), di cui 116 con FLAG DI STATUS S/A – SOCIETA' pari a 1 (10,03%);
- *Validation set*: dalla 5.783 alla 6.938 per un totale di 1.156 osservazioni (1/6), di cui 222 con FLAG DI STATUS S/A – SOCIETA' pari a 1 (19,20%).

5.3 Inizializzazione dei parametri e training della rete

```
%% Set the network
% Transpose samples
Indicatori = Indicatori';
Flag = Flag';

% Configure the network
net = configure(net,Indicatori,Flag);
```

Giunti a questo punto è necessario configurare la rete. La configurazione è il processo di impostazione degli input e degli output della rete che precede l'inizializzazione dei pesi e dei *bias*. Il codice di seguito mostra come la configurazione della rete sia avvenuta attraverso l'utilizzo della funzione `configure`. Essa richiede tre valori in ingresso: il nome della rete, gli input e i flag; quest'ultimi richiesti in forma trasposta rispetto a come sono stati importati.

```
%% Initialize the weights
% Import the weights
Matrice_pesi = readtable('Pesi.xlsx');
Matrice_pesi = table2array(Matrice_pesi);

% Assign the weights
net.IW{1,1} = Matrice_pesi(1:12,1:9);
net.LW{2,1} = Matrice_pesi(1:13,1:12);
net.LW{3,2} = Matrice_pesi(1:11,1:13);
net.LW{4,3} = Matrice_pesi(1:1,1:11);

net.b{1} = zeros(12,1);
net.b{2} = zeros(13,1);
net.b{3} = zeros(11,1);
net.b{4} = zeros(1,1);

Pesi_input = net.IW;
Pesi_hidden_output = net.LW;
bias = net.b;
```

Prima di poter iniziare con la fase di addestramento è necessaria un'ultima operazione di impostazione della rete: l'inizializzazione dei parametri. Tale fase, così come la suddivisione fissa delle osservazioni nei tre campioni, risulta indispensabile per la rimozione dell'elemento di casualità introdotto di default dal software, che assegnerebbe i valori iniziali dei parametri in modo casuale. I valori usati per l'inizializzazione sono stati scritti in un file excel e importarti tramite la stessa funzione usata per gli indicatori ed il flag. In seguito, si sono assegnati i pesi delle connessioni dallo strato di input al primo strato nascosto `net.IW{1,1}` e delle restanti connessioni dallo strato `j` allo strato `i`, `net.LW{i,j}`.

La logica con la quale si è stabilito di inizializzare a 0,1 e non a 0 il peso iniziale di tutte le connessioni è che un'inizializzazione dei pesi a 0 rende la rete neurale equivalente ad un modello lineare eliminando tutta la complessità e quindi il vantaggio di utilizzare una rete neurale. Quando si impostano tutti i valori a 0, allora il processo di aggiornamento dei pesi che avviene attraverso il calcolo delle derivate, illustrato nel paragrafo §3.6.1, si semplifica

eccessivamente annullando il valore Δw_{ij} e quindi impedendo l'aggiornamento del parametro vero e proprio. Inizializzando i pesi ad un valore diverso da 0, come per esempio 0,1, l'algoritmo non si bloccherà e sarà in grado di modificare i pesi fin dalla successiva iterazione. Inoltre, è importante evitare l'assegnazione di valori iniziali troppo piccoli o troppo grandi, in quanto in entrambi i casi si avranno problemi di convergenza della rete. Pesi iniziali troppo piccoli sono responsabili del *vanishing gradient problem*, ossia di un'eccessiva diminuzione del gradiente che tende a "svanire" a 0 procedendo a ritroso nella rete, ciò causa un minor apprendimento nei neuroni dei primi strati per i quali il gradiente risulta minore rispetto ai neuroni degli strati successivi. Al contrario, l'inizializzazione con valori troppo grandi è responsabile dell'*exploding gradient problem*, ossia di una crescita eccessiva del gradiente che tende a "esplodere" all'infinito causando smisurati aggiornamenti dei pesi e quindi una difficile convergenza della rete.

Infine, per ogni strato `i` si è adottata la canonica inizializzazione a 0 dei valori dei *bias* `net.b{i}` associando ad essi vettori colonna di zeri aventi dimensione pari al numero di neuroni presenti nello strato `i`. È opportuno precisare che l'inizializzazione dei *bias* a 0 non crea alcun problema in quanto è sufficiente che i pesi degli input risultino diversi da 0 per dare il via all'aggiornamento dei pesi e anche dei *bias* stessi.

```
%% Train the network
[net,tr] = train(net,Indicatori,Flag);
view(net);
```

Dopo aver inizializzato i parametri, eliminando ogni elemento di casualità, è possibile procedere con l'allenamento della rete. Per farlo, è sufficiente richiamare la funzione `train` fornendole in ingresso la rete configurata, gli input e gli ouput, in questo caso rappresentati rispettivamente dagli indicatori e dal flag.

Definendo in tal modo l'addestramento della rete, si utilizzeranno le funzioni di attivazione, di addestramento e di performance impostate di default del software. La prima è l'*hyperbolic tangent function* (tanh), funzione derivabile, centrata in 0, non binaria e non lineare, tutte qualità che la rendono un'ottima funzione di attivazione. Per l'addestramento si utilizza la *trainscg* che, come suggerisce il nome, aggiorna il valore dei pesi e dei *bias* applicando il metodo *scaled conjugate gradient*. Come illustrato in precedenza, uno dei modi più efficienti

per allenare una rete multistrato è usare il metodo del *gradient descent* con la *backpropagation*, il metodo *scaled conjugate gradient* ne rappresenta una variante. L'algoritmo, nella sua versione più classica, modifica i pesi nella direzione di massima diminuzione della *function loss*. Tuttavia, da molti studi è emerso che nonostante la funzione di perdita diminuisca più rapidamente nella direzione negativa del gradiente, non produrrà necessariamente una convergenza più veloce. Negli algoritmi *conjugate gradient* la ricerca della direzione ottima è effettuata tra le direzioni "coniugate" e si è dimostrato come essa generi una convergenza più rapida rispetto ai classici metodi. Per tali metodi, l'entità dell'aggiornamento dei pesi, anche nota come *step size*, è determinata dal *learning rate* mentre negli algoritmi di *conjugate gradient*, lo *step size* è modificato ad ogni iterazione. I metodi appartenenti a questa famiglia differiscono tra loro per la funzione di ricerca impiegata nel calcolo della direzione del gradiente da seguire per determinare lo *step size* che minimizza la *loss function*. Quest'ultima, anch'essa stabilita di default dal software, è la *cross-entropy*, ottima funzione di perdita i cui vantaggi e le principali proprietà sono stati spiegati nel paragrafo §3.7.2.

5.4 Output della rete

```
%% Predict response
scoreTest = net(Indicatori(:,tr.testInd));

%% Evaluate classification with confusion matrix
FlagTargettest = Flag(tr.testInd);
yT = FlagTargettest;
yP = round(scoreTest);

plotconfusion(yT,yP)

% Determine the accuracy
PerErrRete = 100*nnz(yP ~=
double(FlagTargettest))/length(FlagTargettest);
disp(['L'accuratezza del modello è ',
num2str(PerErrRete),'%']);
```

L'ultima parte del programma riguarda la scrittura di una serie di comandi atti a calcolare i valori di risposta della rete e a riorganizzarli in modo da fornire, fin da subito, indicazioni generali sulla performance del modello costruito. In primis si sono assegnati al vettore `scoreTest` gli output calcolati dalla rete precedentemente addestrata. Tali valori sono stati stimati attraverso la funzione `net`, che ricevendo in input la matrice degli indicatori del *test set* ne calcola i punteggi sulla base dei pesi esplicitati nella fase di addestramento. In seguito, si sono definiti due ulteriori vettori `yT` e `yP` contenenti rispettivamente il FLAG DI STATUS S/A – SOCIETA' delle società appartenenti al campione di test e il valore arrotondato all'intero più vicino degli output calcolati dalla rete. L'operazione di arrotondamento effettuata sul punteggio calcolato dalla rete si rivela necessaria per renderlo confrontabile con il valore binario assunto dal flag. Sulla base dei vettori `yT` e `yP` si costruisce la matrice di confusione, anche nota come tabella di errata classificazione, utilizzata per visualizzare l'accuratezza di una classificazione binaria. Le informazioni contenute nella matrice di confusione forniscono un'idea della performance del modello riassumendo i risultati della classificazione. Le sue proprietà saranno analizzate nel dettaglio nel prossimo capitolo, nel quale si presenteranno gli strumenti impiegati nella valutazione delle reti.

```
%% Organize data
%Training set
Datianatraining = Datianagrafici(tr.trainInd,:);
Indicatoritraining = Indicatori(tr.trainInd,:);
Flagtraining = Flag(tr.trainInd);

%Validation set
Datianavalidation = Datianagrafici(tr.valInd,:);
Indicatorivalidation = Indicatori(tr.valInd,:);
Flagvalidation = Flag(tr.valInd);
```

```
%Test set
Datianatest = Datianagrafici(tr.testInd,:);
Indicatoritest = Indicatori(tr.testInd,:);
Flagtest = Flag(tr.testInd);
scoreTest = scoreTest';
yP = yP';
```

Dopo aver costruito e testato la rete, con le ultime righe di codice del programma si richiamano e si salvano in apposite matrici le informazioni anagrafiche, gli indicatori e i flag dei tre campioni, in modo da poterli esportare più facilmente in vista di una successiva fase di analisi dei risultati che sarà oggetto dei prossimi capitoli.

Capitolo VI

La valutazione e l'evoluzione delle reti

Nel sesto capitolo si presenteranno i principali strumenti valutativi impiegati nella valutazione delle performance delle reti, se ne illustreranno le caratteristiche e si forniranno le linee guida per una corretta interpretazione. Nella seconda parte del capitolo si ripercorrerà l'evoluzione dell'architettura delle reti costruite presentando i problemi, spiegando le soluzioni e motivando le ragioni alla base di ogni cambiamento apportato.

6.1 Gli strumenti valutativi

Prima di procedere con la presentazione delle reti neurali costruite è necessario presentare gli strumenti che hanno consentito di valutare le performance delle singole reti e di coglierne i punti deboli, suggerendo le modifiche da apportare. L'importanza di questi strumenti è tale da renderne universale l'utilizzo infatti, ogniqualvolta risulta necessario valutare le performance di modelli di *machine learning* si ricorre a questi schemi.
Gli strumenti a cui si fa riferimento sono:

- *Confusion matrix*;
- *Receiver Operating Characteristic*;
- *Best Validation Performance*;
- *Gradient and Validation Checks*;
- *Error histogram*.

Data la rilevanza che assumono nel processo decisionale si è deciso di dedicare ad ognuno di essi un paragrafo per illustrarne il significato e fornirne la corretta interpretazione.

6.1.1 Confusion matrix

La matrice di confusione è lo schema di analisi della performance in grado di riassumere al meglio i risultati di un modello di classificazione binaria. La sua struttura, riconducibile a una matrice 3x3, fornisce un'idea immediata della performance del modello semplificando i confronti tra le diverse strutture che operano con lo stesso *dataset*.

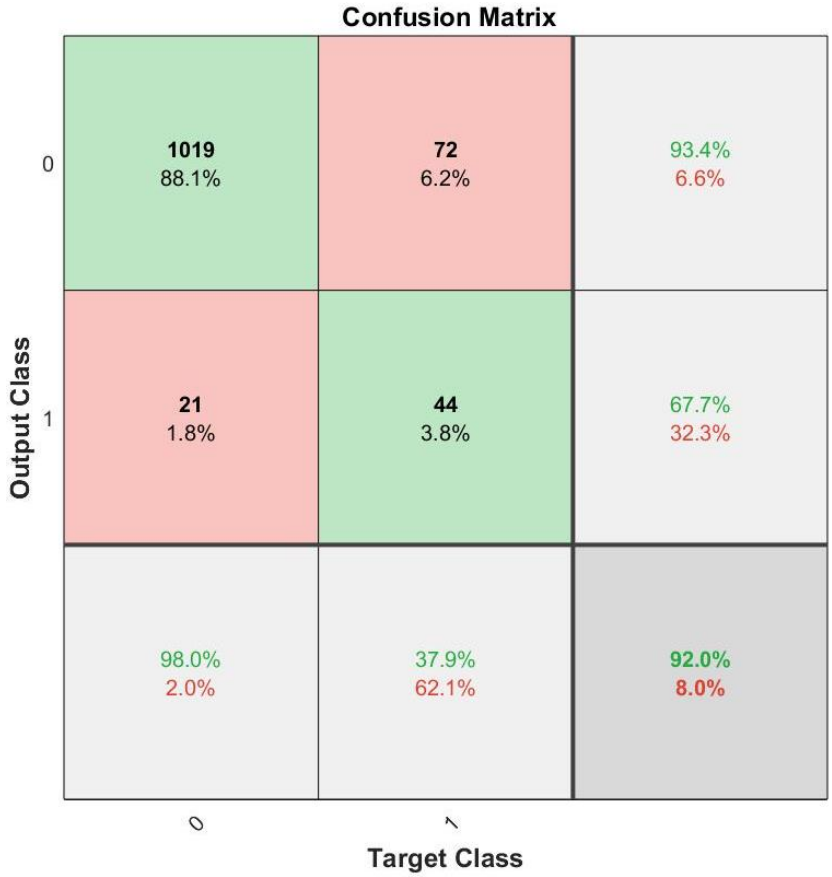

Figura 6.1 *Matrice di confusione generata da una rete formata da 9 neuroni di input, 3 strati nascosti con 11, 14, 10 neuroni rispettivamente e 1 neurone di output*

Nella *confusion matrix* le righe rappresentano le classi predette o *output class*, mentre le colonne rappresentano le classi reali o *target class*. Nella diagonale principale si collocano le osservazioni classificate correttamente, al contrario quelle errate si trovano sull'antidiagonale. In ogni cella sono indicati sia il numero di osservazioni che rientrano in quella casistica sia la percentuale di esse rispetto al numero totale. Notevole importanza assumono la terza colonna e la terza riga, che contengono le percentuali riassuntive dei valori predetti e reali classificati rispettivamente in modo corretto ed errato.

In particolare, considerando la *confusion matrix* come una matrice 3x3, le percentuali più rilevanti sono le due in rosso in posizione [1;3] e [2;3], le due in verde in posizione [3;1] e [3;2] e quella scritta in verde in posizione [3;3]. Ognuna di queste percentuali assume un preciso significato ed è così calcolata:

- Probabilità di mancato allarme:

$$\frac{falsi\ negativi}{veri\ negativi + falsi\ negativi} \tag{6.1}$$

  In statistica è nota con il nome di errore di secondo tipo e corrisponde alla probabilità di considerare erroneamente un'osservazione positiva come negativa.

- Probabilità di falso allarme:

$$\frac{falsi\ positivi}{veri\ positivi + falsi\ positivi} \tag{6.2}$$

  In statistica è nota con il nome di errore di primo tipo e corrisponde alla probabilità di considerare erroneamente un'osservazione negativa come positiva.

- Specificità:

$$\frac{veri\ negativi}{veri\ negativi + falsi\ positivi} \tag{6.3}$$

  Calcola la percentuale delle osservazioni negative classificate correttamente.

