



Dipartimento di Scienze Economiche, Matematiche e Statistiche

Università degli Studi di Foggia

Reti Neurali Artificiali: Teoria ed Applicazioni

Crescenzo Gallo

Quaderno n. 28/2007

“Esemplare fuori commercio per il deposito legale agli effetti della legge 15 aprile 2004 n. 106”

Quaderno riprodotto al
Dipartimento di Scienze Economiche, Matematiche e Statistiche
nel mese di dicembre 2007 e
depositato ai sensi di legge

Authors only are responsible for the content of this preprint.

Dipartimento di Scienze Economiche, Matematiche e Statistiche, Largo Papa Giovanni Paolo II, 1,
71100 Foggia (Italy), Phone +39 0881-75.37.30, Fax +39 0881-77.56.16

Reti Neurali Artificiali: Teoria ed Applicazioni Finanziarie

Crescenzo Gallo

c.gallo@unifg.it

Dipartimento di Scienze Economiche, Matematiche e Statistiche
Centro di Ricerca Interdipartimentale Bioagromed

Università di Foggia

Largo Papa Giovanni Paolo II, 1 - 71100 Foggia (Italy)

Sommario

Le Reti Neurali Artificiali sono diventate uno strumento estremamente efficace nell'analisi di situazioni non “predicibili” analiticamente, e si prestano in maniera estremamente efficace alla modellizzazione di problemi di varia natura. In questo lavoro viene fatta una rassegna sulle origini e lo stato dell'arte delle Reti Neurali Artificiali, con una esposizione sintetica delle possibili applicazioni oggi più promettenti in ambito finanziario.

Keywords: reti neurali artificiali; modellizzazione.

JEL Classification: C45.

1 Introduzione: le Reti Neurali Artificiali

I computer seriali e i relativi programmi tradizionali sono degli strumenti molto potenti per svolgere dei compiti che richiedono la ripetizione di una serie di operazioni ben definite ove l'accuratezza, l'affidabilità e la velocità sono le caratteristiche importanti.

Questi sistemi di elaborazione dell'informazione sono molto utili ma non certo intelligenti: l'unico elemento di intelligenza nell'intero processo è il programmatore che ha analizzato il compito e ha creato il programma. Affinché un sistema artificiale possa definirsi intelligente, esso dovrebbe per lo meno essere in grado di risolvere i problemi che gli esseri umani trovano semplici, banali e naturali. Per questo motivo, sta suscitando notevole interesse una classe di modelli non lineari noti come *Reti Neurali Artificiali* (RNA).

1.1 Che cos'è una rete neurale artificiale?

Le reti neurali artificiali sono dei sistemi di elaborazione dell'informazione che cercano di simulare all'interno di un sistema informatico il funzionamento dei sistemi nervosi biologici che sono costituiti da un gran numero di cellule nervose o neuroni (figura 1) collegati tra di loro in una complessa rete. Ogni neurone è collegato mediamente con una decina di migliaia di altri neuroni. Si hanno quindi centinaia di miliardi di connessioni. Il comportamento intelligente emerge dalle numerose interazioni tra le unità interconnesse.

Alcune di queste unità ricevono informazioni dall'ambiente, altre emettono risposte nell'ambiente e altre ancora - se ve ne sono - comunicano solamente con le unità all'interno della rete (figura 2): esse sono definite rispettivamente unità di ingresso (*input*), unità di uscita (*output*) e unità nascoste (*hidden*). Ciascuna unità svolge un'operazione molto semplice che consiste nel diventare attiva se la quantità totale di segnale che riceve supera una certa soglia di attivazione. Se un'unità diventa attiva, essa emette un segnale che viene trasmesso lungo i canali di comunicazione fino alle altre unità cui essa è connessa; ciascun punto di connessione agisce come un filtro che trasforma il messaggio ricevuto in un segnale inibitorio o eccitatorio aumentandone o diminuendone nel contempo l'intensità a seconda delle proprie caratteristiche individuali.

Il legame input-output, ovvero la funzione di trasferimento della rete, non viene programmato ma è semplicemente ottenuto da un processo di apprendimento basato su dati empirici che può essere:

- **supervisionato** (*supervised learning*), qualora si disponga di un insieme di dati per l'addestramento (o *training set*) comprendente esempi

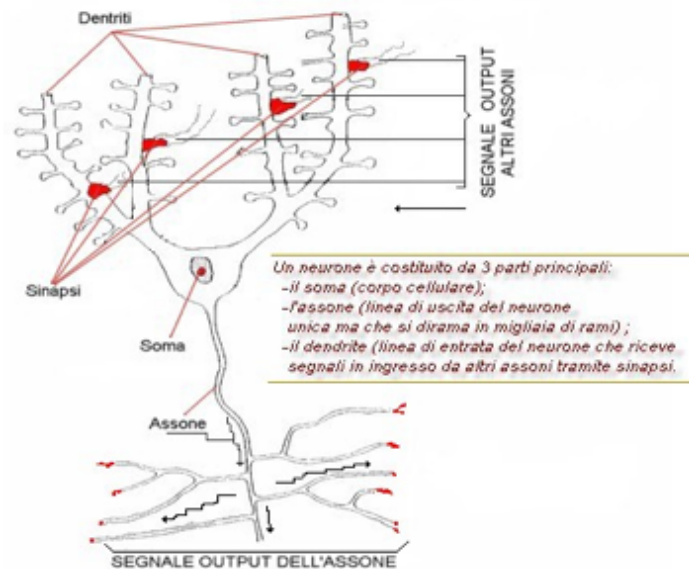


Figura 1: *Il neurone biologico.*

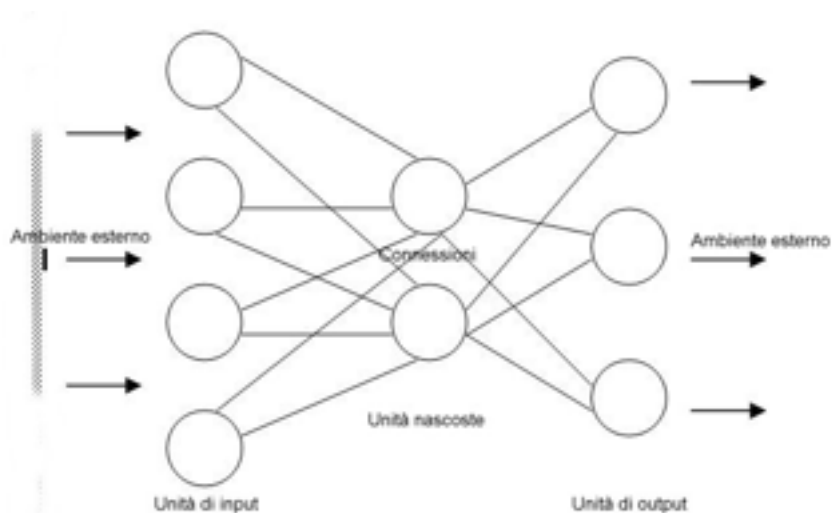


Figura 2: *Schema di una rete neurale artificiale.*

tipici di ingressi con le relative uscite loro corrispondenti: in tal modo la rete può imparare ad inferire la relazione che li lega. Successivamente, la rete è addestrata mediante un opportuno algoritmo (tipicamente il *backpropagation* che è appunto un algoritmo d'apprendimento supervisionato), il quale usa tali dati allo scopo di modificare i pesi ed altri parametri della rete stessa in modo tale da minimizzare l'errore di previsione relativo all'insieme di addestramento. Se l'addestramento ha successo, la rete impara a riconoscere la relazione incognita che lega le variabili d'ingresso a quelle d'uscita, ed è quindi in grado di fare previsioni anche laddove l'uscita non è nota a priori; in altri termini, l'obiettivo finale dell'apprendimento supervisionato è la previsione del valore dell'uscita per ogni valore valido dell'ingresso, basandosi soltanto su un numero limitato di esempi di corrispondenza (vale a dire, coppie di valori *input-output*). Per fare ciò, la rete deve essere infine dotata di un'adeguata capacità di generalizzazione, con riferimento a casi ad essa ignoti. Ciò consente di risolvere problemi di regressione o classificazione.

