

POLITECNICO DI TORINO

Corso di Laurea Magistrale In Ingegneria Gestionale
Percorso Finanza

Tesi di Laurea Magistrale

**Costruzione di modelli di *credit scoring* tramite reti neurali e
macchine a vettori di supporto**



Relatore
Prof. Franco Varetto

Candidato
Alessio Colella

Anno Accademico 2020/2021

Alla mia famiglia

Indice

Introduzione.....	1
1 Il rischio di credito	4
1.1 Generalità sui principali rischi.....	4
1.1.1 Rischio di mercato	4
1.1.2 Rischio di liquidità	5
1.1.3 Rischio operativo	6
1.1.4 Rischio di credito	8
1.2 Il contesto normativo: la Regulation Basilea.....	9
1.2.1 Il Comitato di Basilea	10
1.2.2 Basilea I.....	11
1.2.3 Basilea II.....	12
1.2.4 Basilea III.....	20
1.2.5 Basilea IV	22
1.3 Descrizione del rischio di credito.....	22
1.3.1 Problematiche.....	23
1.3.2 Componenti	23
1.3.3 Modello binomiale.....	29
1.3.4 Rischi insiti nel rischio di credito	30
1.3.5 Sistemi di rating	32
1.3.6 Crediti bancari in sofferenza	34
1.4 Analisi del rischio di credito	36
1.4.1 Linear probability model.....	36
1.4.2 Analisi discriminante	37
1.4.3 Regressione logistica.....	38

1.4.4	Modelli di apprendimento automatico	40
2	Reti neurali.....	42
2.1	Reti neurali biologiche	42
2.2	Cenni storici sulle reti artificiali	44
2.3	Modello generale.....	47
2.3.1	Il neurone di McCulloch-Pitts	47
2.3.2	Componenti di una rete neurale.....	49
2.4	Tipologie di reti.....	57
2.5	Algoritmo di Backpropagation	58
2.6	Funzione di errore	63
2.7	Applicazioni, pregi e limiti	64
3	Macchine a vettori di supporto	66
3.1	Concetti chiave.....	67
3.2	Funzionamento	69
3.3	Modello lineare	70
3.4	Modello non lineare – il Kernel	72
3.5	Applicazioni, pregi e limiti	76
4	Il settore degli elettrodomestici	78
4.1	Cenni storici	78
4.2	Il settore in cifre.....	80
4.2.1	Il contesto europeo	80
4.2.2	Andamento storico in Italia	82
4.2.3	La situazione attuale e gli impatti del Coronavirus.....	83
5	Costruzione dei modelli.....	85
5.1	Preparazione dei dati	85
5.1.1	Il database Aida.....	85
5.1.2	Scarico dei dati	86

5.1.3	Pulizia e controllo dei dati	87
5.2	Software utilizzato per le elaborazioni	89
5.3	Strumenti per la valutazione delle prestazioni	90
5.3.1	Best validation performance	90
5.3.2	Matrice di confusione	96
5.3.3	Istogramma degli errori	99
5.3.4	Curva ROC	100
5.3.5	Gradiente e Validation Check	101
5.4	Modello con reti neurali	103
5.4.1	Modelli di prova e risoluzione dei problemi	103
5.4.2	Architetture testate sperimentalmente	106
5.4.3	Implementazione del modello	107
5.4.4	Presentazione dei tentativi	112
5.5	Prestazioni del miglior modello con reti neurali	119
5.6	Modello con SVM	125
5.6.1	Modelli di prova e risoluzione dei problemi	125
5.6.2	Implementazione del modello	130
5.6.3	Presentazione dei tentativi	132
5.7	Prestazioni del miglior modello con SVM	136
5.8	Confronto dei modelli	139
5.8.1	Prestazione dei migliori modelli	139
5.8.2	Andamento delle prestazioni in base all'evoluzione dei tentativi	141
	Conclusioni	144
	Bibliografia e sitografia	146

Indice delle figure

Figura 1.1 - Rappresentazione grafica di UL ed EL.....	25
Figura 1.2 - Copertura delle perdite.....	26
Figura 1.3 - Crediti bancari in sofferenza	35
Figura 1.4 - Confronto tra un modello lineare e un modello logistico	40
Figura 2.1 - Schema del neurone biologico	44
Figura 2.2 - Schema del Perceptron di Rosenblatt	45
Figura 2.3 - Schema di MLP	46
Figura 2.4 - Confronto tra neurone biologico e artificiale.....	48
Figura 2.5 - Funzione di Heaviside.....	52
Figura 2.6 - Funzione a rampa lineare	53
Figura 2.7 - Funzione sigmoidea logistica	53
Figura 2.8 - Funzione sigmoidea simmetrica	54
Figura 2.9 - Rappresentazione delle principali architetture	57
Figura 2.10 - Rete feedforward	58
Figura 2.11 - Rete feedback	58
Figura 2.12 - Andamento errore	62
Figura 3.1 - Iperpiano in due e tre dimensioni	67
Figura 3.2 - Vettori di supporto	68
Figura 3.3 - Margine	68
Figura 3.4 - Iperpiano	70
Figura 3.5 - Iperpiano e vettori di supporto in due dimensioni	72
Figura 3.6 - Dataset non lineare in due dimensioni	73
Figura 3.7 - Dataset non lineare in tre dimensioni	73
Figura 3.8 - Iperpiano post Kernel in due dimensioni	74
Figura 4.1 - Donna alle prese con una lavatrice non elettrificata	79
Figura 4.2 - Miglioramenti nella produzione di elettrodomestici	81
Figura 4.3 - Installazione di apparecchi intelligenti	82
Figura 4.4 - Elaborazione andamento spesa media familiare mensile	83
Figura 4.5 - Elaborazione focus su spese di riparazione	83
Figura 4.6 - Elaborazione per il valore vendite al dettaglio 2020	84
Figura 5.1 - Esempio di grafico di Best validation performance	91

Figura 5.2 - Underfit di un modello con mancanza di capacità	92
Figura 5.3 - Underfit di un modello che necessita di ulteriore addestramento .	93
Figura 5.4 - Esempio di un modello affetto da overfit	94
Figura 5.5 - Esempio di un modello con good fit	95
Figura 5.6 - Esempio di un modello con training set troppo piccolo rispetto al validation set	95
Figura 5.7 - Esempio di un modello con validation set troppo piccolo rispetto al training set	96
Figura 5.8 - Componenti di una matrice di confusione	97
Figura 5.9 - Esempio di istogramma degli errori	100
Figura 5.10 - Composizione curva ROC	101
Figura 5.11 - Esempio di Gradiente e Validation Check	102
Figura 5.12 - Matrice di confusione di una rete con 4 neuroni in input e 4 e 8 negli strati nascosti	104
Figura 5.13 - Istogramma degli errori di una rete con 4 neuroni in input e 4 e 8 negli strati nascosti	105
Figura 5.14 - Configurazione rete con 4 neuroni in input e 4, 6 e 8 negli strati nascosti	113
Figura 5.15 - Matrice di confusione modello con 4 neuroni in input e 4, 6 e 8 negli strati nascosti	114
Figura 5.16 - Configurazione rete con 4 neuroni in input e 4 nello strato nascosto	115
Figura 5.17 - Matrice di confusione per modello con 4 neuroni in input e 4 nello strato nascosto	116
Figura 5.18 - Architettura modello con 6 neuroni di input e 3 e 6 negli strati nascosti	120
Figura 5.19 - BVP modello con 6 neuroni di input e 3 e 6 neuroni negli strati nascosti	121
Figura 5.20 - Gradiente e validation check modello con 6 neuroni di input e 3 e 6 neuroni negli strati nascosti	122
Figura 5.21 - Istogramma degli errori modello con 6 neuroni di input e 3 e 6 neuroni negli strati nascosti	123

Figura 5.22 - Matrice di confusione modello con 6 neuroni di input e 3 e 6 neuroni negli strati nascosti	124
Figura 5.23 - Curva ROC modello con 6 neuroni di input e 3 e 6 neuroni negli strati nascosti	125
Figura 5.24 - Matrice di confusione di un modello con 4 variabili esplicative e funzione lineare	126
Figura 5.25 - Matrice di confusione di un modello con 8 variabili esplicative e funzione lineare	127
Figura 5.26 - Matrice di confusione per modello SVM con 6 predittori e il 50% di osservazioni anomale nel training set	128
Figura 5.27 - Matrice di confusione per modello SVM con 6 predittori e l'80% di osservazioni anomale nel training set	129
Figura 5.28 - Matrice di confusione per modello SVM con 6 predittori e set con la stessa PD	129
Figura 5.29 - Matrice di confusione per SVM con 6 indicatori	134
Figura 5.30 - Matrice di confusione del miglior modello SVM con 6 indicatori	138
Figura 5.31 - Curva ROC modello SVM con 6 indicatori	139
Figura 5.32 - Modello reti neurali con la stessa architettura del miglior modello, ma con altri indicatori nei 6 neuroni in input	142
Figura 5.33 - Modello SVM con 6 variabili, ma con altri indicatori rispetto al miglior modello	143

Indice delle tabelle

Tabella 1.1 - Coefficienti crediti verso soggetti sovrani	16
Tabella 1.2 - Coefficienti crediti verso banche	17
Tabella 1.3 - Coefficienti crediti verso imprese	18
Tabella 1.4 - Eventi modello binomiale	29
Tabella 1.5 - Scale di rating delle principali agenzie	34
Tabella 1.6 - Valori dei residui	37
Tabella 2.1 - Valori dello stato di attivazione	50
Tabella 2.2 - Schematizzazione delle connessioni	50
Tabella 5.1 - Risultati modello iniziale SVM	133

Introduzione

Posti davanti a un problema possono esserci diversi modi per risolverlo, alcuni più efficienti, altri più veloci, altri ancora più moderni. È proprio il concetto di modernità che individua le tecniche di intelligenza artificiale, non tanto per un discorso cronologico, quanto per una rottura rispetto alle metodologie classiche di derivazione statistica.

Grazie all'intelligenza artificiale è cambiato l'approccio all'analisi e alla risoluzione del problema, pur perseguendo il raggiungimento dello stesso risultato. Il metodo seguito è quanto di più familiare ci sia per un essere umano poiché si tratta della replica dei suoi processi biologici di apprendimento, naturalmente adattati a delle macchine: si parla appunto di *machine learning*.

Il punto critico di questi sistemi è però il dato, il quale diventa una preziosa risorsa da dare in pasto agli algoritmi per renderli sempre più efficaci. Se si guarda ai motori di ricerca e ai *social network*, la questione appare ovvia, ma non è scontato sottolineare come proprio i dati e le informazioni, che da sempre in ambito finanziario costituiscono fonte di competizione, stiano progressivamente divenendo ancora più importanti, alla luce della possibilità di sfruttarli per migliorare gli strumenti decisionali basati sull'apprendimento automatico. Si tratta di un circolo virtuoso in cui chi è in possesso dei dati potrà potenzialmente sviluppare tecniche di analisi più precise rispetto ai concorrenti e guadagnare terreno competitivo.

Negli ultimi anni l'utilizzo di queste tecniche ha subito notevoli sviluppi e sta penetrando in profondità nelle dinamiche finanziarie poiché è in grado di apportare benefici sia in termini di costi che di operatività. Ci sono alcuni campi che più di altri stanno subendo mutamenti, basti pensare al *trading*, alla consulenza tramite i cosiddetti *bot* e all'analisi e gestione dei rischi.

Il mondo dell'intelligenza artificiale e delle applicazioni connesse risulta quindi estremamente vasto, ma il focus della trattazione è il rischio di credito. Infatti, l'obiettivo del presente lavoro è la costruzione di due modelli di *credit scoring*

tramite l'utilizzo di tecniche di intelligenza artificiale, quali le reti neurali e le macchine a vettori di supporto. La base di partenza per il raggiungimento degli obiettivi è un dataset composto dai bilanci delle società attive nel settore degli elettrodomestici. Al termine delle elaborazioni si vogliono raggiungere due classificatori che, utilizzando gli indicatori costruiti a partire dai dati di bilancio, siano in grado di distinguere se la società è anomala, e quindi potenzialmente a rischio default, o sana.

Il percorso seguito per raggiungere gli obiettivi prefissati viene qui illustrato in cinque capitoli, partendo dalla definizione e dalle specificità del rischio credito e arrivando alle valutazioni conclusive sui modelli.

Nel primo capitolo sono citate le principali tipologie di rischio, cui segue un focus sul *credit risk*. In particolare, vengono illustrate le peculiarità di questa categoria di rischio, soffermandosi sul contesto normativo rappresentato dalla Regulation Basilea e illustrandone l'evoluzione dagli anni Settanta ad oggi. Fissato il contesto operativo, si è analizzato nel dettaglio il rischio di credito, formalizzando le sue diverse componenti e citando i principali metodi di stima. Questo passaggio ha permesso alla narrazione di avvicinarsi alle tecniche di intelligenza artificiale utilizzate.

Il successivo capitolo tratta proprio le reti neurali, partendo dal concetto di neurone biologico su cui si fondano le astrazioni matematiche impiegate nei più recenti algoritmi. Ai fini di garantire una miglior comprensione del tema, si è deciso di ripercorrere le tappe attraverso le quali gli studiosi hanno apportato elementi di innovazione per l'avanzamento di questa tecnica. La trattazione prosegue con una dettagliata descrizione delle diverse componenti di una rete neurale, fino a illustrarne applicazioni, vantaggi e svantaggi.

La seconda metodologia in esame è costituita dalle macchine a vettori di supporto. Anche in questo caso si sono specificati gli elementi chiave prima di fornire i dettagli circa il funzionamento. Un concetto fondamentale di questo metodo è il Kernel, una funzione che può assumere forme differenti a seconda delle esigenze.

La parte riguardante il contesto prosegue con la presentazione del settore di riferimento degli elettrodomestici, partendo dalla sua storia e illustrandone le statistiche più rilevanti. Un elemento da menzionare è il focus sui trend di questo comparto, che è sempre più caratterizzato da dispositivi connessi e interattivi. All'interno di questa sezione vengono brevemente segnalati gli impatti che la pandemia ha avuto sul settore nel corso del 2020.

Il quinto capitolo è il nodo principale del lavoro, in quanto contiene le informazioni riguardanti i modelli. Dapprima vengono presentate le operazioni di scarico e di preparazione dei dati per renderli utilizzabili da un algoritmo, le quali sono seguite dalle indicazioni sul software utilizzato per l'analisi e dagli strumenti impiegati per valutare le prestazioni. La trattazione prosegue con il processo di costruzione per i due metodi, il quale si fonda su due percorsi distinti seppur paralleli: i database iniziali sono stati gli stessi per entrambi, ma i processi di implementazione hanno dovuto seguire necessariamente due strade diverse, essendo le tecniche caratterizzate da una natura differente. Pertanto, si illustra prima il percorso per il modello costruito con le reti neurali, cui segue quello per le SVM. Si tratta appunto di un percorso e non di un tentativo *one-shot*, poiché la valutazione delle performance ha messo in luce di volta in volta gli aspetti migliorabili del modello, consentendo di eseguire più tentativi per cercare la soluzione più adatta.

La trattazione prosegue con il confronto delle due tecniche, presentando i punti comuni e le divergenze, riscontrati sia in termini assoluti che durante il percorso evolutivo dei classificatori. Si conclude guardando ai risultati finali dei modelli, illustrandone punti di forza e di debolezza ai fini di una possibile evoluzione futura.

Il perseguimento degli obiettivi ha permesso di lavorare con due tecniche di intelligenza artificiale molto utilizzate. Infatti, tali metodologie, con il passare del tempo, si stanno ritagliando sempre più importanza nelle applicazioni finanziarie e vengono qui impiegate con la finalità della classificazione binaria, ma è bene sottolineare come queste siano molto versatili e possano trovare attuazione anche per scopi differenti.

Capitolo I

1 Il rischio di credito

Questo capitolo tratta i principali rischi che si affrontano solitamente in ambito finanziario per poi entrare nel merito del rischio di credito, illustrando le sue caratteristiche salienti. Si mostra l'incidenza della Regulation Basilea sul rischio di credito, concludendo con l'analisi delle metodologie più diffuse per stimarlo.

1.1 Generalità sui principali rischi

In un sistema economico e finanziario il termine "rischio" ricopre un'importanza non trascurabile. Il significato italiano del termine esprime un'accezione negativa, ma, per catturarne la vera essenza, è utile intendere il rischio come un fenomeno aleatorio che può avere conseguenze sia negative che positive. I rischi vanno individuati, studiati e bisogna capire come ridurre il loro potenziale effetto negativo e come amplificarne il possibile beneficio. È però importante conoscere quali rischi si possono dover fronteggiare. In questi paragrafi vengono descritte le principali caratteristiche del rischio di mercato, del rischio di liquidità, del rischio operativo e del rischio di credito.

1.1.1 Rischio di mercato

Il rischio di mercato (o di prezzo) identifica delle variazioni inattese delle condizioni di mercato, quali: prezzi, tassi d'interesse, tassi di cambio e rispettive volatilità. Il mutamento delle condizioni può portare a una svalutazione significativa di un portafoglio o di un singolo asset finanziario. Per maggior chiarezza, è bene sottolineare come questo rischio, in linea generale, riguardi le posizioni finanziarie indipendentemente dal loro orizzonte temporale, mentre il riferimento che le autorità di regolamentazione considerano è il breve o brevissimo periodo, in virtù della Direttiva europea 93/06/ECC e delle indicazioni del Comitato di Basilea.

La crescita del mercato finanziario e la diffusione di prodotti sempre più innovativi hanno comportato un aumento progressivo della rilevanza di questo rischio. Infatti, la costruzione di strumenti derivati su sottostanti sensibili al tasso d'interesse, al tasso di cambio, ai prezzi delle *commodity* e alle dinamiche relative all'*equity*, oltre alla volatilità, ha portato maggior complessità nei *trading book* delle imprese, con conseguente aumento dell'esposizione alle variazioni inattese delle condizioni di mercato.

Uno degli strumenti più utilizzati – ma non l'unico – per la quantificazione di questo particolare rischio che è necessario citare è il *Value-at-Risk* (VaR). Tramite il calcolo del VaR si identifica, definito un orizzonte temporale, la massima perdita potenziale in modo che la probabilità di conseguire una perdita maggiore sia statisticamente esigua. Il metodo possiede sia punti di forza che di debolezza, ma non saranno qui approfonditi.

1.1.2 Rischio di liquidità

Il rischio di liquidità si concretizza quando esiste uno scostamento temporale tra attività e passività. In particolare, qualora la società si trovasse a dover ripagare dei debiti a breve, potrebbe essere costretta a cedere parte delle sue attività a medio-lungo termine per ottenere la liquidità necessaria. In questo caso subentra il rischio di liquidazione delle attività, in quanto il prezzo di realizzo potrebbe essere inferiore a quello previsto. Il rischio di liquidità si configura, quindi, a fronte di uscite inattese.

Il *funding risk* e il *market liquidity risk* sono due forme del rischio di liquidità. La prima si traduce con la compromissione della gestione aziendale davanti a uscite finanziarie attese e inattese. La seconda innesca un meccanismo che porta, a causa di un'offerta massiva, a influenzare il prezzo delle attività nel tentativo di raccogliere liquidità in breve tempo. Risulta evidente come, nel secondo caso, l'effetto sia tanto più rilevante quanto più piccolo è il mercato di riferimento.

Per la quantificazione di questo rischio sono utilizzati metodi differenti, ad esempio: approccio con dati di consistenza, approccio con flussi di cassa, approccio ibrido e stress test.

1.1.3 Rischio operativo

I rischi operativi non sono riconducibili a fattori di mercato o a eventi creditizi, bensì sono legati alle risorse umane, ai sistemi informativi, alle procedure interne e ad eventi esterni (furti, attacchi informatici, terremoti, ecc.). In quest'ultimo caso, a differenza dei primi tre dove l'organizzazione può attuare investimenti per rinforzare i sistemi operativi e le procedure, il margine di intervento è assai più ristretto e si limita, ad esempio, al miglioramento della sorveglianza, all'implementazione di misure per il *disaster recovery* e la *business continuity* e all'installazione di un gruppo elettrogeno di emergenza.

A partire dagli anni Novanta, l'importanza di questa tipologia di rischio è divenuta notevole. I principali fattori determinanti riguardano: l'impiego crescente dei sistemi informativi nel settore finanziario, la crescita delle vendite elettroniche, le difficoltà di integrazioni sistemiche in seguito a fusioni e acquisizioni, lo sviluppo di strumenti finanziari complessi i cui rischi non sono sempre intelligibili pienamente dai manager e la spinta del Comitato di Basilea per la quantificazione e la gestione di questo particolare tipo di rischio.

Una implicazione da tenere in mente quando si parla di rischio operativo è il danno reputazionale. Come insegnano i trascorsi storici, in ambito finanziario il rapporto fiduciario è fondamentale ed eventi come le truffe legate a dipendenti infedeli o gli attacchi informatici atti a entrare in possesso di informazioni sui clienti potrebbero far venir meno il senso di sicurezza nei confronti della controparte, con ripercussioni su tutto il business.

Ciò che contraddistingue il rischio operativo dalle altre tipologie di rischi è il fatto di essere strettamente legato alla singola realtà e non a fattori di mercato. Costruendo un portafoglio di posizioni finanziarie si decide a quali rischi voler andare incontro, mentre il rischio operativo è dovuto – appunto – alla consueta

operatività. Ad esempio, nel caso di una frode legata a un dipendente infedele, non si arriverebbe a un contagio di carattere sistemico, ma sarebbe con tutta probabilità circoscritto alla singola società. Il discorso è analogo se si guardano le imprese non finanziarie, come ad esempio un'impresa che esegue trivellazioni petrolifere, che avrà tra i suoi rischi operativi quello legato intrinsecamente all'attività svolta: è quindi un rischio inseparabile dalla specifica realtà aziendale.

Un particolare rischio operativo che è necessario citare è il rischio normativo. Esso si verifica al mutare delle condizioni del contesto operativo ad opera di nuove norme a cui adattarsi. Quanto emerge da un'indagine di Moody's del 2018 rivolta ai direttori finanziari delle compagnie assicurative europee è che il 27% di essi ha indicato il rischio normativo e politico come principale preoccupazione, seguito dal rischio di mercato con il 18% [4]. I risultati non stupiscono se si pensa che l'indagine avvenne in un periodo in cui il comparto assicurativo era oggetto di notevoli cambiamenti portati direttamente o indirettamente da normative quali il GDPR e la Mifid II.

Un altro aspetto fondamentale di questo rischio è il suo effetto che può essere esclusivamente negativo. Mentre con i rischi di mercato, di liquidità e di credito è possibile ottenere dei profitti, al manifestarsi del rischio operativo si va incontro a delle perdite.

La quantificazione di questo rischio è complicata anche dalla mancanza di dati storici: si tratta di eventi rari e legati alla singola realtà e non al sistema. Nonostante le difficoltà, si utilizzano, tra gli altri, metriche come l'*exposure indicator*, la *probability of event* e il *loss given event*.

Alle problematiche precedentemente espresse, si aggiungono la mancanza di strumenti di *hedging* e la difficoltà di *pricing* e di trasferimento di questo rischio. Esistono tuttavia dei prodotti assicurativi che proteggono da specifici eventi, ma, come segnalato, le implicazioni non si riconducono a una mera perdita di denaro, infatti possono espandersi e intaccare in modo irreparabile la reputazione. Per questo motivo è bene considerare attentamente le fonti di rischio e cercare, per quanto possibile, di attuare misure preventive.

1.1.4 Rischio di credito

Il rischio di credito, come si intuisce dal termine stesso, riguarda le esposizioni creditizie, un ambito tradizionalmente legato a tutte le realtà bancarie. In particolare, esprime la possibilità di una variazione inattesa del valore di mercato del credito dovuta a una variazione inattesa del merito di credito della controparte. Le cause della variazione del merito di credito possono essere l'insolvenza della controparte o il deterioramento della qualità del credito, con conseguente aumento intrinseco della probabilità di insolvenza. Il deterioramento del credito, non essendo esente dalle logiche di mercato, si traduce con un aumento del premio per il rischio, con conseguente riduzione del prezzo.

Questo meccanismo pone l'attenzione sul fatto che la variazione deve essere inattesa perché, qualora ci si aspetti un deterioramento, questo avrebbe ricadute sul *pricing* al momento della concessione del credito o del suo rinnovo. In altri termini, se ci si aspetta un peggioramento della qualità creditizia, si chiederà un premio per il rischio maggiore; se invece l'evento non è prevedibile, si avranno ricadute sul credito in modo diretto e, nel caso in cui si trovasse un acquirente, questo vorrà un premio maggiore, con conseguente riflesso negativo sul prezzo.

I motivi che portano alla variazione inattesa del valore del credito possono essere una valutazione errata al momento della concessione oppure l'insorgere di problematiche economico-finanziarie riguardanti la controparte che in fase di concessione non erano note. In ogni caso, è bene considerare questo rischio, identificandolo nel modo più accurato possibile e attuando le contromisure più indicate. Ciò che emerge è la necessità di analizzare e selezionare nel modo più puntuale possibile la controparte e il *pricing* del credito.

Per fare ciò è indispensabile ricavare quante più informazioni possibili sul beneficiario del credito, ma ciò si scontra con la realtà, laddove, guardando all'Italia, il modello predominante è la cosiddetta banca per transazioni, in cui l'impresa sceglie sostanzialmente la banca che offre le migliori condizioni. In

questo modo, è difficile creare i presupposti per una conoscenza approfondita dell'impresa, mancando di fatto un rapporto stabile con la banca (banca per relazione). Ovviamente, non si tratta di una mera questione di durata del rapporto con l'istituzione creditizia: se la banca dovesse svolgere un'analisi estremamente approfondita del richiedente prima di concedere il credito, avrebbe dei costi enormi, difficilmente recuperabili con i margini dell'operazione. Ecco perché le informazioni sono sì importanti, ma bisogna porre altrettanta attenzione sul *trade-off* relativo ai costi per la loro acquisizione.

1.2 Il contesto normativo: la Regulation Basilea

Prima di entrare nel merito delle decisioni del Comitato, si specificano i punti di vista riguardo al tema della regolamentazione dei sistemi bancari. Si assiste infatti all'espressione di due correnti di pensiero opposte, che illustrano entrambe delle motivazioni valide, ma, in ottica sistemica, è opportuno considerare le implicazioni di questi due punti di vista e cercare di farli coesistere.

Nel dettaglio, il primo approccio non riconosce la regolamentazione come necessaria, anzi la identifica come superflua per il sistema economico. Secondo questa idea, il mercato dovrebbe essere l'unico *driver*, in quanto sarebbe in grado di autoregolarsi in modo migliore e più velocemente. L'intervento delle autorità di governo viene inteso come potenziale sorgente di danno per il sistema.

Il secondo approccio, invece, si fonda sull'idea che la regolamentazione sia necessaria. L'assenza di un'autorità di controllo graverebbe sul sistema portando condizioni di non concorrenza, instabilità e crisi frequenti. Naturalmente l'introduzione di regole comporta dei costi per gli operatori del settore, la regolamentazione deve quindi arrivare necessariamente ai traguardi che si è prefissata, bilanciando tali costi con i benefici.

Già da questa prima analisi, si evince quando siano importanti la concorrenza e la stabilità all'interno del sistema. Tuttavia, un'impostazione dominata da regole

porta stabilità, ma non garantisce la concorrenza tra gli operatori; al contrario, un sistema in cui c'è la massima concorrenza, potrebbe far insorgere condizioni di instabilità. Occorre gestire questo importantissimo *trade-off* attraverso l'introduzione di regole che indirizzino gli operatori, senza però minare la concorrenza.

A tal proposito, si possono indicare due tipi di regolamentazione: strutturale e prudenziale. La strutturale si configura tramite l'imposizione di restrizioni alla libera operatività degli istituti di credito, così da limitare l'assunzione di rischi. Si parla di regolamentazione strutturale quando questa incide, ad esempio, sull'entrata in un certo mercato, sull'espansione geografica, sull'offerta dei prodotti. La prudenziale, al contrario, non pone limiti all'operatività, ma richiede alcune forme di copertura dei rischi assunti. In questo caso, la banca è libera di operare, tenendo presente che ha l'obbligo di istituire, ad esempio, una riserva obbligatoria o un'assicurazione sui depositi a copertura dei rischi.

Negli scorsi decenni si è assistito al processo di deregolamentazione: si tratta essenzialmente del passaggio da una regolamentazione di tipo strutturale verso un'impostazione prevalentemente prudenziale. Nel complesso, infatti, coesistono aspetti sia strutturali che prudenziali. In concreto, le banche devono soddisfare determinati requisiti di capitalizzazione in relazione ai rischi assunti, così da assorbire le eventuali perdite in caso di scenari sfavorevoli.

1.2.1 Il Comitato di Basilea

Fino agli anni Settanta del secolo scorso, ogni Paese stabiliva autonomamente i criteri per la definizione dell'adeguatezza patrimoniale delle banche. L'assenza di un riferimento centrale aveva portato a marcate condizioni di disparità concorrenziale tra i diversi sistemi bancari su scala internazionale, così, nel 1974, nacque il Comitato di Basilea, un organismo di cooperazione internazionale. Esso prevede ancora oggi la partecipazione delle figure di riferimento delle banche centrali e delle autorità di vigilanza dei Paesi membri del G10. Il nome del Comitato deriva dalla sua sede, in quanto le riunioni si tengono negli uffici svizzeri della Banca dei Regolamenti Internazionali.

Nonostante quanto deciso dal Comitato sia seguito dai Paesi del G10 e sia adottato anche a livello internazionale, manca formalmente un valore giuridico. La composizione attuale del comitato vede la partecipazione di 45 membri provenienti da 28 giurisdizioni, con incontri periodici che si svolgono tre o quattro volte ogni anno [5]. L'obiettivo di questo organismo è proprio colmare le lacune presenti nella vigilanza internazionale in modo opportuno e coerente con le giurisdizioni dei membri.

Nel 1975 si ebbe la prima emissione del Concordato di Basilea, nel quale si stabilivano gli standard internazionali per la regolamentazione delle banche. Di particolare rilevanza sono le pubblicazioni degli accordi sull'adeguatezza patrimoniale, noti come Basilea I, Basilea II e Basilea III.

1.2.2 Basilea I

Come accennato, la situazione precedente agli accordi, era estremamente frammentata in termini di *regulation* bancaria. L'intento di Basilea I era la generazione di un modello di regolamentazione uniforme e in grado di adattarsi alle diverse realtà presenti al tavolo di discussione. Tale lavoro del 1988 riguardava in maniera esclusiva le banche internazionali, le quali venivano riconosciute per la loro efficienza e per la buona gestione in termini di rischi e perdite attese.

Il focus del Comitato erano le perdite inattese, le quali hanno strette implicazioni sui requisiti di capitalizzazione minimi. Originariamente si decise di trattare solamente il rischio di credito, salvo poi aggiungere successivamente il rischio di mercato.

La prima decisione riguardò l'entità del capitale regolamentare, il cui ammontare venne stabilito all'8% delle attività ponderate per il rischio (*Cooke ratio*). Le ponderazioni erano le seguenti:

- 0%: cassa, crediti verso governi 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%: tutti gli altri crediti.

Considerando anche il rischio di mercato, l'espressione della Regulation Basilea I può essere espressa nel modo seguente:

$$TCR = \frac{\text{Patrimonio}}{RWA + 12.5 * RM} \geq 8\% \quad (1.1)$$

Il rapporto del capitale totale (*Total Capital Ratio*) è composto dal patrimonio netto al numeratore e dalla somma dei *Risk Weighted Asset* e del rischio di mercato corretto per il coefficiente 12.5. Tale valore non deve essere inferiore al rapporto di Cooke.

Il principale problema delle ponderazioni fu proprio la loro semplicità, che lasciò spazio ai cosiddetti arbitraggi regolamentari. Pur rispettando le restrizioni imposte da Basilea, era possibile concedere prestiti a un creditore più rischioso per avere un maggior ritorno, senza dover accantonare maggior capitale. Un altro elemento da evidenziare è l'innovazione finanziaria, che ha portato a prodotti sempre più sofisticati e con oneri relativamente bassi in termini di capitale. Quindi, la mancata discriminazione dei crediti all'interno di una stessa classe di ponderazione e l'utilizzo di nuovi strumenti crearono un effetto controproducente, incentivando i comportamenti poco virtuosi delle banche a favore di strategie intente ad aggirare i limiti imposti dal regolatore. Considerata la situazione, si arrivò ad asserire la necessità di revisione del meccanismo per un modello che tenesse conto delle problematiche riscontrate.

1.2.3 Basilea II

La rivisitazione del primo modello di regolamentazione bancaria ha portato il Comitato a un nuovo Accordo nel 2004: Basilea II. Il nuovo approccio si

fondava su tre pilastri: *minimum capital requirement* (requisiti minimi patrimoniali), *supervisory review* (controllo delle Autorità di vigilanza) e *market discipline* (disciplina di mercato).

Il primo pilastro

I requisiti patrimoniali hanno introdotto il rischio operativo come novità, esso non era infatti contemplato nella precedente versione dell'Accordo. Quindi, con questa modifica, i requisiti di capitale riguardavano il rischio di credito, di mercato e operativo. La soglia di riferimento restava il *Cooke ratio*. L'espressione del primo pilastro si ottiene con la formula:

$$\frac{\text{Capitale di vigilanza}}{\text{Rischi di credito} + \text{Rischi di mercato} + \text{Rischi operativi}} \geq 8\% \quad (1.2)$$

Il secondo pilastro

Il controllo delle Autorità sul capitale bancario deve essere continuo. Il motivo di tale scelta è da ricercare nel tentativo per evitare che questo scenda sotto la soglia minima o non sostenga sufficientemente l'assunzione dei rischi da parte dell'istituto. Non si tratta, quindi, di un mero controllo sui livelli di capitale, ma di un vero e proprio monitoraggio ed incoraggiamento affinché le banche sviluppino e rafforzino le tecniche di gestione dei rischi.

Entrando nel concreto dei poteri, la Vigilanza bancaria ha la possibilità di richiedere che vengano rispettati requisiti patrimoniali più stringenti di quelli previsti da Basilea II a seconda dei casi. Gli organismi di Vigilanza possono essere diversi e, nel caso italiano, per le grandi banche è la Banca Centrale Europea a occuparsi di questa funzione, mentre per le banche più piccole il compito è affidato alla Banca d'Italia.

I quattro principi fondamentali a cui si ispira il secondo pilastro sono [6]:

- Le banche dovrebbero disporre di un processo per valutare la loro adeguatezza patrimoniale complessiva in relazione al loro profilo di rischio e di una strategia per mantenere i loro livelli di capitale.
- Le Autorità di vigilanza dovrebbero esaminare e valutare l'adeguatezza del capitale interno delle banche e le rispettive strategie, nonché la loro capacità di monitorare e garantire il rispetto dei coefficienti patrimoniali regolamentari. I supervisori dovrebbero adottare misure appropriate se non sono soddisfatti del risultato di questo processo.
- Le Autorità di vigilanza dovrebbero aspettarsi che le banche operino al di sopra dei coefficienti patrimoniali regolamentari minimi e dovrebbero avere la capacità di richiedere alle banche di detenere capitale in eccesso rispetto al minimo.
- Le Autorità di vigilanza dovrebbero cercare di intervenire in una fase iniziale per prevenire che il capitale scenda al di sotto dei livelli minimi richiesti per supportare le caratteristiche di rischio. Dovrebbero richiedere inoltre un'azione correttiva rapida qualora il capitale non fosse mantenuto o ristabilito.

Il terzo pilastro

Il concetto chiave del terzo pilastro è la trasparenza. Esso ritiene il mercato come il miglior giudice per la valutazione dei rischi assunti da una determinata banca e pertanto si presenta l'esigenza di disponibilità delle informazioni. Le banche sono tenute a fornire pubblicamente una relazione periodica sui rischi e sulle relative coperture. Secondo questa logica, se il mercato valutasse sfavorevolmente le informazioni in suo possesso, la banca sarebbe costretta a rimediare correggendo il suo portafoglio. Per perseguire gli obiettivi del terzo pilastro, le banche dovrebbero adottare formalmente una politica di divulgazione approvata dagli amministratori. Tale documento deve determinare l'entità delle divulgazioni e i relativi controlli sul processo. In aggiunta, le banche dovrebbero attuare delle procedure per la valutazione dell'adeguatezza delle comunicazioni, sia in termini di contenuto che in relazione alla frequenza di pubblicazione delle stesse.

Rischio di credito

Uno dei principali limiti di Basilea I consisteva nella non discriminazione in base al rischio all'interno di una stessa categoria di ponderazione. In altre parole, la concessione di un credito a un'impresa ad alto rischio di insolvenza richiedeva, in termini di capitale, gli stessi accantonamenti di un credito affidato a un'impresa a basso rischio. Questa situazione aveva innescato un meccanismo contrario agli obiettivi del regolatore, lasciando di fatto ampio terreno agli arbitraggi regolamentari.