- Sensibilità:

$$\frac{veri\ positivi}{veri\ positivi + falsi\ negativi} \tag{6.4}$$

  Calcola la percentuale delle osservazioni positive classificate correttamente.

- Accuratezza:

$$\frac{veri\ positivi + veri\ negativi}{veri\ positivi + veri\ negativi + falsi\ positivi + falsi\ negativi} \tag{6.5}$$

  Esprime l'accuratezza del modello calcolando la percentuale di osservazioni classificate correttamente rispetto al totale. Il suo complemento ad 1 è il tasso di errore e, così come suggerisce il nome, calcola la percentuale di previsioni errate.

La Tabella 6.1 riorganizza le informazioni precedenti riproducendo lo schema della matrice di confusione.

Tabella 6.1 *Nomenclatura informazioni contenute nella matrice di confusione*

<table>
<tr><td colspan="5">Matrice di confusione</td></tr>
<tr><td rowspan="4">*output class*
(valori
predetti)</td><td>negativi
(0)</td><td>veri
negativi</td><td>falsi
negativi</td><td>Probabilità
di mancato
allarme</td></tr>
<tr><td>positivi
(1)</td><td>falsi
positivi</td><td>veri
positivi</td><td>Probabilità
di falso
allarme</td></tr>
<tr><td></td><td>Specificità</td><td>Sensibilità</td><td>Accuratezza</td></tr>
<tr><td></td><td>negativi
(0)</td><td>positivi
(1)</td><td></td></tr>
<tr><td></td><td colspan="3">*target class*
(valori reali)</td></tr>
</table>

6.1.2 Receiver Operating Characteristic

Il *Receiver Operating Characteristic* (ROC) è uno schema di analisi della performance utilizzato esclusivamente per i modelli di classificazione binaria. Lungo l'asse delle ascisse è rappresentato il *False Positive Rate* (FPR) ossia la probabilità di falso allarme (6.2) mentre sull'asse delle ordinate si trova il *True Positive Rate* (TPR) ovvero la sensibilità (6.4). Questo strumento si basa sul concetto di accuratezza descritto nel capitolo precedente e calcolato tramite la (6.5) ossia la percentuale di osservazioni classificate correttamente rispetto al totale. In un modello perfetto in grado di classificare correttamente tutte le osservazioni, il suo valore è pari al 100% al contrario ad un modello che attribuisce alle imprese punteggi casuali è associata un'accuratezza del 50% in quanto essendo la classificazione di tipo binario classificherà correttamente in media 1 azienda su 2.

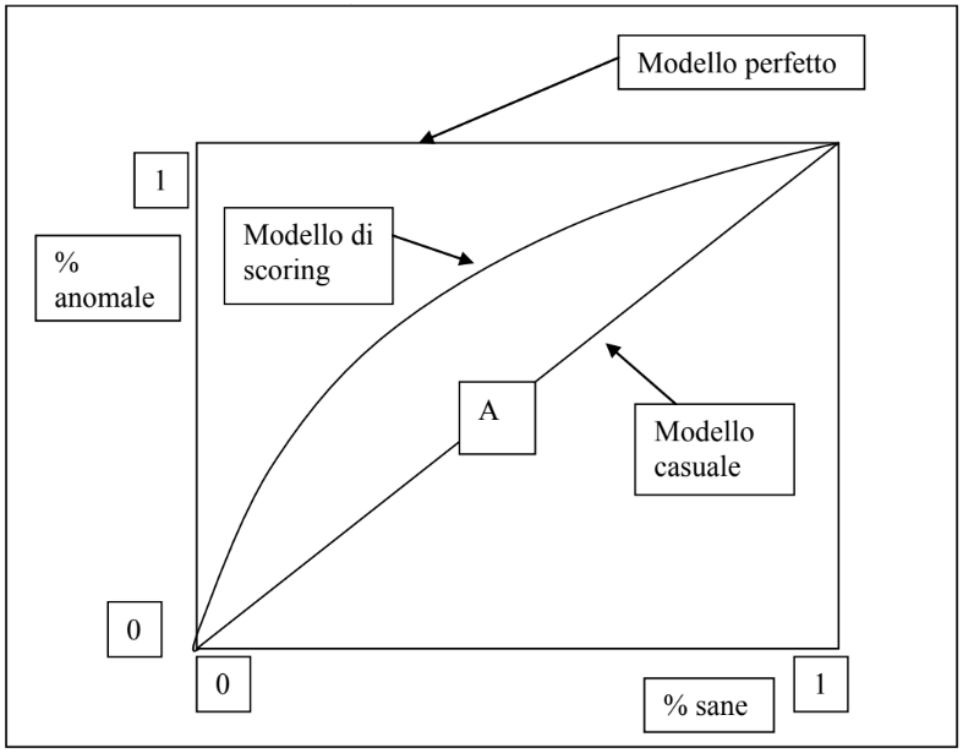

Figura 6.2 *Receiver Operating Characteristic*

La Figura 6.2 mostra un generico schema ROC nel quale è possibile collocare le curve dei modelli descrivendone il comportamento. Il modello casuale è rappresentato dalla diagonale, il segmento che delimita la parte superiore del grafico descrive il modello perfetto, infine nella zona intermedia si collocano i modelli generici. L'accuratezza dei modelli, indicata con *A*, è espressa dal valore dell'area compresa tra il segmento inferiore del grafico e la curva che identifica i modelli. Essa raggiunge un valore massimo di 1 per i modelli perfetti e di 0,5 per quelli casuali mentre nei modelli generici assume un valore compreso tra 0,5 e 1 con valori maggiori per i più performanti.

6.1.3 Best Validation Performance

Tramite questo strumento si analizzano le forme e l'andamento delle curve di training, *validation* e test per identificare il comportamento di un modello di *machine learning* e suggerirne i cambiamenti che possono essere effettuati per migliorarne le performance. Le dinamiche delle curve di apprendimento possono essere ricondotte a tre differenti tipologie:

- *underfit*;
- *overfit*;
- *goodfit*.

L'assunzione di fondo sulla quale si basa la valutazione di una generica curva è che, essendo l'obiettivo finale la minimizzazione della funzione di perdita ossia della *cross-entropy*, valori minori sull'asse delle ordinate denotano un miglior apprendimento.

Con il termine *underfit* si fa riferimento alla difficoltà espressa dal modello nell'apprendere le informazioni rilevanti dal *training set*. In questi casi il modello non è in grado di ottenere un livello di errore sufficientemente basso tale da conferirgli la capacità di cogliere appieno la complessità del *dataset*.

Graficamente, il comportamento dell'*underfit* si manifesta attraverso i due andamenti descritti dalla Figura 6.3 e Figura 6.4. Nel primo caso la *training loss* e la *validation loss* rimangono stabili indicando una pessima capacità da parte del modello nel ridurre il valore delle funzioni di perdita e quindi di imparare e migliorarsi nel corso delle varie *epoch*. Il comportamento descritto dalla Figura 6.4 suggerisce un'interruzione prematura del processo di apprendimento, infatti l'andamento decrescente delle due funzioni di perdita non sembra essersi attestato ad un valore minimo. Altro elemento che contraddistingue l'*underfit* è la distanza minima tra le due curve, in questo caso il grafico segnala ancora una volta l'incapacità da parte del modello di imparare dai dati forniti in input, mantenendo caratteristiche troppo generiche e quindi performando male nella fase di test.

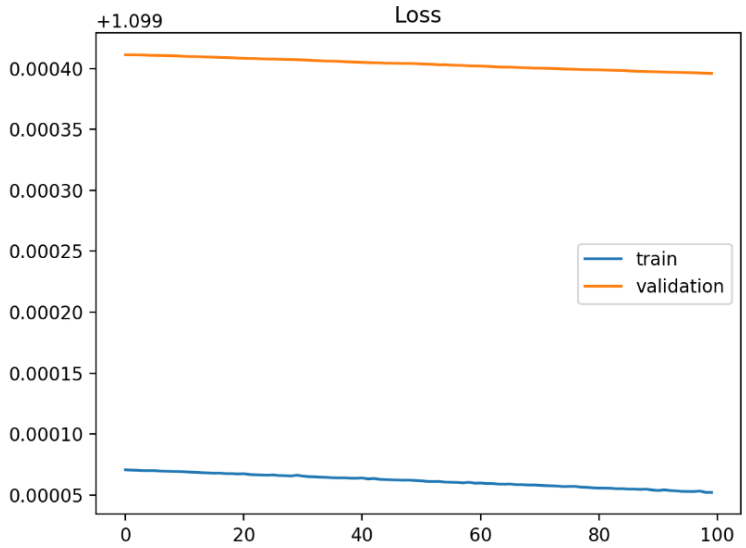

Figura 6.3 *Best Validation Performance: underfit*

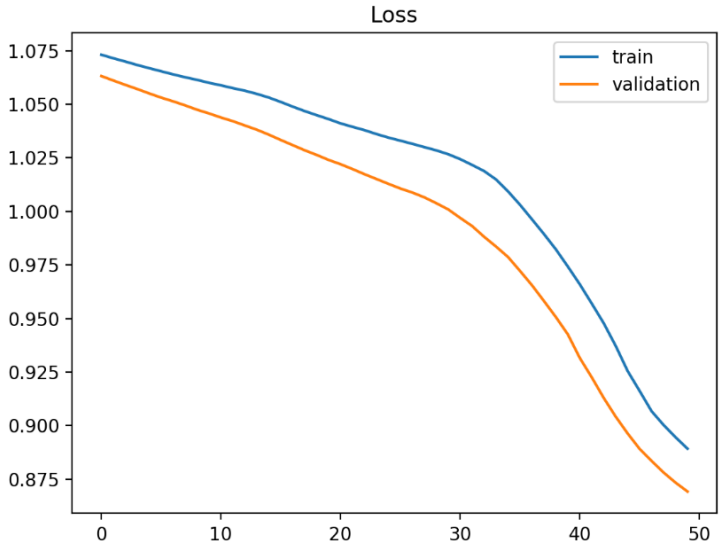

Figura 6.4 *Best Validation Performance: underfit*

Il comportamento opposto è chiamato *overfit* e si riferisce ad un modello che durante il processo di addestramento ha raccolto ogni tipo di informazione dal *training set* includendo anche errori statistici e fluttuazioni casuali. Quest'insieme di informazioni, al contrario rispetto a quanto si possa ritenere, risulta dannoso per la valorizzazione dei parametri del modello, in quanto essi seguiranno anche gli errori ed il rumore (*noise*) presente nei dati. Dunque, il grande problema dell'*overfit* è che più il modello si specializza nei *training data* meno sarà in grado di generalizzare ossia di elaborare delle previsioni su dati diversi rispetto a quelli utilizzati per l'apprendimento. Si tratta chiaramente di una situazione da evitare poiché implicherebbe una limitazione significativa dell'utilizzo del modello rendendolo poco flessibile e incapace di fornire previsioni accurate. Questo errore nel processo di generalizzazione può essere prevenuto valutando la performance della *validation loss* e quindi il grafico del *Gradient and Validation Checks*.

A livello grafico l'*overfit* si traduce nel comportamento illustrato in Figura 6.5, nella quale la curva di *training* continua a diminuire all'aumentare del numero di *epoch* senza raggiungere un punto di minimo. Altro andamento sinonimo di *overfit* è il raggiungimento di un valore minimo della *validation loss* seguito da una nuova crescita significativa. In generale, il fatto che le curve di *training* e *validation loss* siano troppo distanti implica che il modello si è focalizzato su dettagli presenti nei dati infatti, il divario aumenta fisiologicamente dopo le prime *epoch* proprio perché il modello inizia ad apprendere e specializzarsi sui dati.

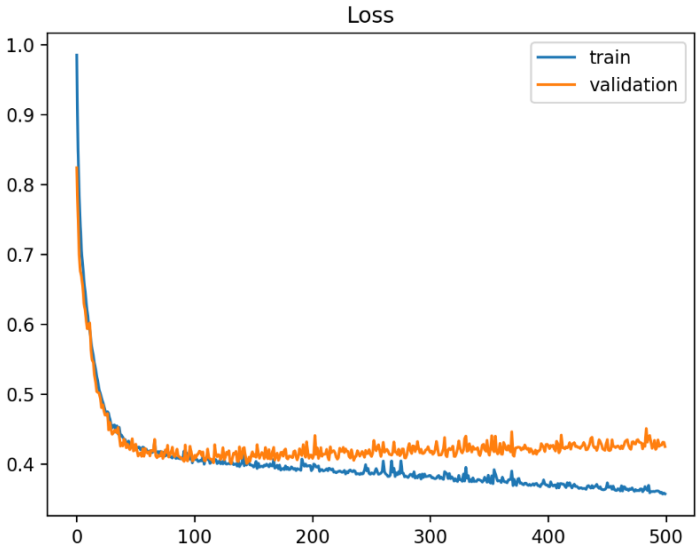

Figura 6.5 *Best Validation Performance: overfit*

Un buon andamento delle funzioni di perdita è chiamato *goodfit* e si tratta di una soluzione intermedia tra le situazioni precedenti. L'andamento di un *goodfit* è quello tipico di Figura 6.6. In questo caso la *training* e la *validation loss* diminuiscono raggiungendo un punto di minimo oltre il quale si stabilizzano lasciando un gap tra i valori finali. Ci si aspetta che tra queste due curve ci sia sempre una distanza fissa chiamata *generalization gap* la cui entità fornisce un'idea della capacità del modello di generalizzare le previsioni.

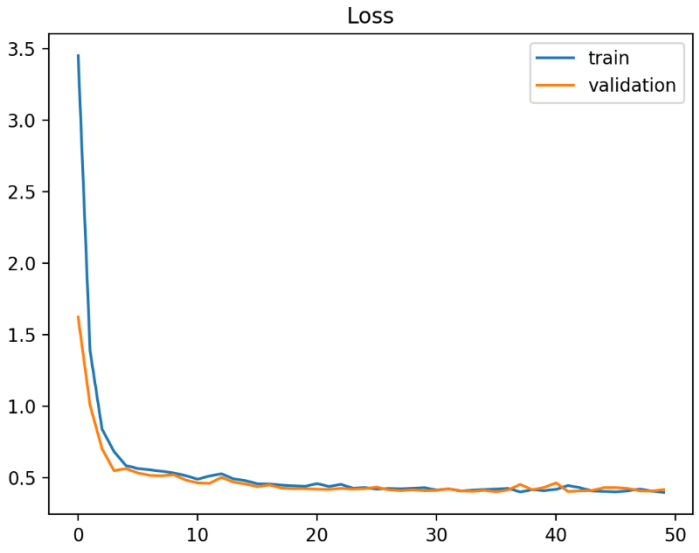

Figura 6.6 *Best Validation Performance: goodfit*

Riassumendo, è necessario evitare un comportamento di *overfit* poiché il modello non avrebbe grandi capacità di generalizzazione ed il suo impiego rimarrebbe fine a sé stesso e, allo stesso modo, occorre sottrarsi da un possibile *underfit* poiché indicherebbe che non si stanno sfruttando appieno le potenzialità del modello. Dunque, l'entità del *generalization gap* deve essere scelta sulla base degli obiettivi che si vogliono raggiungere con il modello. Se si vuole ottenere la sicurezza che il modello sia in grado di gestire molte situazioni diverse si possono accettare gap più grandi correndo il rischio di *underfit* mentre se commettere un errore di tipo falso positivo risulta troppo oneroso in termini di tempi, costi o qualità allora è preferibile un gap di dimensioni ridotte correndo il rischio di *overfit*.