- **non supervisionato** (*unsupervised learning*), basato su algoritmi d'addestramento che modificano i pesi della rete facendo esclusivamente riferimento ad un insieme di dati che include le sole variabili d'ingresso. Tali algoritmi tentano di raggruppare i dati d'ingresso e di individuare pertanto degli opportuni cluster rappresentativi dei dati stessi, facendo uso tipicamente di metodi topologici o probabilistici. L'apprendimento non supervisionato è anche impiegato per sviluppare tecniche di compressione dei dati.
- **per rinforzo** (*reinforcement learning*), nel quale un opportuno algoritmo si prefigge lo scopo di individuare un certo *modus operandi*, a partire da un processo di osservazione dell'ambiente esterno; ogni azione ha un impatto sull'ambiente, e l'ambiente produce una retroazione che guida l'algoritmo stesso nel processo di apprendimento. Tale classe di problemi postula un agente, dotato di capacità di percezione, che esplora un ambiente nel quale intraprende una serie di azioni. L'ambiente stesso fornisce in risposta un incentivo o un disincentivo, secondo i casi. Gli algoritmi per il reinforcement learning tentano in definitiva di determinare una politica tesa a massimizzare gli incentivi cumulati ricevuti dall'agente nel corso della sua esplorazione del problema. L'apprendimento con rinforzo differisce da quello supervisionato poiché non sono mai presentate delle coppie input-output di esempi noti, né si procede alla correzione esplicita di azioni subottimali. Inoltre,

l'algoritmo è focalizzato sulla prestazione in linea, la quale implica un bilanciamento tra esplorazione di situazioni ignote e sfruttamento della conoscenza corrente.

Al di là del carattere schematico di questo modello, è importante sottolineare:

Pregi

- Le reti neurali per come sono costruite lavorano in parallelo e sono quindi in grado di trattare molti dati mentre nei calcolatori tradizionali ciascun dato viene elaborato individualmente e in successione. Malgrado il fatto che ogni singolo neurone sia relativamente lento¹, il parallelismo spiega in parte la maggior velocità del cervello nell'eseguire compiti che richiedono l'elaborazione contemporanea di un elevato numero di dati, come ad esempio il riconoscimento visivo di oggetti. Si tratta in sostanza di un sofisticato sistema di tipo statistico dotato di una buona immunità al rumore; se alcune unità del sistema dovessero funzionare male, la rete nel suo complesso avrebbe delle riduzioni di prestazioni ma difficilmente andrebbe incontro ad un blocco del sistema. I software di ultima generazione dedicati alle reti neurali richiedono comunque buone conoscenze statistiche; il grado di apparente utilizzabilità immediata non deve trarre in inganno, pur permettendo all'utente di effettuare da subito previsioni o classificazioni, seppure con i limiti del caso. Da un punto di vista industriale, risultano efficaci quando si dispone di dati storici che possono essere trattati con gli algoritmi neurali. Ciò è di interesse per la produzione perché permette di estrarre dati e modelli senza effettuare ulteriori prove e sperimentazioni.
- L'elaborazione dei sistemi nervosi è distribuita su molti elementi, ovvero vi sono molti neuroni che si occupano della stessa operazione.
- Ogni dato nella memoria dei calcolatori è identificato da un indirizzo (in pratica un numero) che viene utilizzato dal processore centrale per recuperare le conoscenze necessarie allo svolgimento di un certo compito. Invece gli essere umani accedono alle proprie memorie in base al contenuto: siamo in grado di recuperare un ricordo semplice in base a qualche indizio parziale o a un attributo (un profumo, una voce, ...).

¹Un neurone è in grado di emettere qualche centinaio di impulsi per secondo, mentre l'unità di calcolo di un computer di media potenza elabora parecchi milioni di numeri interi per secondo.

- I sistemi nervosi, al contrario dei calcolatori, non devono essere programmati per svolgere un compito, bensì imparano autonomamente in base all’esperienza o con l’aiuto di un istruttore esterno.

Difetti

- I modelli prodotti dalle reti neurali, anche se molto efficienti, non sono spiegabili in linguaggio simbolico umano: i risultati vanno accettati “così come sono”, da cui anche la definizione inglese delle reti neurali come *black box*. Come per qualsiasi algoritmo di modellazione, anche le reti neurali sono efficienti solo se le variabili predittive sono scelte con cura. Non sono in grado di trattare in modo efficiente variabili di tipo categorico (per esempio, il nome della città) con molti valori diversi. Necessitano di una fase di addestramento del sistema che fissi i pesi dei singoli neuroni e questa fase può richiedere molto tempo, se il numero dei record e delle variabili analizzate è molto grande. Non esistono teoremi o modelli che permettano di definire la rete ottima, quindi la riuscita di una rete dipende molto dall’esperienza del creatore.

Utilizzi

- Le reti neurali vengono solitamente usate in contesti dove i dati possono essere parzialmente errati oppure dove non esistano modelli analitici in grado di affrontare il problema. Un loro tipico utilizzo è nei software di OCR, nei sistemi di riconoscimento facciale e più in generale nei sistemi che si occupano di trattare dati soggetti a errori o rumore. Esse sono anche uno degli strumenti maggiormente utilizzati nelle analisi di Data Mining. Le reti neurali vengono anche utilizzate come mezzo previsionale dell’analisi finanziaria o meteorologica. Negli ultimi anni è aumentata notevolmente la loro importanza anche nel campo della bioinformatica, nel quale vengono utilizzate per la ricerca di pattern funzionali e/o strutturali in proteine e acidi nucleici. Mostrando opportunamente una lunga serie di input (fase di training o apprendimento), la rete è in grado di fornire l’output più probabile.

1.2 Cenni storici sull’evoluzione delle reti neurali

Lo studio delle reti neurali risale ai primi tentativi di tradurre in modelli matematici i principi dell’elaborazione biologica. Le più antiche teorie del cervello e dei processi mentali sono state concepite dai filosofi greci **Platone** (427-347 A.C.) e **Aristotele** (384-322 A.C.). Queste teorie furono riprese molto più tardi da **Cartesio** (1586-1650) e nel XVIII secolo dai filosofi empiristi. Le