Il nuovo modello, nel tentativo di risolvere il problema, adotta una discriminazione più puntuale nella determinazione del rischio di credito. Gli approcci metodologici sono due, uno standard e l'altro basato sui sistemi di rating interni.

Approccio standard

Come per Basilea I, anche in questo caso il primo elemento di valutazione è la categoria del debitore. La vera novità è la differente ponderazione a seconda della classe di rating assegnata dalle Agenzie (*Eligible External Credit Assessment Institution*) alla controparte. Emerge subito una prima criticità, ossia il ricorso ad istituzioni esterne per l'individuazione del merito di credito. Il Comitato ha quindi incaricato le Autorità di vigilanza ad accertare i requisiti delle Agenzie. Nello specifico, i criteri minimi di soddisfacimento sono:

- Obiettività = il metodo deve essere rigoroso e convalidato dall'esperienza storica;
- Indipendenza = l'operatività non deve essere influenzata da pressioni politiche;
- Trasparenza = la metodologia dell'ECAI deve essere accessibile all'esterno, così come le valutazioni devono essere disponibili alle stesse condizioni sia per soggetti nazionali che esterni titolari di un interesse legittimo;

- Pubblicità delle informazioni = metodologia, definizione di default, orizzonte temporale, significato di ciascun rating sono alcune delle principali informazioni che l'ECAI dovrebbe rendere pubbliche;
- Risorse = l'ECAI deve essere in possesso delle risorse necessarie per effettuare valutazioni di elevata qualità;
- Credibilità = è riconducibile all'osservanza dei requisiti precedenti e può essere rafforzata mediante valutazione esterne dell'ECAI da parte di soggetti indipendenti.

Basilea II distingue diverse classi di ponderazione [6]: crediti verso soggetti sovrani, crediti verso enti del settore pubblico non appartenenti alle amministrazioni centrali, crediti verso banche multilaterali di sviluppo, crediti verso banche, crediti verso società di intermediazione mobiliare, crediti verso imprese, attività ricomprese nei portafogli regolamentari al dettaglio, crediti garantiti da ipoteca su immobili residenziali, crediti garantiti da ipoteca su immobili non residenziali, prestiti scaduti, categorie a più alto rischio, altre attività e posizioni fuori bilancio. Vengono di seguito analizzate le categorie principali.

Crediti verso soggetti sovrani

Ai soggetti sovrani e alle relative banche centrali, il Comitato ha assegnato i seguenti coefficienti di ponderazione:

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%

Tabella 1.1 - Coefficienti crediti verso soggetti sovrani [6]

Le indicazioni sulle ponderazioni prevedono per i crediti verso la Banca dei Regolamenti Internazionali e il Fondo monetario internazionale la possibilità di avere un coefficiente nullo. Anche per la Banca centrale europea e la Comunità europea si applicano tali disposizioni.

Crediti verso banche

Per le banche esistono due opzioni applicabili. La prima prevede che:

A tutte le banche con sede in un dato paese sia attribuita la ponderazione di cui alla categoria immediatamente meno favorevole rispetto a quella assegnata ai crediti verso il governo del paese in questione. Tuttavia, i crediti verso le banche appartenenti a paesi con rating da BB+ a B- e a paesi privi di rating saranno soggetti a un coefficiente massimo del 100% [7].

La seconda opzione pondera in base ai rating assegnati alle singole banche dalle ECAI, attribuendo un coefficiente del 50% per quelle prive di rating. Questo secondo metodo favorisce i crediti di durata pari o inferiore ai tre mesi tramite l'assegnazione di coefficienti inferiori.

Il quadro riassuntivo delle ponderazioni è raffigurato nella tabella 1.2.

Opzione 1

Valutazione del soggetto sovrano	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%

Opzione 2

Valutazione della banca	Da AAA ad AA-	Da A+ a 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 termine²⁶	20%	20%	20%	50%	150%	20%

Tabella 1.2 - Coefficienti crediti verso banche [6]

Crediti verso imprese

Il criterio per l'assegnazione dei coefficienti di ponderazione è la classe di rating a cui appartengono le società. Per le società che ne sono sprovviste, la ponderazione è pari al 100% e non può essere più favorevole di quella

assegnata al soggetto sovrano in cui l'impresa è residente. La tabella 1.3 illustra le fasce di rating considerati e le rispettive ponderazioni.

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

Tabella 1.3 - Coefficienti crediti verso imprese [6]

Attività ricomprese nei portafogli regolamentari al dettaglio

Per questa categoria la ponderazione prevista è del 75%. Rientrano in questa casistica i crediti che soddisfano i criteri che seguono:

- Destinazione = il credito deve essere verso persone fisiche o imprese di piccole dimensioni;
- Tipologia di prodotto = l'esposizione deve essere sotto forma di crediti e linee di credito rotativi o prestiti personali e contratti di leasing o facilitazioni e aperture di credito a vantaggio di imprese di piccole dimensioni;
- Frazionamento = è compito dell'Autorità verificare che il portafoglio cosiddetto *retail* sia diversificato efficacemente per la riduzione dei rischi;
- Esposizione unitaria massima consentita = non si può eccedere il milione di euro di esposizione complessiva nei confronti di una singola controparte.

Crediti garantiti da ipoteca su immobili residenziali

Il criterio della destinazione residenziale deve essere verificato dall'Autorità di vigilanza. Per i crediti garantiti in modo completo da un'ipoteca iscritta su un immobile di tipo residenziale, sia esso abitato che destinato alla locazione, la ponderazione ammonta al 35%.

Sistemi di rating interni

Il secondo approccio si basa sull'impiego dei sistemi di rating interni delle banche. Attraverso questi sistemi, le banche possono fare uso delle stime interne per determinare la probabilità di default (PD), l'esposizione al momento del default (EAD), la perdita in caso di inadempienza (LGD) e la scadenza effettiva (M). Il metodo definisce le perdite attese e le perdite inattese, individuando dalle funzioni di ponderazione del rischio i requisiti di capitale a fronte delle perdite inattese.

È bene sottolineare come tali sistemi debbano soddisfare determinati requisiti, inoltre il loro utilizzo deve essere autorizzato dall'Autorità di vigilanza. I requisiti minimi riguardano:

- Valutazione separata di PD e LGD;
- Distribuzione dei crediti tra le varie classi di rating, senza che questi si concentrino in una classe unica;
- Rating attribuito prima della concessione del credito;
- Revisione periodica del rating;
- Rating da impiegare nella gestione dei crediti e nel *pricing* degli affidamenti;
- Necessità di un sistema di validazione dell'accuratezza e della coerenza del sistema di rating;
- Documentazione formale sul sistema di rating e sulle sue modalità di funzionamento.

È presente un'ulteriore distinzione tra i metodi usati per stimare i parametri di rischio. L'approccio base vede la banca stimare la PD, mentre la Vigilanza si occupa di fornire la LGD, la M e la EAD. L'approccio avanzato, invece, prevede che sia la banca a ricavare tutti i parametri.

Limiti di Basilea II

Lo scenario di crisi dei primi anni Duemila ha messo a dura prova i concetti su cui si fondava Basilea II: nonostante le banche rispettassero formalmente i requisiti della *regulation* il capitale risultò insufficiente. Con l'emergere delle criticità l'attenzione si spostò verso la qualità del capitale – che veniva spesso

impovertito della porzione di miglior qualità attraverso la distribuzione di dividendi - e verso il rischio sistemico - aggravato in tempo di crisi dalla sopravvalutazione della bontà dei modelli utilizzati. Proprio quest'ultimo punto mette in luce l'inadeguatezza di una regolamentazione micro-prudenziale nel prevenire una crisi dell'intero sistema.

Le banche operavano, seppur rispettando le imposizioni di Basilea, con livelli di indebitamento molto elevati. L'eccesso di *leverage* e del conseguente *deleverage* al mutare delle condizioni ha contribuito all'instabilità del sistema. Inoltre, la sottovalutazione dell'importanza sistemica delle banche e dell'azzardo morale ha distorto il comportamento di alcuni operatori, i quali, convinti di essere *too big to fail*, hanno agito in modo poco prudente. Infatti, durante la crisi, numerosi istituti vennero salvati grazie all'intervento dei rispettivi governi. Questi episodi hanno messo in luce un problema di difficile risoluzione: anche le grandi banche devono poter fallire senza che l'intero sistema rischi il collasso – si pensi al fallimento (o al mancato salvataggio, a seconda dei punti di vista) della Lehman Brothers e alle notevolissime ripercussioni anche fuori dai confini degli Stati Uniti.

Infine, la pro-ciclicità ha evidenziato la sensibilità del capitale all'andamento del ciclo economico. Quando si è in presenza di condizioni favorevoli uno stesso patrimonio è in grado di sostenere un maggior volume di attività rispetto ai periodi di crisi, ma è proprio in presenza di scenari sfavorevoli che l'economia avrebbe bisogno di essere sostenuta.

1.2.4 Basilea III

Rappresenta il complesso di norme varato nel 2011 - e attualmente in vigore - con l'intenzione di integrare la precedente versione dell'Accordo e superarne i limiti emersi con la crisi. Gli ambiti di intervento riguardano:

- Maggiore qualità del capitale
È stata prevista la revisione del patrimonio di vigilanza, eliminando il *Tier 3* e implementando il *Tier 1* e il *Tier 2*. Contestualmente, è stato deciso

l'innalzamento del requisito di *Core Tier 1* (capitale sociale, riserve da sovrapprezzo e di utili) al 4.5% sui *Risk Weighted Asset*. In questo modo, il *Tier 1* sarebbe in grado di assorbire le perdite in condizioni di continuità aziendale, con il *Tier 2* che interverrebbe in scenari di crisi.

- Introduzione di un *capital conservation buffer*

Si tratta di un cuscinetto di capitale pari al 2.5% dei RWA. Il suo funzionamento prevede un accumulo di risorse in questo buffer sommandosi al *Core Tier 1* e portando il capitale di qualità elevata al 7%, salvo la possibilità di eroderlo nel corso delle fasi negative.

- Introduzione di un *countercyclical buffer*

Si tratta di un ulteriore accumulo di risorse pari fino ad un altro 2.5%, a discrezione delle Autorità di vigilanza. Al fine di evitare uno shock ai mercati e agli intermediari, tale buffer deve essere annunciato con un anno di anticipo. Si tratta di un intervento, insieme al *conservation buffer*, atto a contrastare la prociclicità.

- Introduzione di una soglia massima di leva

Il rapporto tra il patrimonio *Tier 1* e l'attivo contabile, includendo l'attivo fuori bilancio, deve essere maggiore o uguale al valore soglia del 3%. In questo modo, si limiterebbe la leva finanziaria e si arginerebbero eventuali errori di stima della ponderazione dei rischi.

- Introduzione di due coefficienti di liquidità

Il *Liquidity Coverage Ratio* stabilisce che le attività liquide di elevata qualità devono fronteggiare i deflussi di cassa attesi nei 30 giorni successivi, stimati in una condizione di stress. Il *Net Stable Fund Ratio* deve essere superiore ad uno ed è composto dal rapporto tra fonti di finanziamento stabili a medio e lungo termine e fabbisogni a medio e lungo termine.

- Nuovi requisiti per il rischio di mercato e per quello di controparte

Si è intervenuto sui requisiti per i portafogli di *trading*, aggiungendo prudenza in merito ai rischi relativi alle cartolarizzazioni e ai veicoli fuori bilancio. Si è inoltre inasprito il rischio di controparte per quanto riguarda i derivati.

1.2.5 Basilea IV

Al termine del 2017, il Comitato ha reso noto l'Accordo Basilea IV. Si tratta formalmente della revisione finale di Basilea III, svolta con l'intenzione di accrescere l'attendibilità del calcolo dei RWA e di facilitare il confronto degli indicatori patrimoniali delle diverse banche, visto il timore che i modelli interni di valutazione dei rischi potessero essere usati più in ottica di riduzione dei requisiti di capitale che per una quantificazione dei rischi. Le misure varate nel documento entreranno gradualmente in vigore dal 2021 al 2027.

Le novità vedono un affinamento del sistema di ponderazione dei crediti per l'Approccio Standard e una virata verso l'adozione di sistemi interni di rating, secondo il principio che chi si assume il rischio deve essere in grado di valutarlo, non delegando all'esterno questo compito che è fondamentale.

Proprio riguardo ai sistemi di rating interni viene meno la possibilità di impiegare l'approccio avanzato per le esposizioni nei confronti di istituzioni finanziarie e imprese medio-grandi con fatturato maggiore di mezzo miliardo di euro. Così facendo si riduce la variabilità nel calcolo dei RWA dovuta alla LGD e alla EAD.

Per quanto riguarda le banche di importanza sistemica (SIFI), è stata prevista l'introduzione di un ulteriore *leverage ratio buffer* con il fine di limitare l'indebitamento dei colossi e assicurare una maggior stabilità all'intero sistema.

1.3 Descrizione del rischio di credito

Il rischio di credito è il fulcro di questa analisi e, in quando tale, è meritevole di un approfondimento. In questa sezione vengono illustrate le principali problematiche in relazione al contesto italiano, per poi analizzare le singole

componenti del rischio in modo puntuale. Per concludere, si tratteranno i principali rischi insiti nel rischio di credito e si accenneranno i sistemi di rating.

1.3.1 Problematiche

Come precedentemente rilevato, il rischio di credito si riferisce a una variazione inattesa del valore di mercato di una posizione o di un portafoglio (anche dei derivati fuori bilancio) e questo implica un problema notevole. Infatti, occorre valutare il credito al valore di mercato, ma, come è noto, non tutti i crediti sono quotati sul mercato. Il problema si accentua in paesi come l'Italia, laddove il ricorso al finanziamento sul mercato tramite l'emissione di obbligazioni, appunto, quotate, non è un fenomeno largamente diffuso. In questo caso le posizioni sono trattate, per le problematiche sopra esposte, dal punto di vista meramente contabile.

1.3.2 Componenti

La perdita attesa - in lingua inglese *Expected Loss* (EL) - rappresenta il valore atteso della distribuzione delle perdite. Da questa prima definizione emerge subito il rapporto strettissimo tra il rischio di credito e il concetto di perdita. Poiché la EL rappresenta qualcosa di prevedibile, questa viene inclusa nel *pricing* del credito, seguendo la logica di un premio per il rischio più alto a fronte di aspettative di perdite elevate.

Essa è composta dall'*Exposure-At-Default* (EAD), dalla *Default Probability* (PD) e dalla *Loss-Given-Default* (LGD). La EAD rappresenta l'esposizione attesa in caso di insolvenza della controparte, la PD è la probabilità che la controparte giunga allo stato di insolvenza e la LGD esprime la perdita percentuale attesa in caso di insolvenza.

Per comprendere meglio il significato della LGD, può essere d'aiuto specificare che essa è data dal complemento a 1 del *recovery rate* (RR): attesta, quindi, in

termini di punti percentuali, quanto valore del credito non si riuscirebbe a recuperare in caso di insolvenza della controparte. In formule:

$$LGD = 1 - RR \quad (1.3)$$

La perdita attesa si esprime invece tramite la relazione:

$$EL = EAD * PD * LGD = EAD * PD * (1 - RR) \quad (1.4)$$

Per determinare la perdita attesa occorre, quindi, moltiplicare l'esposizione per la probabilità che si verifichi il default, il tutto corretto per il coefficiente 1-RR, ossia quanto di quella esposizione non riuscirei a recuperare. È evidente, quindi, che più grandi saranno EAD e PD e maggiore sarà la perdita attesa, mentre più grande sarà RR e meno la perdita attesa sarà pronunciata. Infatti, nel caso limite in cui riuscissi a recuperare per intero il valore del credito, la EL sarebbe nulla; al contrario, a fronte di un tasso di recupero pari a zero, la EL dipenderebbe in maniera esclusiva da EAD e PD.

La perdita inattesa o in inglese *Unexpected Loss* (UL) è da interpretare come la variabilità della perdita intorno al suo valore medio. Essa rappresenta il vero fattore di rischio, poiché, al contrario della perdita attesa, non esprime una componente di costo stabile, bensì è aleatoria. Essa, dunque, dipende dai singoli crediti che costituiscono il portafoglio o, più precisamente, dal loro grado di correlazione. È facile immaginare come un portafoglio costituito da crediti altamente correlati comporti rischi (e opportunità) maggiori rispetto a un portafoglio ben diversificato. Infatti, qualora la correlazione tra i titoli fosse strettissima, la svalutazione creditizia inciderebbe in maniera estremamente significativa sul portafoglio complessivo. Al contrario, un portafoglio ben diversificato contribuirebbe ad arginare l'effetto di un credito (o di un gruppo di crediti) sul complesso delle posizioni.

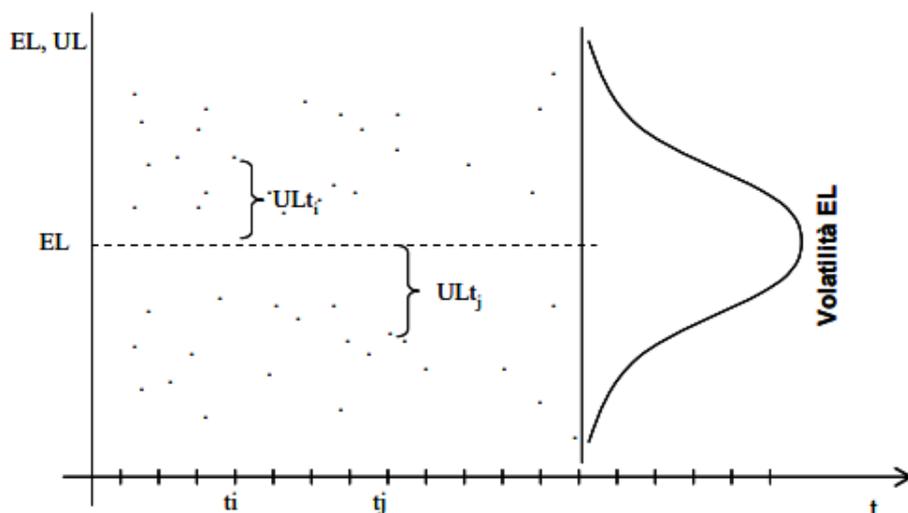


Figura 1.1 - Rappresentazione grafica di UL ed EL [1]

Da questa analisi emerge un elemento chiave di distinzione tra perdita attesa e perdita inattesa: la diversificazione. In riferimento alla UL è stato introdotto il concetto di diversificazione, spiegando come la variabilità della perdita attesa possa essere contenuta attraverso la costruzione di portafogli eterogenei. Quindi, minore è la correlazione tra le perdite non attese e, conseguentemente, minore sarà la perdita non attesa della posizione globale. Il discorso cambia se si guarda alla perdita attesa: essendo insita nella conformazione del credito, essa non è diversificabile, ma è evitabile, mitigabile o ponderabile.

Ad esempio, non concedendo un credito a una controparte con elevata probabilità di insolvenza, è possibile evitare di avere un credito potenzialmente dannoso nel portafoglio. Per quanto riguarda la mitigazione, è possibile chiedere alla controparte garanzie, così da accrescere il *recovery rate* e alleggerire la perdita attesa su quel determinato credito. Infine, la ponderazione può avvenire in ottica di portafoglio, andando a bilanciare crediti rischiosi e crediti meno rischiosi. Queste operazioni, per quanto possano sembrare astratte, sono all'ordine del giorno nell'operatività delle banche, quando decidono di iscrivere nella *blacklist* i cosiddetti "cattivi pagatori" e quando chiedono garanzie come la fideiussione per la concessione di un prestito.

Come si evince dalla figura 1.2, che mostra la distribuzione di probabilità delle perdite, la perdita attesa viene coperta mediante un *pricing* opportuno ed accantonamenti (area A), la perdita inattesa viene coperta fino ad una certa percentuale mediante capitale (area B). Oltre la soglia a un dato percentile, la perdita è caratterizzata da un grosso ammontare, ma da una probabilità esigua che si verifichi. In virtù di queste condizioni, essa viene tenuta sotto controllo con limiti di concentrazione (area C).

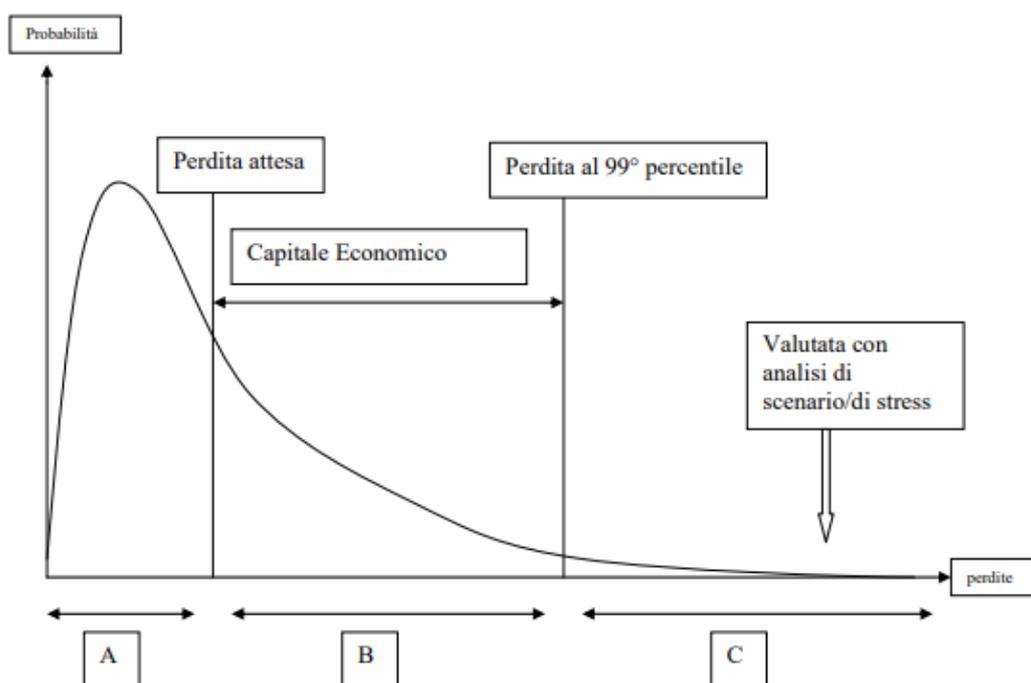


Figura 1.2 - Copertura delle perdite [1]

Exposure-At-Default

L'ammontare che può essere perso dal creditore in caso di default della controparte è l'EAD. L'entità del finanziamento, tuttavia, non è sempre facile da individuare. Nel caso dei mutui, ad esempio, il finanziamento e il piano di recupero non dipendono dalla discrezionalità di colui che si indebita: in questo caso è facile per la banca calcolare l'EAD. Quando invece si parla di fidi è il cliente a decidere se e quando utilizzare le risorse che l'istituto mette a sua disposizione: è qui presente una forte componente aleatoria che non dipende dalla volontà della banca, ma dalle necessità del beneficiario del credito.

Se nel primo caso il valore di EAD è noto a priori alla banca, in quanto in fase di negoziazione si predispongono risorse definite e limitate, nel secondo caso, l'esposizione effettiva sarà nota al concedente soltanto al verificarsi del default. Alcune prassi di comportamento vedono le realtà più in sofferenza fare uso di quante più risorse possibili, talvolta eccedendo e andando in ciò che in gergo si definisce "sconfinamento". Da segnalare come dal 1° gennaio 2021 sia entrata in vigore una nuova normativa europea atta a definire in maniera più stringente il significato di default del cliente, sia esso un'impresa che un privato.

Per quantificare l'EAD in presenza di un finanziamento del primo tipo, basta verificare la quota non ancora restituita dal cliente, mentre per il secondo i calcoli sono leggermente più complicati:

$$EAD = DP + UP * UGD \quad (1.5)$$

La *Drawn Portion* (DP) è la quota effettivamente utilizzata, la *Undrawn Portion* (UP) è la quota non ancora utilizzata. Il coefficiente *Usage Given Default* (UGD) è espresso in termini percentuali e rappresenta quanto della quota disponibile si ritiene che verrà usata da colui a cui è stato concesso, ad esempio, un fido.

Default Probability

La *Default Probability* rappresenta la probabilità che la controparte, ossia colui che riceve il finanziamento, vada incontro a un evento di default. Il verificarsi di questa condizione comporta l'impossibilità di ripagare del tutto o in parte il debito, con conseguente danno a svantaggio del creditore. L'orizzonte di tempo considerato per la valutazione della probabilità di default è generalmente pari a un anno.

Loss-Given-Default

Come evidenziato precedentemente, la LGD è espressa in funzione del *Recovery Rate*. È fondamentale capire come viene determinato RR per

comprendere a fondo la conformazione della perdita attesa. Esso dipende da una moltitudine di fattori, quali:

- Caratteristiche del finanziamento
- Caratteristiche dell'impresa finanziata
- Caratteristiche della banca
- Fattori esterni

In merito alle caratteristiche del finanziamento, è opportuno considerare: l'esistenza di garanzie e la loro efficacia, la *seniority* del credito e il tipo di contenzioso previsto in caso di recupero. Tutti questi aspetti conformano scenari differenti che possono portare a valori di recupero diversi.

Per quanto concerne le caratteristiche dell'impresa finanziata, bisogna valutare la liquidità e la possibilità di recuperare gli asset aziendali con più o meno facilità, oltre alla localizzazione geografica dell'impresa, che incide in modo rilevante su tempi ed efficacia di eventuali procedure legali.

I fattori interni alla banca si esplicitano essenzialmente con l'abilità dell'ufficio legale nel portare avanti i contenziosi in modo sia efficiente che efficace. Bisogna porre attenzione anche alle politiche di recupero crediti che si intendono attuare nel caso insorga qualche problematica con la controparte.

Infine, i fattori esterni da valutare includono il livello dei tassi di interesse e lo stato del ciclo economico.

La definizione del tasso di recupero può avvenire tramite l'utilizzo di dati di obbligazioni quotate delle imprese andate in default, ma anche, qualora l'impresa non avesse debiti quotati sul mercato, impiegando i dati interni dell'istituto.

In termini analitici, il RR è dato da:

$$RR = \frac{\sum_{t=1}^n \frac{ER_t - AC_t}{(1+i)^t}}{EAD} \quad (1.6)$$

Dove: ER rappresenta l'*Expected Recovery*, AC i costi di amministrazione del periodo, i il tasso di attualizzazione, EAD l'esposizione quando si verifica il default ed n la stima del tempo necessario per terminare il recupero.

1.3.3 Modello binomiale

Il modello binomiale è il riferimento principale per il calcolo della perdita inattesa. La caratterizzazione di tale modello è l'esistenza di due singoli stati: default o non default. L'evento default, quindi, si verificherà con probabilità PD , mentre $1-PD$ è la probabilità che non si verifichi. Il riferimento temporale del modello è un anno, mentre la distribuzione alla base della stima del VaR creditizio è la Beta.

Il quadro in tabella 1.4 illustra le casistiche considerate dal modello binomiale, espresse in termini di probabilità e di perdita, a fronte di una EAD pari a 1. L'evento di default si verifica con probabilità PD e in questo caso la perdita è proprio la LGD ; il non default non comporta perdite e si verifica con probabilità $1-PD$.

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

Tabella 1.4 - Eventi modello binomiale [1]

Non rilassando il vincolo e conservando una EAD unitaria per semplicità, a livello di posizioni singole, le formule per EL ed UL sono:

$$EL = PD * LGD \quad (1.7)$$

Se LGD è deterministica:

$$UL = \sigma = LGD * \sqrt{PD * (1 - PD)} \quad (1.8)$$

Dove il termine sotto radice esprime lo scarto quadratico medio del tasso di default.

Se LGD è stocastica, la formula si modifica nel modo seguente:

$$UL = \sigma = \sqrt{PD * (1 - PD) * LGD^2 + PD * \sigma_{LGD}^2} \quad (1.9)$$

A livello di portafoglio, la perdita attesa risulta pari alla media ponderata dei contributi delle singole posizioni. La formula diventa:

$$EL_p = \sum_i a_i * EL_i \quad (1.10)$$

La perdita inattesa è funzione delle perdite inattese delle componenti:

$$UL_p = \sqrt{\sum_i \sum_j a_i a_j UL_i UL_j \rho_{ij}} \quad (1.11)$$

I coefficienti a_i e a_j rappresentano la quota di portafoglio costituita rispettivamente dai crediti i-esimo e j-esimo, mentre ρ_{ij} esprime la correlazione tra la perdita inattesa i-esima e la perdita inattesa j-esima.

1.3.4 Rischi insiti nel rischio di credito

Il rischio di credito può essere scomposto per individuare le sotto-tipologie che lo compongono. Le principali sono: rischio di insolvenza, rischio di migrazione, rischio di esposizione, rischio di *spread*, rischio di recupero, rischio di sostituzione e rischio paese.

Rischio di insolvenza

È strettamente connesso con l'evento di default della controparte e si verifica, ad esempio, in condizioni di bancarotta. Il beneficiario del credito non è in grado di ripagare le obbligazioni assunte e questo fenomeno espone il creditore a delle perdite di entità pari alla EAD in relazione alla LGD.

Rischio di migrazione

Questo rischio è legato al rating pubblico del debitore, in quanto, un peggioramento del suo merito di credito, comporterebbe un declassamento in termini di rating.

Rischio di esposizione

Il rischio di esposizione si riferisce alla quantità di debito risultante nel momento in cui si verifica l'evento default del debitore.

Rischio di spread

Qualora cresca l'avversione al rischio sul mercato, il fenomeno che immediatamente scaturirebbe dalla circostanza comporterebbe un aumento degli *spread*. In queste condizioni, il credito vedrebbe il suo valore ridursi per adattarsi alle circostanze del mercato. È utile sottolineare come l'aumento del premio sia richiesto per una stessa classe di rating – quindi per una data probabilità di default - e non in seguito a un declassamento.

Rischio di recupero

Il verificarsi di questo rischio significa che la liquidazione ha portato a recuperare meno valore del previsto. Le possibilità di questa discrepanza con le aspettative possono essere individuate da una valutazione a priori che non

combacia con quanto registrato a consuntivo o con un eccessivo protrarsi nel tempo della procedura di liquidazione.

Rischio di sostituzione

È un rischio che riguarda i derivati OTC. Nel dettaglio, qualora la controparte di un derivato risultasse insolvente, la banca dovrebbe cercare un suo sostituto a nuove condizioni di mercato. Le possibilità che gli accordi siano di tipo peggiorativo per l'istituto sono da tenere sotto attenta considerazione.

Rischio paese

L'insorgenza di vincoli di cambio o di eventi di carattere politico e legislativo, potrebbero far venir meno la controparte estera all'adempimento dei suoi obblighi contrattuali.

1.3.5 Sistemi di rating

Il rating creditizio esprime una valutazione di sintesi circa la capacità e la volontà di un beneficiario di un credito di rispettare le sue obbligazioni. Esso esprime in modo estremamente intuitivo il profilo di rischio di credito del soggetto sottoposto ad analisi. Si tratta di un vero e proprio giudizio che combina informazioni di tipo quantitativo e qualitativo, ma è bene precisare che viene sottoposto a revisione periodica, al fine di aggiornare il risultato in base al mutamento delle circostanze.

I rating possono essere svolti dalle Agenzie: tra le più famose è bene citare Standard&Poor's, Moody's e Fitch. Le valutazioni vengono diffuse sul mercato e costituiscono informazioni preziose per gli investitori. Un elemento chiave per questo tipo di rating è la reputazione di chi lo ha emesso: una valutazione che non corrispondesse all'evoluzione della situazione dell'impresa comprometterebbe l'immagine del giudicante, così come esso perderebbe credibilità se cambiasse il rating assegnato con eccessiva frequenza. In

quest'ultimo caso, la valutazione passerebbe per temporanea e, quindi, di scarso valore per chi è intenzionato a investire.

I rating vengono assegnati considerando gli scenari pessimistici sul ciclo economico e sulle condizioni del settore in cui l'impresa opera. È buona prassi stressare le condizioni al contorno per verificare la capacità di onorare gli impegni nonostante le circostanze avverse: si parla infatti di rating sostenibile nel ciclo.

Esistono inoltre i rating interni delle banche e delle finanziarie, che servono per assegnare una valutazione alla controparte circa il suo merito di credito. A differenza del rating d'Agenzia, in questo caso non c'è un fine di pubblicità delle informazioni sui mercati. Alla banca servono informazioni precise, puntuali e che rispecchino il più fedelmente possibile la condizione attuale dell'impresa. Vi è dunque la necessità di strumenti in grado di adattarsi immediatamente alle circostanze e si parla in questo caso di rating *point-in-time*.

Per concludere, è opportuno evidenziare come il giudizio espresso dalle Agenzie non specifichi una certa probabilità di insolvenza, ma offra una classificazione per classi di rischio. Tali classi, di natura ordinale, riflettono un parere relativo alla situazione complessiva dell'impresa. Le categorie dei rating assegnati dalle principali Agenzie sono espresse nella tabella 1.5.

Moody's	S&P's	Fitch	Descrizione
Aaa	AAA	AAA	Prime
Aa1	AA+	AA+	High grade
Aa2	AA	AA	
Aa3	AA-	AA-	
A1	A+	A+	Upper medium grade
A2	A	A	
A3	A-	A-	
Baa1	BBB+	BBB+	Lower medium grade
Baa2	BBB	BBB	
Baa3	BBB-	BBB-	
Ba1	BB+	BB+	Non-investment grade speculative
Ba2	BB	BB	
Ba3	BB-	BB-	
B1	B+	B+	Highly speculative
B2	B	B	
B3	B-	B-	
Caa1	CCC+	CCC+	Substantial risks
Caa2	CCC	CCC	
Caa3	CCC-	CCC-	
Ca	CC	CC	Extremely speculative
	C	C	Default imminent
C	RD	DDD	In default
/	SD	DDD	
/	D	D	

Tabella 1.5 - Scale di rating delle principali agenzie [elaborazione]

1.3.6 Crediti bancari in sofferenza

Il mercato creditizio italiano ha subito alcuni cambiamenti negli ultimi anni. In particolare, se si guarda ai cosiddetti *Non-Performing Loan* (NPL), ci si accorge che si sta assistendo a dei mutamenti.



Fonte: segnalazioni di vigilanza e Centrale dei rischi.

(1) Flusso di sofferenze rettificato nel trimestre in percentuale dei prestiti non in sofferenza rettificata alla fine del trimestre precedente, espresso su base annua. Per la definizione di sofferenze rettificate, cfr. nell'Appendice la voce del Glossario: Sofferenze rettificate. Tutte le serie sono depurate dalla componente stagionale, dove presente. Per la definizione delle serie, cfr. nell'Appendice la sezione: Note metodologiche.

Fonte: Banca d'Italia – relazione per l'anno 2013

Figura 1.3 - Crediti bancari in sofferenza [1]

Il grafico in figura 1.3 mostra la situazione al 2013, quando, considerando il totale dell'economia, le nuove sofferenze ammontavano a circa il 2,5% dei prestiti. La situazione oggi è in evoluzione e, a quanto si apprende sulla Relazione della Banca d'Italia per l'anno 2019:

Il flusso di nuovi prestiti deteriorati in rapporto al totale dei crediti è sceso all'1,2 per cento, un valore molto contenuto nel confronto storico e inferiore di quasi un punto percentuale rispetto alla fine del 2007, prima della doppia recessione del periodo 2008-2013. [...] Negli ultimi anni la cessione di crediti deteriorati da parte delle banche italiane è stata molto ingente, in risposta alle sollecitazioni sia delle autorità di vigilanza sia del mercato. Dal 2016, quando le transazioni sul mercato dei crediti bancari deteriorati hanno cominciato ad acquisire rilevanza, le cessioni sono state pari, al lordo delle rettifiche di valore, a 154 miliardi. Le vendite, effettuate sia dalle banche significative sia da quelle meno significative, sono risultate superiori ai piani presentati annualmente dagli intermediari [8].

Quanto si apprende dalla Relazione, è quindi una spinta del Regolatore e del mercato affinché i bilanci delle banche venissero "ripuliti" dai crediti deteriorati. La risposta degli istituti di credito, guardando i numeri, è stata significativa e decisa verso la direzione indicata, ma la situazione è tuttora in divenire, pertanto occorre continuare a monitorarla.

1.4 Analisi del rischio di credito

Si parla di *credit scoring* in relazione ai modelli utilizzati per la valutazione della solvibilità della controparte. Vengono di seguito presentate le principali impostazioni metodologiche, la cui intenzione è stimare un modello per la classificazione delle società in due categorie: sane e anomale.

1.4.1 Linear probability model

È un modello di derivazione statistica basato sulla regressione multipla per stimare la probabilità di default. Per prima cosa, occorre definire la variabile di risposta y_i , che assumerà in valore 0 se la società i -esima è sana o 1 se è anomala. La funzione di regressione è la seguente [1]:

$$y_i = \beta' x_i + \varepsilon_i \quad (1.12)$$

Con:

β = vettore dei coefficienti che determina l'impatto delle variabili esplicative;

x_i = matrice delle variabili osservate sulle imprese dei due insiemi;

ε_i = residui, con $E(\varepsilon_i) = 0$.