6.1.4 Gradient and Validation Checks

Con i prossimi due grafici si riprende in parte il discorso precedente focalizzando l'attenzione sul criterio di interruzione del processo di apprendimento. Il *Gradient* e il *Validation Checks* mostrano rispettivamente l'andamento del gradiente e dell'errore sul *validation set*. Dall'analisi della dinamica del gradiente è possibile cogliere informazioni rilevanti sulle performance della funzione e del metodo di addestramento. Come già illustrato nel paragrafo §5.3 la funzione impiegata è la *trainscg* che aggiorna il valore dei pesi e dei *bias* attraverso il metodo *scaled conjugate gradient* applicato alla *backpropagation*. L'algoritmo rappresenta una variazione rispetto alla versione originale ma basa ugualmente la scelta della migliore direzione di modifica dei pesi sul valore assunto dal gradiente. Nel paragrafo §3.6.1 si era precisato come l'aggiornamento dei parametri del modello avvenga in direzione opposta al gradiente, ossia contrariamente alla direzione di massima crescita della funzione di perdita. Dall'analisi dell'andamento dei valori assunti dal gradiente è quindi possibile valutare se la funzione di apprendimento sta performando bene, ossia se l'addestramento sta aggiornando i pesi e i *bias* diminuendo il valore del gradiente.

Quest'ultimo fornisce inoltre informazioni rilevanti sui valori utilizzati per l'inizializzazione dei parametri. Sempre nel paragrafo §5.3, si sono evidenziate le due problematiche relative all'attribuzione di pesi iniziali troppo piccoli o troppo grandi, entrambi i casi si individuano facilmente tramite l'analisi dell'andamento del gradiente. Il primo caso, *vanishing gradient problem*, si manifesta come una diminuzione repentina del valore del gradiente provocando un minor apprendimento nei neuroni dei primi strati per i quali il gradiente risulta minore rispetto ai neuroni degli strati successivi. L'inizializzazione con valori troppo elevati è l'*exploding gradient problem*, rappresentato da un andamento crescente del gradiente che

causa aggiornamenti dei pesi di maggior entità con il procedere delle *epoch* e di conseguenza una difficile convergenza della rete.

La Figura 6.7 mostra oltre all'andamento del gradiente anche il numero di *validation fail* nelle varie *epoch*, il cui valore è strettamente connesso al processo di interruzione dell'apprendimento. Come già spiegato nel paragrafo §5.2, il modello termina l'aggiornamento dei parametri non appena il numero di *validation fail* consecutivi è pari a 15. Un *validation fail* si verifica ogniqualvolta la *cross-entropy* del *validation set* aumenta invece che diminuire. Fissando un limite a questo valore si previene un possibile *overfit* sul *training set* del modello, il quale continuerebbe ad aggiornare i parametri includendo gli errori e il *noise* presente nei dati. Il criterio di *validation stop* sottolinea in questo modo l'importanza dell'impiego di un *validation set* per la costruzione di un modello di intelligenza artificiale.

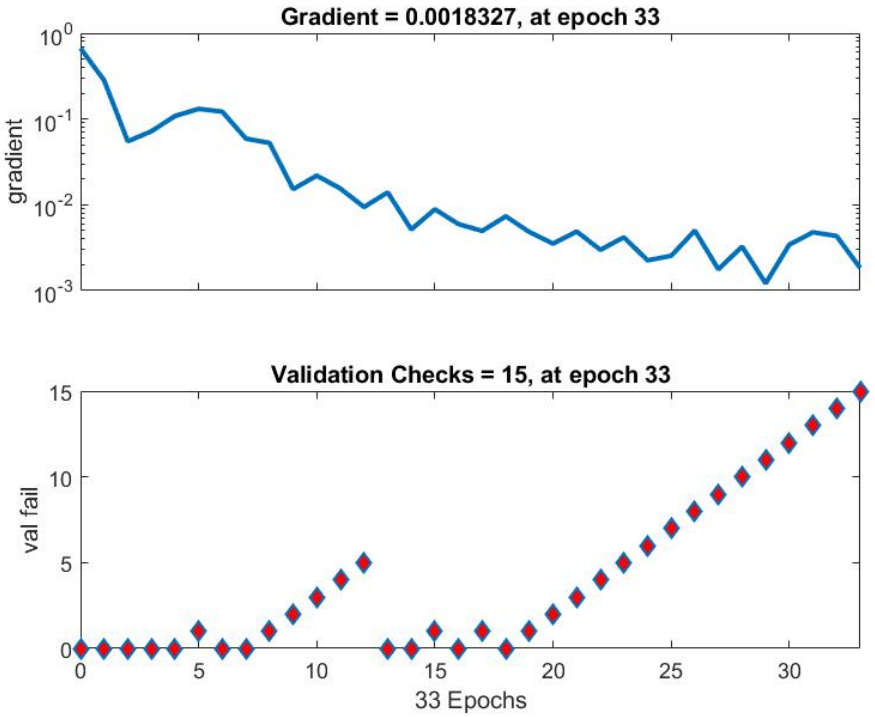

Figura 6.7 *Gradient and Validation Check generati da una rete formata da 9 neuroni di input, 3 strati nascosti con 11, 14, 10 neuroni rispettivamente e 1 neurone di output*

6.1.5 Error histogram

L'ultimo strumento impiegato per analizzare la performance delle reti neurali costruite è l'*Error histogram*. Il grafico in questione assegna le osservazioni a 20 diverse classi sulla base del valore dell'errore calcolato come differenza tra target e output. Sull'asse delle ascisse è indicato il valore dell'errore che rappresenta il centro dell'intervallo della classe mentre sull'asse delle ordinate è riportato il numero delle osservazioni che ricadono in quella classe. Inoltre, l'istogramma di Figura 6.8 divide gli errori nei tre set e segnala con una barra arancione verticale il valore dell'ascissa in cui l'errore è pari a 0. In un *Error histogram* di un buon modello di classificazione gli errori si devono disporre intorno allo zero e non si devono notare particolari concentrazioni in determinate classi in quanto sintomo di valutazioni sistematicamente errate. In particolare, se gli errori sono concentrati maggiormente nelle classi positive significa che il modello tende ad attribuire un punteggio sistematicamente inferiore al target sottostimando il punteggio reale, al contrario se gli errori sono concentrati prevalentemente nelle classi negative significa che il modello è solito attribuire un punteggio superiore al target sovrastimando il punteggio reale.

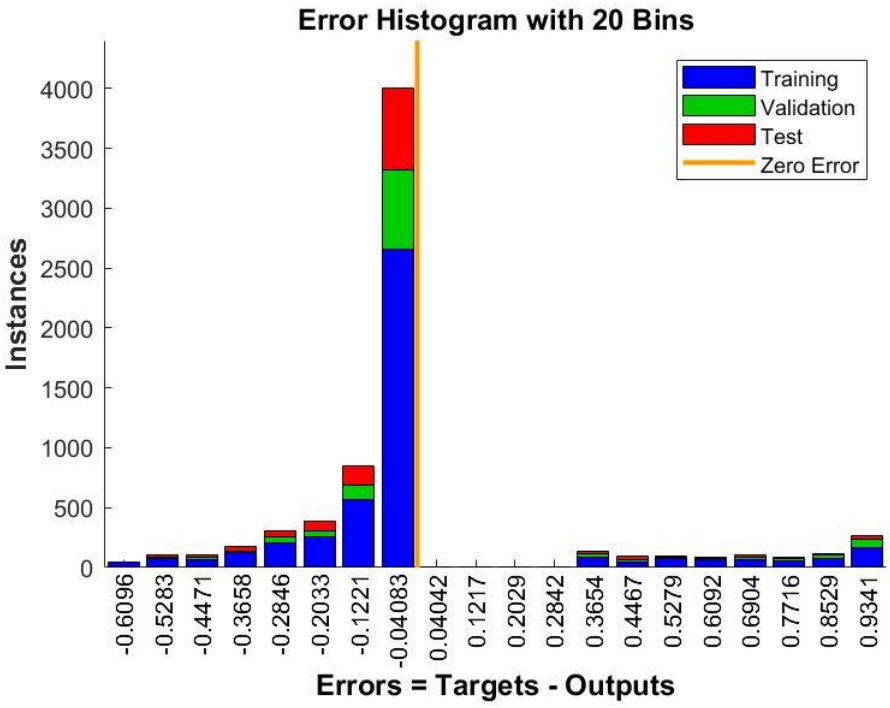

Figura 6.8 *Error histogram generato da una rete formata da 9 neuroni di input, 3 strati nascosti con 11, 14, 10 neuroni rispettivamente e 1 neurone di output*

6.2 Evoluzione della struttura

Il percorso evolutivo dell'architettura delle reti neurali presentato in questo paragrafo è stato caratterizzato da numerose prove in quanto al momento non sono ancora state definite regole o teorie capaci di stabilire a priori la struttura ottimale della rete.

Per prendere confidenza con il software MATLAB si è cominciato costruendo reti base aventi come input due indicatori scelti tra quelli con maggiore correlazione con il FLAG DI STATUS S/A – SOCIETA', uno strato nascosto formato da due neuroni ed un singolo neurone di output. Essendo lo scopo del modello quello di predire lo "status" della società ossia se l'azienda che si sta osservando è sana o anomala, la struttura dello strato di output non ha subito alcuna variazione rimanendo fissa ad un neurone. Infatti, la risposta richiesta dal modello è di tipo binario, pari ad 1 se l'azienda è anomala e 0 se è sana, per questo motivo non essendo richiesta alcuna particolarità costruttiva dell'ultimo strato, questo è rimasto invariato per tutte le prove effettuate. Altro elemento rimasto costante è l'adozione della suddivisione dei dati nei tre set data l'importanza assunta dal *validation set* nel processo di interruzione dell'apprendimento per prevenire i comportamenti di *overfit*.

Una volta superato l'ostacolo relativo alla costruzione della prima rete, si sono inseriti elementi di maggior complessità nella struttura per ottenere prestazioni migliori. Si è iniziato con l'aumentare il numero di neuroni negli strati di input e negli strati nascosti. In particolare, essendo il numero di nodi in ingresso pari agli indicatori forniti al modello, si sono costruite reti formate da 3, 6 e 9 neuroni di input. La scelta di quest'ultimi si è basata sul processo di *feature selection*, presentato in precedenza nel paragrafo §4.5, che consiste nell'individuare gli indicatori che meglio catturano le informazioni in grado di descrivere lo scenario in analisi. Seguendo questo processo si sono scartati gli indicatori maggiormente correlati tra di loro in quanto, non solo incapaci di apportare informazioni aggiuntive alla comprensione del fenomeno in analisi ma anche causa di problemi di collinearità imperfetta. Tale problema, tipico della regressione multipla, emerge quando le variabili esplicative sono altamente correlate tra loro causando un'imprecisione nella stima dei coefficienti di regressione. Infatti, se le variabili esplicative sono altamente correlate ed il valore di una di esse è modificato, la stima dei coefficienti di regressione può variare di molto. Per queste ragioni, si sono selezionati gli indicatori maggiormente correlati con il FLAG DI STATUS S/A – SOCIETA' e allo stesso tempo aventi correlazione minima tra loro.

Gli indicatori, individuati sono in ordine:

- ROE;
- $valore\ aggiunto\ operativo/ricavi$;
- $riserve + utile/attivo\ netto$;
- $passivit\grave{a}\ correnti/ricavi$;
- $risultato\ netto\ rettificato/ricavi$;
- $\ln(ricavi)$;
- $autofinanziamento\ lordo/attivo\ netto$;
- $oneri\ finanziari\ netti/EBIT$;
- $patrimonio\ netto\ tangibile/debiti\ totali + patrimonio\ netto$.

L'aumento dei neuroni in input è stato accompagnato da una crescita del numero di *hidden layers* e di neuroni all'interno di essi a seconda del numero di indicatori utilizzati. In sostanza, partendo dallo strato di input formato da 3, 6 e 9 neuroni si sono aggiunti per ognuna delle tre tipologie 1, 2 e 3 strati nascosti costituiti da 1, 2 o 3 neuroni se il numero di input era rispettivamente pari a 3, 6 o 9, per un totale di 10 strutture diverse.

Le prove sono state effettuate con lo scopo di verificare eventuali miglioramenti nelle performance della rete all'aumentare del numero di indicatori utilizzati come input e all'aumentare del numero di strati nascosti e di neuroni presenti in essi. Data la difficoltà nel cogliere i possibili miglioramenti soprattutto sulla struttura degli strati nascosti, si sono costruite altre due reti con 5 neuroni in ingresso e 3 strati nascosti aventi rispettivamente 5, 5, 5 e 5, 4, 3 neuroni facendo emergere come le performance migliori derivino proprio da quest'ultima tipologia di rete avente un numero di neuroni diverso nei vari strati nascosti.

L'analisi dei risultati delle precedenti reti ha rivelato una certa instabilità nei tassi di errore tra le varie architetture. Per questo motivo si sono studiati nel dettaglio le funzioni e i valori applicati di default da MATLAB nella costruzione delle reti neurali. Approfondendo la conoscenza del software si è osservato che la composizione dei set ovvero l'attribuzione dei dati relativi alle singole società avveniva in modo casuale. Tuttavia, questa procedura di assegnazione casuale delle imprese ai rispettivi set risulta piuttosto fuorviante se si è alla ricerca dell'architettura più adatta a rappresentare il caso di studio, poiché non consente di stabilire a quale fattore attribuire il miglioramento o il peggioramento delle performance della rete.

Trovata la ragione dell'eccessiva variabilità della performance delle reti, si è scritta la parte di codice per indicizzare le aziende ed assegnarle ciascuna al proprio set fissando in tal modo la composizione dei tre set per tutte le prove successive. Apportata la nuova modifica, si sono ricostruite le reti precedenti. Tuttavia, nonostante la variabilità delle performance risultasse significativamente ridotta, i miglioramenti ottenuti non erano monotonici rispetto alla complessità delle architetture. Dalla costruzione di queste prove è emerso che le reti con performance migliori erano quelle aventi un numero di neuroni negli strati nascosti dello stesso ordine di grandezza dei nodi in ingresso. A tal proposito, si sono effettuate due ulteriori prove con 9 e 10 neuroni di input e rispettivamente 8,7,5 e 10,9,8,7 neuroni negli strati nascosti. Nonostante un miglioramento generale delle performance, ciascuna rete persisteva nel generare risultati variabili seppur con varianza minore rispetto a quanto avveniva precedentemente con la composizione casuale dei tre set. Si è cercato quindi di trovare ed eliminare gli elementi responsabili di questa casualità analizzando ancora una volta le funzioni ed i valori di default del software. Questa volta la causa della variabilità rimanente è stata individuata nell'inizializzazione casuale dei pesi e dei *bias*. In sostanza, ogni volta che veniva lanciata una rete, questa iniziava ad aggiornare i propri parametri partendo ogni volta da valori diversi terminando il processo di apprendimento con parametri simili ma differenti. Al fine di ottenere performance più robuste e meno casuali, si sono inizializzati agli stessi valori i pesi e i *bias* delle diverse reti. In particolare, per le motivazioni già spiegate nel paragrafo §5.3 si sono impostati i valori iniziali dei pesi a 0,1 mentre per i *bias* si è optato per un'inizializzazione a 0.