prime realizzazioni di macchine cibernetiche, categoria alla quale appartengono i sistemi neurali, appaiono negli anni quaranta col nascere di una scienza nuova, la cibernetica. La cibernetica è definita come “scienza che studia i processi intelligenti” e fu fondata da **Norbert Wiener** nel 1947. **Ross Ashby**, un altro padre della cibernetica, costruisce nel 1948 l’omeostato, uno dei primi sistemi con connessioni interne regolabili capace di variare la sua configurazione interna adattandola a stimoli esterni. Il neurofisiologo **W.S. McCulloch** e il matematico **W.A. Pitts** (1943) di Chicago sono stati i primi a formulare l’approccio cibernetico fondamentale alla struttura del cervello elaborando il primo modello di rete neurale. **John Von Neumann**, dopo aver formulato l’architettura base dei moderni calcolatori, comincia nel 1948 lo studio delle reti di automi cellulari precursori di nuovi modelli computazionali. Nel 1949 il neurofisiologo **Donald Hebb**, dagli studi sul processo di apprendimento dei neuroni, dedusse la prima regola di apprendimento applicata nelle reti neurali. Contemporaneamente, gli studi di **Lashley** sulla mente umana indicavano che l’organizzazione della conoscenza e la memoria si basavano su rappresentazioni distribuite. Nei primi anni sessanta si costruiscono le prime macchine in grado di presentare primitive forme di apprendimento spontaneo e guidato: sono il *Perceptron* di **Frank Rosenblatt** della Cornell University e l’*Adaline* (Adaptive linear element) di **Bernard Widrow** di Stanford. Il Perceptron è una rete neurale costituita da dispositivi logici in grado di risolvere semplici problemi di riconoscimento di forme. Esso rappresenta un prototipo delle strutture che vennero elaborate più avanti. Anche in Italia si sviluppano iniziative particolarmente importanti. **Eduardo Caianello**, dell’Università di Napoli, sviluppa la sua teoria sui processi e le macchine pensanti sulla base delle idee di **McCulloch**, **Pitts** e **Hebb**. A Genova viene realizzata da **Augusto Gamba** una macchina derivata dal Perceptron. Nel 1969 **Marvin Minsky** e **Seymour Papert**, del Massachusetts Institute of Technology, pubblicano un’analisi molto critica delle macchine del tipo Perceptron. Essi dimostrarono matematicamente le limitazioni delle reti neurali nel risolvere problemi quali la determinazione della parità di un numero binario, il calcolo di una funzione XOR di 2 bit o la classificazione delle immagini in base alla loro connettività. Questi problemi potevano essere risolti solo da reti neurali omniconnesse in cui ogni neurone è connesso con tutti gli altri neuroni della rete: in una simile rete il numero delle connessioni crescerebbe esponenzialmente all’aumentare del numero di neuroni, contrariamente a quanto avviene nei sistemi biologici nei quali le connessioni crescono linearmente. Minsky era uno dei sostenitori di un approccio rivale alle reti neurali, l’Intelligenza Artificiale (A.I.) classica basata su computer tradizionali. In seguito alle tesi di Minsky il campo delle reti neurali fu abbandonato dalla maggior parte degli studiosi, i quali si rivolsero al campo dell’A.I. apparentemente più promettente. Secondo chi scrive

comunque, questo cambiamento di interessi fu causato anche dal fatto che la tecnologia allora disponibile rendeva molto difficoltosa o addirittura impossibile la sperimentazione nel campo delle reti neurali, né vi erano computer abbastanza veloci per simulare reti neurali complesse. Negli anni sessanta e settanta la ricerca continuò con contributi teorici e con poche applicazioni. Alcuni ricercatori come **Shunichi Amari**, **Kunihiko Fukushima** e **Shephen Grossberg** tentarono di simulare il comportamento di neuroni cerebrali con reti di unità di calcolo operanti in modalità parallela. Inoltre formularono teorie matematiche ed architetture per individuare e classificare i tratti caratteristici delle forme da riconoscere e per costruire le prime memorie associative. In queste ultime vengono utilizzate informazioni parziali come chiavi per recuperare dati memorizzati. L'interesse sviluppatosi nei primi anni '80 per i modelli neurali è sicuramente dovuto a diversi fattori, che sono elencati di seguito:

- i progressi compiuti nella comprensione di alcuni fenomeni computazionali biologici, la disponibilità di potenti computer in grado di simulare i nuovi modelli neurali;
- lo sviluppo di tecnologie VLSI e ottiche che si prestano alla costruzione di circuiti di tipo neuronico;
- la determinazione dei limiti dell'A.I., i quali sono strettamente legati ai limiti dei computer seriali di Von Neumann.

John Hopfield del California Institute of Technology propone nel 1982 un modello computazionale basato su concetti energetici e pertanto applicabile in svariati campi. Questo nuovo modello permise di studiare il comportamento globale di reti molto più complesse dei Perceptron, non analizzabili con metodi classici; in particolare, era possibile studiare reti con neuroni nascosti. Con questo risultato termina la “preistoria” dello studio delle reti neurali e inizia la cronaca di un settore in rapida evoluzione (figura 3). Nel 1987 la CEE promuove il progetto **BRAIN** (*Basic Research in Adaptive Intelligence and Neurocomputing*); inoltre si tiene a San Diego la prima conferenza internazionale sulle reti neurali, con 2000 partecipanti provenienti sia dalle università sia dall'industria. Alla luce dei risultati ottenuti, Minsky e Papert rivedono le loro critiche e indicano nuove direzioni di sviluppo nell'area delle reti neurali. I principali settori cui la ricerca attuale è indirizzata sono:

- la visione e il riconoscimento di forme;
- la comprensione del linguaggio naturale;

- le memorie associative, la robotica, la sensoria;
- l'inferenza e la risoluzione di particolari problemi computazionali.

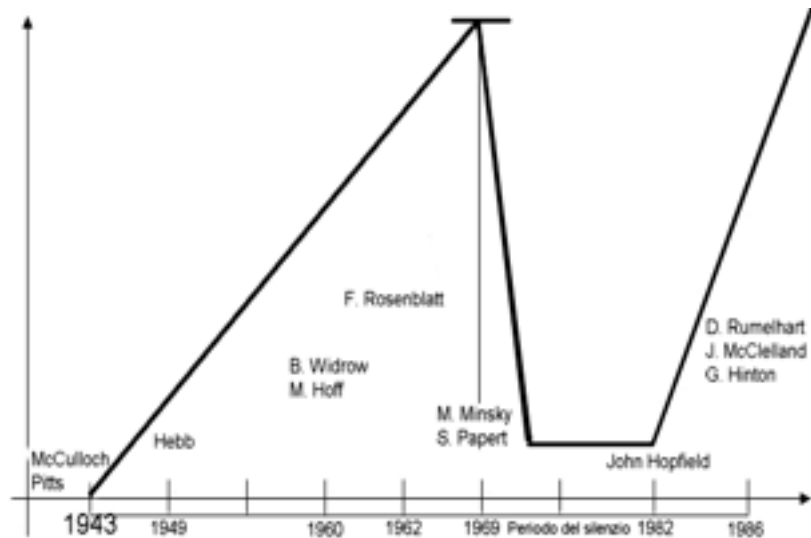


Figura 3: *Evoluzione storica delle reti neurali.*

1.3 Classificazione delle reti neurali

Esistono molti tipi di reti neurali che sono differenziati sulla base di alcune caratteristiche fondamentali:

- tipo di utilizzo;
- tipo di apprendimento;
- algoritmo di apprendimento;
- architettura dei collegamenti.

La divisione fondamentale è quella relativa al tipo di apprendimento che può essere supervisionato o non supervisionato. Dal punto di vista del tipo di utilizzo possiamo distinguere tre categorie basilari:

- **memorie associative:** possono apprendere associazioni tra patterns (insieme complesso di dati come l'insieme dei pixels di una immagine)

in modo che la presentazione di un pattern A dia come output il pattern B anche se il pattern A è impreciso o parziale (resistenza al rumore). Esiste anche la possibilità di utilizzare la memoria associativa per fornire in uscita il pattern completo in risposta ad un pattern parziale in input.

- **simulatori di funzioni matematiche:** sono in grado di comprendere la funzione che lega output con input in base a degli esempi forniti in fase di apprendimento. Dopo la fase di apprendimento, la rete è in grado di fornire un output in risposta ad un input anche diverso da quelli usati negli esempi di addestramento. Ne consegue una capacità della rete di interpolazione ed estrapolazione sui dati del training set. Tale capacità è facilmente verificabile addestrando una rete con una sequenza di dati input/output proveniente da una funzione nota e risulta, invece, utile proprio per il trattamento e la previsione di fenomeni di cui non sia chiaro matematicamente il legame tra input e output. In ogni caso la rete si comporta come una “black box”, poiché non svela in termini leggibili la funzione di trasferimento che è contenuta al suo interno. Di questo tipo fa parte la rete a retropropagazione dell'errore o *error back propagation* che è quella attualmente più utilizzata per efficacia e flessibilità.
- **classificatori:** con essi è possibile classificare dei dati in specifiche categorie in base a caratteristiche di similitudine. In questo ultimo tipo di rete esiste il concetto di apprendimento non supervisionato o “autoorganizzante”, nel quale i dati di input vengono distribuiti su categorie non predefinite. L'algoritmo di apprendimento di una rete neurale dipende essenzialmente dal tipo di utilizzo della stessa, così come l'architettura dei collegamenti. Le reti multistrato prevedono ad esempio l'algoritmo a retropropagazione dell'errore o sono addestrate tramite algoritmi genetici. I classificatori normalmente derivano dall'architettura delle mappe autorganizzanti di **Kohonen**. Esistono diverse regole di base per l'apprendimento ma sono sempre in fase di studio nuovi paradigmi: quello delle reti neurali è un campo in cui c'è ancora molto da inventare e da capire.