La probabilità che si verifichi l'evento default, condizionatamente alla realizzazione di x_i , è l'aspettativa condizionata:

$$E(y_i|x_i) = \beta' x_i \quad (1.13)$$

Il valore calcolato di y_i è $\hat{y}_i = \hat{\beta}' x_i$. Tuttavia, questa probabilità, senza alcun senso economico o statistico, può assumere valori minori di zero e superiori a 1.

I valori dei residui della regressione sono da valutare in relazione ai valori assunti da y_i :

y_i	ε_i	$\text{Probi}[f(\varepsilon_i)]$
0	$-\beta'x_i$	$1-\beta'x_i$
1	$1-\beta'x_i$	$\beta'x_i$

Tabella 1.6 - Valori dei residui [elaborazione]

Dato che $E(\varepsilon_i) = 0$, la varianza di ε_i è data da:

$$\text{var}(\varepsilon_i) = \beta'x_i(1 - \beta'x_i) = E(y_i)[1 - E(y_i)] \quad (1.14)$$

Tale varianza può assumere valori inferiori allo zero e ciò non è ammissibile per una varianza.

Il modello ha ulteriori complicazioni, in quanto la distribuzione dei residui non segue la distribuzione normale e di conseguenza la stima con il metodo dei minimi quadrati non è generalmente efficiente, lasciando prediligere il criterio della massima verosimiglianza. Inoltre, si osserva come la volatilità dei residui dipenda da x_i e sia eteroschedastica. Tutti questi elementi di svantaggio hanno fatto andare il modello in disuso, rendendolo obsoleto.

1.4.2 Analisi discriminante

L'analisi discriminante si fonda su un mondo composto dalla popolazione delle imprese sane e dalla popolazione delle imprese anomale, nettamente distinte. Il fondamento del modello è la costruzione di una funzione lineare che identifichi la distanza di una nuova osservazione dalle due popolazioni. Questa analisi può seguire un approccio non parametrico oppure un approccio parametrico. Nell'approccio non parametrico non si avanzano ipotesi sulla forma della distribuzione dei dati di partenza. L'approccio parametrico, invece, assume che le variabili esplicative seguano una normale multivariata. Inoltre, qualora la matrice varianza-covarianza non fosse uguale nei due insiemi di imprese, la

funzione discriminante perderebbe la caratteristica di linearità. L'obiettivo è trovare una funzione lineare delle variabili esplicative che consenta di ottenere la miglior discriminazione dicotomica - sane e anomale - tra due gruppi di imprese.

Per la costruzione della funzione discriminante, si è soliti modellare la differenza *sane* - *anomale*. Questo accorgimento permette di ottenere i segni dei coefficienti in coerenza con l'influenza economica delle variabili sul default. La formula [1] è:

$$(\mu_S - \mu_A)'V^{-1}X_h + (\mu_S - \mu_A)'V^{-1} \left[- \left(\frac{\mu_A + \mu_S}{2} \right) \right] > 0 \quad (1.15)$$

Dove:

μ_S e μ_A = vettori delle medie delle popolazioni;

X_h = vettore di variabili della nuova osservazione;

V^{-1} = inversa della matrice varianza-covarianza complessiva;

$(\mu_S - \mu_A)'V^{-1}$ = vettore dei coefficienti;

$(\mu_S - \mu_A)'V^{-1} \left[- \left(\frac{\mu_A + \mu_S}{2} \right) \right]$ = costante.

Se questa relazione viene rispettata e la funzione è maggiore di 0, allora la società h appartiene alla categoria delle sane, altrimenti è anomala.

1.4.3 Regressione logistica

Anche questo modello si utilizza quando si intende approfondire la relazione causale tra la variabile di risposta dicotomica e le variabili indipendenti. La y assume il valore di 0, se sana, e di 1, se anomala. La retta di partenza è identificata dalla relazione [1]:

$$y_i = \alpha + \beta'X_i + \tilde{u}_i \quad (1.16)$$

Assumendo che la distribuzione dei residui u_i sia di tipo logistico, si ottengono le seguenti probabilità:

$$Prob(y_i = 1|X_i = x_i) = P_i = \frac{e^{(\alpha + \beta'X_i)}}{1 + e^{(\alpha + \beta'X_i)}} \quad (1.17)$$

$$Prob(y_i = 0|X_i = x_i) = 1 - P_i = \frac{1}{1 + e^{(\alpha + \beta'X_i)}} \quad (1.18)$$

Definite le due probabilità, è possibile determinare il loro rapporto (*odds*):

$$\frac{P_i}{1 - P_i} = e^{(\alpha + \beta'X_i)} \quad (1.19)$$

Applicando il logaritmo naturale a questa relazione si ottiene ciò che viene definito *logit*:

$$logit = \ln\left(\frac{P_i}{1 - P_i}\right) = \alpha + \beta'X_i \quad (1.20)$$

Dall'*odds* è altresì ricavabile la PD attraverso la formula inversa:

$$PD = P_i = \frac{1}{1 + e^{-(\alpha + \beta'X_i)}} = \frac{1}{1 + e^{-logit}} \quad (1.21)$$

Grazie alle PD è possibile determinare l'appartenenza di una società alla categoria delle sane o delle anomale.

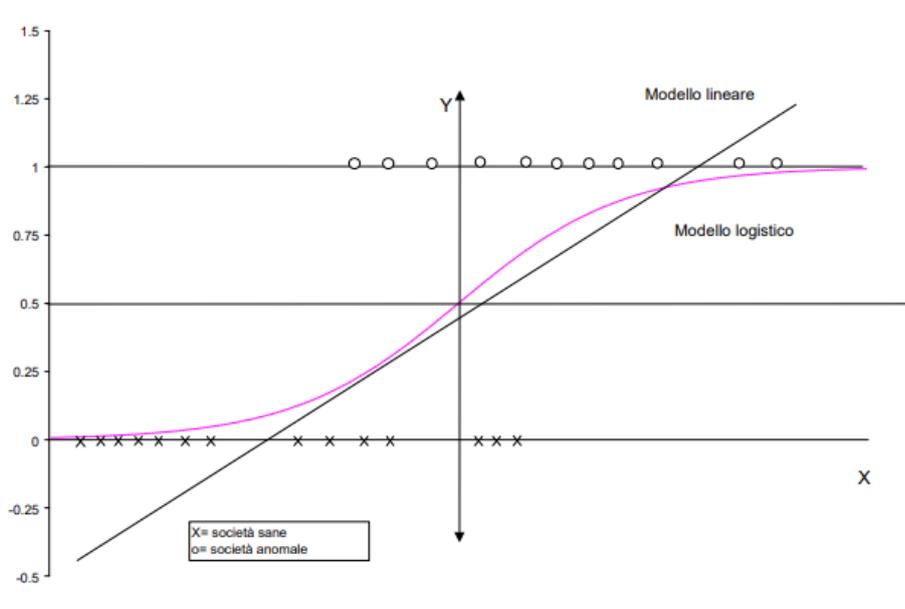


Figura 1.4 - Confronto tra un modello lineare e un modello logistico [1]

1.4.4 Modelli di apprendimento automatico

Un approccio differente alla questione del *credit scoring* è rappresentato dai modelli di apprendimento automatico. Si tratta di tecniche e algoritmi che sfruttano i paradigmi dell'intelligenza artificiale per raggiungere, nel caso del rischio di credito, obiettivi di classificazione. Questi modelli sono estremamente versatili e trovano applicazione in diversi ambiti, dal riconoscimento delle immagini alla finanza, appunto.

A titolo di esempio, è molto utile impiegare le tecniche di apprendimento automatico quando è necessario che il modello si adatti alle condizioni al contorno o quando il sistema va incontro ad esigenze di miglioramento delle prestazioni riguardo allo svolgimento di un compito. Gli svantaggi degli approcci tradizionali riguardano l'impossibilità di formalizzare il problema mediante un algoritmo esatto, la presenza di disturbi nelle informazioni, quali rumore o incertezza e un'elevata complessità nell'identificazione di una soluzione.

Al fine di utilizzare queste tecniche è necessario l'impiego dei dati. Si stabilisce generalmente un set che verrà impiegato per l'addestramento del modello

(*training set*), un secondo set sarà usato per la convalida del modello (*validation set*), infine il *test set* servirà a stimare la bontà della soluzione.

Capitolo II

2 Reti neurali

Prima della costruzione dei modelli basati sulle reti neurali è opportuno capire il funzionamento del neurone biologico, per poi passare alle tappe che hanno caratterizzato l'evoluzione della disciplina nella storia. Una volta compresi gli elementi di base che compongono una rete neurale, saranno illustrate le principali architetture di rete, fino a un breve cenno sulle possibili applicazioni di questa metodologia, mettendo in evidenza i suoi vantaggi e svantaggi.

2.1 Reti neurali biologiche

Il modello di riferimento delle reti neurali artificiali è riconducibile alle reti neurali biologiche. Le reti neurali dell'uomo sono sistemi complessi che ci consentono di recepire stimoli dall'ambiente per fornire risposte dipendenti dalle circostanze. Si tratta quindi, sostanzialmente, dell'elaborazione di un output in relazione agli input ricevuti.

Le reti neurali biologiche sono costituite da cellule nervose interconnesse tra loro. Questa connessione consente di "far viaggiare" lo stimolo lungo il sistema. Ad esempio, avvicinando la mano a una fiamma, il sistema percepisce calore e invia lo stimolo al cervello, il quale elabora e trasmette la risposta che consiste nell'allontanare la mano. Questo processo, in apparenza banale, richiede un sistema ricco di componenti e complesso nelle sue funzionalità. Si noti che il numero di neuroni del sistema nervoso centrale ha un ordine di grandezza pari a 10^{11} . Fanno parte del sistema nervoso umano:

- Soma neuronale = sono i corpi dei neuroni. Sono in grado di ricevere e processare le informazioni. Nel caso in cui il potenziale di azione sia maggiore di una determinata soglia, essi generano degli impulsi che si propagano nel sistema;

- Neurotrasmettitori = sono dei composti biologici sintetizzati all'interno dei somi neuronali. Il loro compito è la modulazione degli impulsi di tipo nervoso;
- Assoni = generalmente se ne trova uno per ogni cellula nervosa e corrisponde alla via di comunicazione in uscita dal neurone;
- Dendriti = sono la via di comunicazione in ingresso e sono molteplici per ogni neurone. L'insieme dei dendriti forma l'albero dendritico;
- Sinapsi = si tratta di siti funzionali altamente specializzati in cui avviene la trasmissione delle informazioni tra i neuroni. Ciascun neurone possiede migliaia di sinapsi. Queste possono svolgere un'azione eccitatoria o inibitoria, rispettivamente facilitando o inibendo la trasmissione dell'impulso nervoso. La connessione avviene tramite il rilascio dei neurotrasmettitori nello spazio inter-sinaptico, così che questi raggiungano il neurone successivo e trasmettano l'impulso in seguito all'alterazione della permeabilità della membrana post-sinaptica.

Il neurone può ricevere contemporaneamente diversi segnali dalle sinapsi. Esso è in grado di misurare il potenziale elettrico dei segnali e di stabilire conseguentemente se è stata raggiunta la soglia di attivazione: in caso positivo, il neurone genera un impulso nervoso.

Un elemento chiave della rete biologica è la dinamicità della rete sinaptica. Infatti, il numero di sinapsi varia di volta in volta a seconda degli stimoli che la rete riceve: tanto più grande è il numero di stimoli, tanto maggiore sarà il numero di connessioni sinaptiche create. La rete è quindi calibrata in modo adattivo a seconda delle circostanze.

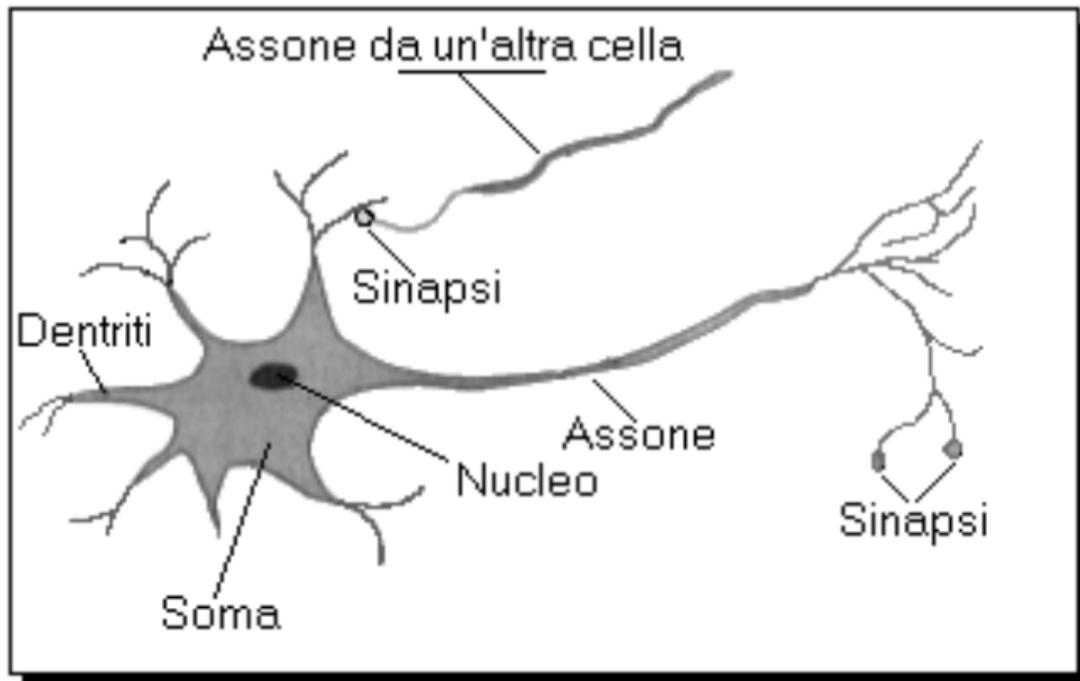


Figura 2.1 - Schema del neurone biologico [11]

2.2 Cenni storici sulle reti artificiali

Per arrivare a comprendere le reti neurali artificiali moderne, è bene ripercorrere le tappe che hanno caratterizzato l'evoluzione della materia. Il primo tentativo di analogia tra un computer e il cervello umano risale al 1936 e fu opera del celebre matematico britannico Alan Turing. La proposta fu la naturale conseguenza degli studi sempre più approfonditi dell'Ottocento sul cervello umano e sul suo funzionamento.

Bisogna attendere il 1943 per vedere Warren McCulloch e Walter Pitts riprodurre una semplificazione di rete neurale con l'utilizzo di circuiti elettrici. Essi dimostrarono l'analogia tra il funzionamento dei neuroni e la macchina di Turing, predecessore dei moderni computer. La rete progettata dalla coppia di scienziati era in grado di ricevere, per ogni elemento della rete, n dati di tipo binario e di produrre un unico dato in uscita. L'ambito di lavoro riguardava esclusivamente le funzioni booleane elementari.

Nel 1949 Hebb ipotizzò la possibilità di istruire la macchina seguendo un principio di apprendimento simile a quello che sta alla base dell'intelligenza umana. Il lavoro dello psicologo si basa su tre ipotesi [13]:

- I neuroni rafforzano le loro connessioni quando risultano frequentemente attivi contemporaneamente (regola di Hebb);
- La corteccia non è altro che una grande memoria associativa in cui il rafforzamento delle sinapsi avviene sia tra neuroni vicini che tra neuroni posti in zone differenti;
- I neuroni legati da un legame forte agiscono come unità funzionale. Quindi, se vengono attivati soltanto alcuni neuroni, si attiva comunque l'intero gruppo a causa delle connessioni.

Nel 1958 Rosenblatt propose il primo modello di rete neurale, denominato *Perceptron*. Tale rete possiede uno strato di neuroni artificiali di input e un nodo di output. La "forza della connessione" è espressa dai pesi sinaptici, che sono di natura dinamica. In questo modo, la macchina può apprendere, così come teorizzato da Hebb, sulla base del funzionamento delle reti neurali biologiche. Il modello è di tipo *feedforward*, con gli impulsi che si propagano esclusivamente in avanti.

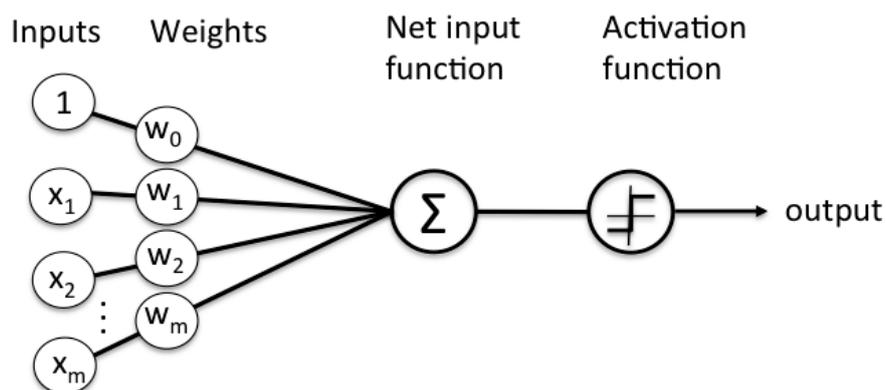


Figura 2.2 - Schema del Perceptron di Rosenblatt [14]

Si arrivò poi a definire il *Multi-Layer Perceptron* (MLP) che si differenzia per l'esistenza di uno strato nascosto (*hidden layer*) tra i nodi di input e quello di output. È proprio in questo strato intermedio che avviene l'elaborazione delle

informazioni che arrivano come input e vanno inviate all'output. Si tratta di un modello di tipo *feedforward* con funzione di attivazione non lineare e con connessioni multiple in ingresso e in uscita da ogni nodo. A lavorare sul modello MLP fu Werbos, quando nella sua tesi di dottorato del 1974 ne spiegò la metodologia di apprendimento.

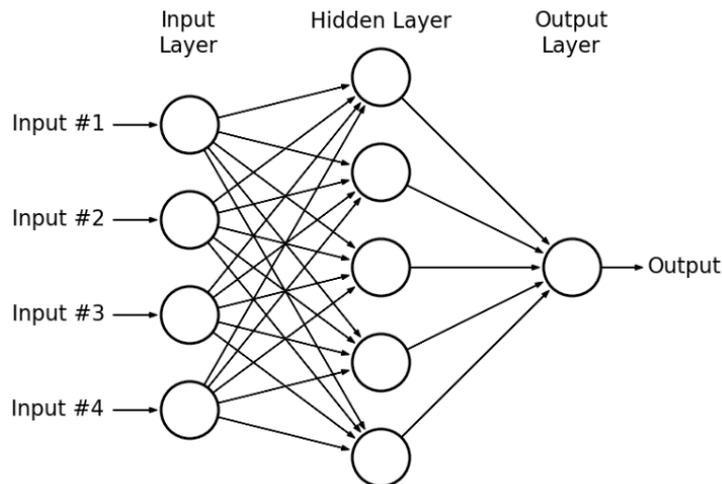


Figura 2.3 - Schema di MLP [15]

Si deve a Rumelhart, Hinton e Williams l'elaborazione dell'algoritmo di retro-propagazione (*Error Back-Propagation*). Tale strumento risale alla seconda metà degli anni Ottanta e viene ancora oggi utilizzato. Il concetto base del suo funzionamento è il perfezionamento dell'apprendimento servendosi di più stadi. Si ricerca l'output ottimale andando di volta in volta a modificare i pesi delle connessioni che non producono tale risultato.

Qualche anno prima dell'EBP si assiste alla nascita delle reti Hopfield: si tratta di reti neurali con architetture *feedback*. L'innovazione sta nel fatto che le informazioni possono viaggiare in qualunque direzione, ossia in avanti, all'indietro e fra i nodi di uno stesso livello.

Kohonen realizzò una *Self-Organizing Maps* (SOM), ossia una rete con struttura sia *feedforward* che *feedback* in grado di cambiare la configurazione dei nodi in relazione al peso che assumono con la graduale introduzione degli input: i nodi con pesi diversi si allontanano, al contrario si avvicinano quando i pesi sono simili.

Nel 1990 Elman progettò un modello di rete ricorrente, la quale aggiunge alla struttura MLP alcuni nodi atti a conservare le informazioni della configurazione precedente di valori della rete. Questo tipo di rete apporta notevoli vantaggi per il calcolo delle sequenze temporali.

2.3 Modello generale

La caratteristica principale delle reti neurali è l'elaborazione delle informazioni parallela e distribuita. Questa peculiarità costituisce anche un elemento di distinzione rispetto ai computer tradizionali, i quali seguono l'architettura di von Neumann. Tale struttura è composta da un'unità di elaborazione centrale (CPU) collegata tramite un canale di transito dei dati ad una memoria. Sebbene questo modello abbia portato allo sviluppo dell'informatica moderna, esso risulta sequenziale e richiede una grande quantità di tempo per processare le informazioni ed accedere alla memoria. Le reti neurali, invece, decentralizzano l'elaborazione verso le unità autonome (nodi) e la memorizzazione nelle connessioni, in maniera simile a quanto avviene nei processi biologici del cervello umano e permettono di apprendere tramite l'esperienza. L'elemento fondamentale delle reti neurali artificiali è appunto il neurone artificiale, il quale convive in un sistema insieme ad altri neuroni per formare un complesso di relazioni.

2.3.1 Il neurone di McCulloch-Pitts

Per prima cosa è importante capire il funzionamento del neurone: esso è una componente importantissima della rete ed è una formalizzazione matematica del funzionamento del neurone biologico. Il neurone j riceve degli input dagli altri neuroni x_1, x_2, \dots, x_n . Gli input corrispondono ai dendriti del neurone biologico. Poiché il neurone tende a rinforzare il legame con i neuroni con cui comunica più di frequente, è opportuno discriminare gli input assegnando dei pesi w_i alle sinapsi. Questo accorgimento permette di differenziare la rilevanza

delle informazioni a seconda della loro provenienza, in pieno accordo con la regola di Hebb.

L'informazione biologica corrisponde a un potenziale elettrico che i dendriti portano al neurone per l'accumulo: analogamente, all'interno del neurone formale di McCulloch-Pitts, si sommano tutti gli stimoli in ingresso tenendo conto dei rispettivi pesi e andando così a creare un potenziale interno:

$$\text{potenziale interno} = E = \sum_{i=1}^N w_i x_i \quad (2.1)$$

A questo punto, l'output è funzione dello stato interno. In particolare, questo viene formulato sulla base di una funzione di attivazione $y = f(E)$ che stabilirà l'informazione da trasmettere. Per il neurone in esame, se il valore di E supera una certa soglia, il neurone fornirà 1 come output, altrimenti y varrà 0, coerentemente con il contesto binario del modello.

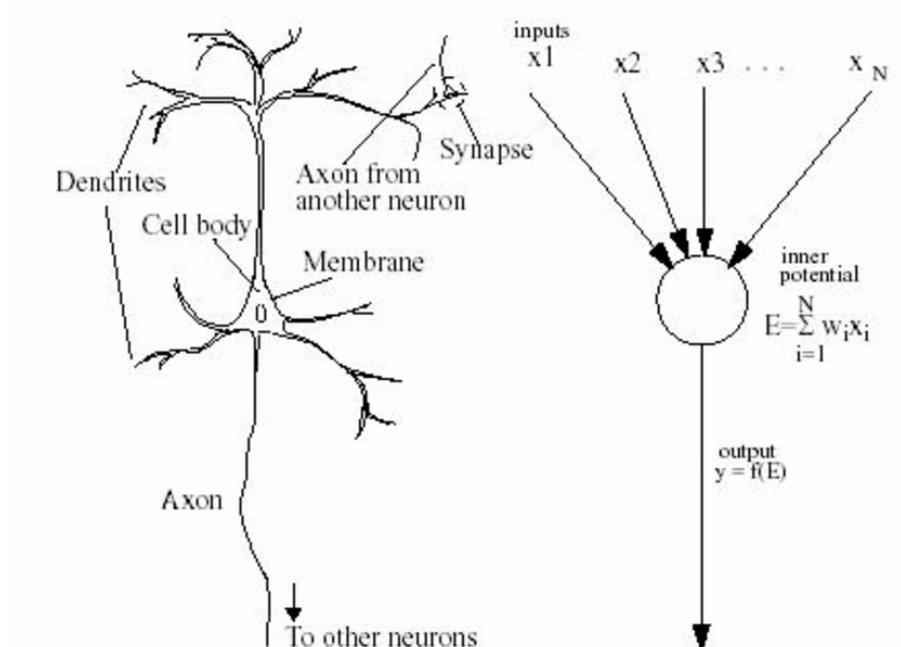


Figura 2.4 - Confronto tra neurone biologico e artificiale [18]

2.3.2 Componenti di una rete neurale

I neuroni non sono entità a sé stanti, ma si relazionano e interagiscono con altri neuroni: si parla infatti di rete neurale. In letteratura esistono differenti versioni di reti neurali, alcune delle quali sono state precedentemente citate, ma, nonostante le differenze teoriche di base, sono accomunate da alcuni elementi:

- un insieme di unità elaborative
- uno stato di attivazione
- una funzione di output per ogni unità
- le connessioni fra le unità
- una regola di propagazione
- una funzione di attivazione
- una regola di apprendimento

2.3.2.1 Unità elaborative

Sono entità semplici e di natura uniforme, in virtù del parallelismo che caratterizza le reti neurali: la computazione, quindi, si realizza in modo distribuito. Ciò che avviene nell'unità è la ricezione di un input da parte di un insieme di unità connesse in ingresso ed il calcolo di un output da mandare a un altro insieme di unità connesse, questa volta, in uscita. Esistono tre differenti tipologie di unità: le unità di input, le unità di output e le unità interne. Queste ultime costituiscono l'*hidden layer*.

2.3.2.2 Stato di attivazione

Lo stato di attivazione è proprio di ogni unità di elaborazione, la quale è caratterizzata da un determinato valore di attivazione al tempo t . I diversi modelli di reti neurali si differenziano, tra i vari elementi, per i valori di stato che le unità possono assumere.

Valori discreti	Esempi
Binari	{0,1}, {+1,-1}

Insieme ristretto	$\{-1,0,+1\}, \{1,2,5,8\}$
Valori continui	Esempi
Limitati	$[0,1]$
Illimitati	$[-\infty, +\infty]$

Tabella 2.1 - Valori dello stato di attivazione [elaborazione]

2.3.2.3 Funzione di output

Il meccanismo di interazione prevede la trasmissione di un segnale da parte di un'unità alle altre unità a essa connesse. La forza di tali segnali esprime il grado d'influenza esercitato da un'unità sullo stato di attivazione delle altre. In particolare, a ogni unità è associata una funzione di output che trasforma il valore dello stato di attivazione nel segnale di output vero e proprio.

Il valore dell'output e della funzione variano a seconda dei modelli. In alcuni l'output ha valore pari allo stato di attivazione, in altri la funzione rappresenta una soglia superata la quale l'unità influenza le altre; nei modelli più raffinati la funzione di output è una funzione stocastica del valore di attivazione.

2.3.2.4 Connessioni

Insieme alla logica di rete, non può venir meno il concetto di connessione. Le unità sono infatti connesse per formare una struttura complessa. Convenzionalmente, si indica con W la matrice delle connessioni, che è composta dai singoli legami w_{ji} . In genere, i rappresenta l'unità di arrivo e j quella di partenza. La forza della connessione può assumere valori maggiori, uguali o minori di zero:

Forza della connessione	Significato
$w_{ji} > 0$	Connessione eccitatoria
$w_{ji} = 0$	Nessuna connessione
$w_{ji} < 0$	Connessione inibitoria

Tabella 2.2 - Schematizzazione delle connessioni [elaborazione]

L'insieme delle connessioni rappresenta la conoscenza della rete ed è essa a determinare la risposta. La memoria risiede proprio nei legami, i quali non vengono stabiliti a priori, ma appresi dalla rete in modo automatico. Tali connessioni possono essere orientate verso un'unica direzione come nel caso di reti *feedforward* o prevedere meccanismi di *feedback*.

2.3.2.5 Regola di propagazione

La regola di propagazione denota il modo in cui gli output delle unità vengono diffusi mediante le connessioni della rete per generare i nuovi input. La propagazione è indipendente per ogni tipo di connessione, siano esse eccitatorie o inibitorie. Generalmente questo principio si concretizza con una somma pesata degli input che arrivano all'unità, tenendo conto della tipologia di connessione.

2.3.2.6 Funzione di attivazione

Si tratta di una funzione che determina il valore del segnale in uscita, ricevendo come parametro la somma pesata degli input (stato interno del neurone). In questo frangente possono entrare in gioco i cosiddetti *bias*, ossia dei termini che si aggiungono ai pesi modificando il risultato della somma degli input e, di fatto, anche la soglia di attivazione. Le funzioni di attivazione più comuni sono a gradino, lineare o sigmoidea.

A gradino

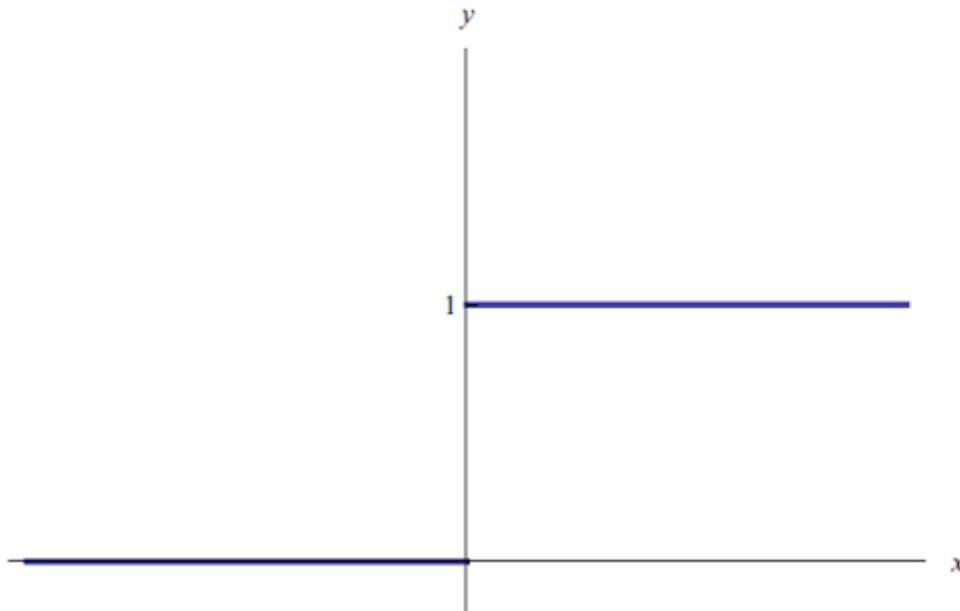


Figura 2.5 - Funzione di Heaviside [20]

Come si vede dalla sua rappresentazione grafica [figura 2.5], la funzione a gradino (o funzione di Heaviside) è discontinua. Il valore della y (output) dipende dal valore della x (stato di attivazione). Secondo questa relazione, quando il valore dell'input è maggiore o uguale a zero, l'output ha valore unitario; quando il valore dell'input è inferiore allo zero, allora l'output è nullo. L'output è quindi di tipo binario. Ha come vantaggi il fatto che è veloce e facile da usare, ma non è invertibile e non ha una zona lineare. La funzione di Heaviside può essere riassunta nel modo seguente:

$$y = H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (2.2)$$

Rampa lineare

La funzione è continua e assume il valore 0 quando la x è inferiore a un certo valore a , mentre vale 1 quando la x è superiore a un valore b , infine vale x nell'intervallo di valori compreso tra le due soglie. Ha una zona lineare ed è relativamente facile da implementare, presenta lo svantaggio di non essere invertibile. In formule:

$$y = R(x) = \begin{cases} 0, & x \leq a \\ x, & a < x < b \\ 1, & x \geq b \end{cases} \quad (2.3)$$

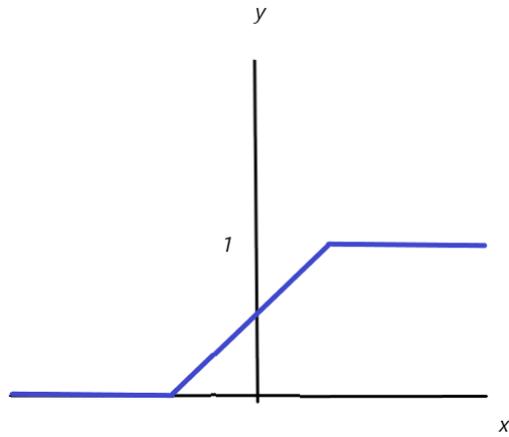


Figura 2.6 - Funzione a rampa lineare [elaborazione]

Sigmoidea logistica

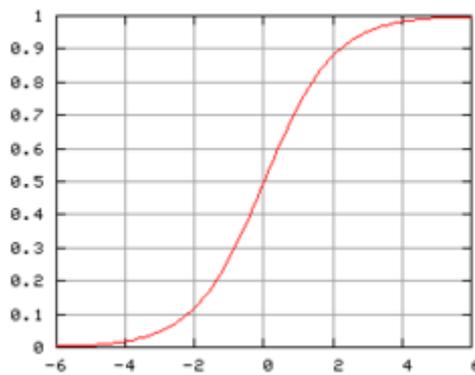


Figura 2.7 - Funzione sigmoidea logistica [21]

La sigmoidea logistica è una funzione espressa dalla relazione:

$$y = S(x) = \frac{1}{1 + e^{-x}} \quad (2.4)$$

Essa restituisce valori tra 0 e 1 in un intervallo di tipo continuo. È invertibile e derivabile, ma la sua implementazione è più complessa delle precedenti funzioni citate e il suo output si limita ai soli valori positivi.

Sigmoidea simmetrica

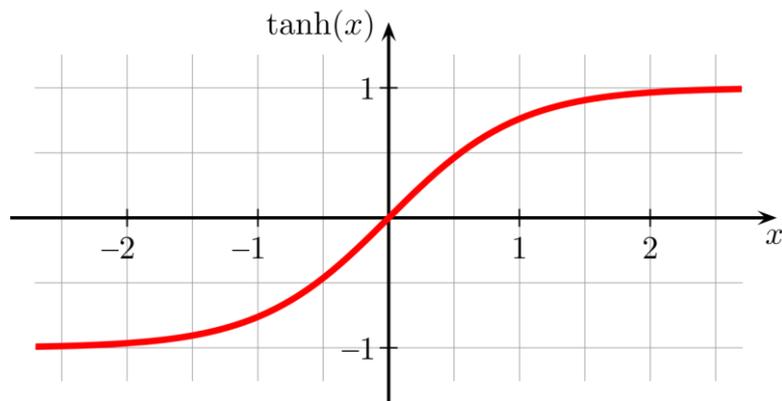


Figura 2.8 - Funzione sigmoidea simmetrica [22]

La sigmoidea simmetrica, a differenza della logistica che può assumere valori tra 0 e 1, ha un codominio che spazia nel range di valori tra -1 e +1. La formula è molto semplice, trattandosi di fatto della tangente iperbolica:

$$y = S(x) = \tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \quad (2.5)$$

È una funzione invertibile e differenziabile e può produrre output sia positivi che negativi. Non è però facile da implementare.

2.3.2.7 Apprendimento

Esistono tre principali paradigmi di apprendimento automatico e sono: l'apprendimento supervisionato, l'apprendimento non supervisionato e, infine, l'apprendimento con rinforzo.

Apprendimento supervisionato

Si basa sulla creazione di una sorta di database di esperienze da fornire alla macchina, così che questa, posta di fronte a un problema, possa attingere da questo bagaglio per identificarne una risoluzione. Esso prevede dei dati in input, una risposta di output codificata e una funzione che associa l'input all'output corretto. Il modello costruito tramite apprendimento supervisionato può poi essere utilizzato per ottenere variabili di output categoriche (classificazione) o continue (regressione).

Apprendimento non supervisionato

L'intento è l'estrazione automatica da un database di conoscenza, senza che ci sia un approfondimento dei contenuti da analizzare. In questo caso si hanno solamente i dati in input e non si fornisce alla macchina nessuna indicazione sull'output desiderato. Mentre per il supervisionato sono presenti degli esempi di utilizzo, in questo caso la macchina ha libertà di scelta circa l'individuazione dei risultati migliori. Gli algoritmi si differenziano in:

- Clustering = hanno l'obiettivo di fornire un raggruppamento dei dati per categorie omogenee;
- Analisi delle associazioni = si desiderano delle regole in grado di descrivere quanti più dati possibile.

Apprendimento con rinforzo

La macchina è dotata di strumenti – sensori, telecamere, ecc. - per migliorare l'apprendimento e per ricevere stimoli al fine di comprendere l'evoluzione dell'ambiente circostante. Si tratta di definire quali scelte rispondano meglio in chiave adattiva rispetto al mutare delle condizioni. Tra i principali algoritmi di questo tipo si trovano alcune applicazioni delle reti neurali.