Eliminato ogni elemento di casualità, si sono ricostruite ancora una volta tutte le reti precedenti in modo da cogliere al meglio le variazioni di performance tra le diverse strutture. Inoltre, si è apportata un'ulteriore modifica invertendo il set di *validation* con quello di *test* al fine di avere una percentuale di società anomale più simile tra il *training set* ed il *test set*. Le ragioni dell'inversione, e quindi della ricerca di una maggiore omogeneità nella composizione dei campioni di training e di test, stanno nel processo di apprendimento. Un modello avrà performance tanto migliori quanto più sono simili le osservazioni che costituiscono i due campioni, poiché partendo dallo studio dei dati del *training set* applicherà i le formule ricercando caratteristiche analoghe nei dati del *test set*. Ricostruite le reti, si sono individuate le più performanti cercando di migliorarle ulteriormente aumentando il numero di neuroni negli strati nascosti ossia di gradi di libertà del modello.

In particolare, si sono costruite le ultime 4 reti così strutturate:

- 5 input e 7, 10, 6 neuroni nascosti;
- 6 input e 8, 11, 7 neuroni nascosti;
- 9 input e 12, 13, 11 neuroni nascosti;
- 10 input e 12, 15, 10 neuroni nascosti.

Tra le reti elencate, la terza e la seconda risultano in ordine le migliori, per questo motivo la loro performance sarà analizzata nel dettaglio nel prossimo capitolo.

Capitolo VII

Analisi delle performance e degli errori

Nel settimo capitolo si applicheranno gli strumenti valutativi precedentemente illustrati alle due reti più performanti costruite. A seguire, si presenterà una valutazione economico-finanziaria degli errori della rete con 9 neuroni di input, 3 strati nascosti con 12, 13, 11 neuroni rispettivamente e 1 neurone di output. Il capitolo proseguirà con l'esposizione di considerazioni riguardanti possibili miglioramenti e si concluderà con la presentazione di una possibile procedura di utilizzo del modello basata sulle performance ottenute.

7.1 Applicazione degli strumenti valutativi

L'analisi delle due reti più performanti è condotta attraverso la descrizione e il commento degli strumenti valutativi introdotti nel paragrafo §6.1. In questa prima parte del capitolo si applicheranno quindi le nozioni teoriche studiate a due casi reali di reti neurali.

7.1.1 La struttura

Procedendo con ordine, la struttura di quella che può essere considerata come la seconda migliore rete neurale, schematizzata in Figura 7.1, è composta da 6 neuroni di input che ricevono in ingresso per ogni azienda i valori dei seguenti indicatori:

- ROE;
- $valore\ aggiunto\ operativo/ricavi$;
- $riserve + utile/attivo\ netto$;
- $passività\ correnti/ricavi$;
- $risultato\ netto\ rettificato/ricavi$;
- $\ln(ricavi)$.

Presenta poi 3 strati nascosti contenenti 8, 11, 7 neuroni rispettivamente e termina con un singolo neurone di output dal quale si ricava la previsione del modello.

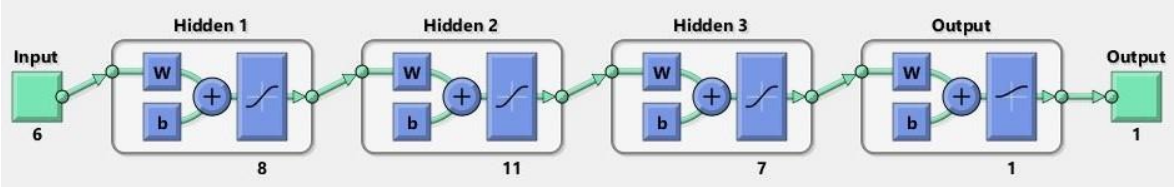

Figura 7.1 *Struttura di una rete formata da 6 neuroni di input, 3 strati nascosti con 8, 11, 7 neuroni rispettivamente e 1 neurone di output*

La struttura miglior rete neurale, illustrata in Figura 7.2, presenta 9 neuroni di input poiché aggiunge ai precedenti 6 i seguenti 3 indicatori:

- $autofinanziamento\ lordo/attivo\ netto$;
- $oneri\ finanziari\ netti/EBIT$;
- $patrimonio\ netto\ tangibile/debiti\ totali + patrimonio\ netto$.

Anch'essa si sviluppa su 3 strati nascosti ma con 12, 13, 11 neuroni rispettivamente e termina con un singolo neurone di output.

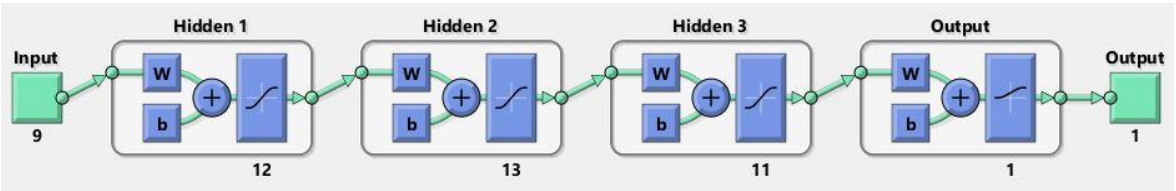

Figura 7.2 *Struttura di una rete formata da 9 neuroni di input, 3 strati nascosti con 12, 13, 11 neuroni rispettivamente e 1 neurone di output*

7.1.2 Confusion matrix

Le matrici di confusione di Figura 7.3 e 7.4 sono lo strumento valutativo che meglio riesce a riorganizzare i risultati della classificazione binaria. Osservando i valori delle due matrici ed in particolare le percentuali collocate nelle celle grigie, il cui significato è spiegato nel paragrafo §6.1.1, si nota come essi migliorino nella seconda rete. L'aumento delle osservazioni classificate correttamente e la conseguente riduzione delle classificazioni errate sono confermati dalla crescita della percentuale in verde nella cella [3;3], che esprime l'accuratezza del modello. Infatti, la rete da 6 neuroni in ingresso ha un'accuratezza del 91.9% mentre la rete migliore raggiunge il 92.3%. Inoltre, dal confronto delle due matrici emerge che l'aggiunta di 3 nuovi indicatori e l'aumento del numero di neuroni nei 3 strati nascosti, ha migliorato il modello consentendo di classificare correttamente 4 aziende sane che erano erroneamente classificate come anomale e 1 società anomala che prima era considerata sana.

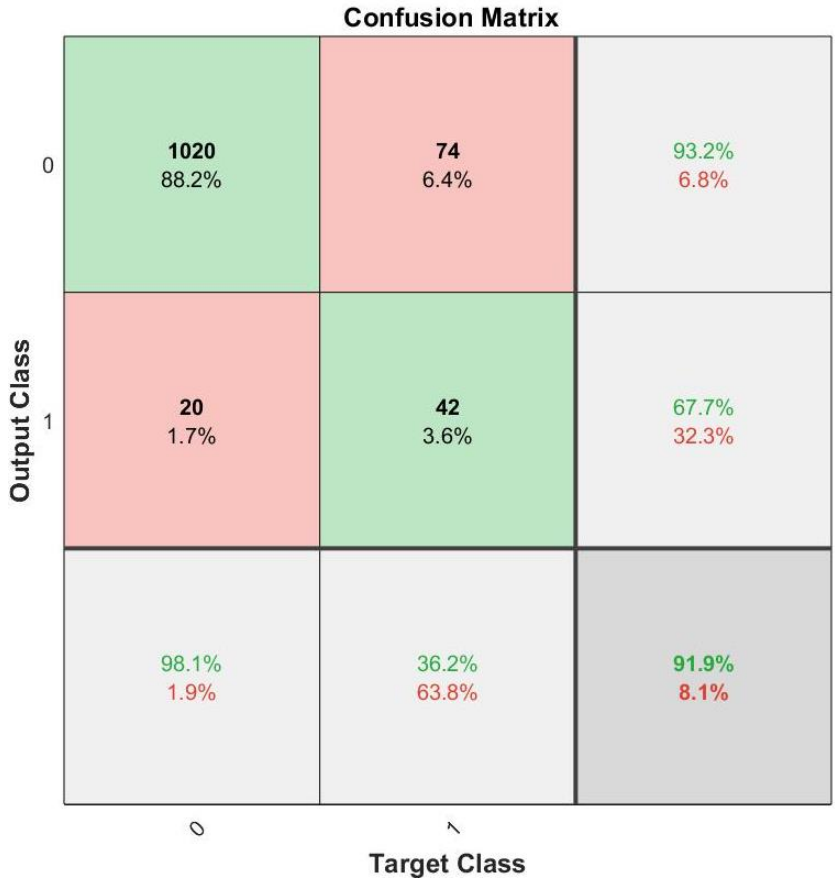

Figura 7.3 *Confusion matrix generata da una rete formata da 6 neuroni di input, 3 strati nascosti con 8, 11, 7 neuroni rispettivamente e 1 neurone di output*

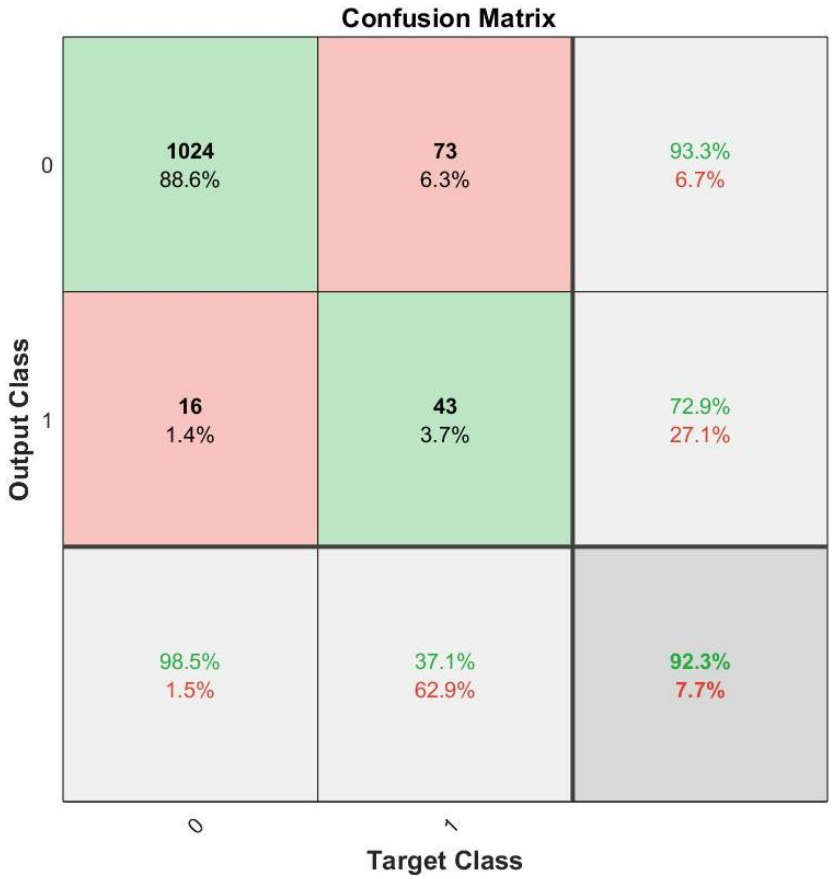

Figura 7.4 *Confusion matrix generata da una rete formata da 9 neuroni di input, 3 strati nascosti con 12, 13, 11 neuroni rispettivamente e 1 neurone di output*

7.1.3 Receiver Operating Characteristic

Le curve ROC dei due modelli sono rappresentate in Figura 7.5 e 7.6. Il riferimento presente nei due grafici è la bisettrice in grigio che rappresenta la performance di un modello casuale senza alcuna capacità di classificazione, in quanto attribuisce a tutte le imprese valori casuali con un'accuratezza media del 50%. In entrambe le reti la curva si colloca al di sopra della bisettrice testimoniando una migliore performance rispetto al modello casuale ma comunque inferiore rispetto al modello perfetto. Inoltre, è possibile osservare andamenti molto simili caratterizzati da una monotonicità generale, anche se con inclinazioni variabili nel campo di esistenza dovuti alla casualità della distribuzione delle imprese sulla base degli output delle reti. A tal proposito, si noti il cambiamento di convessità tra lo 0,3 e lo 0,4 del *False Positive Rate* che si ripresenta in entrambi i grafici in quanto i *test set* delle reti sono composti dalle medesime aziende.

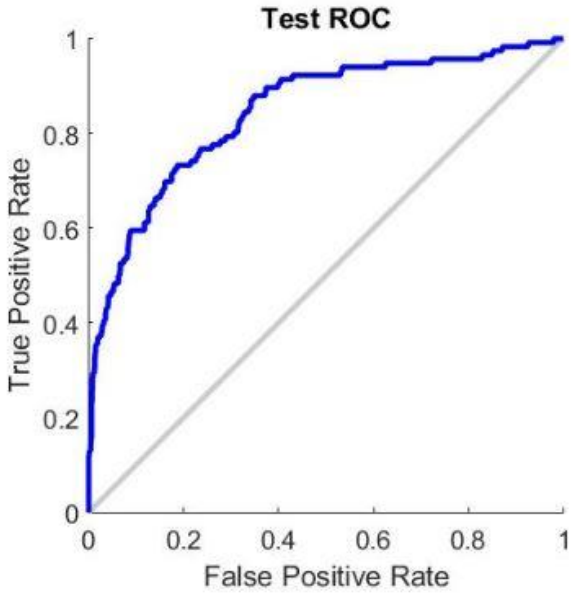

Figura 7.5 *Receiver Operating Characteristic generato da una rete formata da 6 neuroni di input, 3 strati nascosti con 8, 11, 7 neuroni rispettivamente e 1 neurone di output*

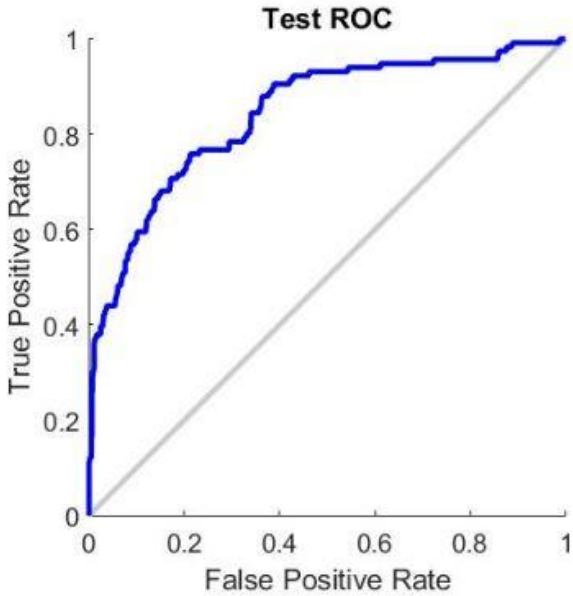

Figura 7.6 *Receiver Operating Characteristic generato da una rete formata da 9 neuroni di input, 3 strati nascosti con 12, 13, 11 neuroni rispettivamente e 1 neurone di output*

7.1.4 Best Validation Performance

I grafici di Figura 7.7 e 7.8 illustrano l'andamento del valore della *cross-entropy* nelle varie *epoch* e individuano il punto di minimo della *validation loss* che determina l'interruzione del processo di apprendimento.