2 Costruzione di una rete neurale per le previsioni finanziarie

Le reti neurali, strumenti di previsione alternativi a quelli tradizionali, consentono di affrontare problemi finanziari altrimenti analizzati con metodologie di natura essenzialmente lineare o linearizzabile. In realtà, i fenomeni, non solo finanziari, solo raramente si manifestano in forma lineare e comunque mai mantengono nel tempo questa regolarità. A maggior ragione le decisioni umane non si affidano a un sistema lineare ma, anzi, risultano da un complesso sistema di connessioni che potenzialmente possono essere emulate dalle reti neurali. Esse “apprendono” come il fenomeno si evolve nel tempo grazie a una serie - talvolta anche prolungata - di tentativi che consentono di modellare in modo ottimale i pesi che collegano gli input con l’output passando per i neuroni degli strati nascosti. Per elaborare una previsione mediante reti neurali è possibile seguire una serie di operazioni che condizionano il risultato finale e l’efficacia complessiva della previsione. Le principali fasi necessarie per minimizzare anche i tempi di costruzione della rete sono:

- l’individuazione dell’obiettivo della previsione;
- la costruzione dell’archivio dei dati su cui attivare l’apprendimento della rete neurale;
- l’apprendimento vero e proprio, con la scelta dell’architettura e dei parametri necessari per la definizione dei pesi di connessione fra neuroni;
- la generalizzazione dell’output per la previsione dei mercati finanziari.

2.1 La definizione dell’ambito di indagine

La prima fase del procedimento di costruzione di una rete neurale applicata all’ambito finanziario è quella che richiede la definizione dell’obiettivo dell’analisi.

Le aree di applicazione finanziaria delle reti neurali sono molto diversificate. Secondo la classificazione proposta dalla *International Conference on Neural Networks in the Capital Markets*, si tratta della gestione dei portafogli, della stima dei modelli di curva dei rendimenti, della valutazione dei titoli obbligazionari e azionari, delle strategie di trading, di copertura e di arbitraggio e, infine, della previsione. In quest’ultimo caso, l’oggetto dell’analisi può riguardare i prezzi azionari, dei titoli obbligazionari o quelli dei contratti futures; i tassi di interesse o i tassi di cambio.

Una volta deciso l'ambito dell'indagine, è necessario stabilire i confini della misurazione del fenomeno. Si consideri la previsione del tasso di cambio: è possibile predirne il valore puntuale, così come può interessare il tasso di variazione su un periodo precedente o anche solo la tendenza di periodo per suggerire la posizione al cambista. La scelta dell'indicatore più utile per la previsione dipende da fattori specifici dell'operatore e da fattori esterni.

Nel primo caso, è possibile che l'intenzione sia quello di utilizzare il risultato della rete neurale per l'implementazione di un trading system; quello che conta è esclusivamente l'indicazione *buy*, *hold* oppure *sell*. Differente è il caso di un analista che voglia individuare le diverse alternative di mercato. In questo caso ciò che conta non è tanto sapere che un mercato è in rialzo, quanto piuttosto avere un'idea dell'entità di quel rialzo per poterlo confrontare con le dinamiche previste sugli altri mercati. Servirà un sistema revisionale in grado di segnalare la futura variazione del prezzo o del tasso.

Fra i fattori esterni all'analista che possono condizionare la scelta vi sono l'eventuale incompletezza o irregolarità di una serie storica, il rumore dovuto a elementi di inefficienza del mercato che possono condizionare negativamente alcuni prezzi.

Un ulteriore aspetto relativo alla definizione dell'ambito di indagine è quello che deve consentire di decidere la frequenza dei dati su cui ottenere l'output. Anche in questo caso i fattori che devono suggerire la scelta dipendono da motivazioni endogene ed esogene.

Le prime vanno individuate nell'orizzonte temporale su cui viene realizzata l'attività di negoziazione: se ai trader risulterà interessante una previsione a brevissima scadenza, per i gestori di patrimoni o di fondi comuni sarà sufficiente un'indicazione meno frequente. Si consideri che poter lavorare su dati a bassa piuttosto che ad alta frequenza consente di ridurre la componente di rumore che risulta frequente nella dinamica di brevissimo periodo. Quanto maggiore è l'incidenza del disturbo sui dati della serie tanto più consigliata è una ridotta periodicità della serie per lavorare su valori medi che annullano fluttuazioni irregolari intorno a un sentiero più facilmente riconoscibile dalla rete neurale.

Un ultimo problema è quello dell'orizzonte temporale della previsione: l'accuratezza delle stime tende in genere a diminuire con l'aumentare del periodo richiesto dall'analista alla rete. I risultati più frequenti mostrano come un'accettabile lunghezza della previsione sia compresa fra i 6 e i 10 anni. La specificità dell'apprendimento delle reti neurali può però far sì che l'accuratezza della stima sia migliore per periodi prolungati che non per periodi molto ravvicinati [1].

2.2 La costruzione dell'archivio dati

Una rilevante componente dello sforzo che richiede la costruzione della rete neurale deve essere necessariamente dedicata alla costruzione del set di informazioni che verranno utilizzate per riconoscere l'eventuale evoluzione della variabile di output. In relazione ai dati, le fasi rilevanti sono:

- a) la raccolta;
- b) l'analisi e la trasformazione;
- c) la selezione delle variabili di input e di output.

La raccolta delle informazioni deve rispondere ad alcuni principi fondamentali: in primo luogo devono essere recuperate da mercati con regolare frequenza di rilevazione, in grado da garantire la continuità della serie storica, almeno con la frequenza scelta per la previsione. Inoltre non è possibile fare apprendere una rete neurale con informazioni la cui rilevazione è avvenuta in periodi fortemente caratterizzati da trasformazioni strutturali di mercato. Si pensi a quanto successo sul mercato primario dei titoli di Stato italiani, con meccanismi di asta che sono mutati radicalmente nel tempo: prevedere oggi il prezzo d'asta utilizzando dati registrati nel periodo in cui esisteva il prezzo base o in cui la Banca d'Italia acquistava tutti i titoli non collocati sopra un certo valore ritenuto corretto, sarebbe altamente improbabile.

Infine, è fondamentale prevedere l'alimentazione periodica dell'archivio dei dati, per aggiornare i parametri della serie: preferibile è quindi individuare degli information provider a cui sia semplice accedere, possibilmente con link automatici.

Una fase particolarmente delicata della costruzione della rete neurale è quella della preparazione dei dati, sia di output sia di input. Alle problematiche tradizionali dell'analisi statistica [2] si aggiunge la necessità di lasciare nelle serie storiche tutti gli elementi necessari per l'apprendimento non lineare della rete ed eliminare quelli che possono condizionare negativamente i risultati.

Un primo problema da risolvere è quello della selezione dei dati. Questo aspetto è intuitivamente decisivo, in quanto l'apprendimento di qualsiasi sistema, biologico o artificiale, dipende dalle informazioni disponibili: nel caso in cui vengano a mancare alcuni di questi dati anche il comportamento non può che risulterne condizionato. Nel caso della previsione, ammessa la possibilità di riconoscere un *pattern* da parte della rete, questa dovrebbe essere teoricamente messa nelle medesime condizioni conoscitive di un analista di mercato.