Regola di apprendimento

Il processo di apprendimento non avviene tramite la programmazione di un essere umano, ma si fonda su un meccanismo automatico. Questo principio si traduce nella modifica delle connessioni seguendo una determinata regola di apprendimento. I cambiamenti possibili consistono essenzialmente nella modifica della forza dei legami, da cui possono scaturire la creazione di nuove connessioni e la perdita di quelle esistenti.

Il pilastro fondamentale delle diverse regole di apprendimento è la regola di Hebb, che è utile tradurre in formule [16] in questo frangente:

$$\Delta w_{ji} = g(a_i(t), d_i(t)) * h(o_j(t), w_{ji}) \quad (2.6)$$

La variazione della connessione da u_j a u_i (Δw_{ji}) è pari al prodotto tra la funzione g e la funzione h , dove g è a sua volta funzione dello stato di attivazione del neurone u_i al tempo t e di un input di addestramento d_i , mentre h dipende dall'output di u_j e dall'entità del legame w_{ji} .

La versione più semplice della relazione vede l'input di addestramento pari a zero, con g e h funzioni lineari soltanto rispettivamente dello stato di attivazione a_i e dell'output o_j . La formula si modifica giungendo alla regola di Hebb classica, dove η rappresenta la velocità di apprendimento:

$$\Delta w_{ji} = \eta a_i o_j \quad (2.7)$$

Un'ulteriore variante significativa della relazione è la *Delta rule*, in cui l'apprendimento dipende dalla differenza tra il valore di attivazione desiderato e il valore corrente:

$$\Delta w_{ji} = g(d_i(t) - a_i(t)) * o_j(t) \quad (2.8)$$

2.4 Tipologie di reti

Durante la panoramica delle tappe storiche che hanno portato all'evoluzione delle reti neurali artificiali si è parlato di reti di tipo *feedforward* e *feedback*. In questa sezione vengono analizzate nel dettaglio le caratteristiche di queste due tipologie di reti, le quali, al loro interno, possono essere suddivise in ulteriori sottocategorie. La figura 2.9 propone una possibile classificazione delle principali configurazioni di rete.

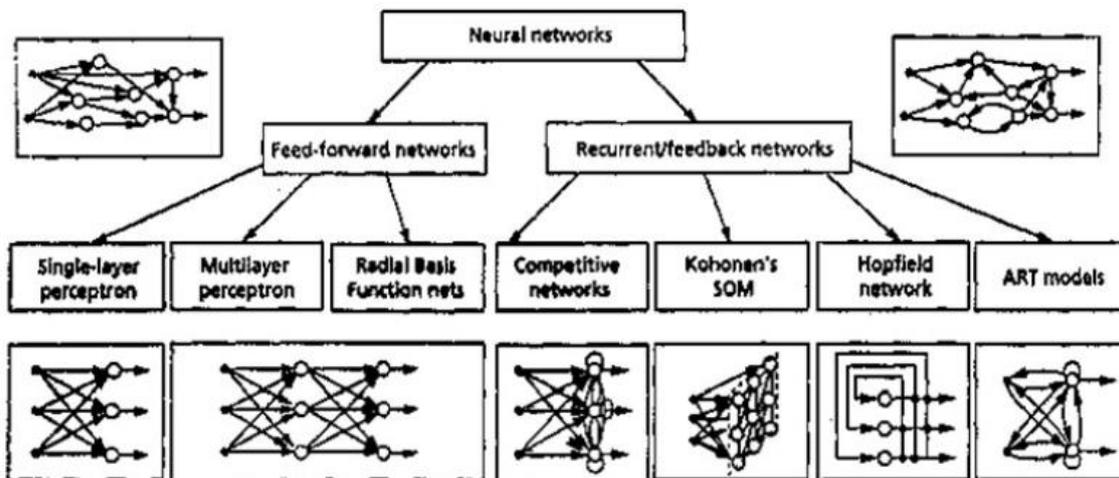


Figura 2.9 - Rappresentazione delle principali architetture [25]

Feedforward

Le reti di tipo *feedforward* sono caratterizzate da connessioni esclusivamente "in avanti". Volendo considerare la rete come un grafo orientato, si osserva come gli archi connettano neuroni di livelli diversi in un'unica direzione e come non ci sia collegamento tra neuroni appartenenti allo stesso *layer*. Questa topografia è molto utilizzata nel contesto applicativo delle reti neurali. Il funzionamento è piuttosto semplice: i neuroni del primo strato forniscono l'input ai neuroni presenti nel livello nascosto, i quali elaborano le informazioni e le trasmettono ai neuroni di output da cui è possibile ottenere la risposta finale.

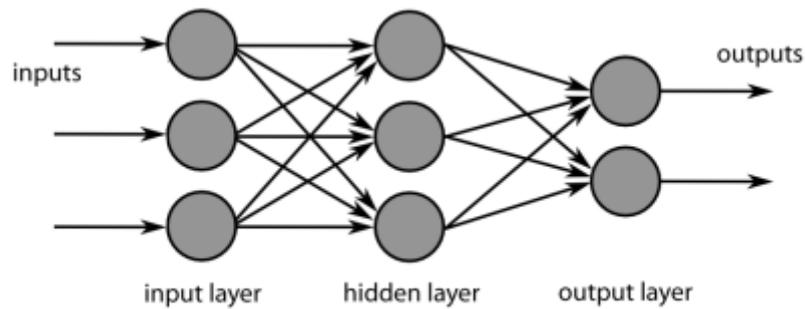


Figura 2.10 - Rete feedforward [12]

Feedback

Le reti di tipo *feedback* prevedono connessioni all'indietro, anche in direzioni di neuroni di livelli precedenti. La conseguenza di questo meccanismo aumenta la complessità della rete e del flusso di informazioni poiché bisogna relativizzare il tutto in base all'istante di tempo di riferimento. Tale rete è particolarmente adatta a elaborare delle sequenze perché la memoria che risiede nei legami rende disponibile l'informazione di istanti di tempo precedenti.

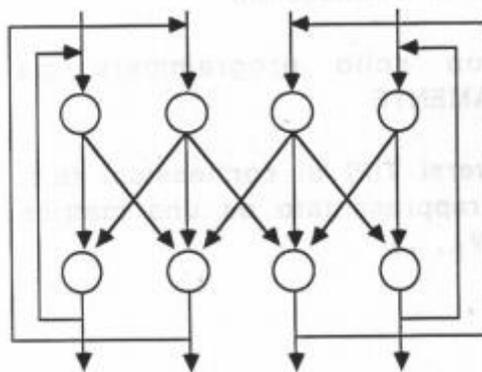


Figura 2.11 - Rete feedback [16]

2.5 Algoritmo di Back-Propagation

L'apprendimento di una rete non differisce in termini generali dall'apprendimento umano: entrambi si basano su un meccanismo di fondo che prevede il miglioramento dovuto all'esperienza nella capacità di esecuzione di un compito. Così come un bambino nasce senza esperienza, allo stesso modo

la rete neurale, creata tramite appositi comandi di programmazione, ha bisogno di un periodo di apprendimento. Essa deve infatti comprendere come rispondere in base agli input che riceve. Una volta generata la rete (topologia, numero e tipo di neuroni, collegamenti), può iniziare il periodo di apprendimento durante il quale vengono modificati i pesi dei legami tra i neuroni in un'ottica di minimizzazione di una funzione di errore. Come già segnalato, esistono tre categorie principali di apprendimento: supervisionato, non supervisionato e con rinforzo. In questa sezione si analizza il funzionamento dell'*Error Back-Propagation* - l'algoritmo più utilizzato per l'apprendimento supervisionato -in relazione a una rete di tipo MLP.

Back-Propagation in formule

L'algoritmo di *Back-Propagation* prevede una fase *forward* in cui si presenta un esempio alla rete, si determina l'output e se ne calcola l'errore e una fase all'indietro in cui l'errore viene propagato a ritroso modificando i pesi in modo progressivo.

Le formule seguenti [12] illustrano i passaggi e le derivate per arrivare alla modifica dei pesi in una rete MLP a tre strati, partendo da quelli w_{jk} dell'*hidden-output* per arrivare poi a quelli w_{ij} dell'*input-hidden*.

Hidden-output

Si comincia con il derivare la funzione d'errore per i pesi w_{jk} :

$$\begin{aligned}
 \frac{\partial J}{\partial w_{jk}} &= \frac{\partial}{\partial w_{jk}} \left(\frac{1}{2} \sum_{c=1, \dots, s} (t_c - z_c)^2 \right) = (t_k - z_k) \frac{\partial (-z_k)}{\partial w_{jk}} = \\
 &= (t_k - z_k) \frac{\partial (-f(\text{net}_k))}{\partial w_{jk}} = -(t_k - z_k) \frac{f(\text{net}_k)}{\partial \text{net}_k} \frac{\partial \text{net}_k}{\partial w_{jk}} = \\
 &= -(t_k - z_k) f'(\text{net}_k) \frac{\partial \sum_{s=1, \dots, n_H} w_{sk} y_s}{\partial w_{jk}} = -(t_k - z_k) f'(\text{net}_k) y_j \quad (2.9)
 \end{aligned}$$

Ponendo:

$$\delta_k = (t_k - z_k)f'(net_k)$$

Allora:

$$\frac{\partial J}{\partial w_{jk}} = -\delta_k y_j \quad (2.10)$$

Così il peso w_{jk} può essere modificato come: $w_{jk} = w_{jk} + \eta \delta_k y_j$, con η che rappresenta il tasso di apprendimento.

Input-hidden

$$\begin{aligned} \frac{\partial J}{\partial w_{ij}} &= \frac{\partial}{\partial w_{ij}} \left(\frac{1}{2} \sum_{c=1, \dots, S} (t_c - z_c)^2 \right) = - \sum_{c=1, \dots, S} (t_c - z_c) \frac{\partial z_c}{\partial w_{ij}} = \\ &= - \sum_{c=1, \dots, S} (t_c - z_c) \frac{\partial z_c}{\partial net_c} \frac{\partial net_c}{\partial w_{ij}} = - \sum_{c=1, \dots, S} (t_c - z_c) f'(net_c) \frac{\partial net_c}{\partial w_{ij}} \end{aligned} \quad (2.11)$$

Ponendo:

$$\begin{aligned} \delta_c &= (t_c - z_c) f'(net_c) \\ \frac{\partial net_c}{\partial w_{ij}} &= \frac{\partial}{\partial w_{ij}} \sum_{r=1, \dots, n_H} w_{rc} y_r = \frac{\partial}{\partial w_{ij}} \sum_{r=1, \dots, n_H} w_{rc} f(net_r) = \frac{\partial}{\partial w_{ij}} (w_{jc} f(net_j)) = \\ &= w_{jc} \frac{\partial f(net_j)}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} = w_{jc} f'(net_j) \frac{\partial}{\partial w_{ij}} \sum_{q=1, \dots, d} w_{jq} x_q = \\ &= w_{jc} f'(net_j) x_i \end{aligned} \quad (2.12)$$

Così:

$$\frac{\partial J}{\partial w_{ij}} = - \sum_{c=1, \dots, S} \delta_c w_{jc} f'(net_j) x_i = -x_i f'(net_j) \sum_{c=1, \dots, S} \delta_c w_{jc} \quad (2.13)$$

Ponendo:

$$\delta_j = f'(net_j) \sum_{c=1, \dots, S} \delta_c w_{jc}$$

Allora:

$$\frac{\partial J}{\partial w_{ij}} = -\delta_j x_i \quad (2.14)$$

Così il peso w_{ij} può essere modificato come: $w_{ij} = w_{ij} + \eta \delta_j x_i$

Back-Propagation step

L'algoritmo esegue un'epoca di apprendimento, ovvero svolge i seguenti passi per tutte le coppie input-output significative del set di addestramento:

- 1- Considera una coppia e calcola la risposta della rete per quel determinato input, previa inizializzazione dei pesi. Il calcolo avanza dal livello di input a quello di output, conteggiando l'attivazione di tutte le unità e propagando in avanti l'errore.
- 2- Calcola l'errore tra output della rete e output desiderato della coppia e calcola i δ_k delle unità di output.
- 3- Propaga l'errore a ritroso verso il livello di input e calcola i δ_i per ogni unità nascosta.
- 4- Esegue la variazione dei pesi.
- 5- Ripete gli step dal passo 1 finché non terminano le coppie.
- 6- Calcola l'errore complessivo e, se questo è ancora alto, ripete l'epoca di apprendimento.

Una importante questione riguarda l'istante in cui occorre fermare l'addestramento della rete. Minimizzare l'errore sul *training set*, infatti, non comporta necessariamente una miglior capacità di generalizzazione. Occorre quindi stoppare l'aggiornamento dei pesi in tempo utile affinché il sistema non impari a memoria il set di addestramento e bisogna utilizzare un set di dati di validazione per analizzare l'errore globale della rete.

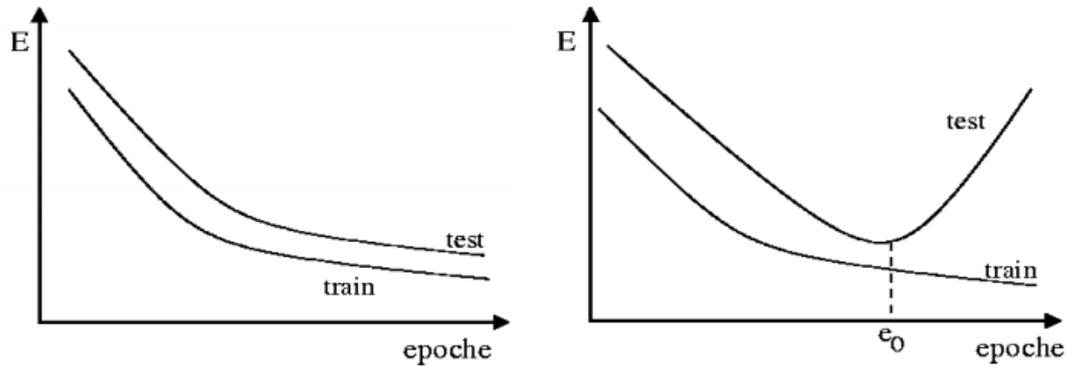


Figura 2.12 - Andamento errore [28]

Nel grafico a sinistra della figura 2.12 si assiste a una buona generalizzazione della rete, con l'errore che cala e assume valori bassi. Nel grafico di destra, invece, si nota come, dopo un numero di epoche pari ad e_0 , l'errore inizia ad aumentare generando *over-training*: la rete perde la sua abilità di generalizzazione, pertanto è necessario bloccare l'addestramento.

Limiti Back-Propagation

L'algoritmo appena mostrato presenta, tuttavia, dei limiti di cui è bene essere consapevoli:

- Lentezza dell'algoritmo, il quale può bloccarsi in un minimo di errore, credendo di aver trovato il valore ottimo dei pesi che rende l'errore minimo.
- La scelta del tasso di apprendimento η rappresenta una criticità poiché un valore troppo basso significa lentezza di apprendimento, mentre un valore troppo grande crea oscillazioni.
- L'algoritmo dipende dal valore iniziale dei pesi e potrebbe significare iniziare le iterazioni da un punto molto vicino a un minimo locale.

2.6 Funzione di errore

La funzione di errore esprime lo scostamento fra l'output prodotto dalla rete neurale e l'output desiderato. È evidentemente dipendente dai pesi delle connessioni, poiché sono proprio essi a determinare l'output.

Definito come $z = [z_1, z_2, \dots, z_s]$ l'output della rete per il pattern $x = [x_1, x_2, \dots, x_d]$ di input, l'output desiderato viene formulato come $t = [t_1, t_2, \dots, t_s]$. Si tratta quindi di una funzione derivabile nel peso in modo tale da trovare il peso tale che l'errore sia minimo. In questo caso, scegliendo come funzione di errore la somma dei quadrati degli errori (MSE), l'errore del pattern x diventa:

$$J(w, x) = \frac{1}{2} \sum_{c=1, \dots, s} (t_c - z_c)^2 \quad (2.15)$$

Tale errore può essere minimizzato attraverso la modifica dei pesi nella direzione opposta al gradiente di J . Infatti, il gradiente mostra la direzione di crescita maggiore per la funzione e, andando nella direzione opposta, si riduce di conseguenza al minimo l'errore. Tale meccanismo è denominato *gradient descent*.

Ciò che emerge è che a ogni iterazione si modificano i pesi sulla base dell'errore. Tale aggiornamento può seguire un approccio di tipo *batch* se i pesi vengono rivisti dopo la presentazione di tutti i campioni del *training set* oppure *on-line* se i pesi subiscono variazioni dopo la presentazione di ogni singolo campione del *training set*.

Un'ulteriore funzione di errore molto utilizzata è la *cross-entropy*. Essa è data dalla formula:

$$H(P, Q) = - \sum_x P(x) * \log(Q(x)) \quad (2.16)$$

Dove $P(x)$ è la distribuzione di probabilità target dell'evento x , mentre $Q(x)$ è la distribuzione dell'output della rete.

2.7 Applicazioni, pregi e limiti

Applicazioni

I modelli statistici sono da sempre utilizzati e largamente diffusi, ma, se prima degli anni Ottanta essi costituivano essenzialmente gli unici metodi, dalla seconda metà dello stesso decennio, i modelli fondati su sistemi di intelligenza artificiale hanno cominciato a diffondersi nell'ambito del rischio di credito, per avere un vero e proprio boom di utilizzo negli anni Novanta [30]. Le reti neurali, inoltre, sono molto versatili e trovano applicazione in moltissimi settori:

- Finanza = previsioni circa l'andamento dei mercati, analisi del rischio di credito, analisi del portafoglio;
- Riconoscimento di immagini;
- Riconoscimento vocale;
- Simulazione di sistemi biologici;
- Diagnosi mediche;
- *Robot steering* = miglioramento delle performance dei veicoli in termini di sterzo;
- Controllo qualità;
- *Data mining*;
- Simulazioni.

Pregi

Tra i vantaggi derivanti dall'utilizzo delle reti neurali si trova la parallelizzazione delle attività di elaborazione, che permette di lavorare su molti dati in tempi piuttosto rapidi. Questo tipo di sistemi è anche tollerante verso i guasti grazie alla decentralizzazione dei centri di elaborazione. Consente inoltre di operare correttamente anche in assenza di input completi e si adatta perfettamente a simulazioni, ottimizzazioni e operazioni di *data mining*. Garantisce, infatti, una elevata accuratezza, riuscendo ad approssimare moltissime informazioni. Dal

punto di vista matematico non hanno necessità di conoscere a priori la distribuzione. Le reti neurali rappresentano anche un ottimo strumento in ragione dei cambiamenti delle circostanze ambientali, vista la possibilità di aggiornarsi automaticamente.

Limiti

Uno dei difetti più importanti è il funzionamento a scatola chiusa della parte computazionale, che non permette di analizzare i singoli stadi di elaborazione. Non è neppure possibile sapere a priori se un problema sarà risolto. Inoltre, gli output spesso non sono la soluzione perfetta per il problema, ma questo aspetto passa molte volte in secondo piano poiché non si sta cercando quel tipo di risposta. Se da un lato l'elaborazione è relativamente rapida, dall'altro si può incorrere in periodi di apprendimento piuttosto lunghi; inoltre, il periodo di apprendimento è molto importante per evitare che la rete incorra in errori di generalizzazione. Infine, le reti neurali, prediligendo informazioni quantitative, non sono adatte a risolvere problemi il cui input è rappresentato da molte variabili categoriche.

Capitolo III

3 Macchine a vettori di supporto

Con il termine *Support Vector Machines* (SVM) si indentificano i modelli di apprendimento supervisionato utili ai fini della regressione e della classificazione. Le SVM, a differenza delle reti neurali, ben si adattano a tecniche di apprendimento di classificatori non lineari. Qualora la tipologia di dati richiedesse un approccio non lineare, il metodo prevede l'utilizzo di funzioni di tipo Kernel per determinare un classificatore non lineare. Storicamente, le SVM hanno seguito un processo di sviluppo opposto rispetto alle reti neurali: mentre queste ultime hanno visto la teoria succedere alle sperimentazioni, le SVM derivano da un iniziale approccio teorico, cui sono seguiti esperimenti e implementazioni.

L'origine di questo algoritmo risale agli anni Sessanta del secolo scorso ad opera di Vladimir Vapnik e Aleksej Červonenkis, ma è stato rinnovato nel tempo dallo stesso Vapnik in collaborazione con altri studiosi. Infatti, il lavoro non riscosse notevole successo appena pubblicato, ma destò particolare interesse a partire dagli anni Novanta.

Essendo un tipo di algoritmo supervisionato, esso necessita di un set di dati di addestramento con appartenenza a una determinata classe, sulla base dei quali sarà calcolato il modello. Partendo da quanto elaborato, lo strumento sarà in grado di associare il nuovo esempio in ingresso a una delle classi.

Si tratta di modelli cosiddetti *non parametrici*, in quanto, a differenza della classica inferenza statistica, i parametri non sono predefiniti e il loro numero dipende dai dati di addestramento utilizzati.

Secondo quanto illustrato da Lipo Wang in *Support Vector Machines: Theory and Applications* [31], le reti neurali e le SVM seguono una strategia differente per la costruzione di un modello con buone capacità di generalizzazione. Infatti, le prime scelgono una struttura appropriata per il modello e, mantenendo fisso l'intervallo di confidenza, minimizzano l'errore di addestramento; le seconde

mantengono fisso il valore dell'errore di addestramento e minimizzano l'intervallo di confidenza.

3.1 Concetti chiave

Prima di entrare nel merito del funzionamento delle SVM, è opportuno considerare alcuni concetti chiave, utili a comprendere meglio le logiche di fondo dell'algoritmo. Si tratta di capire cosa si intende per iperpiano, vettori di supporto e margine. Nel seguito, per semplicità, si fa generalmente riferimento all'identificazione di due classi per dividere il *dataset*.

Iperpiano

Detto anche limite di decisione lineare, è una linea che separa – e quindi classifica - un insieme di dati. Se si aumenta la raffinatezza dell'analisi e si considerano tre dimensioni, l'iperpiano è rappresentato da un piano, mentre aggiungendo ulteriori dimensioni si generalizza con il termine iperpiano. L'obiettivo delle SVM è proprio quello di definire l'iperpiano ottimo.

Nel caso di due dimensioni, è facile intuire come i dati possano trovarsi da una parte o dall'altra della linea: in questo modo si determina in modo visuale l'appartenenza a una classe.

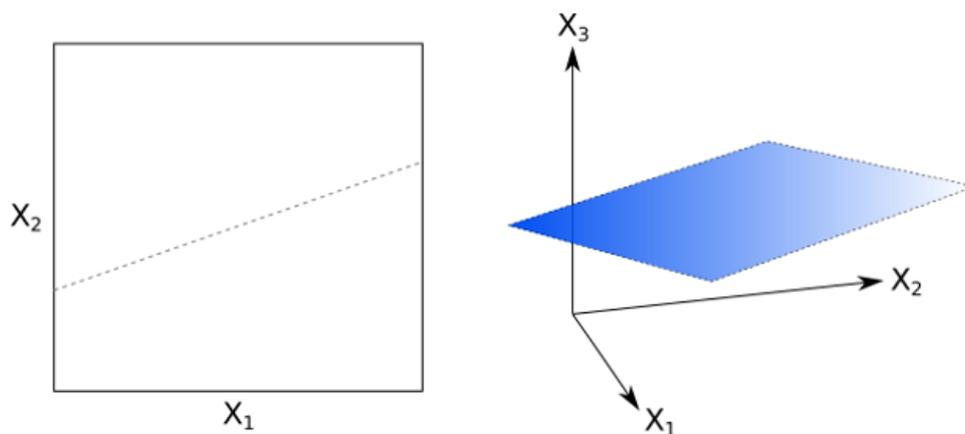


Figura 3.1 - Iperpiano in due e tre dimensioni [32]

Vettori di supporto

I *support vectors* danno il nome all'algoritmo e sono i punti più vicini all'iperpiano. Rappresentano il nodo critico poiché in base ad essi l'iperpiano può variare. È importante specificare che questi punti dipendono dal dataset.

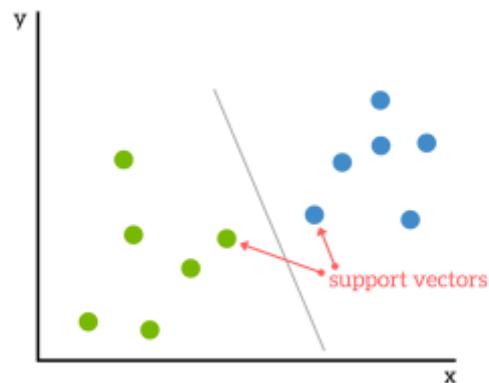


Figura 3.2 - Vettori di supporto [32]

Margine

Il margine identifica la distanza tra i vettori di supporto appartenenti a due classi differenti. L'iperpiano viene tracciato in corrispondenza della metà di tale distanza.

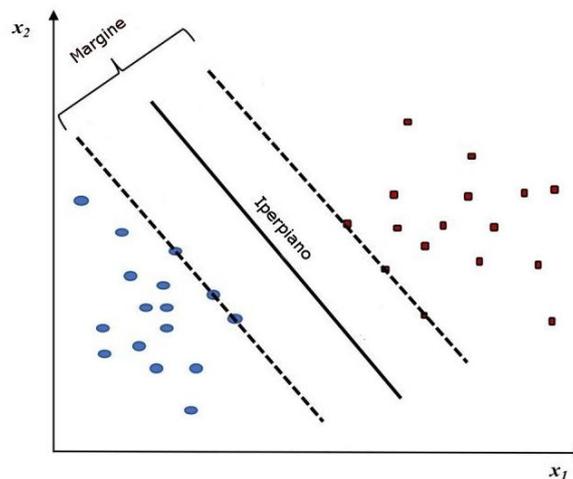


Figura 3.3 - Margine [32]

3.2 Funzionamento

L'utilizzo dell'algoritmo SVM ha lo scopo di definire l'iperpiano migliore per dividere i vettori di supporto in classi. Possono infatti esistere molti iperpiani per dividere due classi, ma questa metodologia non si focalizza sulla ricerca di una generica soluzione, bensì su quella ottima. Per raggiungere questo obiettivo, prima cerca un iperpiano linearmente separabile e, nel caso in cui ne esistesse più di uno, seleziona quello con il margine più ampio per accrescere l'accuratezza del modello (massimizzazione del margine).

È infatti teoricamente possibile con dati in due dimensioni tracciare un numero infinito di rette per separare le due categorie. Un metodo con queste caratteristiche sarebbe pressoché inutilizzabile, ecco perché le SVM si concentrano sulla soluzione ottima, cioè quella che genera il minimo errore nella classificazione di una nuova soluzione. A questo proposito, l'idea di fondo dell'algoritmo è la seguente: più i punti (dati) si trovano lontano dall'iperpiano, più è probabile che siano classificati correttamente. La ricerca è incentrata quindi sui vettori di supporto più lontani dall'iperpiano. La differenza chiave rispetto alle reti neurali sta proprio in questa procedura di ricerca della soluzione poiché non vengono considerati tutti i punti, ma soltanto quelli ritenuti cruciali per la massimizzazione di tale distanza. È però evidente come i vettori di supporto siano una criticità nell'identificazione dell'iperpiano perché una loro rimozione dal training set comporterebbe una modifica dell'iperpiano.

Nella figura 3.4 si osservano tre possibili iperpiani (nero, rosso e verde) per un set di dati di addestramento in due dimensioni. La soluzione ottima sarà la linea nera, in quanto massimizza la distanza tra i vettori di supporto seguendo le prescrizioni dell'algoritmo. Dalla figura è altresì evidente come una rimozione dei *support vectors* comporterebbe una modifica dell'iperpiano.

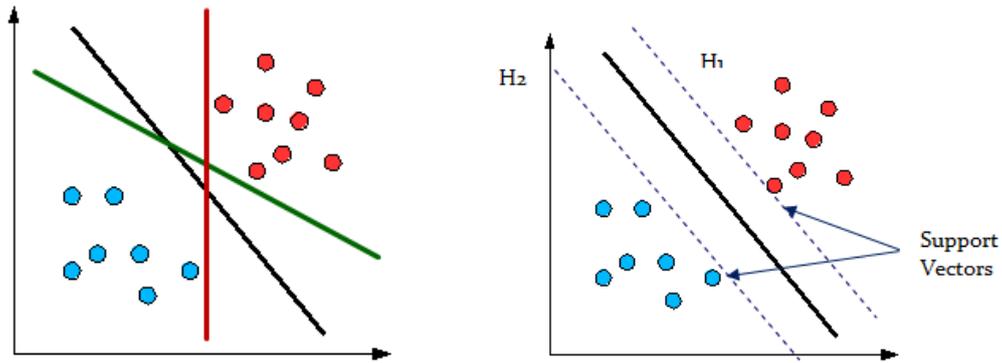


Figura 3.4 - Iperpiano [32]

Se invece non esistesse un iperpiano con queste caratteristiche, allora le SVM impiegherebbero una mappatura non lineare per accrescere le dimensioni dei dati di allenamento. In questo caso, se il *training set* avesse due dimensioni, tale mappatura porterebbe i dati a tre dimensioni, così da rendere sempre possibile la separazione attraverso un iperpiano.

3.3 Modello lineare

Definendo con \vec{w} il vettore dei pesi, con \vec{x} il vettore degli input e con b_0 il bias, si modella l'iperpiano secondo la relazione:

$$\vec{w}\vec{x} + b_0 = w_1x_1 + w_2x_2 + \dots + w_nx_n + b_0 = \sum_{i=1}^n w_ix_i + b_0 = 0 \quad (3.1)$$

Dalla precedente formula si osserva che un iperpiano non è altro che una combinazione lineare delle n dimensioni uguagliata a zero. Pertanto, lavorando in due dimensioni, la relazione diventa:

$$\vec{w}\vec{x} + b_0 = w_1x_1 + w_2x_2 + b_0 = 0 \quad (3.2)$$

Conseguentemente, essendo in questo caso una retta che separa le due categorie, troveremo che i punti appartenenti alla prima e alla seconda classe rispetteranno rispettivamente le condizioni:

$$w_1x_1 + w_2x_2 + b_0 > 0 \quad (3.3)$$

$$w_1x_1 + w_2x_2 + b_0 < 0 \quad (3.4)$$

In sostanza, la prima formula rappresenta i punti che stanno sopra all'iperpiano, mentre la seconda i punti che stanno sotto.

È così possibile definire due classi $y_j = \begin{cases} +1 & \text{se } w_1x_1 + w_2x_2 + b_0 > 0 \\ -1 & \text{se } w_1x_1 + w_2x_2 + b_0 < 0 \end{cases}$

Considerando i confini del margine come i valori +1 e -1 si arriva alle relazioni:

$$w_1x_1 + w_2x_2 + b_0 \geq 1, \text{ per } y = +1 \quad (3.5)$$

$$w_1x_1 + w_2x_2 + b_0 \leq -1, \text{ per } y = -1 \quad (3.6)$$

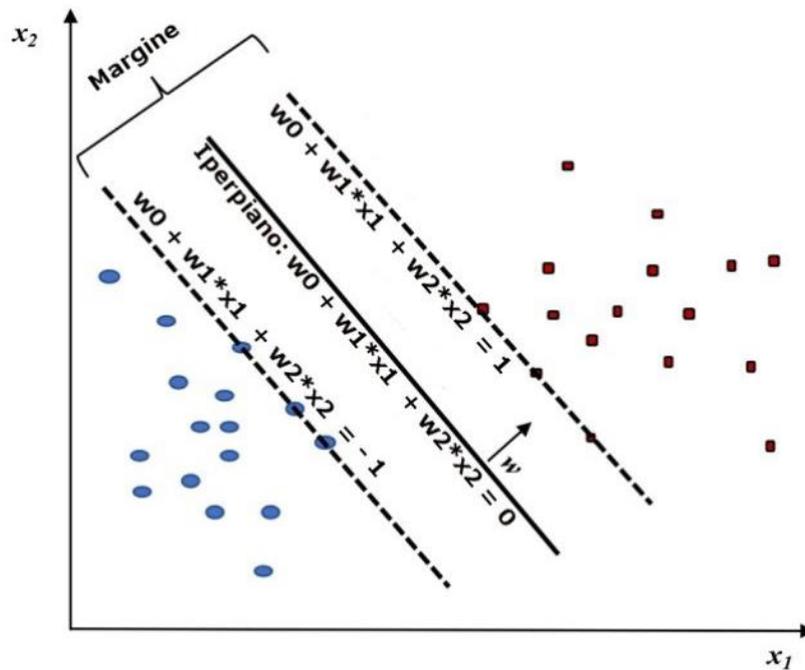


Figura 3.5 - Iperpiano e vettori di supporto in due dimensioni [32]

I dati del *training set* per cui vale la relazione $w_1x_1 + w_2x_2 + b_0 = \pm 1$ sono i vettori di supporto, i quali cadono proprio sui confini del margine. L'iperpiano ottimo è quindi quello per cui il margine di separazione è massimo.

Poiché la massima distanza tra i limiti del margine è pari a $\frac{2}{\|w\|}$, occorre minimizzare la norma del vettore dei pesi per trovare l'iperpiano ottimo.

3.4 Modello non lineare – il Kernel

Come è lecito aspettarsi dalla realtà, non tutti i dataset sono lineari. L'algoritmo SVM nelle sue successive varianti è arrivato a poter trattare anche questo tipo di informazioni, definendo dei limiti di decisione non lineari.

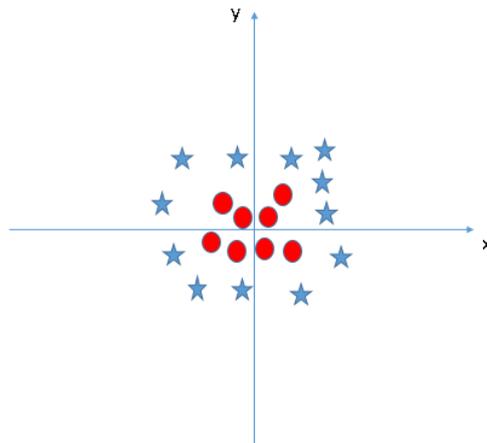


Figura 3.6 - Dataset non lineare in due dimensioni [32]

Guardando la figura 3.6, è evidente come non esista un classificatore lineare, ossia questi dati non possono essere suddivisi in classi tracciando semplicemente una retta, ma appare altresì chiara una distinzione tra le due categorie. In casi come questo si può ovviare al problema dimensionale semplicemente aggiungendo una dimensione: è semplice riconoscere nella distribuzione dei dati nello spazio una sorta di circonferenza, si aggiunge quindi la dimensione $z = x^2 + y^2$.

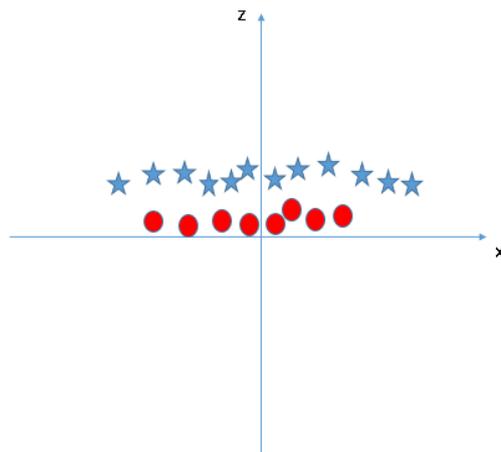


Figura 3.7 - Dataset non lineare in tre dimensioni [32]

Dall'aggiunta della terza dimensione si ricava il disegno in figura 3.7, con il

piano y che è perpendicolare ai piani x e z . In sostanza, è come se si sollevassero i punti a stella sfruttando la dimensione altezza (z) e lasciando inalterati i valori su x e y . Con questa rappresentazione si nota meglio la distinzione ed è a questo punto semplice identificare graficamente un iperpiano, che viene raffigurato nella rappresentazione originale in due dimensioni come una sorta di circonferenza [figura 3.8].

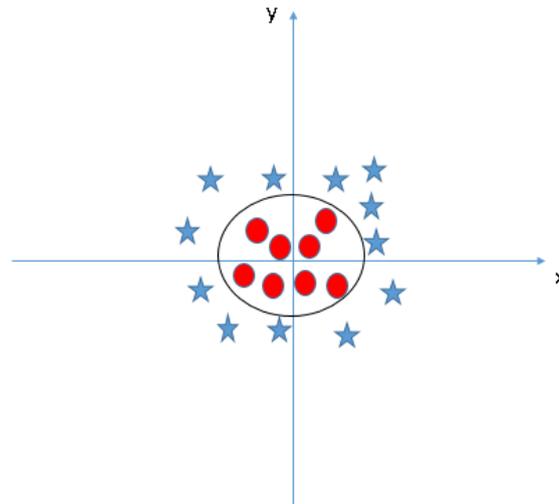


Figura 3.8 - Iperpiano post Kernel in due dimensioni [32]

Lo stratagemma adottato per riuscire a classificare i dati non lineari prende il nome di metodo Kernel. Attraverso questo trucco è possibile trasformare i dati non linearmente separabili in modo più utile al fine di indentificare un iperpiano linearmente separabile.