In generale, l'andamento delle tre curve nei grafici suggerisce un *goodfit* da parte dei due modelli, infatti la *training* e la *validation loss* diminuiscono fino a raggiungere un punto di minimo oltre il quale si stabilizzano lasciando un *generalization gap* tra i valori finali assunti dalle curve. Dunque, è possibile affermare che le dinamiche delle curve non presentano le caratteristiche tipiche di un comportamento di *underfit* o di *overfit*. Infatti, il valore della *cross-entropy* non continua a diminuire per segnalare un'interruzione precoce del processo di addestramento e le curve non risultano né troppo distanti né troppo vicine. In sostanza, l'andamento delle funzioni di perdita denota delle buone capacità di generalizzazione e di apprendimento dai dati forniti in input.

Altro aspetto interessante legato al punto precedente è la dinamica della distanza tra le curve con il procedere delle varie *epoch*. Nella prima parte essa risulta nulla per poi attestarsi a livelli sempre più definiti e costanti. Tale andamento rispecchia quanto descritto dalla teoria infatti, nelle prime *epoch* nelle quali il modello incomincia ad apprendere e specializzarsi sui dati di addestramento riesce a ridurre significativamente il valore delle funzioni di perdita mentre, nelle fasi successive, l'apprendimento incrementale risulta inferiore e il valore finale assunto dalla *cross-entropy* dipende dalla composizione del campione nei tre set, aventi aziende con percentuali di default diverse.

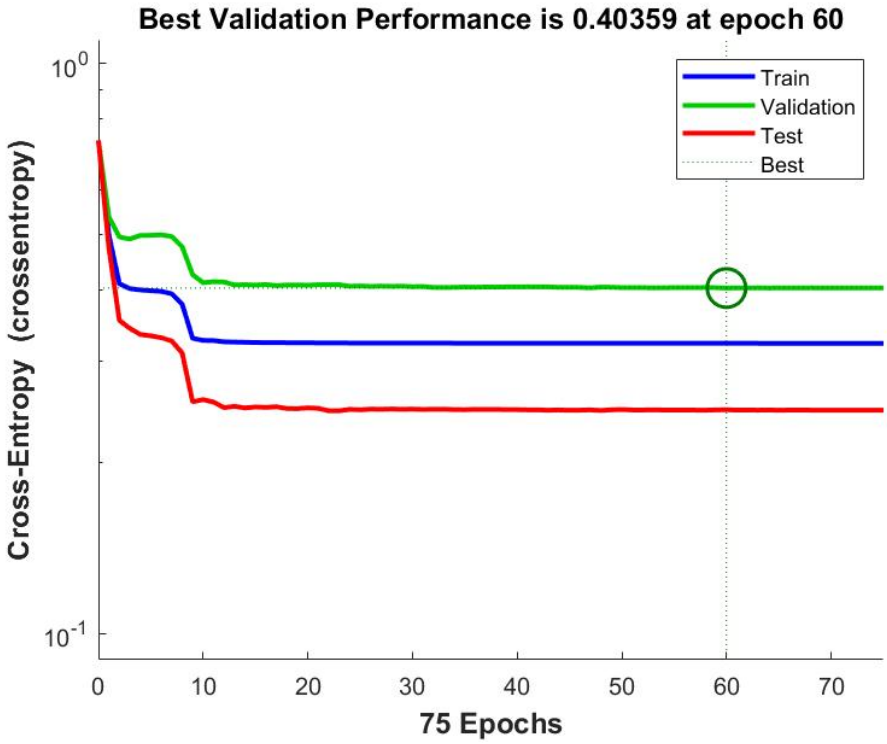

Figura 7.7 *Best Validation Performance generato da una rete formata da 6 neuroni di input, 3 strati nascosti con 8, 11, 7 neuroni rispettivamente e 1 neurone di output*

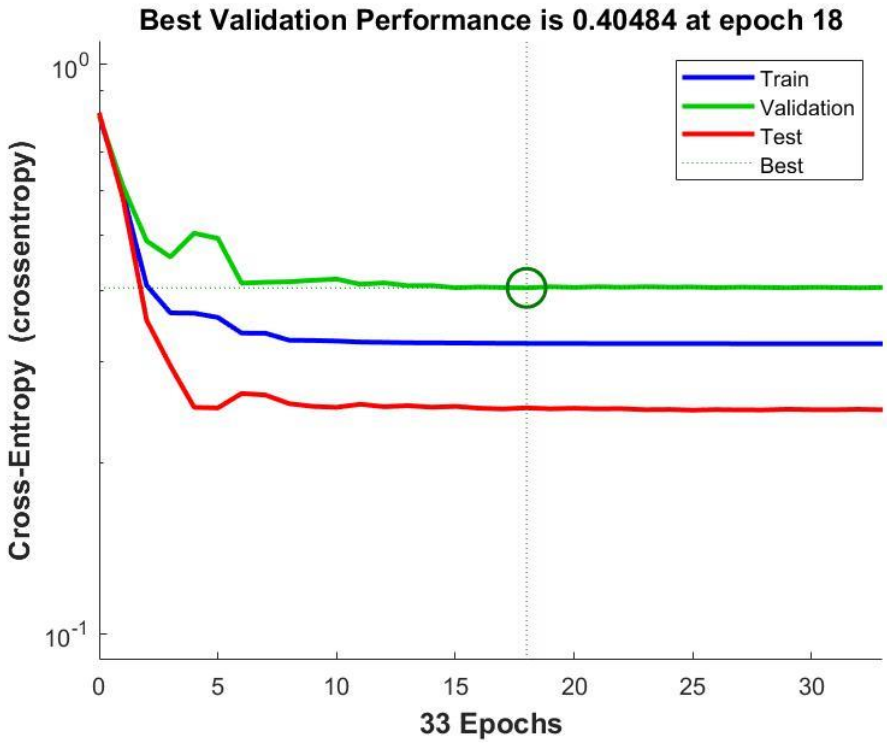

Figura 7.8 *Best Validation Performance generato da una rete formata da 9 neuroni di input, 3 strati nascosti con 12, 13, 11 neuroni rispettivamente e 1 neurone di output*

7.1.5 Gradient and Validation Check

Per esprimere valutazioni sulle performance della funzione di addestramento e sull'inizializzazione dei parametri è necessario analizzare i grafici di Figura 7.9 e 7.10. In entrambi i casi si denota un andamento di diminuzione del valore del gradiente con il procedere delle varie *epoch* confermando un buon funzionamento della funzione di addestramento *trainscg*. Tale funzione aggiorna i valori dei pesi e dei *bias* seguendo il metodo *scaled conjugate gradient* che basa la scelta della direzione di riduzione della funzione di perdita sul valore assunto dal gradiente. La corretta performance della funzione è confermata non solo dalla diminuzione del valore del gradiente ma anche dal fatto che questa riduzione non sia strettamente monotonica. Infatti, nel paragrafo §5.3 si è illustrato come il metodo *scaled conjugate gradient* rappresenti una variante rispetto alla versione più classica, che si limita a modificare i pesi nella direzione di massima diminuzione della funzione di perdita e quindi opposta al suo gradiente, ma ricerchi la direzione ottima tra quelle "coniugate" provocando una convergenza più rapida.

L'andamento del gradiente seppur in diminuzione in entrambi i grafici, manifesta caratteristiche differenti. Osservando la Figura 7.9 è possibile notare come l'andamento del gradiente sia meno lineare rispetto al grafico di Figura 7.10. Questo comportamento è dovuto dal diverso numero di *epoch* che ha caratterizzato l'addestramento dei due modelli. Per ottenere un valore sufficientemente basso della funzione di perdita il modello ricevente in input i valori di 6 indicatori ha dovuto protrarre la fase di allenamento per più del doppio delle *epoch* rispetto al modello da 9 input, facendone contare 75 rispetto alle 33, lasciando così intuire una maggiore difficoltà nel raggiungere un valore adeguato di *cross-entropy*. La conferma di tale difficoltà è riscontrabile anche nelle disparità evidenziate dai rispettivi *Validation Checks*, mentre per il modello con 9 input non si registrano valori superiori ai 5 fallimenti consecutivi, se non nel caso della sequenza che porta all'interruzione, con quello da 6 input si sfiorano 2 volte i 15 *validation fail* che causerebbero un'interruzione precoce dell'apprendimento.

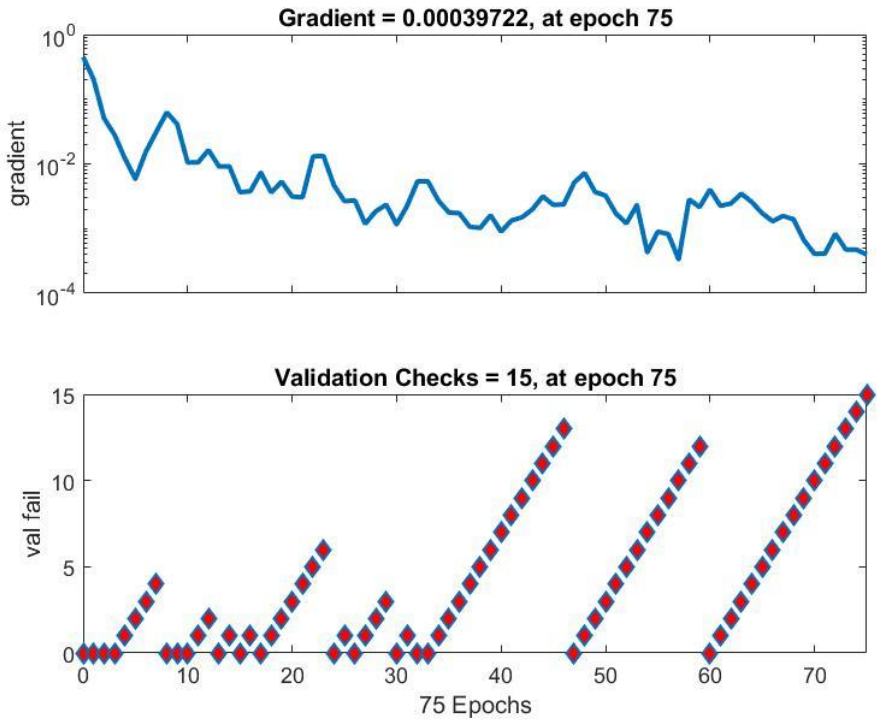

Figura 7.9 *Gradient and Validation Check generati da una rete formata da 6 neuroni di input, 3 strati nascosti con 8, 11, 7 neuroni rispettivamente e 1 neurone di output*

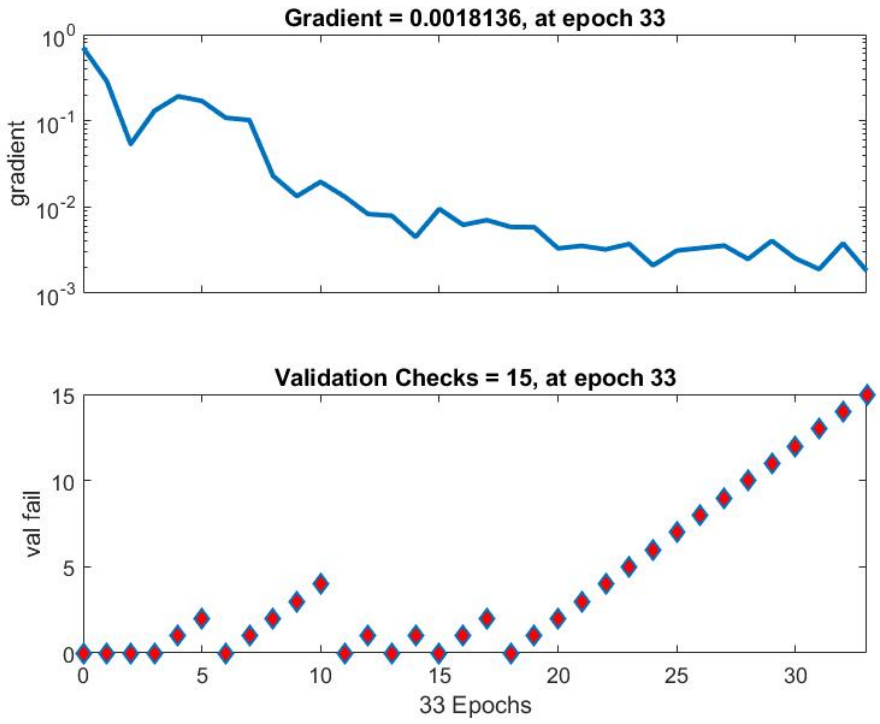

Figura 7.10 *Gradient and Validation Check generati da una rete formata da 9 neuroni di input, 3 strati nascosti con 12, 13, 11 neuroni rispettivamente e 1 neurone di output*

7.1.6 Error histogram

Gli ultimi due strumenti valutativi in analisi sono gli *Error histogram* di Figura 7.11 e 7.12. Dallo studio di questi diagrammi a barre si ottengono informazioni importanti riguardo la distribuzione degli errori calcolati come differenza tra target e output del modello.

In entrambe le reti si evidenzia una maggiore concentrazione degli errori nella parte sinistra del diagramma relativa ai valori negativi denotando una tendenza ad attribuire punteggi sistematicamente superiori al target. Questo comportamento è giustificabile in quanto solamente alle aziende più sane del campione saranno attribuiti punteggi prossimi allo 0 mentre per le aziende sane con peggiori valori di indicatori in ingresso si hanno output piccoli ma meno prossimi allo 0. Considerazioni analoghe possono essere svolte per la parte destra del diagramma relativa ai valori positivi degli errori e quindi alle aziende anomale. In aggiunta, va sottolineato come in questa parte di grafico i valori più estremi risultino in termini assoluti inferiori rispetto a quelli positivi. Questi valori denotano un comportamento da parte dei modelli piuttosto conservativo, infatti entrambe le reti tendono ad attribuire punteggi intermedi sulla base degli indicatori forniti in ingresso, soprattutto quando si tratta di classificare le imprese anomale per le quali evitano output prossimi ad 1. Tale atteggiamento è da tenere fortemente in considerazione durante il processo decisionale di concessione del credito alle aziende con punteggio intermedio, per le quali risultano necessari ulteriori approfondimenti.

Altri elementi a cui è necessario rivolgere particolare attenzione sono i valori inferiori a -0,5 e superiori a 0,5, poiché le reti neurali, per convertire il punteggio in classificazione binaria, arrotondano gli output all'intero più vicino e quindi solo gli errori maggiori in valore assoluto a 0,5 costituiscono reali errori di classificazione.