I fattori che determinano le scelte sui mercati devono potere essere riconosciute dalla rete: rilevanti sono quindi sia i modelli teorici sia le informazioni che condizionano le aspettative economiche dei mercati. Naturalmente, non tutte le informazioni risulteranno, ex-post, rilevanti: si impone, pertanto, una continua calibrazione del database, con l'obiettivo di eliminare i dati che non modificano il contenuto informativo della rete. Sostanzialmente, la procedura dell'analista dovrebbe seguire queste fasi:

1. definizione ampia del primo database;
2. primo apprendimento della rete;
3. valutazione del contenuto informativo delle singole variabili;
4. analisi della matrice di correlazione fra le variabili di input;
5. eliminazione progressiva delle variabili meno significative e più auto-correlate fra loro;
6. successivo apprendimento della rete con il database ridotto.

Questo processo risulta tendere all'individuazione dell'archivio ottimale in funzione del problema analizzato, secondo uno schema iterativo che deve garantire la massima capacità di apprendimento della rete con il minimo sforzo informativo. In particolare, l'osservazione dei coefficienti di correlazione nonché dei relativi grafici consente di valutare la natura dell'eventuale interrelazione, all'interno dell'insieme delle variabili di input e fra queste e la variabile che si intende prevedere (figura 4).

Si nota come la diagonale principale della tabella sia caratterizzata da valori unitari: la correlazione di una variabile con se stessa è infatti massima e positiva. Il coefficiente di correlazione lineare fra due variabili x e y varia in un range $[1, +1]$ e misura la variazione media della variabile dipendente x quando la variabile indipendente y aumenta di un'unità. Le informazioni contenute nella tabella di figura 4 consentono di rimuovere dal set di variabili di input quelle più intensamente correlate, per evitare segnali potenzialmente ambigui o ritardati [3].

Quanto più è ampia la dispersione dei dati intorno alla retta di regressione, tanto maggiore risulta essere la componente di rumore dei dati che possono disturbare l'apprendimento della rete neurale.

La rete neurale per la previsione delle serie storiche finanziarie può essere impostata utilizzando differenti tipologie di informazioni:

- a. anzitutto quelle direttamente legate alla variabile di output (*market data*);

	Prime rate ABI	Riserve bancarie	Tasso interbank	Tasso medio depositi	Tasso p/t	Tasso ufficiale
Prime rate ABI	1,0000	-0,1347	0,9407	0,8341	0,8415	0,9623
Riserve bancarie	-0,1347	1,0000	-0,0386	-0,4875	-0,1231	-0,0012
Tasso interbank	0,9407	-0,0386	1,0000	0,7642	0,8669	0,8593
Tasso medio depositi	0,8341	-0,4875	0,7642	1,0000	0,6773	0,7487
Tasso p/t	0,8415	-0,1231	0,8669	0,6773	1,0000	0,7985
Tasso ufficiale	0,9623	-0,0012	0,8593	0,7487	0,7985	1,0000

Figura 4: *Matrice di correlazione fra tassi di interesse e quantità bancarie (1987-1997, dati mensili).*

- b. quelle che risultano operativamente collegate con la variabile di output (*intermarket data*);
- c. infine, quelle dipendenti dalle componenti fondamentali (*fundamental data*).

Quello descritto può essere considerato il modello generale: non sempre infatti il contributo delle variabili intermarket o quelle fondamentali risultano significative per spiegare il fenomeno: questo può dipendere dalla frequenza delle osservazioni, dall'esistenza di ostacoli al movimento dei capitali o, ancora, da attori di autocorrelazione che impediscono di aggiungere conoscenza alla rete neurale.

Si può però osservare come l'interdipendenza dei mercati sia crescente e come l'analisi tecnica sia sempre più orientata all'osservazione delle dinamiche dei differenti mercati [4].

Lo schema generale dei dati necessari per l'elaborazione delle reti neurali è quello della figura 5.

A loro volta, i dati di mercato possono essere elaborati per mezzo di indicatori di analisi tecnica che consentono di approfondire alcune dinamiche e segnali di inversione della tendenza e del ciclo. Per la rete neurale disporre di questi ulteriori informazioni è rilevante poiché l'obiettivo è quello di ri-



Gli indicatori di analisi tecnica più frequentemente utilizzati [3] possono essere calcolati anche per serie storiche di variabili finanziarie determinate su mercati differenti: se la previsione che si intende fare attiene a variabili molto liquide su cui agiscono numerose forze di mercato, è altamente probabile che la rete debba tenerne conto, così come un analista o trader decide una determinata posizione di mercato sulla base di un segnale emerso su altri mercati.

Un problema associato all’inserimento dei dati fondamentali è quello della frequenza di rilevazione: produzione industriale e inflazione hanno infatti una cadenza mensile e risultano pertanto inadatti per previsioni al di sotto di questo periodo. Due sono le soluzioni preferibili:

16

1. utilizzare non tanto la serie storica dei valori effettivi (i quali sono conosciuti solo ex-post) bensì quelli attesi dal mercato. La previsione è disponibile nei maggiori *information provider* e consente di approssimare la reazione che gli operatori imprimeranno ai prezzi una volta annunciato un valore eventualmente differente;
2. sostituire ai dati fondamentali una variabile *dummy* che misuri solo il momento dell'informazione: in tal modo occorre aggiungere una serie storica cui valori sono nulli nei momenti in cui nessuna informazione rilevante è comunicata al mercato, mentre sono unitari in corrispondenza dell'informazione che può avere un impatto significativo per la dinamica della serie da prevedere.

Questo problema attiene essenzialmente all'individuazione delle relazioni intertemporali fra le variabili. Si è già accennato alla possibilità di utilizzare le serie storiche con lag temporali: questa fase può essere realizzata attraverso alcune ipotesi già testate dall'analista oppure, in modo più scientifico, mediante il confronto dei fattori di contribuzione.

Si ipotizzi di volere inserire, fra le variabili di input, la serie del fenomeno che si intende prevedere, con il lag temporale più significativo; nel caso in cui valesse l'ipotesi di *random walk*, il lag da utilizzare sarebbe quello di un periodo, ma se la serie rispondesse a elementi di differente natura i lag accettabili potrebbero essere numerosi e diversi da 1.

Il metodo più semplice è quello di costruire nel foglio di lavoro che contiene variabili di input le serie con i differenti lag temporali; i software di rete neurale e i pacchetti statistici permettono di misurare i contributi delle diverse serie storiche, mediante regressioni sia univariate sia multivariate.

La prima colonna della tabella di figura 6 definisce sostanzialmente il grado di importanza (decrescente) attribuibile a ciascuna variabile esplicativa del fenomeno. La scelta può dunque cadere sui lag che contribuiscono maggiormente alla comprensione del fenomeno che si intende prevedere con la rete neurale. Esistono alcuni fenomeni finanziari che sono caratterizzati da *mean reversion*, cioè dalla tendenza a non mantenere determinati trend crescenti o decrescenti per un periodo di tempo troppo prolungato. È il caso tipico dei tassi di interesse e dei tassi di cambio, i quali tendono, nel medio termine, a mantenersi sufficientemente stabili intorno a un valore medio di equilibrio. Livelli molto elevati dei tassi rispetto alla media di periodo riflettono una tendenza successiva a ritornare verso la media (*mean reversion*, appunto)³.

³Con riferimento ai tassi di interesse si veda [6], mentre per i tassi di cambio si rimanda a [7].