In linea generale, il metodo Kernel può essere descritto da:

$$K(x, y) = \langle f(x), f(y) \rangle \quad (3.7)$$

Dove K è la funzione del Kernel - nell'esempio precedente era la formula della circonferenza -, x e y sono i vettori dei dati di input con dimensione n . La funzione è usata per aumentare le dimensioni da n a un numero superiore. Si tratta essenzialmente di eseguire un prodotto scalare per la nuova dimensione:

$$x * y = x_1 * x_2 + y_1 * y_2 + z_1 * z_2 =$$

$$= x_1 * x_2 + y_1 * y_2 + (x_1^2 + y_1^2) * (x_2^2 + y_2^2) \quad (3.8)$$

È evidente come la scelta della funzione del Kernel rappresenti una criticità poiché è in grado di influenzare in modo determinante la definizione dell'iperpiano. Poiché non esiste un modo per scegliere tale funzione indipendentemente dal problema, occorre provare diverse funzioni e confrontarne i risultati prima di effettuare la scelta.

Tra i diversi tipi di Kernel troviamo il lineare, il polinomiale e il *Radial Basis Function* (RBF).

Kernel lineare

$$K(x_i, y_j) = x_i * y_j \quad (3.9)$$

Si tratta di una tipologia di funzione basilare, che è molto diffuso per la classificazione del testo.

Kernel polinomiale

$$K(x_i, y_j) = (x_i * y_j + c)^d \quad (3.10)$$

In questo tipo di Kernel si trovano la costante c e il grado di libertà d . È evidente che se $d = 1$ il Kernel polinomiale si riconduce al lineare, mentre per valori diversi da uno aumenta la complessità dell'iperpiano con conseguenti possibili problemi di *overfitting*.

Kernel RBF

$$K(x_i, y_j) = e^{(-\gamma |x_i - y_j|^2)} \quad (3.11)$$

In questo caso il limite decisionale assume una forma complessa. In modo simile al Kernel polinomiale, si trova il parametro γ : per valori bassi il modello si comporta come un Kernel lineare, mentre per valori alti sarà maggiore l'influenza sul modello dei vettori di supporto, con conseguente pericolo di *overfitting*.

3.5 Applicazioni, pregi e limiti

Applicazioni

L'algoritmo SVM è molto utilizzato nei seguenti ambiti:

- Rilevamento dei volti;
- Classificazione del testo;
- Classificazione delle immagini;
- Bioinformatica;
- Riconoscimento della scrittura a mano.

Pregi

Tra i punti a favore di questo metodo si trova sicuramente la sua efficacia in dimensioni spaziali numerose poiché si presta bene ad analisi con dimensioni dell'ordine di 10^n . La memoria è efficiente perché bisogna considerare soltanto un sottoinsieme dei punti del *training set* per assegnare la classe ai nuovi dati. Inoltre, un vantaggio derivante dall'impiego del Kernel è che questo strumento è molto versatile, potendo cambiare la funzione a seconda delle esigenze.

Limiti

Uno dei principali problemi dell'SVM è la difficoltà di interpretazione, mancando infatti di trasparenza nei risultati. Tuttavia, è possibile arginare questa situazione mediante il ricorso alla visualizzazione grafica. Si tratta inoltre di un metodo non probabilistico: la classificazione avviene posizionando gli oggetti

sopra o sotto l'iperpiano, senza che ci sia un'interpretazione in termini di probabilità. Anche in questo caso si può ovviare con un'interpretazione di efficacia attraverso la distanza del punto da classificare dal margine, in quanto più esso è lontano dal margine più è probabile che sia classificato correttamente.

Capitolo IV

4 Il settore degli elettrodomestici

In riferimento al rischio di credito, la valutazione avviene relativamente al settore degli elettrodomestici in Italia, considerando tutti gli attori che prendono parte alla filiera. L'analisi prevede una prima parte dedicata alla storia del settore, dal boom economico ai giorni nostri, passando poi alla descrizione delle specificità e dei trend sia a livello europeo che nazionale. La parte conclusiva è dedicata alla quantificazione delle informazioni di settore, con un breve cenno circa l'impatto della pandemia del 2020 sui risultati.

4.1 Cenni storici

Gli elettrodomestici sono ritenuti oggi indispensabili, ma la loro diffusione è avvenuta in tempi relativamente recenti. Tuttavia, sin dagli albori della civiltà, l'uomo si è servito di strumenti per la semplificazione di attività quotidiane, così da risparmiare tempo ed energie. Si trattava di dispositivi rudimentali e soprattutto non elettrificati.

Il punto chiave per l'invenzione e la diffusione degli elettrodomestici è appunto l'elettricità. L'arrivo della corrente elettrica nelle abitazioni e il suo utilizzo per sostituire l'energia dell'uomo hanno preparato il terreno agli elettrodomestici come oggi li conosciamo. Infatti, alcuni prodotti come le lavatrici esistevano già in versioni non elettrificate.



Figura 4.1 - Donna alle prese con una lavatrice non elettrificata [35]

Prendendo come esempio la lavatrice, dalla sua invenzione al suo consumo di massa sono passati diversi anni: il primo prototipo elettrico risale ai primi del Novecento ad opera di Alva J. Fisher, ma bisogna attendere il secondo Dopoguerra affinché lo strumento diventi un fenomeno diffuso su ampia scala.

È proprio a partire dagli anni Cinquanta del secolo scorso, nel periodo del cosiddetto boom economico, che gli elettrodomestici iniziano ad entrare nelle case italiane: frigorifero, aspirapolvere e la lavatrice stessa hanno rivoluzionato lo stile di vita familiare, semplificando lo svolgimento delle operazioni giornaliere.

Il fenomeno di crescita ha portato nel giro di pochi decenni a una saturazione del mercato. Se si guarda all'indice di saturazione relativo alle lavatrici in Italia - calcolato come il rapporto tra il numero di elettrodomestici e il numero di famiglie - si vede una netta crescita, poiché è passato da circa il 5% degli anni Sessanta [37] a pressoché il 100% di oggi [38]. La saturazione si traduce con un mercato prevalentemente di sostituzione e non di primo acquisto. In queste condizioni e con una competizione essenzialmente basata sul prezzo, molte imprese hanno dovuto fronteggiare periodi di crisi o fallimenti.

Per ovviare a questo problema, si è assistito a un fenomeno di delocalizzazione della produzione al fine di ridurre i costi e di avvicinarsi ai nuovi mercati in via di sviluppo, essendo gli elettrodomestici particolarmente appetibili per le fasce di reddito medio-basse.

Il trend del settore punta verso un'evoluzione delle apparecchiature mediante l'integrazione di servizi innovativi in ottica IoT. Infatti, tra i prodotti che in Italia trainano il settore ci sono gli elettrodomestici definiti *smart*: nel 2018 per quanto riguarda i piccoli elettrodomestici, la quota di mercato è cresciuta del 75% arrivando a valere 1,2 miliardi di euro [40].

4.2 Il settore in cifre

In questa sezione si passano in rassegna le specificità del contesto, cominciando con l'Europa. Una volta definiti i trend e le peculiarità, si analizzerà il caso italiano mediante l'esposizione di alcuni dati storici fino al 2020 e all'impatto che ha avuto il Coronavirus durante il primo semestre.

4.2.1 Il contesto europeo

A livello continentale, il comparto degli elettrodomestici è molto importante: esso contribuisce per circa 72 miliardi al Pil della regione, dando lavoro a quasi 1 milione di persone [41]. I due punti focali del settore per il presente e il futuro perseguono obiettivi in termini di sostenibilità degli stili di vita e di connettività delle apparecchiature.

La sostenibilità è un fenomeno di fondamentale importanza all'interno dell'Unione Europea, viste le crescenti pressioni politiche e sociali che puntano alla decarbonizzazione dell'area per completare la transizione verso le fonti rinnovabili. La filiera può incidere in questo senso considerando non soltanto l'aspetto relativo ai consumi di energia, ma l'intero ciclo di vita del prodotto, inclusi il suo smaltimento e la sua produzione.

Il consumo di energia è un aspetto estremamente rilevante per i consumatori, i quali tendono a prediligere dispositivi meno energivori. Alla riprova di ciò, si osserva come nel 2018 il 90% circa delle lavatrici e delle lavastoviglie vendute appartenesse alla classe energetica A+++, sulla scia di un trend di crescita cominciato già prima del 2010.

Sullo smaltimento è opportuno considerare i dati: per 6,3 milioni di tonnellate di materiale finito sul mercato, 3,7 tonnellate sono state recuperate. Questa evidenza esplicita l'impegno da parte dei Paesi e di riflesso di tutti gli attori coinvolti nella filiera affinché il settore diventi più sostenibile.

La figura 4.2 mostra gli impatti negli anni dal 2011 al 2018 dei miglioramenti apportati per quanto concerne l'aspetto produttivo dei beni in oggetto: emerge un netto calo nel consumo di acqua, insieme a un significativo decremento degli scarti e dell'utilizzo di energia.



Figura 4.2 - Miglioramenti nella produzione di elettrodomestici [41]

Il processo di digitalizzazione porterà 31 milioni di europei a utilizzare elettrodomestici smart entro il 2024 [41]. Da questa previsione si evince l'importanza per questo settore di internet e dei servizi digitali. Infatti, il numero delle case *smart* sta crescendo ed è ipotizzato che passeranno dai 18 milioni del 2017 ai 72 milioni nel 2024, vista la comodità di affidarsi ad apparecchi in grado di semplificare ulteriormente la quotidianità.

Dal grafico in figura 4.3 si assiste a una proiezione che illustra un trend di crescita pressoché esponenziale per quanto riguarda il numero di dispositivi connessi, fino al raggiungimento dei 75 miliardi nel 2025. Un'indagine del 2016 mostra come i consumatori prediligano l'acquisto di dispositivi connessi, quali termostati e impianti di illuminazione.

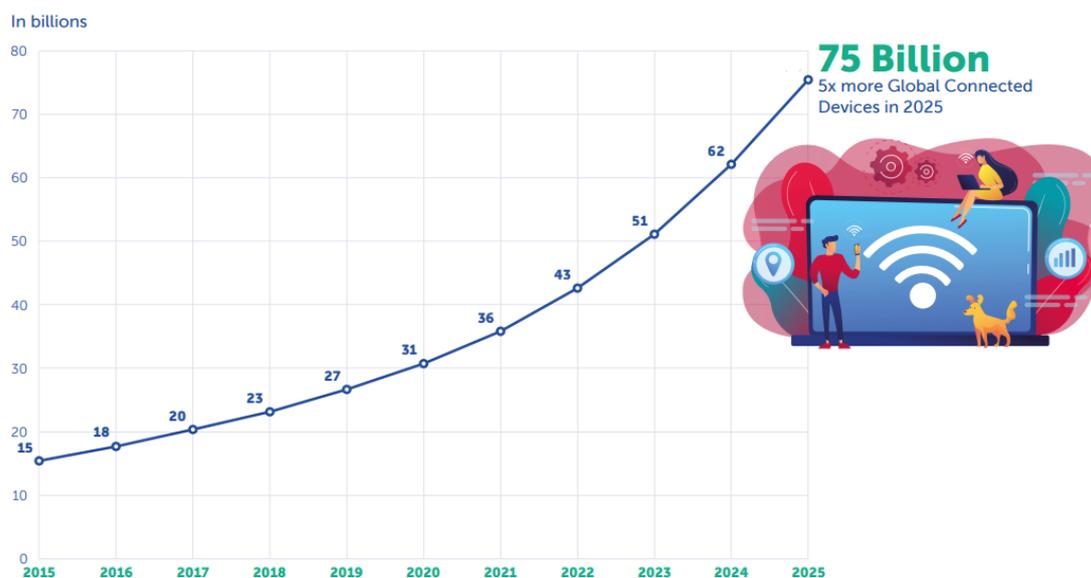


Figura 4.3 - Installazione di apparecchi intelligenti [41]

Dal punto di vista economico, è inevitabile un aumento dei ricavi legati agli apparecchi *smart*, con un tasso di crescita medio annuale del 20% fino al 2024. L'Italia sembrerebbe essere un mercato particolarmente promettente visto che già oggi contribuisce per il 6,5% alle entrate complessive europee di questa specifica tipologia di prodotti.

4.2.2 Andamento storico in Italia

L'andamento della spesa media mensile delle famiglie italiane considera le tre categorie: grandi apparecchi, piccoli elettrodomestici e riparazione. Il grafico in figura 4.4 mostra l'andamento complessivo delle voci di spesa in termini di spesa media mensile per ciascuna famiglia nel periodo compreso tra il 2014 e il 2019. La voce più significativa è quella per i grandi elettrodomestici, seguita dai piccoli elettrodomestici; è esigua la quota che interessa la riparazione.

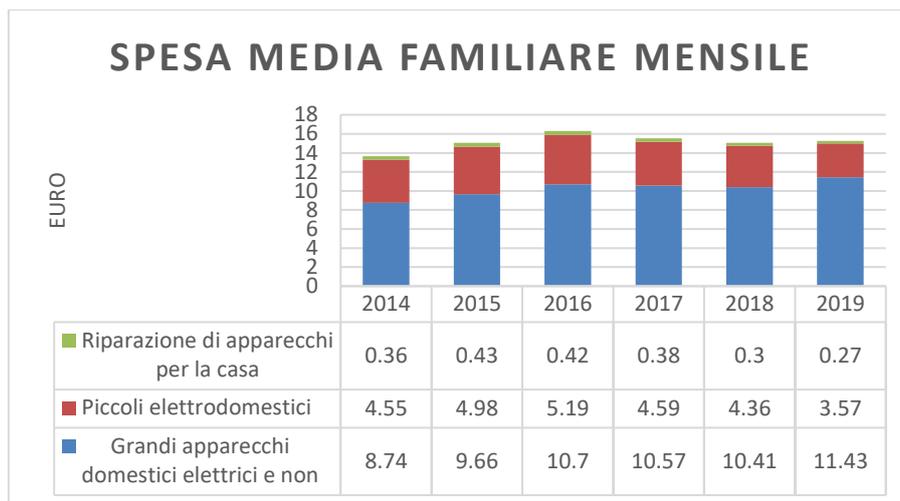


Figura 4.4 - Elaborazione andamento spesa media familiare mensile [42]

Il focus sulla riparazione [figura 4.5], infatti, mostra come dal 2015 le famiglie spendano sempre meno, sia in termini assoluti che percentuali sul totale delle uscite per elettrodomestici: le spese di riparazione non superano mai il 3% del totale e mostrano un trend di sostanziale decrescita.

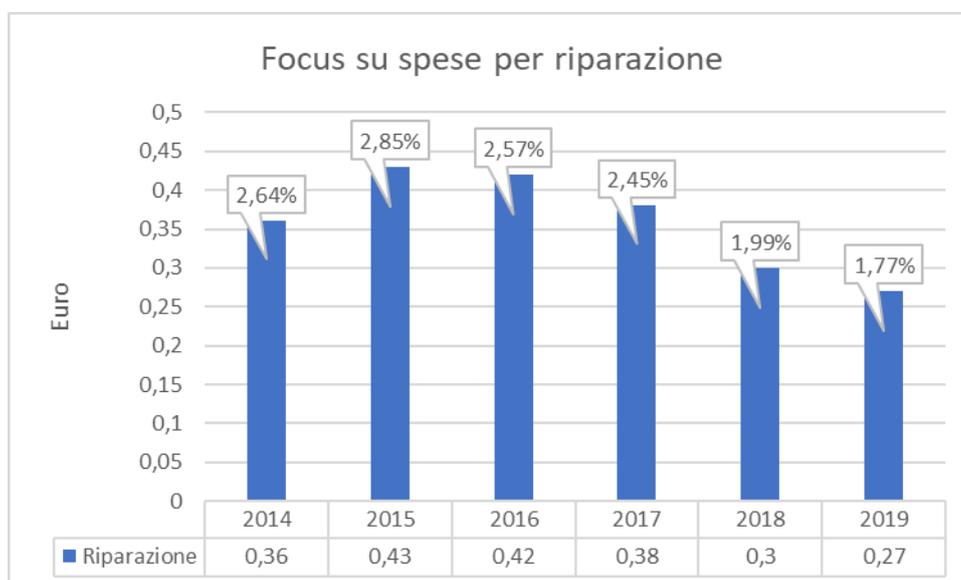


Figura 4.5 - Elaborazione focus su spese di riparazione [42]

4.2.3 La situazione attuale e gli impatti del Coronavirus

Secondo i dati di Statista [43], il settore degli elettrodomestici in Italia vale circa 4,8 miliardi di euro in termini di ricavi. Gli analisti riconoscono, in accordo con lo

scenario europeo, un cambiamento del contesto, che è spinto sempre più dalle novità tecnologiche verso la realizzazione di apparecchiature innovative e attente ai consumi energetici. Vi è, inoltre, una dualità all'interno del comparto, con gli elettrodomestici più grandi che vengono cambiati meno spesso, mentre i più piccoli hanno una sostituzione più frequente che contribuisce in modo importante ai volumi complessivi. Un ulteriore aspetto rilevante riguarda la vendita su internet: si stima un apporto dell'online al fatturato totale del settore intorno al 25%.

L'impatto del Coronavirus

La condizione di epidemia del 2020 ha costituito un importantissimo fattore di stress sia socio-sanitario che economico. La contrazione dei consumi a causa del *lockdown* è stata inevitabile, così come la riduzione della fiducia dei consumatori. Il settore degli elettrodomestici ha dimostrato di essere in grado di riprendersi mediamente meglio rispetto agli altri non alimentari. Ciò trova una certa coerenza con il contesto, in quanto è stata precedentemente citata la crescente importanza delle vendite online, elemento che ha subito un'ulteriore accelerata proprio in tempo di pandemia con un aumento nel secondo trimestre del 2020 di oltre il 40% delle vendite al dettaglio tramite commercio elettronico.

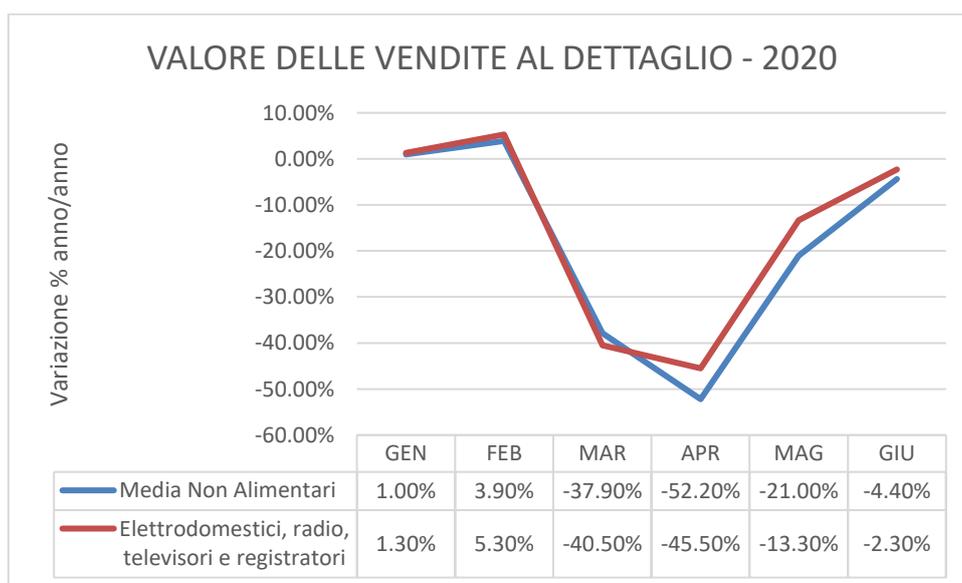


Figura 4.6 - Elaborazione per il valore vendite al dettaglio 2020 [44]

Capitolo V

5 Costruzione dei modelli

In questa sezione viene mostrato il processo di costruzione dei modelli, cominciando dalla presentazione degli strumenti utilizzati e dall'elaborazione dei dati. La trattazione proseguirà con l'illustrazione dei passaggi che hanno portato al miglior classificatore per entrambe le metodologie adottate, fino a spiegare e confrontare i risultati ottenuti.

5.1 Preparazione dei dati

La prima fase della costruzione dei modelli ha previsto la raccolta delle informazioni necessarie e la loro elaborazione. Nelle seguenti righe saranno illustrati i passaggi del processo di preparazione dei dati, partendo dalla presentazione dello strumento usato per la collezione.

5.1.1 Il database Aida

La banca dati utilizzata per l'estrazione del campione è Aida, realizzata e distribuita da Bureau Van Dijk. Essa raccoglie informazioni anagrafiche e finanziarie sulle società che operano in Italia, con uno storico massimo di dieci anni. Il metodo di costruzione del database si basa sulla raccolta di dati dai principali Information Provider del mondo, così da poter fornire agli utenti una conoscenza dettagliata ai fini della loro ricerca.

Aida è molto versatile e, tra le numerose funzionalità, consente di impostare i criteri di ricerca: particolarmente utile per il lavoro in oggetto è stata la possibilità di svolgere l'esplorazione delle società legando in modo booleano le condizioni da tenere in considerazione, come i codici Ateco e il valore dell'attivo positivo. È altresì possibile personalizzare quali informazioni delle società si ha intenzione di visualizzare e in quale ordine. Un'altra funzione utile ai fini dell'analisi è la possibilità di esportare i risultati della ricerca in un file Excel.

L'accesso a questa risorsa è reso disponibile gratuitamente dal Politecnico di Torino grazie a una convenzione.

5.1.2 Scarico dei dati

I dati scaricati riguardano le società operanti nella filiera degli elettrodomestici in Italia. In particolare, sono state considerate le società attive nella fabbricazione (codice Ateco 275100), nella riparazione (codice Ateco 952201), nel commercio all'ingrosso (codice Ateco 464310), nel commercio al dettaglio in esercizi specializzati (codice Ateco 475400) e nel commercio al dettaglio in esercizi non specializzati (codice Ateco 471920).

Il download è avvenuto considerando l'insieme delle società anno per anno, per poi aggregare i dati in un unico database e ordinarli alfabeticamente per ragione sociale e, per la stessa società, dal bilancio meno recente al più recente.

Sono stati poi calcolati indicatori di: sviluppo, redditività, produttività e liquidità.

Infine, sono stati assegnati dei flag alle società:

- 0 per le società senza particolari segnalazioni;
- 1 per le società anomale;
- 2 per le società in condizioni di liquidazione;
- 3 per le società in condizioni di fusione o incorporazione.

Sono state così inserite delle informazioni *ad hoc*, in particolare per tutti i bilanci di tutte le società che si trovano nelle condizioni dei flag 0, 1, 2 e 3. Inoltre, si è deciso di segnalare l'evento di tipo 1 soltanto nell'anno in cui si è verificato.

5.1.3 Pulizia e controllo dei dati

La pulizia dei dati è funzionale al loro utilizzo per l'elaborazione del modello. Per prima cosa, una volta scaricati e ordinati i dati, è stato necessario sostituire i dati non numerici – indicati con n.d. ed n.s.– con lo zero.

Il passo successivo ha previsto la cancellazione dei bilanci inesistenti, ossia quelli il cui attivo non è strettamente maggiore di zero, così da eliminare una possibile fonte di distorsione delle informazioni. In tal senso, nonostante il filtro sul download dei dati, un bilancio è risultato essere inesistente poiché ha aggirato il fatto di avere un attivo positivo grazie a crediti verso soci, i quali sono però dedotti dal patrimonio netto.

In seguito, si è verificata la correttezza dei dati esportati attraverso dei semplici controlli sui valori di bilancio. Tale verifica ha messo in luce alcune discrepanze che, caso per caso, sono state analizzate e corrette. Si è deciso, per semplicità, di trascurare gli errori di entità meno rilevante fino a circa 3000 euro.

Per i controlli sull'attivo dello stato patrimoniale il principale intervento ha riguardato i ratei e risconti, in quanto tali valori, talvolta omessi su Aida, non venivano inclusi nella somma dei parziali per costituire il totale dell'attivo.

Per quanto riguarda il passivo, si sono analizzati i singoli bilanci per comprendere le fonti di errore. Le più ricorrenti sono state: altre riserve, copertura parziale perdita di esercizio, utile di esercizio, utile a nuovo, oltre a errori nella somma dei parziali dei debiti a breve e a oltre. Anche in questo caso si sono modificati i dati originali in modo da correggere le inesattezze.

Nel conto economico le principali sorgenti di errore sono state le somme dei parziali, in particolare per ammortamenti e variazioni delle rimanenze di prodotto finito. È stata trovata, seppur con meno frequenza, qualche incongruenza circa i costi del lavoro, i costi per servizi e le imposte. Un altro intervento eseguito ha riguardato la correzione delle voci mancanti con l'aggiunta del delta in partite straordinarie.

Gli indicatori hanno richiesto attività correttive differenti a seconda della categoria di appartenenza. In tutti i casi, l'errore era dovuto al denominatore

pari a zero. Si è allora deciso di intervenire alterando in maniera minima i dati originali e ponendo uguale a 1 la sorgente di errore. Tale modifica ha avuto necessariamente un impatto sui precedenti controlli, ma tali valori sono rimasti entro la soglia di tolleranza.

Gli indicatori di sviluppo hanno riscontrato errori su ricavi, debiti totali, debiti finanziari e valore aggiunto operativo. Per gli indicatori di redditività c'è stato un intervento sui ricavi, mentre per quelli di produttività si è deciso di porre uguale a 1 i consumi. Gli indicatori di liquidità hanno richiesto un intervento sulle passività correnti e sui debiti finanziari.

Assegnato un valore a ogni indicatore di ogni società, è stato necessario svolgere un'analisi di tipo statistico sui dati al fine da individuare e allineare gli *outlier*. A tal proposito, per ogni indicatore sono stati calcolati i percentili a 5% e al 95%, dopodiché ogni singolo valore è stato comparato con queste due soglie e quando è risultato inferiore al primo limite è stato posto uguale ad esso, mentre quando è risultato superiore al secondo limite è stato posto pari a quest'ultimo.

A questo punto, i dati sono stati normalizzati, così da avere valori oscillanti tra lo 0 e l'1. Si è deciso di operare nel seguente modo per ogni singolo indicatore:

1. Calcolo del minimo
2. Calcolo del massimo
3. Normalizzazione del valore secondo la relazione $x_n = \frac{x_i - x_{min}}{x_{max} - x_{min}}$

Il passo successivo ha previsto l'inizializzazione del processo di selezione degli indicatori da utilizzare nel modello. In linea generale, al fine di evitare problemi di collinearità, è bene non includere nel modello due o più variabili altamente correlate e che spiegano lo stesso comportamento. L'obiettivo dell'analisi è stato la formulazione di un'approssimazione del comportamento della variabile risposta in relazione alle variabili indipendenti e l'inserimento di elementi ridondanti avrebbe rischiato di nuocere al modello.

Conseguentemente, la *feature selection* ha avuto inizio con il calcolo di due matrici: la prima per la correlazione tra gli indicatori e la seconda per la

correlazione tra gli indicatori e il flag di status. Il procedimento, che generalmente viene ripetuto più volte, deve gestire il *trade-off* tra maggiore correlazione con il flag e minore correlazione tra gli indicatori.

Il criterio scelto per le iterazioni ha previsto l'esclusione progressiva degli indicatori che superavano una certa soglia di correlazione con gli altri indicatori. Nello specifico, se la correlazione tra l'indicatore *A* e l'indicatore *B* era superiore a un valore limite predeterminato, si è mantenuto l'indicatore con la correlazione maggiore con il flag e si è eliminato l'altro. Si sono così creati alcuni dataset, con la cancellazione e graduale dei valori nel seguente ordine [47]:

- Indicatori con correlazione tra lo 0,8 e 1 (molto alta);
- Indicatori con correlazione tra lo 0,6 e 0,8 (alta);
- Indicatori con correlazione tra lo 0,4 e 0,6 (regolare);
- Indicatori con correlazione tra lo 0,2 e 0,4 (bassa).

Da questi dataset si sono selezionati gli indicatori sulla base della correlazione con il flag, così da inserirli nei primi modelli di prova. L'analisi ha messo in luce risultati interessanti per i dataset con correlazioni maggiori di 0,2 e inferiori a 0,6. Sulla base di questa evidenza empirica è stato costruito un dataset con le variabili esplicative candidate del modello.

5.2 Software utilizzato per le elaborazioni

Il software utilizzato per la costruzione dei modelli, sia tramite reti neurali che SVM, è Matlab. Esso, realizzato dalla MathWorks, è reso disponibile agli studenti del Politecnico di Torino gratuitamente mediante un accordo tra le parti nella forma di una licenza per un utilizzo a fini didattici.

Si tratta di un programma progettato per il calcolo numerico e l'analisi di tipo statistico, ma attraverso alcune estensioni scaricabili in fase di installazione del programma, è possibile lavorare anche con applicazioni nel campo dell'intelligenza artificiale.

Matlab fornisce strumenti per gestire grandi set di dati e *tool* specifici per l'utilizzo delle tecniche del presente lavoro, oltre alla possibilità di personalizzare gli algoritmi attraverso la scrittura di codice. L'offerta consiste essenzialmente in due punti: da un lato dei *toolbox* preconfigurati e dall'altro la possibilità di generare modelli estremamente personalizzati.

Se da una parte i *toolbox* garantiscono facilità d'uso, essendo di fatto degli applicativi grafici, dall'altra mantengono una certa rigidità e non lasciano molto spazio a variazioni. Si è quindi optato per una maggior flessibilità del programma, preferendo l'implementazione tramite righe di codice. Nel software sono tuttavia presenti alcune funzioni che permettono di lavorare con relativa semplicità ad algoritmi altrimenti complessi da realizzare partendo da zero. Ad esempio, basta richiamare la funzione di addestramento, dopo aver settato le opportune personalizzazioni dei parametri, per permettere alla rete di apprendere.

5.3 Strumenti per la valutazione delle prestazioni

Per la valutazione delle performance si è fatto uso dei seguenti strumenti:

- Best validation performance;
- Matrice di confusione;
- Istogramma degli errori;
- Curva ROC;
- Gradiente e Validation Check.

5.3.1 Best validation performance

Questo strumento si configura come un grafico in cui le ascisse rappresentano le epoche di apprendimento e le ordinate i valori della funzione di *cross-entropy* selezionata per il modello. L'obiettivo dell'addestramento è la minimizzazione di tale funzione, pertanto occorre monitorare il suo andamento al fine di prevedere

eventuali miglioramenti da apportare all'approssimazione del comportamento delle variabili.

In particolare, il grafico mostra l'andamento dell'apprendimento al variare delle epoche per i set di *training*, *validation* e test. Esso fornisce una sintesi grafica di come il modello sta imparando e di come generalizza, oltre a dare un'indicazione sul minimo valore di errore e sulla relativa epoca in cui esso è raggiunto. La figura 5.1 è tratta da una configurazione con 6 neuroni in input e due strati nascosti, ognuno con 6 neuroni.

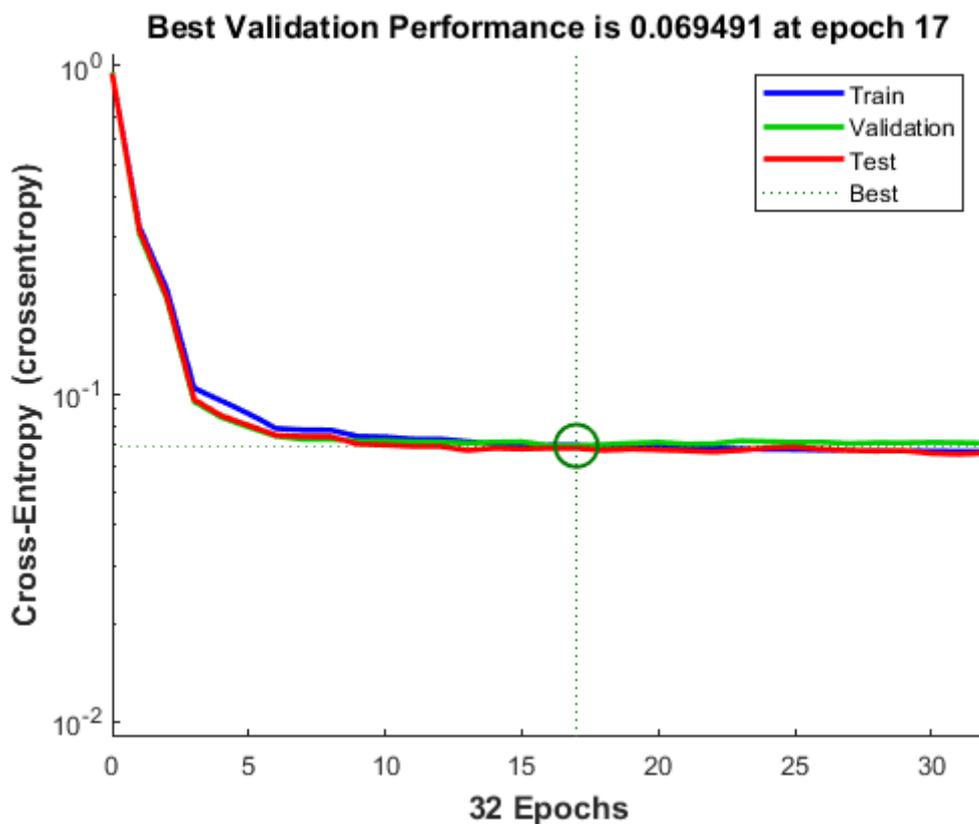


Figura 5.1 - Esempio di grafico di Best validation performance [elaborazione]

Tale grafico può mettere in luce alcuni comportamenti delle curve, alcuni anomali come l'*underfit* e l'*overfit*, altri regolari come il *good fit*.

Underfit

Questo comportamento evidenzia l'incapacità del modello di apprendere dal training set. Secondo la definizione che Goodfellow, Bengio e Courville riportano in *Deep Learning*:

“Si verifica underfit quando il modello non è in grado di ottenere un valore di errore sufficientemente basso sul training set” [50].

Questa affermazione si traduce graficamente in una funzione di errore costante nonostante l'addestramento [Figura 5.2] o che continua a decrescere fino al termine dell'addestramento [Figura 5.3]. Il primo caso rivela che il modello non è dotato di sufficiente capacità, mentre il secondo indica come il modello sia in grado di apprendere ulteriormente e suggerisce il verificarsi di un'interruzione prematura dell'addestramento. In questo caso è utile ripetere l'operazione di *training* allungando il ciclo di addestramento.

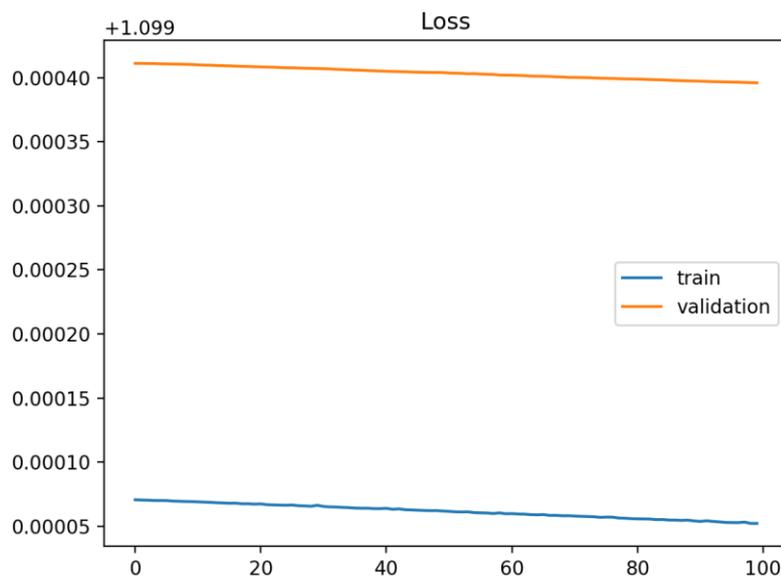


Figura 5.2 - Underfit di un modello con mancanza di capacità [49]

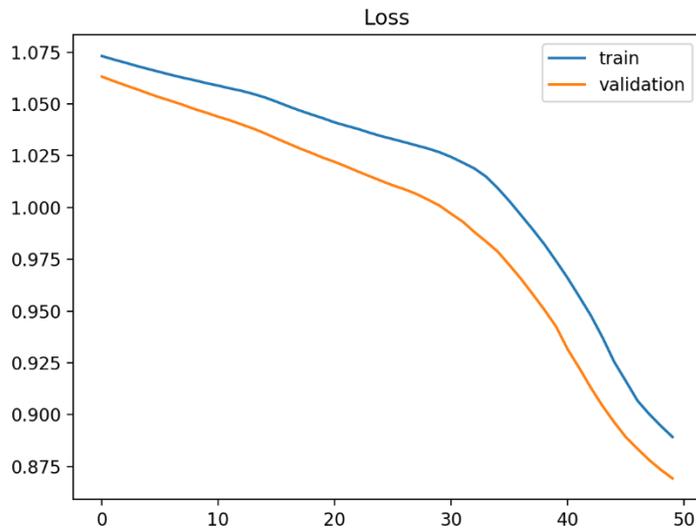


Figura 5.3 - Underfit di un modello che necessita di ulteriore addestramento [49]

Overfit

Quando un modello apprende eccessivamente dal *training set*, includendo fluttuazioni causali e rumore statistico, si verifica *overfit*. Questo problema causa l'incapacità di adattarsi e predire una risposta corretta per un nuovo set di osservazioni. Per analogia, un modello di questo tipo è come uno studente che ha imparato a memoria le risposte dell'esame e non è in grado di rispondere correttamente quando la domanda è posta in maniera differente.