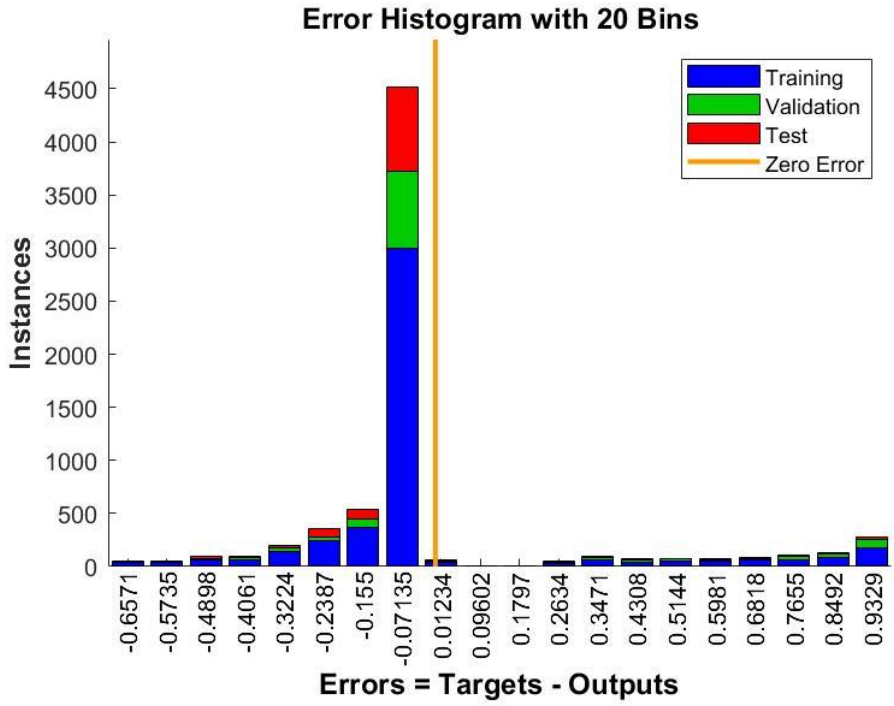

Figura 7.11 *Error histogram generato da una rete formata da 6 neuroni di input, 3 strati nascosti con 8, 11, 7 neuroni rispettivamente e 1 neurone di output*

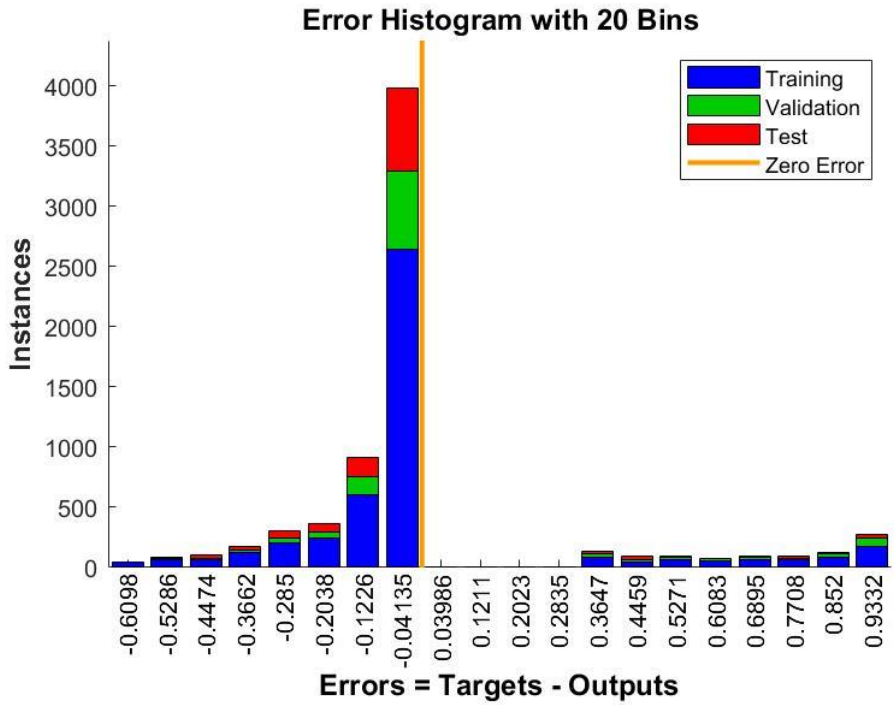

Figura 7.12 *Error histogram generato da una rete formata da 9 neuroni di input, 3 strati nascosti con 12, 13, 11 neuroni rispettivamente e 1 neurone di output*

7.2 Analisi economico-finanziaria degli errori

Un'attività rilevante del processo di valutazione delle performance del modello è l'analisi degli errori. È importante fornire un'interpretazione economico-finanziaria degli errori per poter valutare le logiche di funzionamento di un modello. Tale valutazione assume notevole importanza per le reti neurali la cui logica di elaborazione dei dati è spesso paragonata ad una *black box*, ossia una scatola nera che nasconde al proprio interno i meccanismi che ne regolano il funzionamento. Al fine di svelare il contenuto della *black box* si approfondisce l'analisi degli errori della rete più performante avente 9 neuroni di input, 3 strati nascosti con 12, 13, 11 neuroni rispettivamente e 1 neurone di output, riprendendo i valori della *confusion matrix* di Figura 7.4 e presentando quattro confronti, divisi in due gruppi, tra i valori medi degli indicatori delle aziende appartenenti al *test set*.

7.2.1 Confronto target class

I primi due confronti sono stati raggruppati nella categoria *target class* in quanto analizzano le differenze tra i valori degli indicatori delle classificazioni corrette ed errate appartenenti alla stessa *target class*, ossia tutte le società aventi stesso target ma alle quali la rete neurale ha assegnato output differenti.

7.2.1.1 Target 0: veri negativi e falsi positivi

Nel confronto target 0 rientrano le osservazioni delle società sane. Tali aziende possono essere classificate in due modi:

- veri negativi: società sane classificate correttamente, 1.024 su 1.040 (98,5%).
- falsi positivi: società sane classificate erroneamente, 16 su 1.040 (1,5%).

La percentuale di osservazioni che ricade nel campione dei veri negativi rappresenta la specificità del modello, concetto già introdotto nel paragrafo §6.1.1. I valori della matrice di confusione suggeriscono una buona capacità della rete nel riconoscere le società sane. Tale capacità deriva dall'elevato numero di società di questo tipo presenti nel *training set*, pari a 3.992 su 4.626. La rete, avendo a disposizione numerosi esempi di valori di indicatori di società sane, riesce a coglierne le logiche e replicare il ragionamento sul campione del *test set*.

È possibile analizzare nel dettaglio la composizione dei due campioni valutando le differenze tra i valori medi dei 9 indicatori usati come input della rete. Tali indicatori sono stati suddivisi in due tabelle sulla base della correlazione con il FLAG DI STATUS S/A – SOCIETA' per facilitarne il confronto. Gli indici di Tabella 7.1 sono negativamente correlati con il flag, ossia più elevati sono i valori di tali indicatori migliore è la condizione economico-finanziaria dell'azienda. Al contrario, la Tabella 7.2 mostra gli indicatori con correlazione positiva rispetto al flag e quindi maggiori sono questi valori peggiore è la situazione economico-finanziaria della società.

Procedendo con ordine, la Tabella 7.1 mostra importanti differenze tra i valori degli indicatori dei veri negativi e dei falsi positivi. Le società sane classificate correttamente sono caratterizzate da valori più elevati rispetto alle imprese che il modello non è stato in grado di riconoscere. In particolare, si riscontrano importanti differenze in termini di ROE e di $valore\ aggiunto\ operativo/ricavi$ pari rispettivamente a 0,7512 e 0,6674. Tali differenze e quindi tali indicatori hanno un impatto negativo sulla classificazione in quanto suggeriscono al modello che si tratta di osservazioni appartenenti a gruppi distinti mentre in realtà sono tutte le osservazioni di società sane. Al contrario, gli indicatori aventi una differenza inferiore tra i valori medi sono quelli che segnalano una maggior somiglianza tra le due classi. Tali indici, nel confronto target 0, sono quelli che se avessero avuto maggior peso all'interno del modello sarebbero stati in grado di ridurre gli errori. In questo caso le differenze minori si registrano con l'$autofinanziamento\ lordo/attivo\ netto$ a 0,3633 e il $patrimonio\ netto\ tangibile/debiti\ totali + patrimonio\ netto$ a 0,289.

Tabella 7.1 *Valori medi degli indicatori correlati negativamente con il* FLAG DI STATUS S/A – SOCIETA': *confronto target 0*

	ROE	val agg oper / ricavi	riserve + utile / AN	risultato netto rettif / ricavi	ln(ricavi)	autof lordo / AN	patr netto tan / debiti tot + PN
veri negativi	0,6824	0,7515	0,6984	0,8932	0,5947	0,6494	0,6329
falsi positivi	0,0150	0,0004	0,1927	0,3739	0,0693	0,2861	0,3438
differenza	0,6674	0,7512	0,5056	0,5192	0,5254	0,3633	0,2890

Entrambi gli indicatori di Tabella 7.2 segnalano differenze rilevanti tra le osservazioni delle imprese sane classificate correttamente e quelle errate. In particolare, la differenza per le $passività\ correnti/ricavi$ vale -0,5333 mentre per gli $oneri\ finanziari\ netti/EBIT$ è pari a -0,6647. Per quest'ultimo indicatore è importante notare che, il valore medio dei falsi positivi ossia delle società sane classificate erroneamente, è uguale ad 1, ciò implica che, essendo gli input del modello normalizzati ad 1, tutte le osservazioni sane classificate erroneamente hanno un valore di $oneri\ finanziari\ netti/EBIT$ pari al valore massimo.

Tabella 7.2 *Valori medi degli indicatori correlati positivamente con il* FLAG DI STATUS S/A – SOCIETA': *confronto target 0*

	PC / ricavi	OFN / EBIT
veri negativi	0,0563	0,3353
falsi positivi	0,5896	1,0000
differenza	-0,5333	-0,6647

7.2.1.2 Target 1: veri positivi e falsi negativi

Nel confronto target 1 si studiano le osservazioni delle società anomale. Anche in questo caso le aziende possono essere classificate in due modi:

- veri positivi: società anomale classificate correttamente, 43 su 116 (37,1%).
- falsi negativi: società anomale classificate erroneamente, 73 su 116 (62,9%).

La prima percentuale, riferita alle osservazioni che ricadono nel campione dei veri positivi, rappresenta la sensibilità del modello. Al contrario rispetto a quanto avveniva con le società sane, si denota una certa difficoltà da parte della rete nel riconoscere le società anomale. Questa problematicità costituisce il vero limite del modello ed è causata dal minor numero di aziende anomale fornite in input alla rete. Infatti, solamente 634 delle 4.626 osservazioni del *training set* sono aziende anomale ed è per questo motivo che il modello avrà maggiore difficoltà nel riconoscere quali sono i valori degli indicatori che rendono un'azienda anomala.

Si approfondisce la valutazione dei due campioni analizzando le differenze tra i valori medi dei 9 indicatori usati come input della rete. Tra gli indicatori di Tabella 7.3 quelli che segnalano una maggiore differenza tra i veri positivi ed i falsi negativi sono le $riserve + utile/attivo\ netto$ con -0,5323 e il $valore\ aggiunto\ operativo/ricavi$ con -0,4862. Tali indicatori forniscono un contributo negativo al modello in quanto segnalano un'importante differenza tra le osservazioni che si sarebbero invece dovute classificare tutte come anomale. In senso contrario operano il $\ln(ricavi)$ e l'$autofinanziamento\ lordo/\ attivo\ netto$ che con le differenze di -0,3207 e -0,2776 suggeriscono al modello possibili somiglianze tra i due campioni.

Tabella 7.3 *Valori medi degli indicatori correlati negativamente con il* FLAG DI STATUS S/A – SOCIETA': *confronto target 1*

	ROE	val agg oper / ricavi	riserve + utile / AN	risultato netto rettif / ricavi	ln(ricavi)	autof lordo / AN	patr netto tan / debiti tot + PN
veri positivi	0,0349	0,1184	0,0397	0,3353	0,1164	0,1830	0,1511
falsi negativi	0,4058	0,6046	0,5721	0,7974	0,4370	0,4606	0,5602
differenza	-0,3710	-0,4862	-0,5323	-0,4621	-0,3207	-0,2776	-0,4091

Per quanto riguarda gli indicatori positivamente correlati con il flag, la Tabella 7.4 evidenzia il loro contributo opposto. Mentre le $passività\ correnti/ricavi$ indicano una rilevante differenza di 0,6316 tra le società anomale, gli $oneri\ finanziari\ netti/EBIT$ individuano una differenza minore pari a 0,3548. Si noti che, come per il confronto target 0, le imprese anomale classificate correttamente sono tutte caratterizzate da un valore di $oneri\ finanziari\ netti/EBIT$ uguale al valore massimo in quanto il valore medio di quest'indicatore è 1.

Tabella 7.4 *Valori medi degli indicatori correlati positivamente con il* FLAG DI STATUS S/A – SOCIETA': *confronto target 1*

	PC / ricavi	OFN / EBIT
veri positivi	0,7112	1,0000
falsi negativi	0,0796	0,6452
differenza	0,6316	0,3548

7.2.2 Confronto output class

I due confronti rimanenti sono stati raggruppati nella categoria *output class* in quanto analizzano le differenze tra i valori degli indicatori delle classificazioni corrette ed errate appartenenti alla stessa *output class*, ossia tutte le società alle quali la rete neurale ha assegnato lo stesso output ma aventi target differenti.

7.2.2.1 Output 0: veri negativi e falsi negativi

Nel confronto output 0 rientrano le osservazioni delle società che il modello classifica come sane. Tali aziende possono essere classificate in due modi:

- veri negativi: società sane classificate correttamente, 1.024 su 1.097 (93,3%).
- falsi negativi: società anomale classificate erroneamente, 73 su 1.097 (6,7%).

La percentuale di osservazioni che ricade nel campione dei falsi negativi rappresenta la probabilità di mancato allarme, ossia la possibilità che la rete non riconosca l'azienda anomala classificandola come sana. I valori della matrice di confusione sottolineano ancora una volta la capacità della rete di riconoscere le aziende sane in linea con quanto avviene con il confronto target 0. Anche in questo caso il modello, avendo a disposizione un elevato numero di società sane sulle quali imparare, riesce a identificare i valori degli indicatori che rendono un'azienda sana. Il mancato allarme può essere considerato come una tipologia di errore più costosa da commettere in quanto, considerare una società sana come anomala implica non solo una concessione del credito ad un'azienda che difficilmente riuscirà a ripagarlo ma anche l'applicazione di un tasso di interesse che non rispetta il rischio dell'operazione. Nel caso peggiore l'azienda fallirà non ripagando i debiti provocando una perdita pari all'ammontare di credito concesso, al contrario, nel caso in cui l'azienda riesca ugualmente a ripagare il credito la perdita sarà la mancata retribuzione del rischio corso.

L'analisi della composizione dei due campioni, attraverso la valutazione delle differenze tra i valori medi dei 9 indicatori usati come input della rete, si fa più complessa rispetto alla categoria target. Infatti, nei confronti output, le differenze tra i vari indicatori risultano inferiori poiché si tratta di osservazioni a cui il modello assegna lo stesso output. Tale comportamento trova riscontro nei valori di Tabella 7.5 ed in particolare con le differenze minime di $risultato\ netto\ rettificato/ricavi$ 0,0958 e $patrimonio\ netto\ tangibile/debiti\ totali + patrimonio\ netto$ 0,0727. Al contrario, il ROE tra gli indicatori

negativamente correlati con il flag è l'unico che segnala una maggior differenza tra i valori medi dei due campioni pari a 0,2766.