Contribuzione alla rete	Correlazione con l'output	Indicatore di analisi tecnica ottimizzata
0,05837	0,8156	LinRegSlope(6) of DEM/USD
0,05778	0,8180	LWgtMvAvgDiff(5,6) of DEM/USD
0,05735	0,7128	EMAAofEMADiff(0.55,0.525,0.575) of DEM/USD
0,05653	0,8156	LinRegChange(6,2) of DEM/USD
0,05531	0,6847	ExpMvAvgDiff(0.3500001,0.5000001) of DEM/USD
0,04500	0,5778	%D(6,2) of DEM/USD
0,04499	0,7101	MvAvgDiff(2,7) of DEM/USD
0,03804	0,5594	BBand%B(9,0.5) of DEM/USD
0,03758	0,9956	AvgChange(6) of DEM/USD
0,03473	0,5492	WildcRSI(4) of DEM/USD
0,03208	0,8468	LagMvAvgDiff(5,2,4) of DEM/USD
0,02811	0,5165	NegChangeSum(6) of DEM/USD
0,02496	0,5861	CCI(9,0.5) of DEM/USD
0,02220	0,8224	LwgtLagMvAvgDiff(2,5,2) Of DEM/USD
0,01881	0,7528	MvAvgofDiff(4,2,3) of DEM/USD0,01875
0,01875	0,5187	Slow%D(4,2,2) of DEM/USD

Figura 6: *Fattori di contribuzione delle variabili esplicative e correlazione con l'output.*

La previsione di queste variabili può dunque risentire in modo distorsivo della dinamica storica, poiché il comportamento tipico della serie è appunto quello di invertire la tendenza. Si può in questi casi suggerire di prevedere non tanto il livello assoluto del dato, quanto la sua variazione rispetto a un precedente periodo. La scelta del lag della variazione dipende dalla possibilità che la serie sia affetta da rumore, che risulta tanto maggiore quanto più è esasperata la brevità della rilevazione.

Una soluzione spesso utilizzata passa per la normalizzazione della serie storica [8]: una prima tecnica si realizza sottraendo al dato effettivo quello corrispondente della media mobile (figura 7). In tal modo si elimina il trend e si riporta tutto su un piano che risulta maggiormente stabile e più prevedibile.

Periodo	Dato originale	Media mobile (3giorni)	Dato trasformato
1	25,3	----	----
2	22,6	23,20	-0,60
3	21,7	20,63	1,07
4	17,6	20,90	-3,30
5	23,4	20,77	2,63
6	31,3	24,40	-3,10
7	28,5	25,40	3,10
8	26,4	29,10	-2,70
9	32,4	32,10	0,30
10	37,5	----	----

Figura 7: Trasformazione per le serie storiche con mean reversion.

Un secondo metodo per la normalizzazione è quello di rendere la serie compresa fra un valore minimo e un valore massimo. Diverse possibilità si offrono, generalmente legate alla distanza dell'osservazione rispetto ai valori minimi e massimi. Se interessa far variare la serie normalizzata nell'intervallo $[0, 1]$ la formula da adottare è la seguente:

$$I_i = \frac{V_i - V_{min}}{V_{max} - V_{min}} \quad (1)$$

La formula generale che permette di far variare la serie nell'intervallo $[L_{min}, L_{max}]$ è la seguente [9]:

$$I_i = L_{min} + [L_{max} - L_{min}] \cdot \frac{V_i - V_{min}}{V_{max} - V_{min}} \quad (2)$$

Si ipotizzi di avere una serie storica di cui si conosca il valore massimo (V_{max}) e il valore minimo (V_{min}). Si consideri che è poi possibile fissare un limite oltre il quale il dato è considerato *outlier*: sia detto L_{max} il limite massimo e L_{min} il limite minimo. La tabella di figura 8 mostra come una serie storica possa essere normalizzata.

Un altro metodo utilizzato per la normalizzazione della serie è quello che si basa sulla media e sullo scarto quadratico medio della serie storica stessa. La formula della normalizzazione statistica è la seguente:

$$I_i = \frac{V_i - M(V_i)}{\sigma(V_i)} \quad (3)$$

V_i	V_{\min}	V_{\max}	L_{\min}	L_{\max}	I_i
456	132	765	200	500	353,6
342	132	765	200	500	299,5
365	132	765	200	500	310,4
765	132	765	200	500	500,0
354	132	765	200	500	305,2
132	132	765	200	500	200,0
367	132	765	200	500	311,4
398	132	765	200	500	326,1
421	132	765	200	500	337,0

Figura 8: *Normalizzazione con minimo e massimo prefissato.*

dove $M(V_i)$ è la media della serie storica analizzata e $\sigma(V_i)$ è il suo scarto quadratico medio. La serie storica dei dati I_i trasformati in questo modo presenta la caratteristica di avere media nulla e scarto quadratico medio unitario (tabella figura 9).

Queste trasformazioni sono ugualmente valide nel caso in cui si voglia analizzare un fenomeno caratterizzato da elevata stagionalità (si pensi, in particolare, al prezzo delle commodity e dei relativi contratti derivati). In tal caso la media mobile deve avere la frequenza della stagionalità. In alternativa è possibile assegnare ai singoli giorni della settimana (tabella figura 10), ai mesi o alle stagioni dei punteggi differenti, in modo da consentire alla rete di riconoscere l'effetto che si produce sulla variabile analizzata al trascorrere del tempo.

Quanto la normalizzazione condizioni l'efficacia dell'apprendimento della rete neurale è argomento analizzato in numerosi contributi [10]: in generale si ritiene che il beneficio sia rilevante nella fase di classificazione e in termini di errore medio quadratico, mentre incerto è l'effetto quando la dimensione del dataset aumenta; inoltre si registra un significativo rallentamento del processo di training.

V_i	$M(V_i)$	$\sigma(V_i)$	I_i	$M(I_i)$	$\sigma(I_i)$
456	400	164,48	0,34	0	1
342	400	164,48	-0,35	0	1
365	400	164,48	-0,21	0	1
765	400	164,48	2,22	0	1
354	400	164,48	-0,28	0	1
132	400	164,48	1-,63	0	1
367	400	164,48	-0,20	0	1
398	400	164,48	-0,01	0	1
421	400	164,48	0,13	0	1

Figura 9: *Standardizzazione statistica.*

Lunedì	(1,0,0,0,0)
Martedì	(1,1,0,0,0)
Mercoledì	(1,1,1,0,0)
Giovedì	(1,1,1,1,0)
Venerdì	(1,1,1,1,1)

Figura 10: *Assegnazione valori per il riconoscimento della stagionalità.*

Ulteriori particolarità delle serie storiche previste possono essere date da valori eccezionali (*outlier*): il trattamento di questi dati può avvenire in due modi. Se si vuole istruire la rete per consentire un riconoscimento degli shock analoghi futuri è necessario lasciare la serie storica inalterata, sebbene il rischio sia quello di forzare l'intero modello all'adattamento su questi valori. Se, al contrario, si desidera stimare i parametri di una rete neurale maggiormente adatta a riconoscere il pattern della serie storica nelle fasi "normali", si può trasformare la dinamica intorno alla media del fenomeno applicando ai dati una distribuzione normale. A questo proposito si rammenta la forma generale della distribuzione normale che consente di tagliare le osservazioni che caratterizzano le code (tabella figura 11).

$n\sigma$	Area della distribuzione
0,67	50%
1,00	68%
1,96	95%
2,58	99%

Figura 11: *Forma generale della distribuzione normale.*

Se si recupera l'esempio della tabella di figura 9, i valori minimi e massimi vengono trasformati in relazione all'ampiezza della distribuzione normale prescelta (tabella figura 12).

È ovvio che l'incidenza degli outlier deve essere minima, e il limite di variazione sarà pertanto condizionato dalla volatilità del fenomeno. Un sistema spesso applicato dagli analisti è quello che accetta valori minimi e massimi collegati ad un predefinito coefficiente di deviazione standard. L'ipotesi di normalità è spesso solo un'approssimazione della reale distribuzione del fenomeno: in particolare, se la numerosità della serie è ridotta, si suggerisce di valutare direttamente la numerosità e la significatività dei fenomeni che vengono scartati mediante lo scarto quadratico medio.

In generale, questa fase deve essere ispirata dalla finalità di ridurre il rumore della serie, mediante l'utilizzo di trasformazioni che non alterino la dinamica del fenomeno. Fra quelle più diffuse si citano: le trasformazioni logaritmiche; l'analisi di Fourier; le regressioni statistiche.