Questa circostanza può verificarsi quando il modello possiede troppa flessibilità e si adatta in maniera eccessiva al set di addestramento o se è addestrato troppo a lungo. Graficamente è riconoscibile quando la curva riferita al *training set* continua a decrescere o quando la curva del *validation set* decresce per poi cominciare a salire [Figura 5.4].

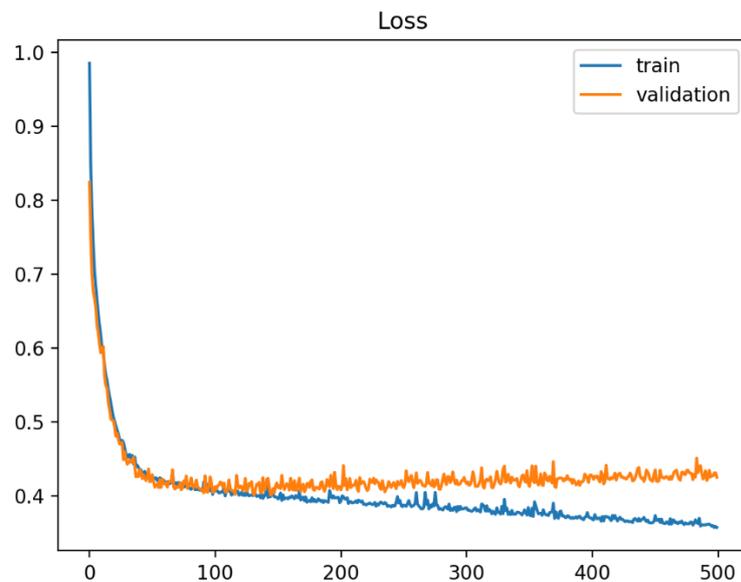


Figura 5.4 - Esempio di un modello affetto da overfit [49]

Good fit

Avere un buon adattamento è sicuramente l'obiettivo di ogni modello. In questa casistica, le curve dapprima decrescono fino a un punto di minimo, salvo poi stabilizzarsi. Generalmente, l'errore del *training set* è minore rispetto a quello del set di validazione. Tale differenza è definita *gap di generalizzazione*.

Quindi, in caso di *good fit* si assiste a una decrescita delle curve del *training set* e del *validation set* fino a un punto in cui diventano stabili, mantenendo tuttavia un piccolo gap tra le due [Figura 5.5].

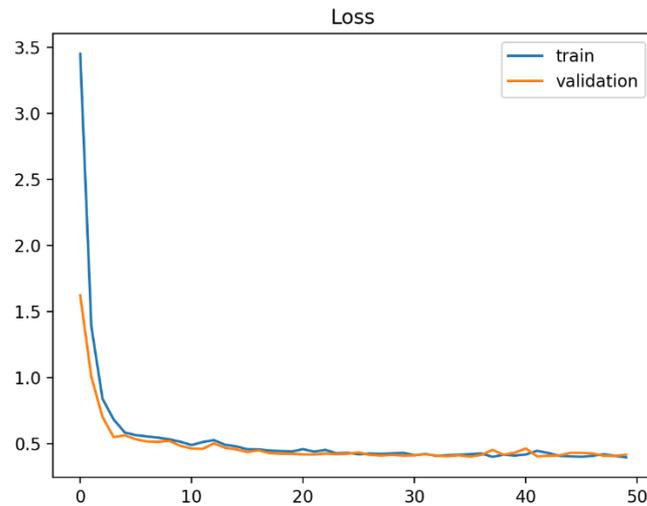


Figura 5.5 - Esempio di un modello con good fit [49]

Altri comportamenti

Si può assistere ad altri comportamenti anomali, i quali vengono spiegati dalla poca rappresentatività dei dati di addestramento e validazione.

Quando i dati di *training* sono troppo pochi rispetto a quelli di validazione si assiste a un grande gap tra le due curve [Figura 5.6].

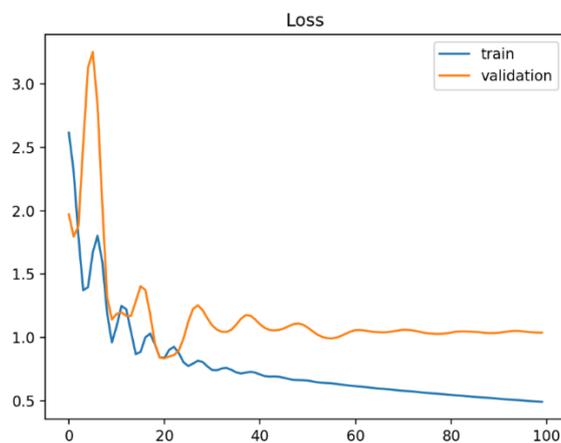


Figura 5.6 - Esempio di un modello con training set troppo piccolo rispetto al validation set [49]

Quando è invece il set di validazione a essere poco rappresentativo, viene meno l'informazione circa la capacità di generalizzazione del modello. Per

questa fattispecie, il grafico della curva relativa ai dati di validazione si presenta come molto distorto intorno ai valori della curva del *training set* [Figura 5.7].

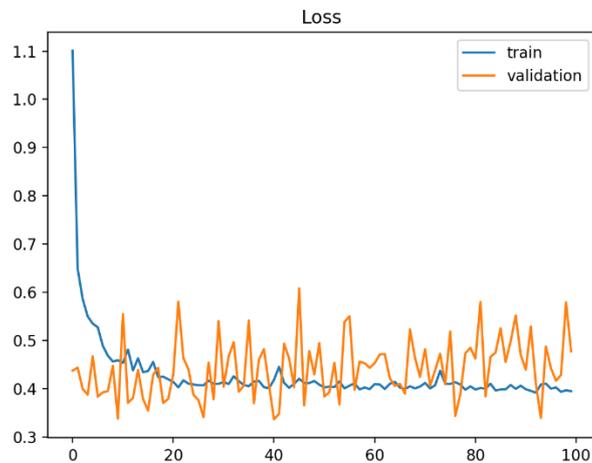


Figura 5.7 - Esempio di un modello con validation set troppo piccolo rispetto al training set [49]

5.3.2 Matrice di confusione

La matrice di confusione è una tabella molto utilizzata per la valutazione delle performance di un modello di classificazione binario. Nello specifico, essa presenta sulle righe la classe predetta dal modello e sulle colonne la classe target, tuttavia non è raro trovare questa disposizione invertita. Le celle frutto dell'incrocio tra righe e colonne forniscono indicazioni utili sulla capacità predittiva. Utilizzando la matrice di confusione, si possono pertanto valutare l'accuratezza e l'efficacia del modello. In figura 5.8 sono rappresentate le componenti di una matrice di confusione.

Output class	0	TN	FN	NPV
	1	FP	TP	Precisione
		Specificità	Sensitività	Accuracy
		0	1	
	Target class			

Figura 5.8 - Componenti di una matrice di confusione [elaborazione]

Assumendo di identificare con 0 la classe negativa (società sana) e con 1 la classe positiva (società anomala), possiamo definire i seguenti parametri:

- True negative (TN)= valori classificati dal modello come negativi ed effettivamente negativi;
- False negative (FN)= valori classificati dal modello come negativi, ma positivi;
- True positive (TP)= valori classificati dal modello come positivi ed effettivamente positivi;
- False positive (FP)= valori classificati dal modello come positivi, ma negativi.

Utilizzando la notazione appena esposta è possibile definire alcune metriche per la valutazione delle performance.

Accuracy

Indica l'accuratezza del modello ed è data dal numero di previsioni corrette sul totale delle previsioni. Il valore massimo è 1, mentre un'Accuracy tendente allo 0 non identifica un buon modello.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (5.1)$$

Tasso di errore

Rappresenta il numero di previsioni errate sul totale delle previsioni. È il complemento a uno dell'*Accuracy*, per cui più è vicino allo 0 più il modello è buono.

$$\text{Tasso di errore} = 1 - \text{Accuracy} = \frac{FP + FN}{TP + FP + TN + FN} \quad (5.2)$$

Specificità

La specificità rappresenta la capacità del modello di identificare come negative le istanze effettivamente negative.

$$\text{Specificità} = \frac{TN}{TN + FP} \quad (5.3)$$

Sensitività

Detta anche *recall*, la sensitività esprime la capacità del modello di identificare come positive le istanze effettivamente positive.

$$\text{Sensitività} = \frac{TP}{TP + FN} \quad (5.4)$$

Precisione

Questa metrica indica la capacità del modello di classificazione a etichettare correttamente le osservazioni positive.

$$\text{Precisione} = \frac{TP}{TP + FP} \quad (5.5)$$

Negative predictive value

Questa metrica indica la capacità del classificatore a etichettare correttamente le osservazioni negative.

$$NPV = \frac{TN}{TN + FN} \quad (5.6)$$

5.3.3 Istogramma degli errori

L'istogramma degli errori è un utile strumento per visualizzare graficamente gli errori tra i valori target e i valori previsti dalla rete neurale. In riferimento all'istogramma di figura 5.9, si nota come i risultati siano suddivisi in 20 classi, che corrisponde al valore di default del software utilizzato per l'analisi. L'asse delle ordinate rappresenta il numero di istanze appartenenti a una data classe, mentre sulle ascisse sono rappresentate le classi. Ogni barra può contenere istanze relative al *training*, al *validation* e al *test set*, mentre la linea arancione verticale rappresenta il livello di zero errori. Un buon modello dovrebbe veder concentrata la maggior parte degli errori proprio intorno al valore nullo, in quanto una concentrazione verso valori lontani dallo zero potrebbe suggerire una distorsione sistematica.

Per comprendere meglio questo aspetto è utile riferirsi alla composizione dell'errore, ossia alla differenza tra target ed output. Una concentrazione degli errori verso valori positivi potrebbe suggerire un'attribuzione del punteggio sottostimata rispetto alla realtà, poiché i valori di output sarebbero costantemente inferiori a quelli di target. Ovviamente una simile distorsione potrebbe avvenire anche in senso opposto, con una sovrastima dei valori di output e un conseguente errore concentrato nelle classi che occupano la parte sinistra del grafico.

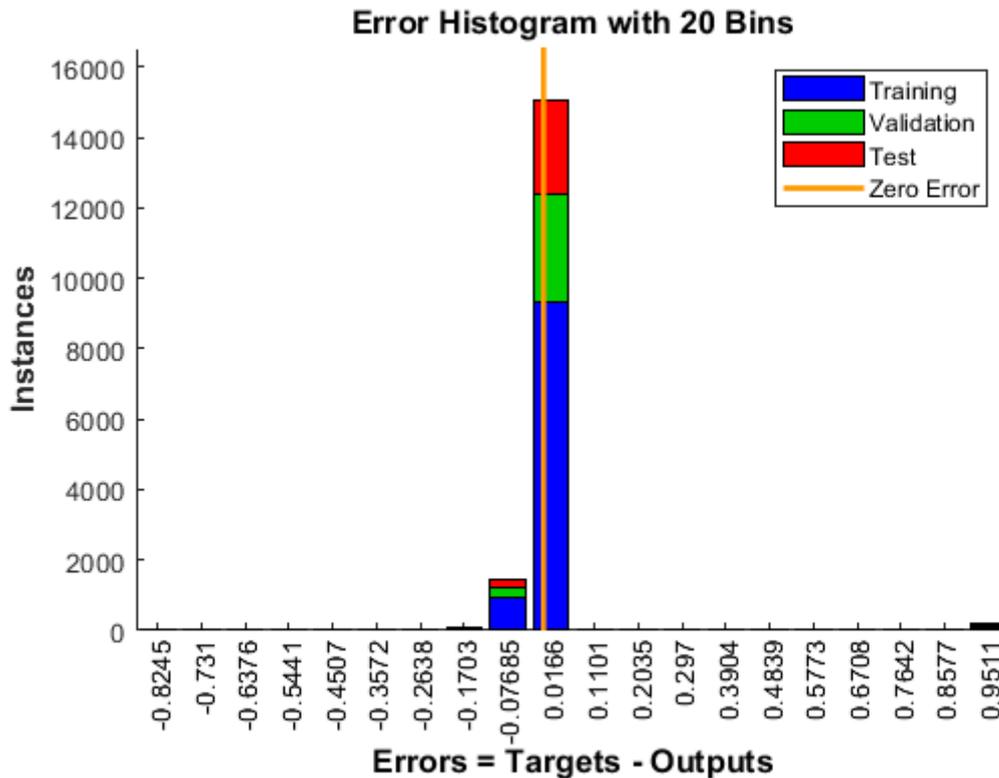


Figura 5.9 - Esempio di istogramma degli errori [elaborazione]

5.3.4 Curva ROC

La curva *Receiver Operating Characteristic* è uno schema utilizzato generalmente per un classificatore binario. Esso presenta sull'asse delle ascisse il *False Positive Rate*, definito come $FPR = \frac{FP}{FP+TN} = 1 - Specificità$ e sull'asse delle ordinate il *True Positive Rate*, ossia $TPR = \frac{TP}{TP+FN} = Sensitività$.

Una curva ROC fornisce indicazioni sui valori di FPR e TPR per diverse soglie di classificazione. Infatti, l'abbassamento della soglia di classificazione definisce più elementi come positivi, aumentando così sia i falsi positivi che i veri positivi.

Il grafico in figura 5.10 mostra la composizione della curva ROC. L'area sottesa alla curva (A) è definita come *Area Under the Curve* (AUC) ed è utile per paragonare modelli di classificazione differenti. Infatti, un modello con AUC maggiore indica generalmente un classificatore migliore. La diagonale rappresenta un modello casuale, in cui il valore di AUC si attesta allo 0,5,

mentre in un modello perfetto, identificato dai segmenti superiori della rappresentazione, tale numero sale a 1. Ci si aspetta quindi che un modello di classificazione generico faccia registrare un'area compresa tra lo 0,5 ed 1.

La curva ROC è strettamente connessa con il concetto di *Accuracy* precedentemente presentato. Esiste infatti una semplice relazione per legare i valori dell'AUC con la metrica:

$$Accuracy = 2 * AUC - 1 \quad (5.7)$$

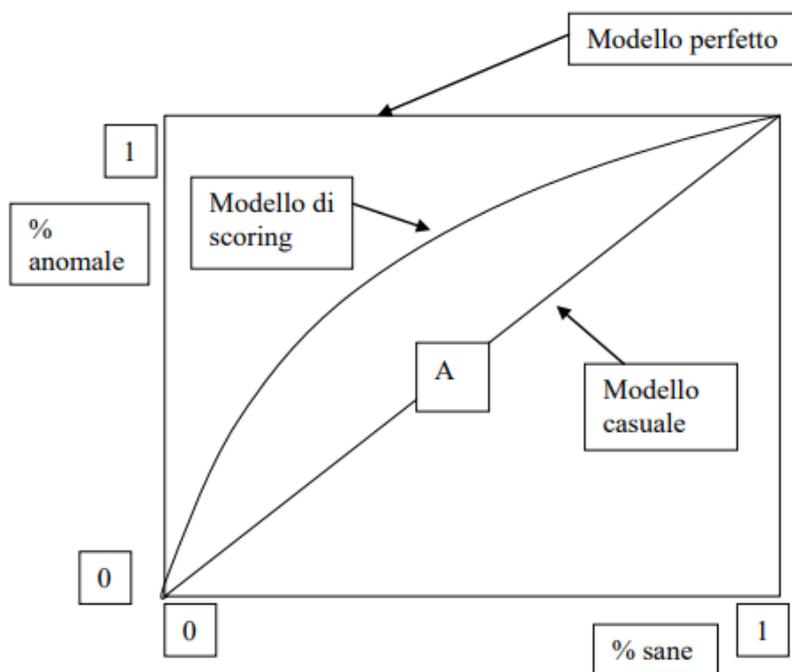


Figura 5.10 - Composizione curva ROC [1]

5.3.5 Gradiente e Validation Check

Utilizzando la funzione di addestramento riferita al metodo del *scaled conjugate gradient*, ci si aspetta che l'apprendimento avvenga modificando i valori dei pesi in modo da diminuire progressivamente il valore gradiente. Il grafico del gradiente può mettere in luce comportamenti anomali del modello come il problema della scomparsa del gradiente e quello dell'esplosione del gradiente.

Il primo problema si riferisce a un modello che vede il gradiente ridursi esclusivamente durante le prime fasi di apprendimento, mentre nel secondo caso si assiste a una crescita dello stesso. Per ovviare a questo problema è possibile mettere in atto alcune soluzioni, la prima delle quali prevede di ridurre il numero di strati della rete per abbassarne la complessità. Un ulteriore intervento prevede una modifica nell'inizializzazione dei pesi. Infine, un accorgimento da impiegare nel caso di *exploding gradient* per le reti molto profonde è il cosiddetto *clipping*, ossia un particolare ridimensionamento del valore del gradiente: quando questo supera una certa soglia predeterminata, esso viene riportato entro i limiti, evitando così una crescita incontrollata.

Il secondo grafico in figura 5.11 si riferisce al numero di *validation fail* a cui il modello è andato incontro nel corso delle epoche di addestramento. Si parla di *fail* quando il valore della funzione da minimizzare - *cross-entropy* in questo caso – aumenta anziché diminuire. Esso fa parte del processo più generale di *validation stop*, che serve per evitare che la rete abbia performance scadenti: se le prestazioni di convalida peggiorano per un numero consecutivo e predeterminato di epoche, il processo viene interrotto al fine di preservare l'integrità del modello.

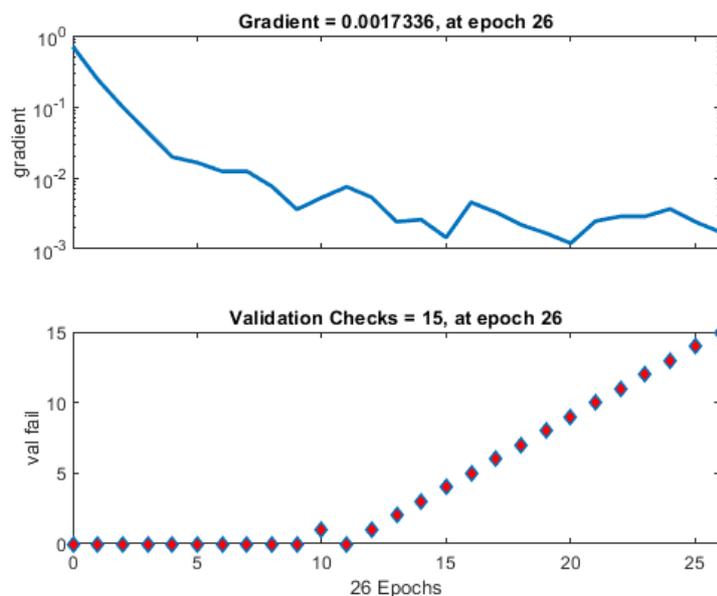


Figura 5.11 - Esempio di Gradiente e Validation Check [elaborazione]

5.4 Modello con reti neurali

Questa sezione tratta del procedimento che ha portato alla costruzione della rete, passando in rassegna le problematiche riscontrate e la loro risoluzione. Le architetture testate sperimentalmente hanno consentito l'individuazione del miglior modello, ma prima di valutare le performance di quest'ultimo è opportuno presentare la costruzione della rete tramite linee di codice.

5.4.1 Modelli di prova e risoluzione dei problemi

Il modello realizzato tramite reti neurali ha previsto l'utilizzo del dataset di variabili candidate e, per poter giungere a un classificatore che approssimasse in modo accettabile le osservazioni, si sono svolte alcune prove. Nel corso di queste sperimentazioni sono emerse alcune problematiche relative all'implementazione di tale modello, in particolare: il bilanciamento delle correlazioni, l'inizializzazione dei pesi e dei *bias*, l'inizializzazione delle funzioni di addestramento e di errore e la suddivisione dei dati nei set.

Bilanciamento correlazioni

Tra gli indicatori candidati, la selezione con il criterio di massima correlazione con il flag ha portato a modelli che mal approssimavano il comportamento generale, in quanto non in grado di predire essenzialmente quando una società fosse da ritenere anomala. Approfondendo l'analisi, si è notato come la selezione degli indicatori maggiormente correlati con il flag portasse a uno sbilanciamento della somma delle correlazioni verso il comportamento opposto a quello del flag (risultato della somma < 0). Per ovviare a questo problema, oltre a considerare gli indicatori con maggiore correlazione, si è tenuto conto della somma delle correlazioni con il flag degli indicatori inclusi, in modo da tenerla il più vicino possibile allo zero.

Di seguito, la matrice di confusione [Figura 5.12] e l'istogramma degli errori [Figura 5.13] di un modello con 4 neuroni di input non bilanciati e con 4 e 8 neuroni nei due strati nascosti. Dalla matrice di confusione è evidente come lo sbilanciamento della somma delle correlazioni non permetta al classificatore di

identificare le osservazioni come anomale. Infatti, esso non riesce a predire l'appartenenza a tale classe. È altresì importante notare come l'errore predittivo della rete si attesti solamente all'1,6%. Quindi, un'analisi superficiale che consideri esclusivamente questa metrica potrebbe indurre in errore: questo risultato è dovuto al fatto che le osservazioni appartenenti alla classe delle sane (classe 0) siano molte di più rispetto alle anomale (classe 1), per cui tale disproporzione contribuisce a tenere basso l'errore. Anche l'istogramma degli errori presenta dei segnali di allarme: l'eccessiva concentrazione intorno a valori positivi indica che il modello fornisce degli output sistematicamente inferiori ai valori target.

Test Confusion Matrix

Output Class	0	2956 98.4%	49 1.6%	98.4% 1.6%
	1	0 0.0%	0 0.0%	NaN% NaN%
		100% 0.0%	0.0% 100%	98.4% 1.6%
		0	1	
		Target Class		

Figura 5.12 - Matrice di confusione di una rete con 4 neuroni in input e 4 e 8 negli strati nascosti [elaborazione]

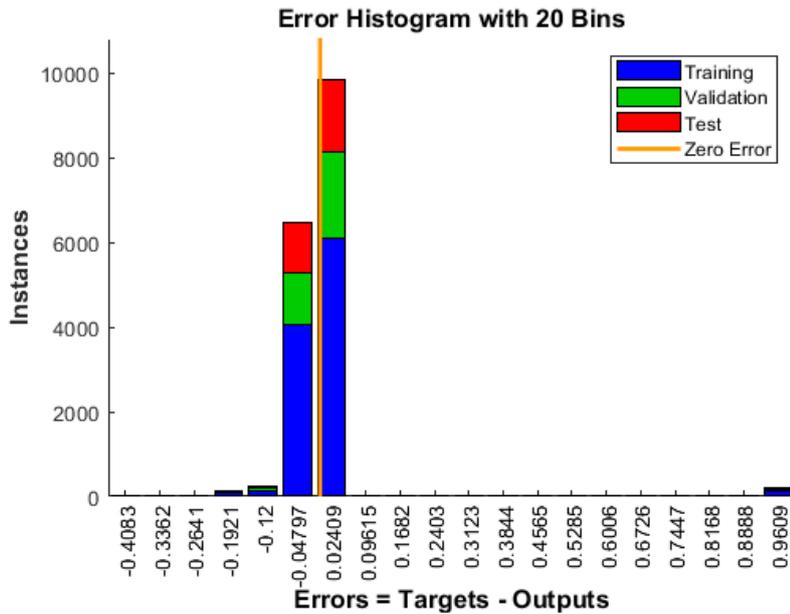


Figura 5.13 - Istogramma degli errori di una rete con 4 neuroni in input e 4 e 8 negli strati nascosti [elaborazione]

Inizializzazione pesi e bias

L'inizializzazione di pesi e *bias* avviene di default in modo casuale all'interno dell'ambiente di sviluppo. Si è quindi deciso di eliminare questa fonte di variabilità, inizializzando i *bias* a zero e settando i pesi a valori sempre uguali. Infatti, è stata costruita in un file Excel una matrice di pesi utilizzando la funzione di creazione di valori casuali compresi tra 0,1 e 0,9, così da non avere numeri eccessivamente alti o bassi ed evitare una distorsione dei risultati. Dopodiché, visto che la natura della funzione faceva variare i numeri ad ogni apertura del file, si è copiata e incollata questa matrice mantenendo soltanto i valori. In questo modo, si è simulato il comportamento di casualità proprio del software, rendendo però confrontabili le performance dei modelli, non avendo valori di default differenti per ogni addestramento.

Inizializzazione funzioni di addestramento e di errore

Per le funzioni di addestramento e di errore si è cercato nella letteratura disponibile e per la prima si è scelta la funzione *trainscg*, mentre per la seconda la *cross-entropy*.

Suddivisione dei dati nei set

La suddivisione dei dati nei set da parte del software avviene di default in modo casuale e questa particolarità avrebbe reso i modelli difficilmente confrontabili tra loro. Si è deciso di settare inizialmente la divisione dei set secondo una precisa indicizzazione.

Sulla base di questa suddivisione sono state implementate le prime reti neurali. Tale organizzazione ha previsto una prevalenza di osservazioni anomale nel set di addestramento poiché è da esso che il modello apprende, pertanto è opportuno che abbia le informazioni necessarie per garantire i risultati che si stanno cercando. Sarà altresì evidente dai paragrafi successivi come il set di validazione e di test abbiano valori anomali con ordini di grandezza confrontabili. Il primo, infatti, permette di convalidare quanto la rete ha appreso, mentre con il secondo si presentano delle nuove osservazioni per valutare l'effettiva capacità predittiva del classificatore.

Successivamente, sono state effettuate ulteriori prove dividendo i set in base alla stessa probabilità di default: in questo modo le osservazioni anomale sono state bilanciate tra i diversi set. Sono stati fatti altri tentativi inserendo nei tre set lo stesso numero di anomalie per poter ampliare il materiale utile per la scelta del miglior modello.

5.4.2 Architetture testate sperimentalmente

Il passo successivo ha previsto la costruzione dei modelli veri e propri. Non essendoci una *regola d'oro* per la selezione dell'architettura ottimale, si è deciso di procedere iterativamente testando diverse strutture. Si è infatti variato il numero di neuroni di input, il numero di strati nascosti e il numero di neuroni appartenenti agli strati nascosti.

Per il numero di neuroni in input, si sono costruite reti con 4, 6, 8 e 10 neuroni. Tali neuroni rappresentano di fatto gli indicatori e, come segnalato in

precedenza, la somma delle correlazioni non deve allontanarsi troppo dallo zero, pena la perdita di capacità di classificazione.

Si è deciso di testare i risultati di reti via via più profonde (*deep learning*), partendo da un *hidden layer* e arrivando a tre strati nascosti. Occorre considerare che l'aggiunta progressiva di livelli porta a un incremento della complessità computazionale.

Per formalizzare i tentativi, si sono costruite le architetture seguendo un semplice schema matematico in base al numero di neuroni in input e qui presentate per ciascuno strato nascosto:

- 1 hidden layer $\rightarrow n ; n + \frac{n}{2} ; 2n$
- 2 hidden layer $\rightarrow \frac{n}{2}, n ; n, n ; n, 2n$
- 3 hidden layer $\rightarrow \frac{n}{2}, n, n + \frac{n}{2} ; n, n, n ; n, n + \frac{n}{2}, 2n$

Di seguito sono esplicitate le 36 differenti architetture:

- | | | |
|-----------------|--------------------|-----------------------|
| 1) 4 input-4 | 13) 4 input-2-4 | 25) 4 input-2-4-6 |
| 2) 4 input-6 | 14) 4 input-4-4 | 26) 4 input-4-4-4 |
| 3) 4 input-8 | 15) 4 input-4-8 | 27) 4 input-4-6-8 |
| 4) 6 input-6 | 16) 6 input-3-6 | 28) 6 input-3-6-9 |
| 5) 6 input-9 | 17) 6 input-6-6 | 29) 6 input-6-6-6 |
| 6) 6 input-12 | 18) 6 input-6-12 | 30) 6 input-6-9-12 |
| 7) 8 input-8 | 19) 8 input-4-8 | 31) 8 input-4-8-12 |
| 8) 8 input-12 | 20) 8 input-8-8 | 32) 8 input-8-8-8 |
| 9) 8 input-16 | 21) 8 input-8-16 | 33) 8 input-8-12-16 |
| 10) 10 input-10 | 22) 10 input-5-10 | 34) 10 input-5-10-15 |
| 11) 10 input-15 | 23) 10 input-10-10 | 35) 10 input-10-10-10 |
| 12) 10 input-20 | 24) 10 input-10-20 | 36) 10 input-10-15-20 |

5.4.3 Implementazione del modello

Viene di seguito presentato il codice utilizzato per la realizzazione dei modelli tramite reti neurali.

Per prima cosa, è importante chiudere ogni file aperto dal software, eliminare ogni variabile presente nel *workspace* e pulire la linea di comando. In questo modo, eseguendo il comando *run* dopo ogni modifica all'architettura della rete, si è sicuri di partire effettivamente da zero.

```
%% Pulizia workspace, variabili e linea di comando
clear;
close all;
clc;
```

Il secondo step concerne l'importazione del dataset. A questo proposito, è necessario dare il comando di lettura del file. In alternativa, è possibile caricare manualmente il database tramite l'apposito comando della barra degli strumenti, ma, dovendo operare ripetutamente sulla struttura, è stato conveniente automatizzare l'operazione.

```
%% Importare il dataset
Dataset=readtable('C:\Percorso\Dataset.xlsx');
```

Nonostante i controlli dei dati, per essere sicuri di non lavorare con un database affetto da valori mancanti, si è deciso di effettuare un'ulteriore operazione di pulizia tramite il richiamo di un'apposita funzione. Si è poi impostato il valore della variabile Dataset al contenuto ripulito dai *missing values*.

```
%% Pulizia dei dati dai missing values
DatiClear=rmmissing(Dataset);
Dataset=DatiClear;
```

Il passo successivo ha previsto il salvataggio dei valori della tabella di Excel in appositi vettori adatti agli scopi del lavoro. È importante sottolineare che i flag rappresentano la variabile risposta nonché l'unico neurone di output, mentre gli indicatori i possibili predittori. L'indicizzazione delle colonne è 4 per i flag, in quanto tali valori sono registrati nella quarta colonna, mentre i valori degli indicatori variano da 5 a 14 a seconda del numero di neuroni di input. Nel codice seguente, il modello è costruito con 4 neuroni di input.

```
%% Salvataggio dei dati in vettori
%Flag
flag=table2array(Dataset(:,4));
```

```
%Indicatori
indicatori=table2array(Dataset(:,5:8));
```

L'inizializzazione del numero di neuroni nei *layer* è avvenuta facendo ricorso a dei vettori, uno per ogni strato, e includendo tali valori in un ulteriore vettore denominato `hiddenLayerSize`. In questo modo, la variazione dell'architettura della rete ha potuto configurarsi tramite la modifica dei vettori, talvolta anche variando la dimensione di `hiddenLayerSize` per rispecchiare il numero di strati nascosti utilizzati.

La funzione di addestramento selezionata è stata `trainscg`, ossia con la metodologia dello *scaled conjugate gradient*.

Con il metodo `patternnet` si è configurata la rete secondo i parametri appena citati.

```
% Inizializzazione layer e funzione di addestramento
hiddenLayerSize1=[4];
hiddenLayerSize2=[6];
hiddenLayerSize3=[8];

hiddenLayerSize =[hiddenLayerSize1 hiddenLayerSize2,
hiddenLayerSize3];

trainFcn='trainscg';

net=patternnet(hiddenLayerSize,trainFcn);
```

Il numero dei fallimenti consecutivi accettabili è stato portato dal valore di default 6 a 15 per permettere alla rete di apprendere più a lungo, ma non incappando nella possibilità di generare *overfitting*.

La funzione per il calcolo delle performance più adatta per le esigenze del lavoro è la `crossentropy`.

```

%% Altri parametri della rete
%Max fail
net.trainParam.max_fail=15;
%Funzione per il calcolo delle performance
net.performFcn='crossentropy';

```

La suddivisione dei dati nei tre set di riferimento è avvenuta modificando l'impostazione di default, la quale opera secondo una indicizzazione casuale, e definendo manualmente la composizione degli insiemi. In questo esempio si segue l'indicizzazione iniziale, ma i valori di riferimento dei diversi set sono stati modificati nel corso delle prove a seconda delle esigenze precedentemente illustrate.

```

%% Suddivisione dei dati nei set
%Metodo di divisione indicizzato e non casuale
net.divideFcn='divideind';
%Training set
net.divideParam.trainInd=1:10600;
%Validation set
net.divideParam.valInd=10601:14000;
%Test set
net.divideParam.testInd=14001:17005;

```

Per costruire la rete è stato necessario calcolare le matrici trasposte degli indicatori e dei flag. Si tratta di un'operazione convenzionale per evitare errori nel richiamo della funzione di configurazione.

```

%% Costruzione rete
%Trasposta dei vettori
indicatoriTrasp=indicatori';
flagTrasp=flag';
%Configurazione della rete
net=configure(net,indicatoriTrasp,flagTrasp);

```

L'inizializzazione dei pesi e dei *bias* ha previsto l'importazione di un file Excel contenente dei valori randomizzati (per riprodurre ciò che il software fa di default) tra 0,1 e 0,9.

I *bias* hanno assunto valore nullo, mentre i pesi a seconda del livello, hanno attinto i valori dalla matrice importata. Per quanto riguarda i *bias*, nel codice seguente sono presenti tre righe di inizializzazione trattandosi di una rete con tre strati nascosti. I pesi, invece, rappresentano la "forza" dei legami tra il livello di input formato da 4 neuroni e il primo strato nascosto, tra il primo e il secondo strato nascosto, tra il secondo e il terzo strato nascosto e infine tra il terzo strato nascosto e il livello di output, per un totale di quattro righe di codice dedicate.

```
% Inizializzazione dei pesi e dei bias
%Importo una matrice di numeri casuali tra 0,1 e 0,9
Pesi=readtable('C:\Users\aless\Documents\Università\Tesi\Mo
dello rete neurale\Pesi.xlsx');
Pesi=table2array(Pesi);
%Inizializzazione bias
net.b{1}=zeros(hiddenLayerSize1,1);
net.b{2}=zeros(hiddenLayerSize2,1);
net.b{3}=zeros(hiddenLayerSize3,1);
%Inizializzazione pesi
netIW{1,1}=Pesi(1:hiddenLayerSize1,1:4);
netLW{2,1}=Pesi(1:hiddenLayerSize2,1:hiddenLayerSize1);
netLW{3,2}=Pesi(1:hiddenLayerSize3,1:hiddenLayerSize2);
netLW{4,3}=Pesi(1:1,1:hiddenLayerSize3);
```

Una volta importati i dati e impostati i parametri della rete, è stato possibile procedere con l'addestramento.

```
% Addestramento
[net,tr]=train(net,indicatoriTrasp,flagTrasp);
```

L'esecuzione delle linee di codice fino a qui illustrate ha consentito al software di creare e allenare la rete neurale. I risultati di tale allenamento sono presentati nello strumento denominato *nntraintool* che viene aperto di default al lancio delle istruzioni. Da qui è possibile osservare graficamente l'architettura della rete, oltre a visualizzare i grafici relativi alle performance interagendo con l'interfaccia grafica.

5.4.4 Presentazione dei tentativi

Ai fini della ricerca del miglior modello tramite reti neurali sono stati svolti alcuni tentativi, variando la composizione dei set e gli indicatori da includere. Le prove hanno previsto:

- Modello iniziale con gli indicatori scelti tramite *feature selection*;
- Suddivisione dei set in base alla stessa percentuale di default;
- Suddivisione dei set in base allo stesso numero di default;
- Creazione di un dataset di indicatori candidati tenendo presente correlazione con il flag, somma delle correlazioni e tipologia di indicatore (sviluppo, redditività, produttività e liquidità);
- Creazione di dataset cambiando l'ordine dei 10 indicatori inclusi nel dataset di partenza (es. *A-B-C-D-E-F-... → C-D-E-F-...-A-B → ecc.*).
- Creazione di un dataset che non contenesse indicatori adatti ai bilanci in forma ordinaria poiché, siccome non tutti i bilanci sono stati presentati in questa forma, i modelli avrebbero potuto subire distorsione.

5.4.4.1 Modello iniziale

Suddivisione dati

- *Training set* = righe da 1 a 10600 (62%), di cui 191 anomale (PD=1.8%);
- *Validation set* = righe da 10601 a 14000 (20%), di cui 48 anomale (PD=1.4%);
- *Test set* = righe da 14001 a 17005 (18%), di cui 49 anomale (PD=1.6%).