Tabella 7.5 *Valori medi degli indicatori correlati negativamente con il* FLAG DI STATUS S/A – SOCIETA': *confronto output 0*

	ROE	val agg oper / ricavi	riserve + utile / AN	risultato netto rettif / ricavi	ln(ricavi)	autof lordo / AN	patr netto tan / debiti tot + PN
veri negativi	0,6824	0,7515	0,6984	0,8932	0,5947	0,6494	0,6329
falsi negativi	0,4058	0,6046	0,5721	0,7974	0,4370	0,4606	0,5602
differenza	0,2766	0,1470	0,1263	0,0958	0,1576	0,1888	0,0727

Gli indicatori positivamente correlati con il flag in Tabella 7.6 manifestano due comportamenti opposti. Le $passivit$à $correnti/ricavi$, in linea con gran parte degli altri indicatori, segnalano una differenza minima di -0,0232 evidenziando la somiglianza tra i due campioni. Invece, l'indicatore $oneri\ finanziari\ netti/EBIT$ con una differenza di -0,31 tra i valori medi è l'indice che più di tutti ha colto una possibile differenza tra i due campioni.

Tabella 7.6 *Valori medi degli indicatori correlati positivamente con il* FLAG DI STATUS S/A – SOCIETA': *confronto output 0*

	PC / ricavi	OFN / EBIT
veri negativi	0,0563	0,3353
falsi negativi	0,0796	0,6452
differenza	-0,0232	-0,3100

Dal confronto output 0 emerge che i valori medi degli indicatori delle 73 osservazioni anomale classificati erroneamente dalla rete neurale come sane sono economicamente peggiori rispetto a quelli delle imprese sane classificate correttamente. Questa osservazione suggerisce un'interpretazione più severa dei punteggi attribuiti dal modello alle aziende ossia una maggiore attenzione alla concessione del credito alle società con punteggi prossimi al valore di soglia 0,5 soprattutto se possiedono un ROE e degli $oneri\ finanziari\ netti/EBIT$ che discostano di molto dalle aziende che si considerano sicuramente più sane.

7.2.2.2 Output 1: veri positivi e falsi positivi

Nel confronto output 1 si studiano le osservazioni delle società che il modello classifica come anomale. Tali aziende possono essere classificate in due modi:

- veri positivi: società anomale classificate correttamente, 43 su 59 (72,9%).
- falsi positivi: società sane classificate erroneamente, 16 su 59 (27,1%).

La percentuale di osservazioni che ricade nel campione dei falsi positivi costituisce la probabilità di falso allarme, ossia la possibilità che la rete non riconosca l'azienda sana classificandola come anomala. I valori della matrice di confusione confermano l'affidabilità degli output della rete anche quando la risposta è 1, infatti, si conta un maggior numero di società classificate correttamente rispetto a quelle errate. Tuttavia, rispetto al caso precedente, la percentuale di output errati aumenta rimarcando ancora una volta la necessità di aumentare gli esempi di società anomale da fornire in input al modello per identificazione più precisa. In concreto l'errore di falso allarme si traduce nella mancata concessione del credito ad un'azienda poiché la si ritiene ingiustamente rischiosa. Anche a questo errore, seppur meno grave rispetto al precedente, sono associati dei costi quantificabili attraverso la valutazione dei mancati ricavi che derivano dalla rinuncia alla concessione del prestito. Considerazioni riguardanti la differenza di importanza degli errori previsionali commessi dalla rete consentono di pesare le conseguenze e di prendere decisioni relative alle società da analizzare con maggior consapevolezza.

L'analisi delle differenze tra i valori medi dei 9 indicatori usati come input della rete relative ai campioni del confronto output 1 è la più complessa in assoluto. Tale difficoltà deriva sia dai bassi valori delle differenze tra i due campioni sia dal loro segno. Essendo la differenza di Tabella 7.7 calcolata come veri positivi meno falsi positivi, l'operazione dovrebbe restituire valori negativi in quanto le società sane sono caratterizzate da valori medi di tali

indicatori superiori rispetto alle società anomale. Questa situazione non si verifica per il ROE 0,0199, il $valore\ aggiunto\ operativo/ricavi$ 0,118 e il $\ln(ricavi)$ 0,0471, per i quali le aziende anomale possiedono valori medi superiori rispetto alle sane in contrapposizione con il loro significato economico. Tra gli indicatori i cui valori sono in accordo con le teorie economiche quello che evidenzia una possibile differenza tra i due campioni è il $patrimonio\ netto\ tangibile/debiti\ totali + patrimonio\ netto$ che segna un -0,1928.

Tabella 7.7 *Valori medi degli indicatori correlati negativamente con il* FLAG DI STATUS S/A – SOCIETA': *confronto output 1*

	ROE	val agg oper / ricavi	riserve + utile / AN	risultato netto rettif / ricavi	ln(ricavi)	autof lordo / AN	patr netto tan / debiti tot + PN
veri positivi	0,0349	0,1184	0,0397	0,3353	0,1164	0,1830	0,1511
falsi positivi	0,0150	0,0004	0,1927	0,3739	0,0693	0,2861	0,3438
differenza	0,0199	0,1180	-0,1530	-0,0387	0,0471	-0,1031	-0,1928

Per quanto riguarda gli indicatori positivamente correlati con il flag di Tabella 7.8, i segni delle differenze tra i valori medi dei due campioni risultano coerenti con l'interpretazione economica. Tuttavia, mentre per le $passività\ correnti/ricavi$ si registra una differenza positiva di 0,1216 tra i due campioni, per gli $oneri\ finanziari\ netti/EBIT$ tale differenza si annulla. Infatti, entrambi i campioni sono caratterizzati da un valore medio di quest'ultimo indicatore pari ad 1, ciò implica che il modello ha classificato come anomale tutte le società aventi $oneri\ finanziari\ netti/EBIT$ pari al valore massimo.

Tabella 7.8 *Valori medi degli indicatori correlati positivamente con il* FLAG DI STATUS S/A – SOCIETA': *confronto output 1*

	PC / ricavi	OFN / EBIT
veri positivi	0,7112	1,0000
falsi positivi	0,5896	1,0000
differenza	0,1216	0,0000

Contrariamente rispetto a quanto accadeva con il confronto output 0 non si può affermare che le 16 osservazioni sane classificate erroneamente dalla rete neurale come anomale abbiano tutti gli indicatori che suggeriscono una situazione economica migliore rispetto alle anomale. In questo senso gli errori del modello sono giustificabili in quanto i valori risultano in contrapposizione con le teorie economiche. Inoltre, quest'osservazione fa riflettere sul possibile futuro di queste aziende che, nonostante non abbiano ancora subito alcuna procedura e quindi siano per definizione sane, la loro situazione economica molto simile alle società anomale indica che potrebbero subirne una in futuro. Dati i molteplici dubbi sulla situazione economico-finanziaria di questo tipo di aziende, il modello suggerisce di non concedere il credito.

7.3 Osservazioni per ulteriori miglioramenti

L'analisi dei risultati delle reti costruite ha fatto emergere una maggiore difficoltà da parte dei modelli nel classificare correttamente le aziende anomale rispetto a quelle sane. Una tra le principali cause di tale problematicità è la differenza di numerosità tra le aziende sane e anomale fornite in input al modello. In particolare, il numero di società anomale analizzato risulta significativamente inferiore rispetto a quello delle sane. La scarsa disponibilità di dati di società anomale si converte in una maggiore difficoltà da parte del modello nel riconoscere i valori degli indicatori che caratterizzano questo status. Il campione in analisi rappresenta però uno scenario reale nel quale verosimilmente il numero di imprese che hanno subito una procedura nel corso del loro periodo di attività è inferiore rispetto a quelle considerate come

sane. Per migliorare le performance dei modelli, una tecnica che potrebbe rivelarsi efficace è il *bootstrapping*, ossia un metodo statistico di ricampionamento basato sulla sostituzione per reimmissione delle osservazioni nel campione di partenza per approssimare al meglio la distribuzione di probabilità di una data statistica. Al contrario del campionamento casuale semplice, in ogni estrazione *bootstrap*, le osservazioni possono essere estratte più di una volta per comporre il campione finale in modo che ogni osservazione ha sempre la stessa probabilità di essere estratta. La reimmissione dei dati delle società anomale potrebbe rendere comprensibili al modello alcuni comportamenti che prima non lo erano in quanto poco rappresentati. Tuttavia, la modifica del campione in analisi rimane uno strumento assai pericoloso poiché distorce l'originalità dei dati con il rischio che il campione non sia più rappresentativo della situazione reale.

Un'altra strategia che potrebbe apportare ulteriori migliorie consiste nell'individuare all'interno del campione di partenza, gruppi di società con caratteristiche simili di fatturato o regione di provenienza dedicando ad ognuno di essi la costruzione di una rete neurale. Il campione di partenza dal momento che include tutte le imprese italiane appartenenti al settore del noleggio e leasing di autoveicoli risulta molto ampio e annovera al suo interno un gran numero di realtà differenti. Costruire una rete neurale per ogni gruppo in cui si divide il campione consente di ottenere modelli capaci di cogliere anche gli aspetti più specifici che altrimenti andrebbero persi. L'applicazione di questa strategia va limitata solamente ai casi in cui il campione in analisi risulta piuttosto eterogeneo a causa dell'elevato numero di aziende che comprende al suo interno. Inoltre, costruire più modelli con un minor numero di osservazioni risulta controproducente se l'obiettivo finale è ottenere un modello con una buona capacità di generalizzazione.

Le osservazioni precedenti sono solo alcune dei possibili miglioramenti applicabili alle reti. Infatti, per la costruzione di qualsiasi modello di *machine learning* non sono ancora state espresse regole in grado di garantire un sicuro miglioramento delle performance. Questa mancanza è la causa delle numerose prove finalizzate alla ricerca della struttura ottimale, delle funzioni più performanti, dell'inizializzazione dei parametri adeguata e della migliore composizione del campione. L'assenza di regole precise è mitigata dalla presenza di numerose indicazioni e linee guida maturate dall'esperienza sul campo e confermate dai risultati ottenuti. Il notevole successo delle reti neurali come strumento di analisi dei casi di studio deriva dalla loro caratteristica principale: la versatilità. Attraverso la personalizzazione dei singoli elementi che compongono la rete è possibile adattare il

modello alle proprie esigenze rendendolo uno strumento in grado di fornire una risposta a qualsiasi necessità.

7.4 Come utilizzare il modello

Con lo studio degli errori della rete si sono concluse le analisi sul funzionamento e sulle performance del modello. Dal momento che si conoscono tutti gli aspetti della migliore rete costruita è possibile formulare dei ragionamenti riguardo un suo eventuale impiego. Al momento, i principali fruitori dei modelli di *machine learning* sono le imprese che devono prendere decisioni analizzando grandi quantità di dati. In particolare, il modello in questione troverebbe applicazione nel processo decisionale di concessione del credito nei confronti delle aziende che si occupano di leasing e noleggio di autoveicoli.

Inquadrato il contesto nel quale potrebbe operare il modello non resta che analizzare operativamente il suo utilizzo. Il primo step di valutazione dello status di un'azienda consiste nel consultare l'output del modello. Fermarsi a questo punto significherebbe affidarsi totalmente alla performance del modello sapendo che la sua accuratezza è del 92.3%. Essendo consapevoli che la rete non è esente dal commettere errori e avendo preziose informazioni sulla distribuzione di quest'ultimi è possibile procedere alla mitigazione del rischio di credito attraverso ulteriori analisi. Uno di questi studi consiste nell'associare agli output binari i punteggi calcolati dalla rete fornendo un maggiore livello di dettaglio alla lettura dei risultati. Nel paragrafo §5.4 si è mostrato come i punteggi calcolati dalla rete siano arrotondati all'intero più vicino per poterli confrontare con i target binari, per quest'analisi è necessario fare un passo indietro e risalire ai valori di risposta originali. Tali punteggi, compresi tra 0 e 1, sono da interpretare non più come lo status certo di un'azienda ma come la probabilità di aver subito, o di stare per subire, una procedura. Questa modalità di lettura del dato permette di individuare le osservazioni la cui valutazione da parte del modello è incerta, ovvero quelle aventi punteggio in un intorno di 0,5. Individuate queste società, si può approfondirne l'analisi impiegando i tradizionali strumenti valutativi di analisi di bilancio. Applicando questo metodo alla rete costruita si sarebbero potute evitare 19 classificazioni errate, di cui 11 sane e 8 anomale.

Altre importanti valutazioni possono essere effettuate dopo aver analizzato i risultati della *confusion matrix*, studiate le inclinazioni del modello nel formulare le risposte e compreso che non tutti gli errori hanno lo stesso costo. Non concedere il credito ad un'azienda che si pensa sana ma in realtà è anomala ha un costo inferiore rispetto a concedere un credito ad un'azienda anomala che si riteneva sana. Il costo del caso iniziale, o errore di primo tipo, è il mancato profitto derivante dalla concessione del credito mentre l'errore di secondo tipo, ovvero la concessione del credito ad un'azienda anomala che si considerava sana, può provocare una perdita massima pari all'ammontare di credito concesso. Considerazioni di questo tipo consentono di restringere il campo delle società da analizzare tramite gli strumenti tradizionali e quindi i tempi e i costi ad essi associati.

Applicando entrambe le considerazioni precedenti alla migliore rete costruita si ottiene la seguente procedura decisionale riguardante la valutazione di concessione del credito ad un nuovo gruppo di aziende dello stesso settore:

- Valutare il nuovo gruppo di società mediante la rete neurale;
- Per le società classificate come anomale (output = 1) non concedere il credito;
- Per le società classificate come sane (output = 0) se l'impresa ha un punteggio in un intorno di 0,5 approfondire l'analisi mediante strumenti tradizionali altrimenti concedere il prestito.

Questa procedura può essere considerata come soluzione intermedia tra le metodologie di analisi più classiche e quelle che prevedono l'utilizzo esclusivo della nuova tecnologia. La scelta della metodologia più adatta è da effettuarsi sulla base dei vincoli di tempi, costi e qualità della previsione. In particolare, l'adozione della procedura proposta in questo paragrafo è da preferire nei casi in cui è richiesta un'elevata qualità dei risultati avendo a disposizione sia un budget in grado di coprire i costi del lavoro di analisti di bilancio e di programmatori sia tempi sufficienti per la raccolta e la gestione dei dati, la costruzione del modello e la valutazione economica delle società incerte.

Conclusioni

Il presente lavoro di tesi ha messo in luce le potenzialità dell'approccio di *machine learning* delle reti neurali nella costruzione di modelli di rischio di credito. Attraverso una prima trattazione teorica, si sono illustrati il contesto storico e le motivazioni concettuali che rendono la valutazione del rischio di credito un'attività imprescindibile nel mondo finanziario. L'obiettivo finale della costruzione di una rete neurale capace di riconoscere lo status di un'impresa attraverso l'analisi dei suoi indicatori economico-finanziari principali è stato perseguito non solo tramite la costruzione del modello utilizzando il software MATLAB ma anche attraverso lo studio degli elementi che ne regolano il funzionamento, l'accurata lettura degli strumenti valutativi e l'analisi degli errori.