$n\sigma$	Minimo	Massimo
Dati effettivi	365,00	765,00
0,67	289,80	510,20
1,00	235,52	564,48
1,96	77,62	722,38
2,58	-24,35	824,35

Figura 12: *Forma generale della distribuzione normale.*

2.3 Le fasi di apprendimento

La costruzione della rete neurale passa necessariamente per alcune fasi che devono consentire di fissare i parametri utili per l'apprendimento idoneo alla soluzione del problema.

La prima fase dell'analisi è quella della scelta dell'architettura più adatta per l'apprendimento. La delicatezza di questa fase dipende dal fatto che il meccanismo di connessione degli input fra loro e fra questi e l'output, passando per gli strati nascosti, risulta un elemento decisivo per il successo dell'operazione.

Esistono alcuni algoritmi proposti in letteratura per ottimizzare la scelta dell'architettura (*pruning algorithm*, *polynomial time algorithm*, *canonical decomposition technique*, *network information criterion*), ma si ritiene che l'esperienza e la fase di test siano criteri più affidabili, soprattutto se non esistono riferimenti diretti al problema che si intende risolvere.

Le caratteristiche delle architetture possono inoltre costituire un ulteriore elemento di scelta [11]: ognuna di queste opzioni contiene dei parametri che devono essere definiti dall'analista, e le scelte possono comportare anche numerosi tentativi di affinamento per l'individuazione del risultato migliore.

In generale si può ritenere che i parametri da determinare nella definizione dell'architettura siano:

1. la suddivisione temporale del database;
2. il numero degli strati nascosti e dei neuroni da inserire in ciascuno strato;
3. i meccanismi connettivi fra i differenti strati;

4. la funzione di attivazione;
5. le regole di apprendimento e il momentum;
6. l'aggiornamento dei pesi di connessione dei neuroni.

2.4 Le fasi temporali dell'apprendimento e della previsione

Una volta definita la forma dell'output e il contenuto del dataset da cui estrarre le variabili di input, l'analista deve suddividere la serie storica in sotto-periodi, i quali delimitano l'ambito di apprendimento (*training set*) e di valutazione. Quest'ultimo viene in genere distinto in *test set* e *generalisation set*⁴.

In sostanza, la rete impara cercando di riconoscere la dinamica del training set, verifica come si adatta sul test set e poi si applica su un insieme di dati (*generalisation*) che non ha mai potuto osservare. Il periodo di generalizzazione può anche essere ignorato, nel caso in cui l'analista avesse a disposizione dei dati esterni al data set originario e solo successivamente volesse applicare i risultati della rete su quei dati: questa soluzione implica però un tempo più ampio per la verifica della bontà della rete.

I criteri di estrazione dei sottoperiodi sono assai vari:

- a. casuale, nel quale l'analista definisce solo la proporzione percentuale dei tre sottoperiodi;
- b. definito in termini quantitativi ma i dati di training test e generalisation risultano essere in posizione variabile;
- c. i dati di training e di test sono fissati ma la rete estrae in modo casuale: la serie di generalisation è invece tenuta alla fine della serie storica;
- d. la serie storica viene rigidamente divisa in un periodo di training test e uno di generalisation, secondo questo ordine cronologico. Quest'ultimo criterio è quello che viene suggerito nell'ambito delle reti applicate alla previsione dei mercati finanziari. Non esistono regole universalmente valide per la suddivisione della serie storica da analizzare: si va dal suggerimento di utilizzare $\frac{2}{3}$ del totale della serie per il training [12] fino all'indicazione di un numero di dati utilizzati per l'apprendimento non inferiore a $\frac{4}{5}$ del totale della serie: le soluzioni più adottate sono, rispettivamente (60%; 20%; 20%) e (60%; 30%; 10%) [13].

⁴In letteratura e in alcuni programmi applicativi, il test set è chiamato "validation set" mentre il generalisation set è chiamato "production set".

2.5 Il numero degli strati e dei neuroni

In relazione al numero di strati, si trovano numerosi contributi metodologici secondo i quali un singolo strato nascosto sarebbe sufficiente ad approssimare le funzioni non lineari più ricorrenti con elevato grado di accuratezza [14]. Per questo motivo sono numerosi i lavori empirici previsionali che stimano la rete con un solo strato nascosto; il limite di questo approccio è legato alla necessità di dover utilizzare un elevato numero di neuroni, che limita il processo di apprendimento. L'utilizzo di reti con due strati nascosti risulta più efficace per problemi di previsione su dati ad alta frequenza [15].

Con riferimento alla numerosità dei neuroni che devono essere assegnati a ciascuno strato nascosto, il criterio cui affidarsi nella scelta è quello della minimizzazione del rischio di *overlearning* che si corre quando si decide di inserire un numero eccessivo di neuroni, i quali disegnano quasi perfettamente il pattern della serie storica ma non sono in grado di generare una previsione affidabile perché riducono il contributo degli input. Viceversa, il rischio di assegnare un numero troppo basso di neuroni è quello di ridurre il potenziale di apprendimento della rete neurale. Occorre pertanto trovare una soluzione di compromesso tra un numero troppo elevato o troppo basso di neuroni. Le formule proposte in letteratura sono assai varie e, in alcuni casi, contraddittorie:

$$n_{hl} = 2 \cdot n_{input} + 1 \quad (4)$$

$$n_{hl} = 2 \cdot n_{input} \quad (5)$$

$$n_{hl} = n_{input} \quad (6)$$

$$n_{hl} = \frac{n_{input} + n_{output}}{2} + \sqrt{n_{training}} \quad (7)$$

dove:

n_{hl} è il numero dei neuroni nascosti;

n_{input} è il numero di input;

n_{output} è il numero di output;

$n_{training}$ è il numero di osservazioni contenute nel training set.

I risultati empirici dimostrano come nessuna di queste regole risulta generalizzabile ad ogni problema previsionale, sebbene un numero non irrilevante di risultati positivi sembri preferire la (6).

2.6 Le connessioni dell'architettura

In relazione al problema delle connessioni fra strati di input, nascosti e di output, si consideri il caso dell'architettura *Back Propagation*, ampiamente

utilizzata e diffusa anche a livello di software per la capacità che questo insieme di modelli ha di generalizzare i risultati per un ampio numero di problemi finanziari. Si tratta di un'architettura supervisionata, il cui apprendimento risulta condizionato sia da variabili di input che da output [16]. Esistono numerose possibilità di connessione: standard; a salto e ripetute.

Le *connessioni standard* prevedono connessioni fra input e output che passano attraverso uno o più strati nascosti, ai quali risultano essere stati assegnati dei neuroni che facilitano l'apprendimento della rete. Le connessioni, a prescindere dal numero degli strati nascosti che può essere elevato a piacere, sia pure a costo di un più elevato tempo di apprendimento, sono dirette e non prevedono né salti né ritorni su se stessi (figura 13).

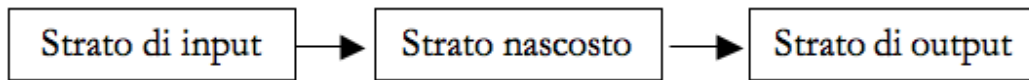


Figura 13: *Architettura back-propagation a uno strato nascosto con connessioni standard.*

Le *connessioni a salti* prevedono invece che la rete assegni dei pesi connettivi anche fra neuroni presenti in strati non adiacenti: nel caso più semplice lo strato di input presenta una serie di connessioni non solo con lo strato nascosto ma anche con quello di output (figura 14). Questa ramificazione connettiva si articola ulteriormente se il numero di strati nascosti aumenta.