Variabili incluse

- $(Riserve+Utile)/Attivo\ netto$;
- $Ln(Attivo\ netto)$;
- $Oneri\ Finanziari/Ricavi$;
- $Liquidità/Attivo\ netto$;
- $(Rimanenze + Crediti\ vs\ clienti\ entro\ esercizio - Debiti\ vs\ fornitori\ entro\ esercizio)/Ricavi$;
- $Debiti\ finanziari\ (stimati)/Valore\ aggiunto$;
- $Oneri\ finanziari/Debiti\ Finanziari$;
- $Ammortamenti\ materiali/Costi\ operativi$;
- $Costo\ lavoro/Costi\ operativi$;
- $Debiti\ vs\ fornitori\ entro\ esercizio*360/Acquisti$.

Risultati

Al termine delle elaborazioni il miglior modello è risultato essere quello con 4 neuroni in input e una rete composta da tre strati nascosti, rispettivamente con 4, 6 e 8 neuroni [Figura 5.14].

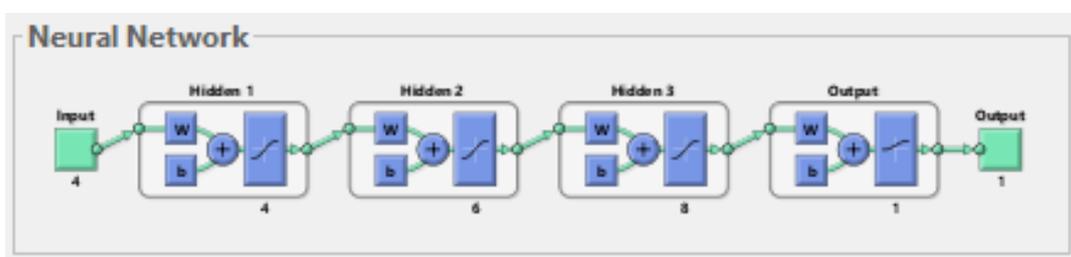


Figura 5.14 - Configurazione rete con 4 neuroni in input e 4, 6 e 8 negli strati nascosti [elaborazione]

La matrice di confusione ha evidenziato una buona capacità predittiva del modello. L'errore si è attestato intorno all'1,5%, mentre i livelli di precisione e sensibilità sono 63,2% e 24,5% [Figura 5.15]. Tuttavia, alcuni indicatori inclusi possono introdurre distorsione in quanto non adatti a bilanci non in forma ordinaria.

Test Confusion Matrix

Output Class	0	1	2
	0	1	2
0	2949 98.1%	37 1.2%	98.8% 1.2%
1	7 0.2%	12 0.4%	63.2% 36.8%
	99.8% 0.2%	24.5% 75.5%	98.5% 1.5%
	0	1	2
	Target Class		

Figura 5.15 - Matrice di confusione modello con 4 neuroni in input e 4, 6 e 8 negli strati nascosti [elaborazione]

5.4.4.2 Modello con set con la stessa probabilità di default

Suddivisione dati

- *Training set* = righe da 1 a 6999 (41%), di cui 123 anomale (PD=1.76%);
- *Validation set* = righe da 7000 a 10299 (19%), di cui 58 anomale (PD=1.76%);
- *Test set* = righe da 10300 a 17005 (40%), di cui 107 anomale (PD=1.6%).

Variabili incluse

- *(Riserve+Utile)/Attivo netto;*
- *Ln(Attivo netto);*
- *Oneri Finanziari/Ricavi;*
- *Liquidità/Attivo netto;*
- *(Rimanenze + Crediti vs clienti entro esercizio – Debiti vs fornitori entro esercizio)/Ricavi;*
- *Debiti finanziari (stimati)/Valore aggiunto;*
- *Oneri finanziari/Debiti Finanziari;*
- *Ammortamenti materiali/Costi operativi;*
- *Costo lavoro/Costi operativi;*

- *Debiti vs fornitori entro esercizio*360/Acquisti.*

Risultati

Il miglior modello per questa sezione di test è stato quello con 4 neuroni in input (i primi quattro indicatori nell'elenco appena citato) e 4 nell'unico strato nascosto, come è evidenziato dalla figura 5.16.

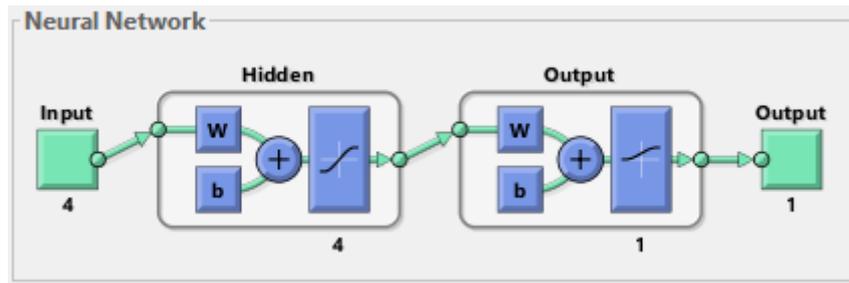


Figura 5.16 - Configurazione rete con 4 neuroni in input e 4 nello strato nascosto [elaborazione]

Nonostante il nuovo modello possa definirsi in estrema sintesi come accettabile, le performance non hanno fatto registrare dei miglioramenti, lasciando preferire il precedente modello. Infatti, quanto emerge dalla matrice di confusione [Figura 5.17] è, a parità di errore, la registrazione di valori di precisione e sensibilità al di sotto di quelli del modello iniziale. Pertanto, l'attribuzione ai set di osservazioni aventi complessivamente lo stesso tasso di default, non ha consentito di apportare miglioramenti al classificatore.

Test Confusion Matrix

Output Class	0	1	
	0	1	2
0	6584 98.2%	86 1.3%	98.7% 1.3%
1	15 0.2%	21 0.3%	58.3% 41.7%
	99.8% 0.2%	19.6% 80.4%	98.5% 1.5%
	Target Class		

Figura 5.17 - Matrice di confusione per modello con 4 neuroni in input e 4 nello strato nascosto [elaborazione]

È stato anche svolto un tentativo suddividendo i dataset in base al numero di default (e non al tasso), così che di 288 osservazioni anomale, ogni set ne contenesse 96. Anche in questo caso, però, i risultati non hanno messo in luce miglioramenti nella classificazione.

5.4.4.3 Modello con altre variabili incluse

Suddivisione dati

- *Training set* = righe da 1 a 6999 (41%), di cui 123 anomale (PD=1.76%);
- *Validation set* = righe da 7000 a 10299 (19%), di cui 58 anomale (PD=1.76%);
- *Test set* = righe da 10300 a 17005 (40%), di cui 107 anomale (PD=1.6%).

Variabili incluse

- *EBIT/Ricavi*;
- *Debiti vs fornitori entro esercizio*360/Acquisti*;
- *Costo lavoro/Costi operativi*;

- $\ln(\text{Attivo netto})$;
- $(\text{Rimanenze} + \text{Crediti vs clienti entro esercizio} - \text{Debiti vs fornitori entro esercizio})/\text{Ricavi}$;
- $\text{Debiti finanziari (stimati)}/\text{Valore aggiunto}$;
- $\text{Var\% Attivo netto}$;
- $\text{Risultato ante oneri finanziari}/(\text{Patrimonio netto} + \text{Debiti finanziari})$;
- $\text{Var\% Valore aggiunto operativo}$;
- $\text{Ammortamenti materiali}/\text{Costi operativi}$.

Risultati

In questo caso non è opportuno parlare di miglior modello, in quanto l'inclusione di nuove variabili precedentemente considerate marginali dalla *feature selection* ha portato a una perdita di capacità predittiva. Infatti, in molti casi il modello non è stato in grado di predire la classe delle anomalie. Pertanto, con questo tentativo, non si è assistito alla scoperta di un modello più performante, ma si è consolidata la bontà del processo che ha portato alla selezione dei migliori indicatori da includere. Alla riprova di ciò sono stati eseguiti ulteriori test utilizzando i 10 indicatori del modello iniziale, ma variandone l'ordine di inclusione. Nello specifico, codificati con *A-B-C-D-E-F-G-H-I-J* gli indicatori originali del paragrafo 5.4.4.1, si è modificato il database in modo da includere indicatori appartenenti a famiglie differenti nelle prove con 4, 6 e 8 neuroni di input secondo i seguenti schemi:

- *C-D-E-F-G-H-I-J-A-B*;
- *E-F-G-H-I-J-A-B-C-D*;
- *G-H-I-J-A-B-C-D-E-F*;
- *I-J-A-B-C-D-E-F-G-H*.

Il risultato di queste prove non ha portato a un miglioramento della capacità predittiva del modello, ma i risultati sono stati tendenzialmente migliori rispetto a quelli ottenuti con gli indicatori sopraelencati.

5.4.4.4 *Miglior modello*

L'ultimo tentativo ha previsto la sostituzione degli indicatori il cui calcolo, per bilanci non in forma ordinaria, è suscettibile di distorsione. Rispetto al modello iniziale sono stati sostituiti (*Rimanenze + Crediti vs clienti entro esercizio – Debiti vs fornitori entro esercizio*)/Ricavi, *Oneri finanziari/Debiti Finanziari e Oneri finanziari/Debiti Finanziari* con gli indicatori *Valore aggiunto/Immobilizzazioni materiali, Debiti totali/Valore aggiunto e Oneri finanziari netti/Attivo netto*.

Anche per questo nuovo dataset si è operato con una diversa impostazione dei set, assegnando la stessa probabilità di default e la divisione in base al modello iniziale. Tale seconda configurazione ha portato a risultati migliori.

Suddivisione dati

- *Training set* = righe da 1 a 10600 (62%), di cui 191 anomale (PD=1.8%);
- *Validation set* = righe da 10601 a 14000 (20%), di cui 48 anomale (PD=1.4%);
- *Test set* = righe da 14001 a 17005 (18%), di cui 49 anomale (PD=1.6%).

Variabili incluse

- *(Riserve+Utile)/Attivo netto*;
- *Ln(Attivo netto)*;
- *Oneri Finanziari/Ricavi*;
- *Liquidità/Attivo netto*;
- *Costo lavoro/Costi operativi*;
- *Debiti finanziari (stimati)/Valore aggiunto*;
- *Valore aggiunto/Immobilizzazioni materiali*;
- *Debiti totali/Valore aggiunto*;
- *Ammortamenti materiali/Costi operativi*;
- *Oneri finanziari netti/Attivo netto*.

Tali metriche appartengono essenzialmente a due famiglie, ossia liquidità e produttività e struttura operativa. In particolare, il quinto, il settimo ed il nono

indicatore esprimono la produttività, mentre i restanti forniscono indicazioni sulla liquidità della società.

5.5 Prestazioni del miglior modello con reti neurali

Il passo successivo alla definizione delle architetture e alla scrittura del codice per l'implementazione dell'algoritmo è stata l'analisi dei risultati trovati. Di tutti i modelli è stato individuato un miglior modello sulla base delle performance che ha fatto registrare e sul fatto che gli indicatori selezionati, a differenza del modello iniziale, ad esempio, non introducono distorsione a fronte di capacità predittiva equivalente.

Proprio per quest'ultimo motivo si è preferito selezionare il miglior modello considerando gli indicatori non distorti, a fronte di performance relativamente simili rispetto al modello iniziale. Tale scelta ha comportato la preferenza del classificatore con sei indicatori anziché quattro. Di riflesso, l'inclusione di ulteriori due predittori ha consentito di inserire nel miglior modello anche una metrica di produttività, a fronte di una predominanza di liquidità. In questo modo, il classificatore può tener conto di aspetti rilevanti riguardanti la struttura operativa della società e non unicamente fenomeni legati alla liquidità, una sfumatura in grado di impreziosire il modello, davanti, come già segnalato, a performance comparabili con quelle ottenute utilizzando quattro variabili soltanto.

Il miglior modello individuato al termine dei tentativi effettuati possiede un'architettura con 6 neuroni di input e 2 strati nascosti con rispettivamente 3 e 6 neuroni [Figura 5.18]. L'output rappresenta l'esito della classificazione binaria, ovvero se l'osservazione è da ritenersi sana o anomala. I neuroni di input sono costituiti dagli indicatori:

- $(Riserve+Utile)/Attivo\ netto;$
- $Ln(Attivo\ netto);$
- $Oneri\ Finanziari/Ricavi;$
- $Liquidità/Attivo\ netto;$

- *Costo lavoro/Costi operativi;*
- *Debiti finanziari (stimati)/Valore aggiunto.*

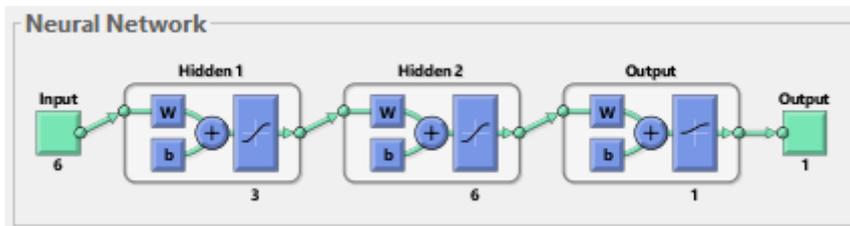


Figura 5.18 - Architettura modello con 6 neuroni di input e 3 e 6 negli strati nascosti [elaborazione]

Best Validation Performance

Il grafico Best Validation Performance [Figura 5.19] illustra l'andamento della *cross-entropy* al variare delle epoche di addestramento per i set di addestramento, validazione e test. L'andamento ottimale della curva dovrebbe presentare un punto di minimo per poi stabilizzarsi intorno a quel valore, mantenendo un delta di generalizzazione tra la curva di validazione e quella di addestramento.

La configurazione delle curve descrive ciò che viene definito come *good fit*, ossia un comportamento ottimale e non distorto del modello. Infatti, si osserva come il punto di minimo sia raggiunto nel corso della ventesima epoca e le curve rimangono stabili per i successivi periodi. È comunque osservabile il *generalization gap* tra le curve. Si nota che all'inizio dell'apprendimento tale delta è pari a zero, ma aumenta con il procedere dell'addestramento.

Tale comportamento consente di escludere che il modello sia afflitto da *overfit* o *underfit*, in quanto non vengono evidenziate dal grafico anomalie riconducibili a queste problematiche. Si può pertanto concludere che, dal punto di vista dell'analisi del *Best validation performance*, il modello è buono.

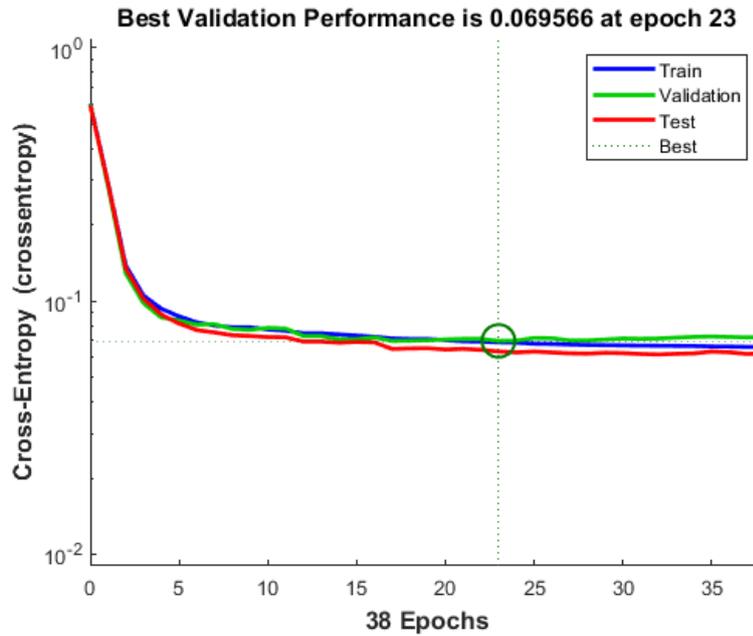


Figura 5.19 - BVP modello con 6 neuroni di input e 3 e 6 neuroni negli strati nascosti [elaborazione]

Gradiente e validation check

La porzione superiore del grafico in figura 5.20 raffigura l'andamento del gradiente al variare delle epoche di addestramento. Esso serve per determinare la bontà della funzione di addestramento, che si ricorda essere la *trainscg*, la quale si pone come obiettivo la minimizzazione del valore del gradiente. Tale minimizzazione si configura mediante l'aggiornamento dei pesi nelle varie epoche, così da ridurre gradualmente il gradiente.

È proprio nella gradualità che risiede una possibile insidia, in quanto, come già illustrato nel paragrafo 5.3.5, una riduzione repentina potrebbe indicare il problema del *vanishing gradient*. In questo caso, il gradiente vede diminuire il suo valore con l'avanzare delle epoche in modo non improvviso, ma graduale.

Un altro elemento su cui è necessario focalizzarsi è la non monotonicità dell'andamento: il metodo utilizzato cerca la direzione ottima per la riduzione del gradiente, percorrendo strade differenti. Una linea strettamente decrescente avrebbe evidenziato dei potenziali problemi nel meccanismo di addestramento.

Infine, il *validation check* specifica il numero dei *validation fail* nel corso delle epoche. Nel modello in questione, tale numero raggiunge la soglia preimpostata

durante la trentottesima epoca di addestramento, ovvero quando il modello ha subito l'interruzione dell'addestramento.

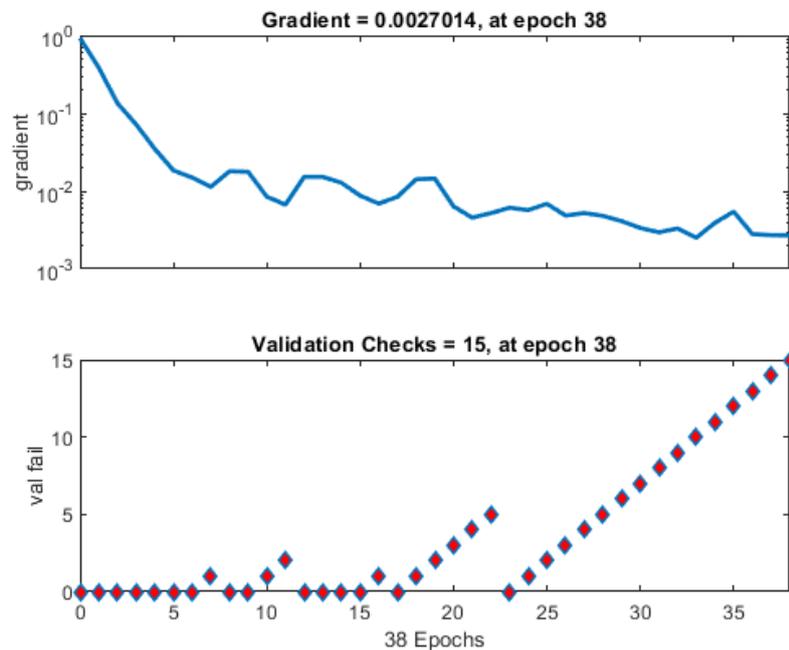


Figura 5.20 - Gradiente e validation check modello con 6 neuroni di input e 3 e 6 neuroni negli strati nascosti [elaborazione]

Istogramma degli errori

L'istogramma degli errori fornisce indicazioni sulla distribuzione dei delta tra target e output. Dal grafico si nota come la maggior parte degli errori ricada nella classe più prossima allo zero, non lasciando intravedere alcuna anomalia.

Guardando alle classi meno popolate, si evidenzia come quelle a sinistra dello zero rientrino intorno a valori che superano di poco il -0.2 e ciò significa che una parte delle osservazioni ottiene un punteggio di output superiore al target. Tuttavia, tenendo conto che nell'attribuzione della classificazione il software approssima all'intero più vicino, tali società vengono categorizzate correttamente.

Per quanto riguarda gli errori nella classe di destra, essi costituiscono il vero deficit del modello, in quanto l'output risulta inferiore al target per un valore di circa 0,96. Nonostante ciò, tali osservazioni rappresentano una piccola

percentuale del totale e possono essere interpretate come errore fisiologico di un modello di previsione.

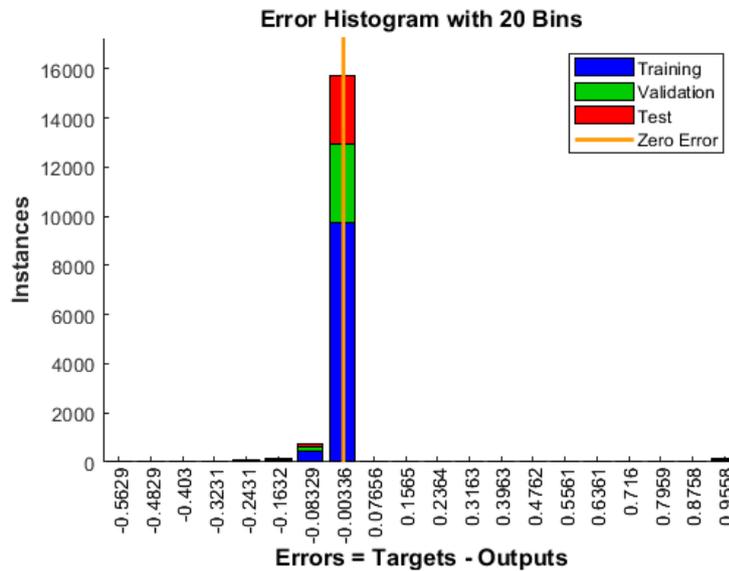


Figura 5.21 - Istogramma degli errori modello con 6 neuroni di input e 3 e 6 neuroni negli strati nascosti [elaborazione]

Matrice di confusione

La matrice di confusione è in grado di riassumere sotto forma tabellare l'esito della classificazione. Quanto si osserva nella figura 5.22 è la prevalenza di osservazioni appartenenti alla classe delle sane e categorizzate come tali. L'errore complessivo del modello si è attestato all'1.5%, conseguentemente l'accuratezza è pari al 98.5%.

Ulteriori valori su cui è opportuno soffermarsi sono la precisione e la sensibilità. Per quanto riguarda la prima metrica, il modello è stato in grado di classificare correttamente circa il 64% delle osservazioni che ha indicato come anomale. Il valore della sensibilità, invece, è pari a circa il 14%, che è inferiore a quanto trovato nel modello iniziale, ma, poiché quest'ultima architettura non include indicatori in grado di distorcere le elaborazioni, si è deciso di indicarlo come miglior modello.

Test Confusion Matrix

Output Class	0	1	2
	0	1	2
0	2952 98.2%	42 1.4%	98.6% 1.4%
1	4 0.1%	7 0.2%	63.6% 36.4%
	99.9% 0.1%	14.3% 85.7%	98.5% 1.5%
	0	1	2
	Target Class		

Figura 5.22 - Matrice di confusione modello con 6 neuroni di input e 3 e 6 neuroni negli strati nascosti [elaborazione]

Curva ROC

L'andamento della curva in figura 5.23 mostra il classico grafico di un modello di classificazione, pertanto è evidente come le prestazioni del classificatore siano migliori rispetto al modello casuale, qui indicato dalla linea grigia e a cui corrisponde una capacità di interpretazione dei dati casuale con il 50% di probabilità che le osservazioni siano classificate correttamente.

Allo stesso modo, la linea blu non aderisce al modello perfetto: si tratta di una sintesi di informazioni dotata di capacità interpretativa, è pertanto logico attendersi un andamento simile a quello riportato nel grafico piuttosto che l'aderenza all'andamento perfetto.

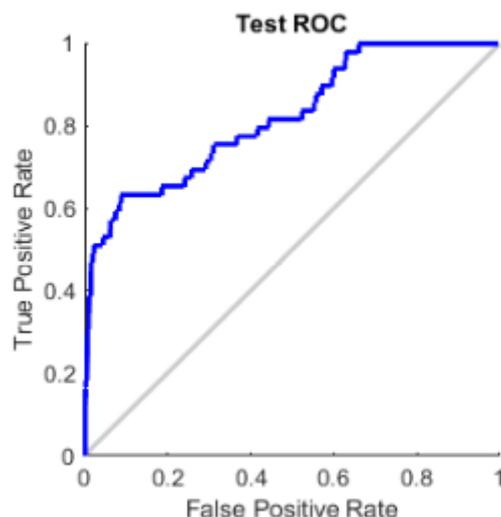


Figura 5.23 - Curva ROC modello con 6 neuroni di input e 3 e 6 neuroni negli strati nascosti [elaborazione]

5.6 Modello con SVM

Nei seguenti paragrafi viene trattata la metodologia che ha previsto l'impiego dell'algoritmo SVM per la stima di un modello di classificazione. Sono qui presentate le problematiche riscontrate e la loro risoluzione per individuare il miglior modello. Infine, anche in questo caso, come per le reti neurali, è illustrato il codice che ha permesso di lavorare con questo strumento di intelligenza artificiale.

5.6.1 Modelli di prova e risoluzione dei problemi

Il punto di partenza per i modelli costruiti mediante SVM è lo stesso delle reti neurali, ossia l'individuazione delle variabili candidate ad essere utilizzate nel modello. Si è deciso di iniziare dallo stesso dataset di variabili per uniformità di trattazione e per rendere i due metodi più facilmente confrontabili.

I punti critici per la realizzazione del modello sono: selezione della *Kernel function*, suddivisione dei dati nei set e numero di variabili da utilizzare.

Selezione della funzione Kernel

Poiché tale funzione è deducibile soltanto empiricamente, si è iniziato con la relazione più semplice, ossia la lineare. Sono stati costruiti dei modelli sulla base di questo parametro, ma i risultati non sono stati soddisfacenti. Infatti, utilizzando questa particolare tipologia di funzione, i modelli non erano in grado di prevedere la classe delle anomale, fornendo esclusivamente la categoria sane come output. Tale comportamento è indipendente dal numero di variabili in input, ma è dovuto all'inadeguatezza del settaggio dei parametri. Dunque, affinando il modello e introducendo ulteriori variabili esplicative, il risultato rimaneva lo stesso. Tale evidenza è osservabile nelle figure successive, dove vengono presentate le matrici di confusione per due modelli con 4 [Figura 5.24] e 8 variabili incluse [Figura 5.25].

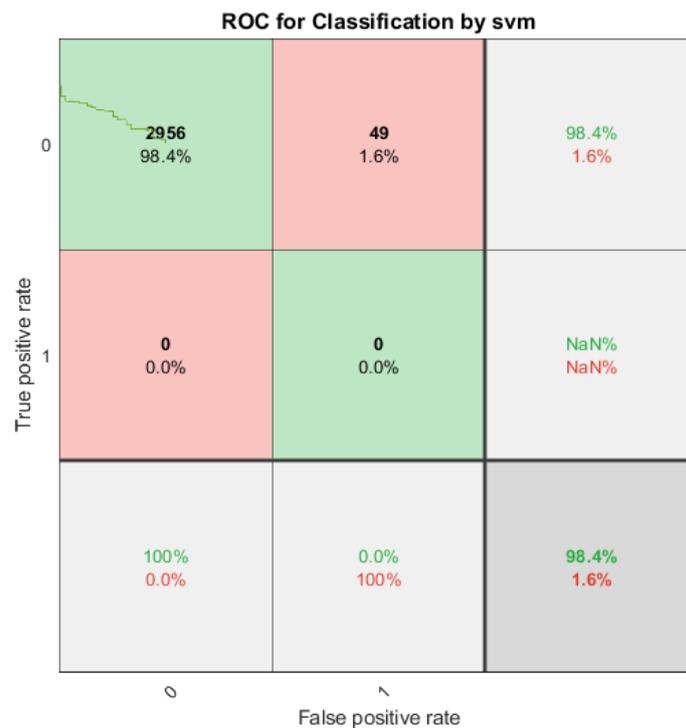


Figura 5.24 - Matrice di confusione di un modello con 4 variabili esplicative e funzione lineare [elaborazione]

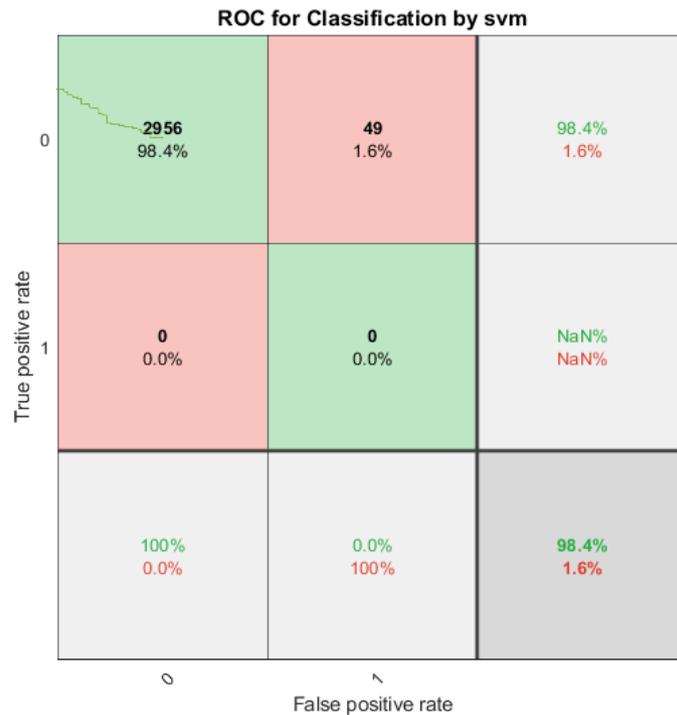


Figura 5.25 - Matrice di confusione di un modello con 8 variabili esplicative e funzione lineare [elaborazione]

Si è deciso allora di proseguire testando una funzione radiale come Kernel (*radial basis function*) e i risultati sono cambiati considerevolmente, a dimostrazione dell'inadeguatezza dell'approssimazione lineare. La *rbf* è stata scelta come *Kernel function* per la costruzione del modello.

Suddivisione dei dati nei set

Per la costruzione tramite SVM si è operato con un *training set* e con un *test set*. Sono stati fatti alcuni tentativi per meglio comprendere la miglior suddivisione dei dati, provando dapprima a suddividere il 50% delle osservazioni anomale in ciascun set e successivamente incrementando il numero di osservazioni anomale nel *training set*; si sono infine analizzati i casi con stessa probabilità di default nei set.

L'esito conclusivo, che sarà illustrato successivamente, ha mostrato come, a parità di input, sia da preferire uno sbilanciamento delle osservazioni anomale

in favore del *training set*. Infatti, il modello gode di un'accuratezza maggiore e possiede anche una miglior capacità di discernere le osservazioni anomale. Questo accorgimento ha apportato vantaggi alla capacità predittiva del modello, come dimostrano le figure seguenti per due modelli con 6 variabili predittive e con rispettivamente il 50% [Figura 5.26], l'80% [Figura 5.27] delle osservazioni anomale all'interno del set di addestramento, oltre alla suddivisione in base alla stessa probabilità di default [Figura 5.28].

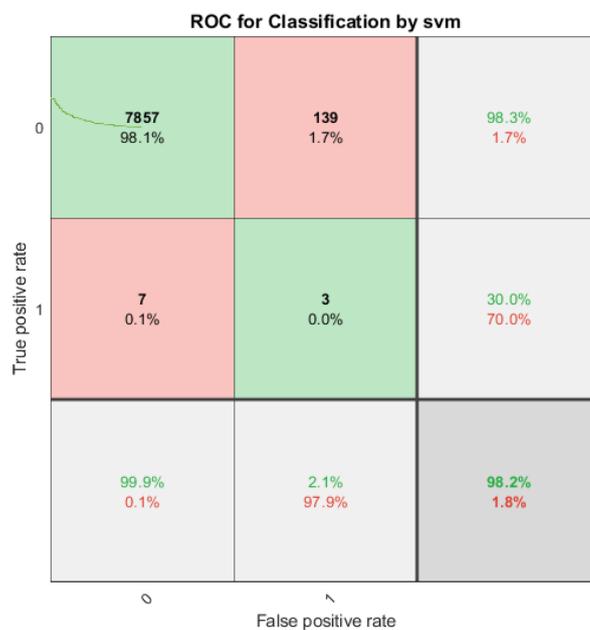


Figura 5.26 - Matrice di confusione per modello SVM con 6 predittori e il 50% di osservazioni anomale nel training set [elaborazione]

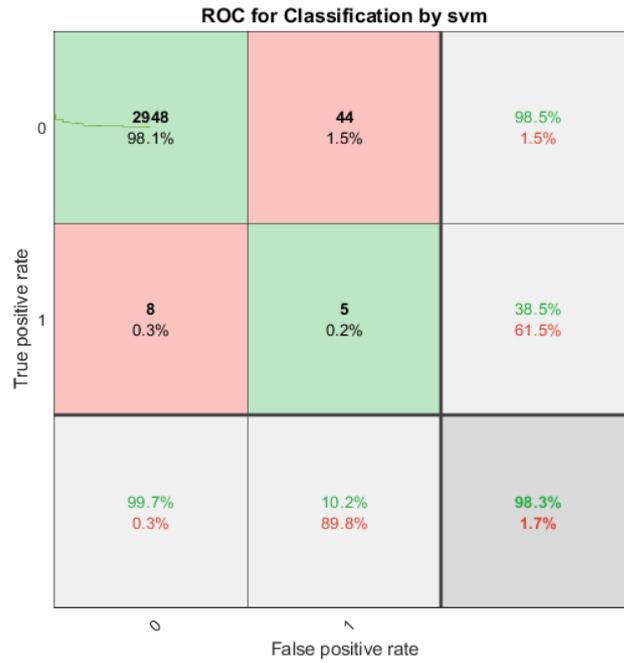


Figura 5.27 - Matrice di confusione per modello SVM con 6 predittori e l'80% di osservazioni anomale nel training set [elaborazione]

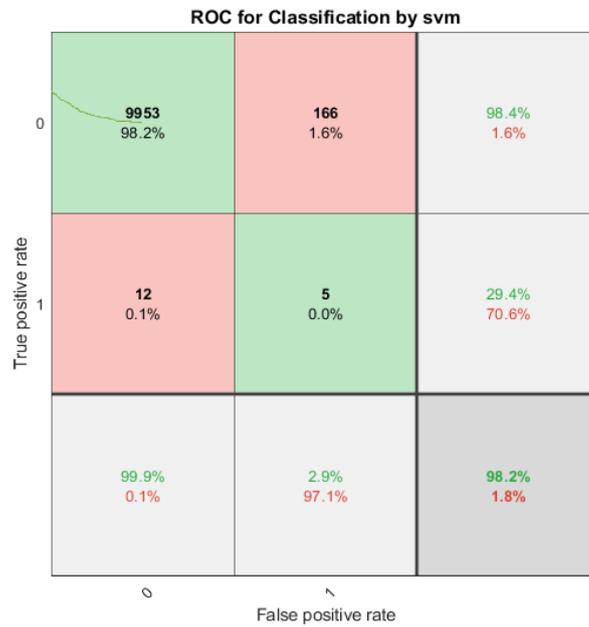


Figura 5.28 - Matrice di confusione per modello SVM con 6 predittori e set con la stessa PD [elaborazione]

Selezione del numero di variabili

L'ultimo step per la definizione del modello è la selezione del numero di variabili da includere. Il processo seguito ha previsto l'analisi di alcune ipotesi, partendo da 2 variabili e arrivando progressivamente a testarne 10. La composizione delle variabili è stata fatta variare seguendo lo stesso percorso impostato con le reti neurali.

5.6.2 Implementazione del modello

Viene di seguito presentato il codice utilizzato per la realizzazione dei modelli tramite SVM. Le parti di pulizia del *workspace*, importazione del dataset e salvataggio delle variabili sono identiche al modello delle reti neurali, pertanto si rimanda al paragrafo 5.4.3 per la loro implementazione dettagliata.

I dati dovevano essere divisi nei due set di riferimento, per cui si è deciso di assegnare inizialmente al *training set* le osservazioni da 1 a 14000 e nel *test set* quelle da 14001 a 17005. Tali valori sono stati modificati in base alle esigenze del tentativo da riprodurre.

```
% Suddivisione dei dati nei set
%Training set
X_train = X(1:14000,:);
Y_train = Y(1:14000,:);
%Test set
X_test=X(14001:17005,:);
Y_test=Y(14001:17005,:);
```

Successivamente, il modello è stato addestrato tramite il richiamo della funzione `fitcsvm`, che chiede come parametri obbligatori le variabili da includere e la risposta. Si sono anche impostate la funzione di Kernel e la standardizzazione dei predittori.

```
% Addestramento del modello
```

```

SVMModel =
fitsvm(X_train,Y_train,'KernelFunction','rbf','Standardize
',true);

```

Sono state poi calcolate le performance del modello per il *test set* ed è stata realizzata la matrice di confusione.

```

%% Calcolo performance
GroupTest = predict(SVMModel,X_test);
perfTest=sum(GroupTest==Y_test)/size(GroupTest,1);
%% Plot matrice di confusione
plotconfusion(Y_test',GroupTest');

```

In ottica performance, è stato necessario calcolare il grado di accuratezza del modello e il conseguente errore.

```

%% Calcolo accuracy ed errore
accuracy=sum(GroupTest==Y_test)/length(Y_test)*100;
errore=100-accuracy;

```

Infine, è stato realizzato il grafico della curva ROC, per consentire una visualizzazione grafica della metrica.

```

%% Grafico curva ROC
%Settaggio parametri
[~,score]= predict(SVMModel,X_test);
[Xsvm,Ysvm,Tsvm,AUCsvm] = perfcurve(Y_test,score(:,1),'0');
%Grafico
plot(Xsvm,Ysvm);
xlabel('False positive rate') ;
ylabel('True positive rate');
title('ROC for Classification by svm');

```

L'esecuzione del codice presentato ha permesso di costruire e valutare i diversi modelli relativi alle SVM.