Il carattere innovativo degli algoritmi di *machine learning* unito alla versatilità delle reti neurali sono stati i motivi che hanno guidato la scelta dell'argomento del presente lavoro di tesi. Procedendo con lo studio e l'analisi dei risultati si sono riscontrate alcune importanti caratteristiche che rendono le reti neurali un valido strumento di analisi dei dati. La principale qualità emersa è la notevole capacità di adattamento, le reti infatti sono costituite da diversi parametri la cui personalizzazione ne conferisce capacità descrittive superiori rispetto ai classici modelli di classificazione. Nel presente lavoro, si è sfruttata questa qualità caratterizzando ogni elemento del modello in modo da rappresentare al meglio il problema della valutazione dello status di un'azienda del settore del noleggio e leasing di autoveicoli. La personalizzazione più importante riguarda l'architettura della rete conclusa con la scelta di 9 neuroni di input, 3 strati nascosti con 12, 13, 11 neuroni rispettivamente e 1 neurone di output. Oltre alla struttura, sempre per necessità rappresentative, le altre personalizzazioni hanno riguardato: la scelta dell'*hyperbolic tangent function* per l'attivazione, della *scaled conjugate gradient* per l'addestramento, della *cross-entropy* come *loss function*, la divisione del campione di analisi in *training*, *validation* e *test set*, l'inizializzazione dei pesi a 0,1 e dei *bias* a 0.

Inoltre, si è potuto osservare come le reti rappresentino uno strumento di analisi più evoluto rispetto ai modelli statistici di regressione permettendo una classificazione più efficiente dal punto di vista della gestione del dato. A tal proposito, si è evidenziata una procedura di impiego della rete costruita che velocizza il processo di selezione delle imprese da analizzare con maggior attenzione.

La procedura si basa sulle capacità predittive della migliore rete costruita e suggerisce una corretta lettura dei risultati del modello e i successivi passi del processo decisionale di valutazione del merito creditizio. L'impiego del modello come step iniziale del processo decisionale costituisce il vero vantaggio della scelta delle reti neurali. L'uso di questi modelli aiuta l'analista finanziario riducendone i tempi di lavoro senza sostituirlo ma supportandolo in una scrematura iniziale delle aziende da analizzare. Infatti, nonostante la rete costruita classifichi correttamente 1.067 su 1.156 aziende del *test set* facendo registrare un'accuratezza del 92,3%, esistono ulteriori margini di miglioramento che solo l'analista finanziario è in grado di ridurre tramite accurate valutazioni del merito creditizio. In tal senso i modelli possono essere intesi come strumenti di affiancamento per il controllo dell'uniformità e della correttezza delle valutazioni formulate sulle diverse società dagli analisti finanziari.

L'adozione delle reti neurali come strumento d'ausilio per la valutazione del rischio di credito ha evidenziato alcuni limiti e complicazioni soprattutto di natura operativa. Procedendo con ordine, la prima attività a richiedere un grande impiego di tempo e quindi di costi è l'operazione di predisposizione dei dati. Tale attività rappresenta il punto di partenza per la costruzione di ogni modello e comprende tutte le operazioni di *data cleaning* ovvero di pulizia, completamento e correzione dei dati in analisi. La sua importanza è dovuta al fatto che un possibile errore in questa fase è difficilmente recuperabile e si ripercuote sull'accuratezza e sulle performance del modello. Per queste ragioni è opportuno dedicarvi il giusto tempo svolgendo accurati controlli in modo da avere la certezza di star operando con un *dataset* completo e corretto. A seguire, un altro processo che necessita di tempo per essere portato a termine è la *feature selection* ovvero la scelta delle migliori variabili economiche capaci di cogliere le relazioni trai i dati e la variabile di risposta. Tale scelta è effettuata sulla base dei valori assunti dalla matrice di correlazione tra indicatori e flag e dei risultati ottenuti in risposta alle variazioni degli indicatori forniti in input. La determinazione delle variabili in ingresso si rivela un'operazione lunga ma indispensabile per individuare il minor numero di indicatori capaci di fornire una visione completa del contesto da descrivere. L'ultimo processo che richiede un notevole dispendio di tempo è la definizione dell'architettura della rete. La mancanza di regole teoriche riguardanti la costruzione di una rete neurale rende necessarie numerose prove per individuare la struttura ottimale. Questi tentativi costituiscono l'attività più dispendiosa in termini di risorse temporali in quanto è possibile costruire reti aventi un qualsiasi numero di *hidden layers* e di neuroni in essi

contenuti. Inserendo un maggior numero di strati e neuroni nascosti si aumentano i gradi di libertà del modello rischiando un comportamento di *overfit*, al contrario con un numero esiguo di strati e neuroni nascosti si ottengono scarse performance. Riassumendo, il principale ostacolo all'adozione delle reti neurali come modello previsionale è rappresentato dal considerevole impiego di risorse temporali associato alle attività di sviluppo del modello quali il *data cleaning*, la *feature selection* e la definizione della composizione degli *hidden layers*.

Oltre che per le limitazioni di natura temporale, le reti neurali sono criticate per il loro funzionamento a *black box*. Nonostante questi modelli siano in grado di produrre output con elevati livelli di accuratezza, difficilmente si riesce a ricostruire la logica dietro l'elaborazione dei dati. In questo senso le reti neurali, ma più in generale i modelli di intelligenza artificiale, sono da considerarsi al pari di scatole nere che nascondono il contenuto al loro interno. Nonostante l'efficienza ormai accertata degli algoritmi di *machine learning*, rimane fondamentale non dedicarsi esclusivamente all'analisi degli input e output ma è necessario addentrarsi, per quanto possibile, nella complessità strutturale dei modelli cercando di comprendere le ragioni delle previsioni in modo da sviluppare una coscienza nell'utilizzo di questi strumenti. A tal proposito, l'analisi delle differenze tra i valori medi dei 9 indicatori utilizzati come input dalla rete ha messo in luce il contributo degli indicatori nella determinazione dell'output della rete neurale verificando la correttezza del funzionamento del modello da un punto di vista economico. Infatti, si riscontra una coerenza generale, derivante dal fatto che gli indicatori negativamente correlati con il flag delle osservazioni classificate come sane sono caratterizzate da valori medi maggiori rispetto alle aziende considerate come anomale mentre gli indicatori positivamente correlati con il flag delle aziende classificate come sane possiedono valori medi inferiori rispetto alle società identificate come anomale.

Al termine di questo studio risulta evidente come le reti neurali grazie alle loro capacità di apprendimento, classificazione e generalizzazione rappresentino uno strumento di analisi del rischio di credito con grandi potenzialità. Nonostante siano necessari ulteriori studi per superare i diversi limiti che le caratterizzano, le reti neurali artificiali si confermano un ottimo strumento di valutazione del merito creditizio, infatti, attraverso l'analisi delle loro performance, si è fatto prova della capacità di cogliere i principali elementi che determinano lo status delle aziende appartenenti al settore del noleggio e leasing di autoveicoli. Le necessità di gestione di grandi quantità di dati, la crescente esperienza maturata sul campo,

l'evoluzione della tecnologia e della potenza di calcolo giocheranno un ruolo fondamentale nel superamento dei limiti che oggi affliggono questi modelli, oltrepassati i quali, gli algoritmi di *machine learning*, ed in particolar modo le reti neurali, vivranno da protagonisti la rivoluzione digitale che sta coinvolgendo il settore finanziario.

Bibliografia e sitografia

[1] Andrea Resti Andrea Sironi, *"Rischio e valore nelle banche. Risk management e capital allocation"*, EGEA, Milano, II edizione, 13 agosto 2008.

[2] Comitato di Basilea per la vigilanza bancaria, *"Nuovo accordo di Basilea sui requisiti patrimoniali"*, Banca dei regolamenti internazionali, Basilea, aprile 2003, https://www.bis.org/bcbs/cp3fullit.pdf.

[3] Banca d'Italia, *"Nuove disposizioni di vigilanza prudenziale per le banche"*, Circolare n. 263 del 27 dicembre 2006, 15° aggiornamento, 2 luglio 2013, https://www.bancaditalia.it/compiti/vigilanza/normativa/archivio-norme/circolari/c263/Circ_263_annotata_V03.pdf.

[4] Comitato di Basilea per la vigilanza bancaria, *"Basilea 3 – Schema di regolamentazione internazionale per il rafforzamento delle banche e dei sistemi bancari"*, *"Basilea 3 – Riforme del Comitato di Basilea per la vigilanza bancaria*", Banca dei regolamenti internazionali, Basilea, dicembre 2010 (aggiornamento al giugno 2011), https://www.bis.org/publ/bcbs189_it.pdf, https://www.bis.org/bcbs/basel3/b3summarytable_it.pdf.

[5] Marco Ferfoglia, *"Basilea4: il framework normativo"*, Risk & Compliance, 15 luglio 2019, http://www.riskcompliance.it/news/basilea4-il-framework-normativo/.

[6] Dispense dei corsi di Economia degli intermediari finanziari e Analisi finanziaria e creditizia per l'impresa tenuti dal professore Franco Varetto nell'anno accademico 2018/2019.

[7] Vincenzo Paolo Senese, *"Regressione Multipla e Regressione Logistica: concetti introduttivi ed esempi"*, Dipartimento di Psicologia Università degli Studi della Campania, I edizione, Caserta, ottobre 2006, http://psiclab.altervista.org/MetTecPsicClinica2017/2.1.RegressioneMultipla_Logistica2016.pdf.

[8] Ayyüce Kızrak, *"Comparison of Activation Functions for Deep Neural Networks"*, Towards Data Science, 9 maggio 2019, https://towardsdatascience.com/comparison-of-activation-functions-for-deep-neural-networks-706ac4284c8a.

[9] Sovit Ranjan Rath, *"Activation Functions in Neural Networks"*, Degbugger cafe, 1 aprile 2019, https://debuggercafe.com/activation-functions-in-neural-networks/.

[10] Vishnu Kakaraparthi, *"Activation Functions in Neural Networks – What are they? How they work? Where to use them?"*, Towards Data Science, 8 febbraio 2019, https://medium.com/@prateekvishnu/activation-functions-in-neural-networks-bf5c542d5fec.

[11] Dispense del corso di Machine Learning dell'Università degli Studi di Bologna tenuto dal professore Davide Maltoni nell'anno accademico 2019/2020. http://bias.csr.unibo.it/maltoni/ml/DispensePDF/8_ML_RetiNeurali.pdf, http://bias.csr.unibo.it/maltoni/ml/DispensePDF/9_ML_DeepLearning.pdf, http://bias.csr.unibo.it/maltoni/ml/DispensePDF/10_ML_DeepLearning.pdf.

[12] Jason Brownlee, *"How to Choose Loss Functions When Training Deep Learning Neural Networks"*, Machine Learning Mastery, 30 gennaio 2019, https://machinelearningmastery.com/how-to-choose-loss-functions-when-training-deep-learning-neural-networks/.

[13] ANIASA, *"18° Rapporto ANIASA sul noleggio veicoli 2018"*, Sumo Publishing – Fleet Magazine, Roma, maggio 2019, https://www.aniasa.it/uploads/allegati/Rapporto%20Aniasa2018_definitivo.pdf.

[14] Dario Russo, *"Mercato auto, in Italia rallenta anche il noleggio"*, Linkiesta, 11 marzo 2019, https://www.linkiesta.it/it/blog-post/2019/03/11/mercato-auto-in-italia-rallenta-anche-il-noleggio/27833/.

[15] Bureau Van Dijk, *"Descrizione database Aida"*, Bureau Van Dijk edizioni elettroniche S.P.A, Milano, https://www.sba.unipi.it/sites/default/files/aida_descrizione_banca_dati.pdf.

[16] MATLAB Deep Learning Toolbox, https://it.mathworks.com/products/deep-learning.html.

[17] Jason Brownlee, *"How to use Learning Curves to Diagnose Machine Learning Model Performance*", Machine Learning Mastery, 27 Febbraio 2019, https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/.

[18] Franco Varetto e Giancarlo Marco, *"Diagnosi delle insolvenze e reti neurali. Esperimenti e confronti con l'Analisi Discriminante Lineare"*, Bancaria Editrice, Roma, luglio 1994.

Ringraziamenti

A conclusione di questo elaborato è doveroso impiegare queste ultime pagine per ringraziare tutte le persone che hanno contribuito a rendere il percorso universitario una crescita personale oltre che professionale. Ci tengo a nominare ogni singola persona affinché il ringraziamento possa risultare il più personale possibile e per evidenziare l'importanza di ognuno nell'aver apportato qualcosa di speciale nella mia vita.

In primis, vorrei ringraziare il professore e relatore Franco Varetto senza il quale questo lavoro di tesi non avrebbe potuto prendere vita. Lo ringrazio non solo per avermi fornito tutti gli strumenti necessari allo svolgimento dell'elaborato ma anche per i preziosi consigli e per essere riuscito ad accrescere in me la passione per le materie finanziarie indirizzando in questo verso il mio percorso professionale.

Ringrazio ogni singolo elemento della mia famiglia. I miei genitori Michele e Valeria, che mi sono sempre stati d'esempio attraverso i loro comportamenti e non mi hanno mai fatto mancare nulla. Li ringrazio per il loro amore e per l'educazione impartita, a loro devo tutto, ogni tipo di supporto: morale, decisionale ed economico. È grazie a loro che posso affermare con orgoglio di essere la persona che sono. Mia sorella Francesca, caratterialmente molto distante ma emotivamente sempre molto vicina, la ringrazio per la spensieratezza e la complicità che deriva dal nostro legame fraterno che ci unisce. Infine, la mia piccola nipotina Isabel per la sua tenerezza e le soddisfazioni che dà il vederla crescere.

Ringrazio gli amici, la mia seconda famiglia, con i quali ho condiviso momenti indimenticabili della mia vita. Inizio dagli amici che posso affermare di conoscere da sempre, che tutto sanno di me e con i quali sono cresciuto e continuerò a farlo sia fisicamente che caratterialmente: Eddy, Adri, Mazza, Momo, Paolo, Deco, Marco e Mattia. Gli amici del liceo: Lago, Mauri, Tony, Raffo, Ire, Defa, Vecchio, Alice e Roberta con i quali, nonostante il passare del tempo e le differenti strade intraprese, continuo a frequentarmi perché "una volta al mese ci sta". Gli amici e colleghi gestionali: Vise, Pietro, Prove, Schiuch, Simo, Gogo, Alberto, Angelo, Nicola e Federico con i quali ho condiviso ogni aspetto del percorso universitario fatto di lezioni, progetti, esami, infinite ore di studio ma anche giornate e serate di svago. Con loro ho condiviso questo percorso indimenticabile, con la speranza di aver creato un legame solido capace di durare nel tempo.

Ringrazio gli amici che non rientrano nelle categorie precedenti ma con i quali ho trascorso momenti ugualmente memorabili quali vacanze, eventi, concerti oppure semplicemente passato piacevoli giornate in compagnia: Fabio, Michela, Ale, Riky, Dega e Ila.

Ringrazio ancora tutte le persone che non ho citato ma con cui desidero ugualmente condividere il raggiungimento di questo traguardo in particolare tutti quelli che hanno sempre creduto in me, i parenti e chi mi guarda e protegge da lassù.

Infine, vorrei dedicare questo importante traguardo a me stesso per i sacrifici, la costanza e dedizione che non mi sono mai mancati e che mi hanno permesso di arrivare fino a qui, con la speranza che questo traguardo possa rappresentare l'inizio di una lunga e brillante carriera professionale.

Grazie infinite a tutti voi.