5.6.3 Presentazione dei tentativi

Come per le reti neurali, anche in questo caso sono stati svolti ulteriori tentativi seguendone la stessa impostazione, già presentata al paragrafo 5.4.4.

5.6.3.1 Modello iniziale

Suddivisione dati

- *Training set* = righe da 1 a 14000 (82%), di cui 239 anomale (PD=1.71%);
- *Test set* = righe da 14001 a 17005 (18%), di cui 49 anomale (PD=1.63%).

Variabili incluse

- *(Riserve+Utile)/Attivo netto;*
- *Ln(Attivo netto);*
- *Oneri Finanziari/Ricavi;*
- *Liquidità/Attivo netto;*
- *(Rimanenze + Crediti vs clienti entro esercizio – Debiti vs fornitori entro esercizio)/Ricavi;*
- *Debiti finanziari (stimati)/Valore aggiunto;*
- *Oneri finanziari/Debiti Finanziari;*
- *Ammortamenti materiali/Costi operativi;*
- *Costo lavoro/Costi operativi;*
- *Debiti vs fornitori entro esercizio*360/Acquisti.*

Risultati

Nella tabella seguente sono sintetizzati i risultati.

# Variabili	Errore	Precisione	Sensitività	AUC
2	1,53%	55,60%	30,60%	0,711
3	1,56%	54,50%	24,50%	0,7276
4	1,46%	69,20%	18,40%	0,6447
5	1,70%	43,80%	14,30%	0,6395
6	1,73%	38,50%	10,20%	0,6697
8	1,76%	25%	4,10%	0,6731
10	1,76%	25%	4,10%	0,7238

Tabella 5.1 - Risultati modello iniziale SVM [elaborazione]

Dalla tabella si assiste a un generale trend peggiorativo dell'errore, fatta eccezione per il modello che include 4 variabili. L'andamento della precisione, invece, aumenta fino a 4 variabili, per poi scendere fino a 10. La sensitività ha un trend decrescente, mentre l'area sottesa alla curva ROC registra valori più variegati.

5.6.3.2 Modello con set con la stessa probabilità di default

Suddivisione dati

- *Training set* = righe da 1 a 6869 (62%), di cui 117 anomale (1.70%);
- *Test set* = righe da 6870 a 17005 (18%), di cui 171 anomale (1.69%).

Variabili incluse

- *Riserve+Utile/Attivo netto*;
- *Ln(Attivo netto)*;
- *Oneri Finanziari/Ricavi*;
- *Liquidità/Attivo netto*;
- *(Rimanenze + Crediti vs clienti entro esercizio – Debiti vs fornitori entro esercizio)/Ricavi*;
- *Debiti finanziari (stimati)/Valore aggiunto*;
- *Oneri finanziari/Debiti Finanziari*;
- *Ammortamenti materiali/Costi operativi*;
- *Costo lavoro/Costi operativi*;
- *Debiti vs fornitori entro esercizio*360/Acquisti*.

Risultati

Il modello con i primi 6 indicatori è risultato essere il migliore, ma le performance sono lontane dal miglior modello identificato con l'impostazione originale dei dati. Infatti, guardando alla matrice di confusione in figura 5.29, si osserva come i valori di precisione e sensitività siano di gran lunga inferiori rispetto a quelli trovati in precedenza, in aggiunta tali evidenze non consentono di definire come buono il classificatore.

Inoltre, è stata svolta una prova attribuendo ai dataset lo stesso numero di default ed il modello che include i primi 2 indicatori dell'elenco ha fatto registrare performance simili all'impostazione originale dei dati, ma senza miglioramenti.

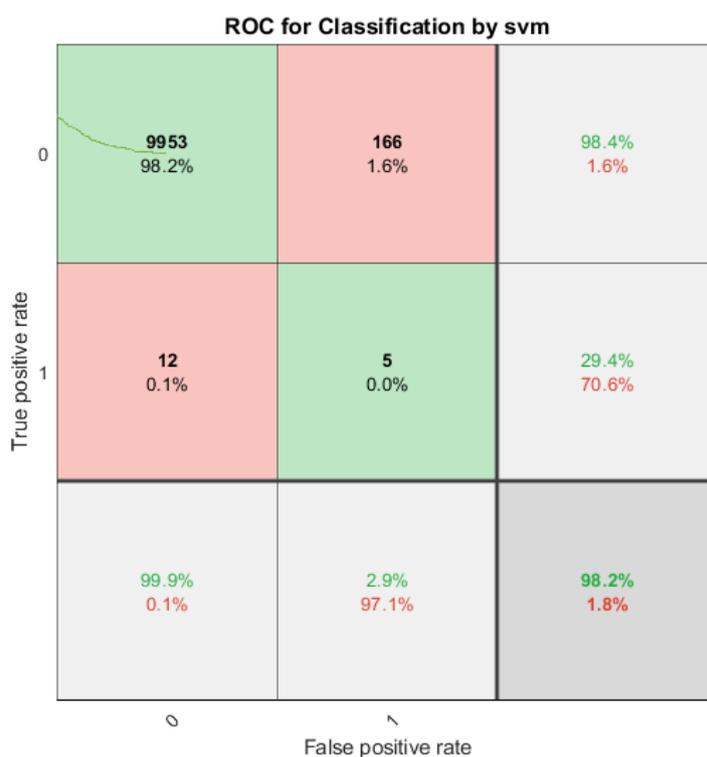


Figura 5.29 - Matrice di confusione per SVM con 6 indicatori [elaborazione]

5.6.3.3 Modello con altre variabili incluse

Suddivisione dati

- *Training set* = righe da 1 a 6869 (62%), di cui 117 anomale (1.70%);
- *Test set* = righe da 6870 a 17005 (18%), di cui 171 anomale (1.69%).

Variabili incluse

- *EBIT/Ricavi*;
- *Debiti vs fornitori entro esercizio*360/Acquisti*;
- *Costo lavoro/Costi operativi*;
- *Ln(Attivo netto)*;
- *(Rimanenze + Crediti vs clienti entro esercizio – Debiti vs fornitori entro esercizio)/Ricavi*;
- *Debiti finanziari (stimati)/Valore aggiunto*;
- *Var% Attivo netto*;
- *Risultato ante oneri finanziari/(Patrimonio netto + Debiti finanziari)*;
- *Var% Valore aggiunto operativo*;
- *Ammortamenti materiali/Costi operativi*.

Risultati

Anche in questo caso non è opportuno parlare di miglior modello, in quanto l'inclusione di nuove variabili ha comportato a una perdita di capacità predittiva, rendendo il classificatore di fatto inutilizzabile. Infatti, nella maggior parte delle prove il modello non è stato in grado di predire la classe delle anomale.

Sono state eseguite ulteriori prove utilizzando i 10 indicatori del modello iniziale e mutandone l'ordine di inclusione, secondo gli schemi illustrati precedentemente. I risultati non hanno portato a miglioramenti della capacità predittiva del modello, ma sono stati tendenzialmente migliori rispetto a quelli ottenuti con gli indicatori sopraelencati.

5.6.3.4 Miglior modello

L'ultimo tentativo ha previsto la sostituzione degli indicatori il cui calcolo è suscettibile di distorsione per i bilanci scritti non in forma ordinaria. Rispetto al modello iniziale, al pari delle reti neurali, sono stati sostituiti *(Rimanenze + Crediti vs clienti entro esercizio – Debiti vs fornitori entro esercizio)/Ricavi*,

Oneri finanziari/Debiti Finanziari e Oneri finanziari/Debiti Finanziari con gli indicatori Valore aggiunto/Immobilizzazioni materiali, Debiti totali/Valore aggiunto e Oneri finanziari netti/Attivo netto.

Si è operato assegnando ai diversi set la stessa probabilità di default e la divisione in base al modello iniziale. Come per le reti neurali, anche in questo caso la seconda configurazione ha portato a risultati migliori.

Suddivisione dati

- *Training set* = righe da 1 a 14000 (82%), di cui 239 anomale (PD=1.71%);
- *Test set* = righe da 14001 a 17005 (18%), di cui 49 anomale (PD=1.63%).

Variabili incluse

- *(Riserve+Utile)/Attivo netto;*
- *Ln(Attivo netto);*
- *Oneri Finanziari/Ricavi;*
- *Liquidità/Attivo netto;*
- *Costo lavoro/Costi operativi;*
- *Debiti finanziari (stimati)/Valore aggiunto;*
- *Valore aggiunto/Immobilizzazioni materiali*
- *Debiti totali/Valore aggiunto;*
- *Ammortamenti materiali/Costi operativi;*
- *Oneri finanziari netti/Attivo netto.*

Essendo tali indicatori gli stessi del modello delle reti neurali, è nuovamente possibile riconoscere l'appartenenza delle metriche alle categorie degli indicatori di liquidità e di produttività.

5.7 Prestazioni del miglior modello con SVM

Dopo aver definito la scrittura del codice per l'implementazione dell'algoritmo e la costruzione dei tentativi si è affrontata l'analisi dei risultati trovati. Di tutti i

modelli, è stato individuato un miglior modello sulla base delle prestazioni che ha fatto registrare e sul fatto che gli indicatori selezionati non introducessero distorsione davanti a un'abilità di classificazione piuttosto simile.

Il miglior modello individuato al termine dei tentativi effettuati è stato costruito con 6 indicatori:

- $(Riserve+Utile)/Attivo\ netto$;
- $Ln(Attivo\ netto)$;
- $Oneri\ Finanziari/Ricavi$;
- $Liquidità/Attivo\ netto$;
- $Costo\ lavoro/Costi\ operativi$;
- $Debiti\ finanziari\ (stimati)/Valore\ aggiunto$.

Matrice di confusione

Il modello costruito tramite SVM ha fatto registrare un errore dell'1.4%, con un'accuratezza del 98.6% [Figura 5.30]. Il grado di sensibilità è pari al 14% circa, con una precisione di oltre l'87%. La maggior parte delle osservazioni appartiene alla categoria delle sane: la matrice di confusione mette in luce questa caratteristica del set utilizzato per testare la classificazione.

Il confronto con il modello iniziale mostra un miglioramento dell'accuratezza e conseguentemente l'errore di classificazione risulta minore; sono altresì evidenti i progressi riguardo alla precisione. Per quanto concerne la sensibilità si è registrato un leggero calo, ma la selezione del modello più performante non ha considerato una metrica soltanto, ma un insieme di indici di prestazione, per questo motivo il classificatore qui presentato è stato indicato come il migliore.

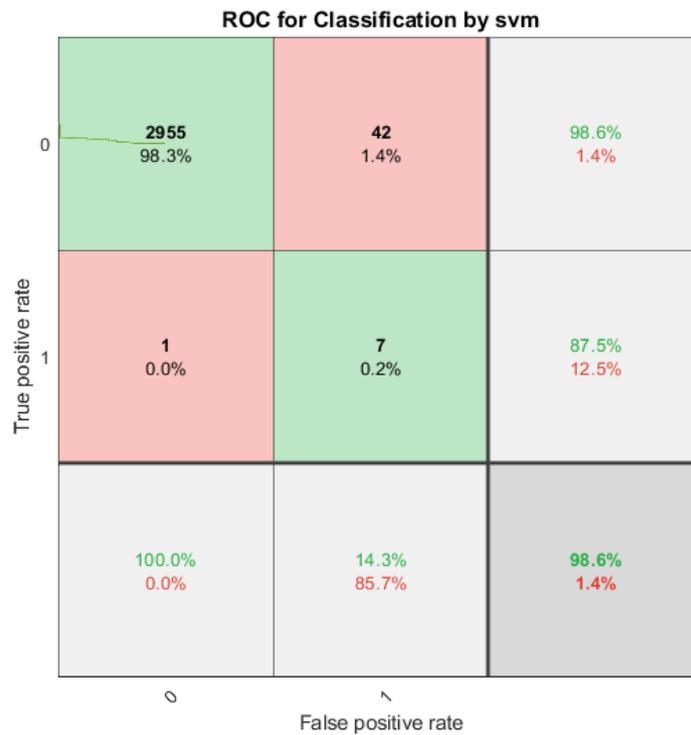


Figura 5.30 - Matrice di confusione del miglior modello SVM con 6 indicatori [elaborazione]

Curva ROC

L'andamento della curva illustra la capacità interpretativa del classificatore e risulta chiaro come le sue prestazioni siano migliori rispetto al modello casuale, pur rimanendo inferiori al modello perfetto.

Tale configurazione è tipica dei modelli di classificazione, in quanto ci si aspetta proprio un comportamento intermedio tra la casualità e la perfezione: essendo il classificatore costruito a partire da una selezione di informazioni, è lecito aspettarsi una capacità di previsione che sia superiore al 50%, ma inferiore al 100%. La curva in figura 5.31 mostra queste caratteristiche, pertanto non si evincono anomalie da questo grafico.

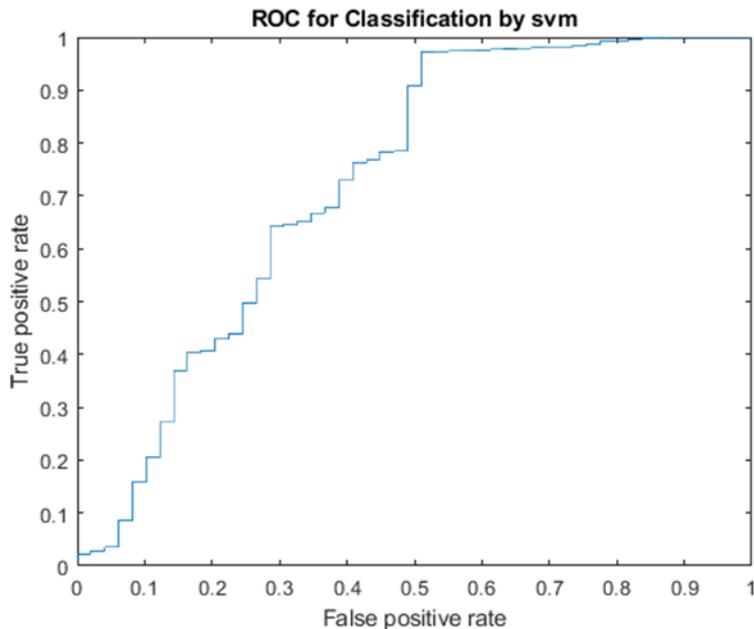


Figura 5.31 - Curva ROC modello SVM con 6 indicatori [elaborazione]

5.8 Confronto dei modelli

Dopo aver costruito e selezionato i migliori modelli mediante i due metodi è opportuno confrontarne le prestazioni, così come le variazioni mentre si effettuavano i diversi tentativi. In questo modo è possibile evidenziare analogie e discrepanze tra reti neurali ed SVM. Si ricorda che sono stati utilizzati gli stessi dataset per entrambi i metodi, così da rendere il confronto possibile e agevole.

5.8.1 Prestazione dei migliori modelli

Riprendendo quanto indicato al paragrafo 5.4.4.4, per quanto riguarda le reti neurali il miglior modello identificato è composto da 6 neuroni in input; il classificatore più performante per le SVM indicato nel paragrafo 5.6.3.4 è dotato di 6 indicatori come predittori del comportamento.

Le variabili dei due modelli, considerando le premesse sopracitate, sono naturalmente le stesse:

- $(Riserve+Utile)/Attivo\ netto;$
- $Ln(Attivo\ netto);$
- $Oneri\ Finanziari/Ricavi;$
- $Liquidità/Attivo\ netto;$
- $Costo\ lavoro/Costi\ operativi;$
- $Debiti\ finanziari\ (stimati)/Valore\ aggiunto.$

Quindi, il primo punto in comune riguarda proprio la selezione degli indicatori da utilizzare: avendo testato architetture differenti e combinazioni di variabili diverse, è possibile affermare che i due metodi abbiano raggiunto il miglior risultato attraverso l'utilizzo degli stessi sei indicatori. Questo risultato non è scontato a priori poiché, come evidenziato nei rispettivi paragrafi, le reti neurali e le *support vector machines* si fondano su approcci differenti per la classificazione: aver trovato una corrispondenza tra le performance dei due metodi è indice della validità di entrambi.

Tuttavia, ciò che cambia è la capacità di classificazione seppur utilizzando gli indicatori appena elencati. Richiamando le matrici di confusione in figura 5.22 per le reti neurali e in figura 5.30 per le SVM, si osserva in entrambi i casi una netta prevalenza delle osservazioni cosiddette sane, con livelli di errore simili e pari all'1,5% per le reti neurali e all'1,4% per le SVM. Anche la sensibilità è paragonabile: in entrambi i casi i valori registrati sono pari al 14,3%. La differenza si nota guardando il grado di precisione, con le SVM che si attestano all'87,5% e le reti neurali al 63,6%: in base a questo dato, il metodo con i vettori di supporto è in grado di categorizzare meglio le osservazioni positive.

Questa differenza si percepisce guardando i valori assoluti di classificazione: a fronte di uno stesso numero di classificazioni vere positive (7) e false negative (42), ciò che cambia sono i valori dei falsi positivi (4 per le reti neurali e 1 per le SVM) e conseguentemente dei veri negativi (2952 per le reti neurali e 2955 per le SVM). Quindi, la differenza di performance, che pur sembra rilevante guardando alla precisione, è in realtà dovuta all'errata classificazione di soltanto 3 osservazioni.

Si può pertanto concludere che i due modelli hanno un comportamento predittivo simile e che le differenze di prestazione risultano relativamente

marginali, essendo influenzate dal fatto che la maggior parte delle osservazioni appartenga alla classe delle sane.

5.8.2 Andamento delle prestazioni in base all'evoluzione dei tentativi

La costruzione di diversi modelli ha permesso di edificare uno storico di dati, utile a comprendere l'evoluzione dei modelli e l'individuazione dei margini di miglioramento. Come si è già specificato, per rendere confrontabili i risultati ottenuti dalle due tecniche si sono utilizzati gli stessi dati di partenza, inoltre, il processo di ricerca del miglior modello ha seguito gli stessi passi procedurali.

Per entrambi i metodi si è deciso di cominciare da un modello iniziale, da cui si sono ricavati i primi dati analitici. Successivamente sono state introdotte alcune modifiche nella composizione dei set, assegnando a ognuno di essi le osservazioni tali per cui questi avessero la stessa probabilità di default. Si è poi provato a includere variabili appartenenti a famiglie differenti, per poi sostituire gli indicatori iniziali che potevano introdurre distorsione con altri più robusti.

Ciò che si vuole illustrare nelle prossime righe è l'andamento delle prestazioni al variare delle condizioni testate. In particolare, si è osservato come i modelli facessero registrare un macro-andamento pressoché parallelo verso il miglioramento o il peggioramento a seconda del mutamento dei parametri.

Quando sono state redistribuite le osservazioni nei set affinché questi avessero la stessa probabilità di default, il modello delle reti neurali ha fatto registrare buone performance con 4 indicatori, mentre le SVM con 6. A fronte di un non miglioramento complessivo rispetto ai modelli iniziali, il modello con SVM si è dimostrato meno resiliente a questo cambiamento, con prestazioni che si sono allontanate di molto rispetto alla precedente versione della classificazione.

Il tentativo di inclusione di variabili differenti e appartenenti a famiglie diverse non ha avuto buon esito in nessuno dei casi considerati. Entrambi i classificatori hanno perso capacità predittiva divenendo essenzialmente inaffidabili nelle previsioni. Questa evidenza mostra il ruolo del processo di inclusione delle

variabili: esso rappresenta uno step fondamentale ed estremamente delicato nella costruzione di modelli di predittivi. Infatti, realizzare una classificazione con variabili “non adatte” rischia di causare danni che potrebbero compromettere la bontà del modello. Fortunatamente, gli strumenti di valutazione delle performance consentono di confermare l’adeguatezza della categorizzazione, rilevando quando il problema non è ben approssimato dalle variabili scelte.

Nelle figure 5.32 e 5.33 si osservano le matrici di confusione derivanti da dataset con stessa probabilità di default e con indicatori differenti rispetto a quelli del modello iniziale: con nessuno dei due metodi tali classificatori permettono di ottenere buone performance.

In questo caso, si è partiti da un modello migliore e lo si è peggiorato, ma il procedimento sarebbe potuto avvenire nel senso opposto: partendo da quest’ultimo come modello iniziale, avremmo constatato l’essenziale incapacità di previsione. Tale indicazione sarebbe stata un monito affinché si includessero variabili differenti in ottica *feature selection*.

Test Confusion Matrix

Output Class	0	1	
	6598 98.4%	107 1.6%	98.4% 1.6%
1	1 0.0%	0 0.0%	0.0% 100%
	100.0% 0.0%	0.0% 100%	98.4% 1.6%
	0	1	
	Target Class		

Figura 5.32 - Modello reti neurali con la stessa architettura del miglior modello, ma con altri indicatori nei 6 neuroni in input [elaborazione]



Figura 5.33 - Modello SVM con 6 variabili, ma con altri indicatori rispetto al miglior modello [elaborazione]

Infine, gli ultimi modelli sono stati costruiti sostituendo gli indicatori specifici per i bilanci in forma ordinaria. In questo caso il grado di parallelismo tra le prestazioni è nuovamente marcato. Infatti, la suddivisione dei set in base alla stessa probabilità di insolvenza ha portato a risultati peggiori rispetto a uno sbilanciamento di osservazioni anomale nel set di addestramento. Proprio quest'ultimo tentativo è risultato essere il migliore, con performance simili a quelle del modello iniziale e soprattutto con indicatori non passibili di distorsione.

Anche in quest'ultimo tentativo, così come per l'impostazione iniziale, le variabili incluse sono in numero uguale. Si può quindi concludere che, secondo quanto registrato empiricamente, i due metodi portino a conclusioni equivalenti e che la selezione delle variabili da includere è estremamente delicata e può variare di molto la capacità interpretativa di un modello.

Conclusioni

La presente trattazione ha previsto l'applicazione di due tecniche differenti di intelligenza artificiale per la realizzazione di un modello di *credit scoring* nell'ambito del settore degli elettrodomestici. La costruzione del classificatore con le reti neurali e le SVM in parallelo ha consentito di differenziare la metodologia di lavoro, fornendo dati utili per confrontare i due metodi.

Il primo punto da sottolineare è come l'utilizzo delle due tecniche abbia condotto a risultati simili: nonostante il differente principio di funzionamento, gli sviluppi hanno mostrato risposte analoghe alle variazioni che sono state di volta in volta introdotte.

Una peculiarità da evidenziare è l'importanza della selezione delle variabili da includere nel modello. Infatti, qualora queste ultime non fossero adatte a predire il comportamento della risposta, il classificatore risulterebbe inutilizzabile indipendentemente dalla metodologia utilizzata. Tuttavia, riguardo alla questione delicata della selezione degli indicatori, le reti neurali hanno fatto registrare una maggior robustezza rispetto alle SVM.

Un elemento comune è la difficoltà nel prevedere le osservazioni anomale. In questo caso si tratta di una questione meramente numerica, in quanto le società anomale sono di gran lunga inferiori rispetto alle sane. Infatti, la composizione del database fornito agli algoritmi ha causato una difficoltà di interpretazione degli indicatori, divenendo per certi versi un impedimento per la corretta identificazione dei *pattern* che contraddistinguono una società anomala.

Tuttavia, la selezione dei dati è tratta dai bilanci ufficiali, pertanto rappresenta un caso reale, oltretutto verosimile: un settore le cui imprese operanti sono caratterizzate da un'alta probabilità di default costituisce perlopiù un'eccezione poiché altrimenti sarebbe indice di problemi strutturali o di shock esogeni. Inoltre, è bene ricordare che l'orizzonte temporale considerato è di dieci anni, per cui, a rigor di logica, è legittimo aspettarsi che le società anomale siano numericamente inferiori rispetto alle sane. Per ovviare a questa problematica è

possibile utilizzare tecniche di ricampionamento, in modo da garantire al modello una più facile interpretazione delle informazioni.

Con la consapevolezza dei limiti della classificazione presentata, è possibile individuare il suo campo di applicazione, ossia il processo decisionale che porta alla concessione del credito. Si parla di processo poiché affidare la decisione unicamente alla previsione del presente modello potrebbe risultare una mossa incauta: il classificatore, come evidenziato, non è esente dal commettere errori, può però indicare un percorso da seguire per approfondire l'analisi.

Nell'ottica di un miglioramento continuo, il classificatore potrebbe essere arricchito con nuove informazioni. Si pensi, a titolo di esempio, allo scenario eccezionale di pandemia del 2020: l'inclusione, con gli opportuni accorgimenti, dei dati derivanti da una rara situazione di stress economico globale potrebbe impreziosire il modello.

L'aspetto fondamentale delle tecniche di *machine learning* è il dato, coniugato nella sua disponibilità e nella capacità di elaborarlo. L'innovazione tecnologica consente sempre più di estrarre valore dalle informazioni, permettendo di discernere la particolare casistica e di individuare la risposta più adatta, sia per la concessione del credito che per, ad esempio, le scelte di investimento. La sensazione è che tali tecnologie acquisiranno sempre più importanza nel tempo fino a condurre verso una *finanza sartoriale*.

Bibliografia e sitografia

[1] Varetto F., *Dispense del corso di Economia degli intermediari finanziari del Politecnico di Torino*, A.A. 2019/2020.

[2] Varetto F., *Dispense del corso di Analisi finanziaria e creditizia per l'impresa del Politecnico di Torino*, A.A. 2019/2020.

[3] Resti A., Sironi A., *Risk management and shareholders' value in banking*, Wiley Finance, 2007.

[4] Redazione, *Moody's: il rischio normativo la maggiore preoccupazione delle compagnie europee*, Assinews, 29 maggio 2018, <https://www.assinews.it/05/2018/moodys-rischio-normativo-la-maggiore-preoccupazione-delle-compagnie-europee/660052620/>.

[5] Sito ufficiale del Comitato di Basilea, *The Basel Committee – overview*, <https://www.bis.org/bcbs/index.htm>.

[6] Basel Committee on Banking Supervision, *International Convergence of Capital Measurement and Capital Standards*, Bank for International Settlements, Novembre 2005, <https://www.bis.org/publ/bcbs118.pdf>.

[7] Comitato di Basilea per la vigilanza bancaria, *Convergenza internazionale della misurazione del capitale e dei coefficienti patrimoniali*, Banca dei Regolamenti Internazionali, giugno 2006, p. 23, <https://www.bis.org/publ/bcbs128ita.pdf>.

[8] Banca d'Italia, *Relazione annuale - anno 2019*, 29 maggio 2020, p. 169, https://www.bancaditalia.it/pubblicazioni/relazione-annuale/2019/rel_2019.pdf.

[9] Sperduti A., *Dispense del corso di Elementi di Apprendimento Automatico dell'Università di Padova*, <https://www.math.unipd.it/~sperduti/AI10/app.pdf>.

[10] Redazione, *Reti neurali*, Intelligenza artificiale – Il portale dedicato all'intelligenza artificiale, <https://www.intelligenzaartificiale.it/reti-neurali/>.

- [11] Sciarrone F., *Dispense del corso di Sistemi Informatici dell'Università degli Studi Mediterranea di Reggio Calabria*, dicembre 2005, http://www.unirc.it/documentazione/materiale_didattico/599_2008_91_1578.pdf.
- [12] Maltoni D., *Dispense del corso di Reti neurali dell'Università di Bologna*, http://bias.csr.unibo.it/maltoni/ml/DispensePDF/8_ML_RetiNeurali.pdf.
- [13] *Regola di Hebb*, in Enciclopedia della Scienza e della Tecnica, Treccani, 2008, https://www.treccani.it/enciclopedia/regola-di-hebb_%28Enciclopedia-della-Scienza-e-della-Tecnica%29/.
- [14] Singh S., Srivastava A., Mi L., Chen K., Wang Y., Caselli R., Goradia D., Reiman E., *Deep Learning based Classification of FDG-PET Data for Alzheimers Disease Categories*, Proceedings of SPIE--the International Society for Optical Engineering, 2017, p. 7, https://www.researchgate.net/publication/321141639_Deep_Learning_based_Classification_of_FDG-PET_Data_for_Alzheimers_Disease_Categories.
- [15] Hassan H., Negm A., Zahran M., Saavedra O., *Assessment of artificial neural network for bathymetry estimation using high resolution satellite imagery in shallow lakes: case study El Burullus Lake*, International Water Technology Journal, 2015, p. 252, https://www.researchgate.net/publication/303875065_ASSESSMENT_OF_ARTIFICIAL_NEURAL_NETWORK_FOR_BATHYMETRY_ESTIMATION_USING_HIGH_RESOLUTION_SATELLITE_IMAGERY_IN_SHALLOW_LAKES_CASE_STUDY_EL_BURULLUS_LAKE.
- [16] *Dispense del corso di Apprendimento mimetico del Politecnico di Torino*, https://areeweb.polito.it/didattica/gcia/Apprendimento_Mimetico/Lucidi/Lucidi_Reti_Neurali1.pdf.
- [17] McCulloch W.S., Pitts W., *A logical calculus of the ideas immanent in nervous activity*, Bulletin of Mathematical Biology, 1990, pp. 99-115, <https://www.cs.cmu.edu/~.epxing/Class/10715/reading/McCulloch.and.Pitts.pdf>.

- [18] Ruini F., *La motivazione come determinante del comportamento di organismi artificiali: una simulazione di Artificial Life*, 2007, <https://www.researchgate.net/publication/229091518> La motivazione come determinante del comportamento di organismi artificiali una simulazione di Artificial Life/.
- [19] Medici P., *Dispense del corso di Reti neurali dell'Università degli Studi di Parma*, 2017, <http://www.ce.unipr.it/people/medici/geometry/node108.html>.
- [20] *Funzione gradino di Heaviside*, in Wikipedia, https://it.wikipedia.org/wiki/Funzione_gradino_di_Heaviside#/media/File:Dirac_distribution_CDF.svg.
- [21] *Funzione sigmoidea*, in Wikipedia, https://it.wikipedia.org/wiki/Funzione_sigmoidea.
- [22] *Hyperbolic Tangent*, in Wikimedia, https://commons.wikimedia.org/wiki/File:Hyperbolic_Tangent.svg.
- [23] Redazione, *Machine learning*, Intelligenza artificiale – Il portale dedicato all'intelligenza artificiale, <https://www.intelligenzaartificiale.it/machine-learning/>.
- [24] Redazione, *Che cos'è l'unsupervised learning?*, Ionos Digital Guide, 14/09/2020, <https://www.ionos.it/digitalguide/online-marketing/marketing-sui-motori-di-ricerca/unsupervised-learning/>.
- [25] Marmo R., *Dispense del Laboratorio di Visione Artificiale dell'Università degli Studi di Pavia*, A.A. 2002/2003, <https://slideplayer.it/slide/534847/1/images/47/>.
- [26] Valentini G., *Dispense del Corso di Bioinformatica dell'Università degli Studi di Milano*, https://homes.di.unimi.it/valenti/SlideCorsi/Bioinformatica06/PercMLP_BP.pdf.
- [27] Grippo L., Sciandrone M., *Metodi di ottimizzazione per le reti neurali*, <http://www.dis.uniroma1.it/~lucidi/didattica/Reti-Neurali-1.pdf>.

- [28] Bicego M., *Dispense del corso di Riconoscimento e recupero dell'informazione per bioinformatica dell'Università di Verona*, <https://www.di.univr.it/documenti/OccorrenzaIns/matdid/matdid833568.pdf>.
- [29] Brownlee J., *A Gentle Introduction to Cross-Entropy for Machine Learning*, Machine Learning Mastery, 21 ottobre 2019, <https://machinelearningmastery.com/cross-entropy-for-machine-learning/>.
- [30] D'Annunzio N., Falavigna G., *Modelli di analisi e previsione del rischio di insolvenza. Una prospettiva delle metodologie applicate*, Ceris-Cnr, W.P. N° 17/2004, 2004, p. 12, <https://core.ac.uk/download/pdf/6524031.pdf>.
- [31] Wang L., *Support Vector Machines: Theory and Applications*, Singapore, Springer, 2005, pp. 2, 3, <https://books.google.it/books?id=uTzMPJjVjsMC&pg=PA219&dq=svm&hl=it&sa=X&ved=2ahUKEwiY4lrXrYLuAhUPCuwKHZP1ArEQ6AEwAnoECAAQAg#v=onepage&q=svm&f=true>.
- [32] Govoni L., *Algoritmo Support Vector Machine*, <https://lorenzogovoni.com/support-vector-machine/>.
- [33] Ivanciuc O., *Applications of Support Vector Machines in Chemistry*, in *Reviews in Computational Chemistry*, 2007, 23, pp. 291-400, [http://www.ivanciuc.org/Files/Reprint/Ivanciuc SVM CCR 2007 23 291.pdf](http://www.ivanciuc.org/Files/Reprint/Ivanciuc_SVM_CCR_2007_23_291.pdf).
- [34] Berwick R., *Dispense del corso di Intelligenza artificiale del MIT*, <http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf>.
- [35] Giannetti A., *La prima lavatrice italiana fu progettata a Napoli*, *Storie di Napoli*, 06 marzo 2018, <https://storienapoli.it/2018/03/06/lavatrice-italiana-progettata-napoli/>.
- [36] Redazione, *Tutti gli elettrodomestici innovativi che ci hanno cambiato la vita*, *SkyTg24*, 23 ottobre 2018, <https://tg24.sky.it/tecnologia/approfondimenti/elettrodomestici-innovativi>.

[37] Stime della Associazione Nazionale delle Industrie elettromeccaniche, Milano, 1960.

[38] Guidi P., *Elettrodomestici Made in Italy: l'export vola, lavatrici in panne*, First online, 27 febbraio 2020, <https://www.firstonline.info/elettrodomestici-made-in-italy-lexport-vola-lavatrici-in-panne/>.

[39] Rassegna stampa, *Le innovazioni tecnologiche e i mercati emergenti guidano la crescita del Grande Elettrodomestico*, Gfk, Norimberga, 30 agosto 2018, <https://www.gfk.com/it/stampa/le-innovazioni-tecnologiche-e-i-mercati-emergenti-guidano-la-crescita-del-grande-elettrodomestico>.

[40] Rassegna stampa, *Piccolo Elettrodomestico: in crescita i dispositivi smart, quelli per il benessere e quelli che semplificano la vita*, Gfk, Norimberga, 30 agosto 2018, <https://www.gfk.com/it/stampa/piccolo-elettrodomestico-in-crescita-i-dispositivi-smart-quelli-per-il-benessere-e-quelli-che-sempli>.

[41] Applia, *BY THE NUMBERS: The Home Appliance Industry in Europe, 2018-2019*,

<https://www.applia-europe.eu/statistical-report-2018-2019/files/applia-statistical-report-2019.pdf>.

[42] *Spese per consumi*, Istat.

[43] *Household Appliances*, Statista,

<https://www.statista.com/outlook/16000000/141/household-appliances/italy?currency=eur>.

[44] GS1 Italy, *Lo shock del Covid-19*, estratto da OSSERVATORIO NON FOOD 2020, 2020,

https://osservatori.gs1it.org/media/cms/2020/09/15/estrattononfood_covid_settembre20.pdf.

[45] Bureau Van Dijk, *AIDA Analisi Informatizzata Delle Aziende Italiane*, https://www.univaq.it/include/utilities/blob.php?table=banca_dati_biblioteca&id=233&item=info_file.

- [46] Scarpa B., *Dispense del corso di Statistica per l'analisi organizzativa dell'Università di Padova*, <http://homes.stat.unipd.it/bruno/bocconi/multreg.pdf>.
- [47] Rossi G., *Dispense del corso di Psicometria dell'Università di Padova*, p. 4-4, <https://www.germanorossi.it/mi/file/disp/correlaz.pdf>.
- [48] Matlab, *Help Center*, <https://it.mathworks.com/help/>.
- [49] Brownlee J., *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/>.
- [50] Goodfellow I., Bengio Y., Courville A., *Deep Learning*, The MIT Press; 18 novembre 2019, p. 111.
- [51] Govoni L., *Matrice di confusione: cos'è e come funziona?*, <https://lorenzogovoni.com/matrice-di-confusione/>.
- [52] Google developers, *Classification: ROC Curve and AUC*, Machine Learning Crash Course, <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>.
- [53] Pykes K., *The Vanishing/Exploding Gradient Problem in Deep Neural Networks*, towardsdatascience.com, 17 maggio 2020, <https://towardsdatascience.com/the-vanishing-exploding-gradient-problem-in-deep-neural-networks-191358470c11>.
- [54] Wong W., *What is Gradient Clipping?*, towardsdatascience.com, 03 marzo 2020, <https://towardsdatascience.com/what-is-gradient-clipping-b8e815cdfb48>.