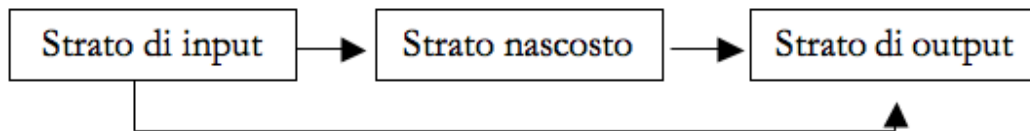


Figura 14: *Architettura back-propagation a uno strato nascosto con connessioni a salto.*

Infine, le reti neurali a *connessioni ripetute* risultano utili per l'analisi delle serie storiche finanziarie in quanto riconoscono le sequenze che si creano nell'ambito dei mercati. Queste connessioni prevedono la possibilità che i neuroni assegnati agli strati nascosti possano ritornare sulle variabili di input con processi iterativi così da quantificare in modo preciso il peso connettivo (figura 15).

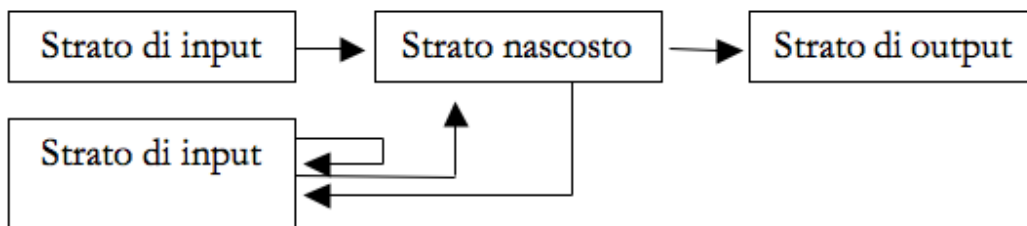


Figura 15: *Architettura back-propagation a uno strato nascosto con connessioni ripetute.*

2.7 Le funzioni di attivazione

Il quarto problema che si pone al momento della costruzione della rete neurale è quello della funzione di attivazione che condiziona il legame dei neuroni; le soluzioni funzionali sono riportate nella tabella di figura 16.

Non esiste una regola teoricamente accettabile per definire la funzione di attivazione dei vari strati. Sebbene molti studi presentino funzioni differenti per i singoli strati, alcuni adottano la medesima forma per input, hidden e output.

La funzione lineare viene in genere utilizzata per lo strato che contiene l'output della rete neurale; la ragione è che questa funzione, pur essendo più rigida delle alternative, evita che il risultato tenda verso il minimo o il massimo. Meno efficace è, invece, l'utilizzo della funzione lineare negli strati nascosti, soprattutto se questi sono caratterizzati da un elevato numero di neuroni, che risulterebbero così connessi proprio su una base funzionale che si vuole superare con l'utilizzo della rete stessa. Il limite rilevante della funzione lineare è quello di non consentire un fitting adeguato per serie storiche caratterizzate trend persistente.

La funzione logistica e quella logistica simmetrica presentano la caratteristica di variare nell'ambito di un range, rispettivamente $[0, 1]$ e $[-1, 1]$. La prima risulta essere particolarmente utile negli strati nascosti delle reti applicate alle serie storiche finanziarie. Alcuni problemi presentano caratteristiche dinamica che risultano essere colte in misura più precisa dalla funzione simmetrica, soprattutto nello strato di input e in quelli nascosti. La maggior parte della letteratura empirica presenta l'utilizzo di questa funzione nello strato nascosto, sebbene senza che vi sia una motivazione teorica robusta.

La funzione tangente iperbolica consente di adattare la rete in modo affidabile negli strati nascosti (in particolare nelle reti a tre strati), soprattutto nel caso in cui l'analista abbia scelto una funzione logistica o lineare per

Funzione	Formula
Lineare	$f(x) = x$
Logistica	$f(x) = \frac{1}{1 + e^{-x}}$
Logistica simmetrica	$f(x) = \left(\frac{2}{1 + e^{-x}} \right)^{-1}$
Tangente iperbolica	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
Tangente corretta (coeff. = c)	$f(x) = \tanh(c * x)$
Sinusoidale	$f(x) = \sin(x)$
Gaussiana	$f(x) = e^{-x^2}$
Gaussiana inversa	$f(x) = 1 - e^{-x^2}$

Figura 16: *Funzioni di attivazione delle reti neurali.*

l'output.

La funzione sinusoidale viene generalmente adottata nell'ambito della ricerca e si suggerisce di normalizzare input e output all'interno di un range $[-1, +1]$.

La funzione gaussiana si presta all'individuazione di particolari processi dinamici colti con architetture a due strati nascosti paralleli, con una funzione tangente nel secondo strato.

2.8 L'algoritmo di apprendimento e il *momentum*

Un ulteriore problema consiste nella scelta della regola di apprendimento: in particolare è necessario decidere con quale tasso di cambiamento la rete deve modificare la definizione dei pesi dei neuroni rispetto alla significatività dell'errore commesso.

Si consideri la figura 17: a sinistra (a) si può osservare un progresso sia di apprendimento sia di oscillazione, con una frequenza che raggiunge il risultato atteso più velocemente; a destra (b), invece, la rete apprende modificando il proprio sentiero con un tasso inferiore e raggiunge in un tempo superiore l'obiettivo.

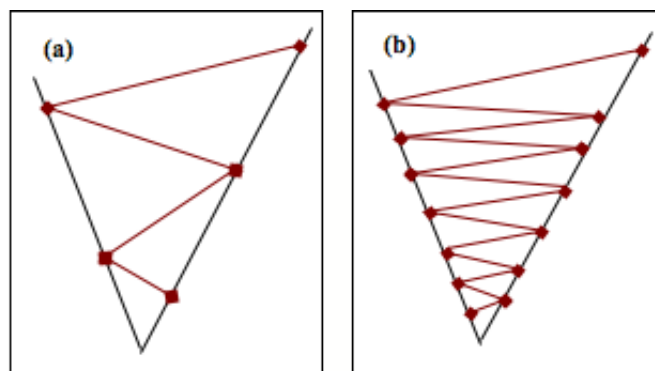


Figura 17: *Sviluppo della rete con tassi di apprendimento differenti.*

Il rischio nel decidere un tasso di apprendimento troppo elevato è quello di condizionare la rete a oscillare in modo troppo ampio e provocare così la perdita dell'evoluzione corretta della serie storica. In questo modo si potrebbe verificare un fenomeno di non convergenza che deve essere corretto riducendo adeguatamente il *learning rate*.

In alternativa, è possibile assegnare un valore corretto al *momentum*. Si tratta della proporzione con cui si aggiunge al nuovo peso la variazione dell'ultimo peso raggiunto dalla rete neurale. In questo modo la rete può imparare anche a tasso elevato ma non rischia di oscillare perché recupera un'elevata quota dell'ultimo peso raggiunto.

L'analista deve poi decidere il livello iniziale del peso da dare alla connessione fra neuroni e valutare se le osservazioni sono caratterizzate da un elevato tasso di rumore. In questo caso sarà opportuno mantenere elevato il valore del *momentum*.

Il processo di apprendimento della rete passa naturalmente dall'individuazione progressiva del valore più adatto di questi parametri. Si impone

pertanto una serie numerosa di tentativi che possono generare risultati sensibilmente differenti. Il suggerimento è quello di evitare radicali variazioni dei parametri.

2.9 Gli indicatori di errore della rete neurale

Una volta selezionate le caratteristiche iniziali della rete neurale è possibile definire i criteri di arresto dell'apprendimento. Questi criteri possono essere anzitutto collegati al training set oppure al test set: se l'obiettivo è quello di descrivere adeguatamente il fenomeno studiato, è preferibile bloccare l'apprendimento sul training set, mentre se si vuole costruire una rete neurale con finalità previsionali è preferibile bloccare l'apprendimento sul test set.

In secondo luogo, i parametri di apprendimento sono generalmente legati agli indicatori di errore commessi dalla rete: errore medio; errore massimo; numero di epoche senza miglioramento dell'errore.

Fissando un valore a questi parametri la rete si bloccherà una volta raggiunto il valore desiderato. In genere è più semplice fissare un elevato numero di epoche (che sono condizionate dalla numerosità delle osservazioni del training set) che la rete analizza senza migliorare l'errore. Si può ritenere che un valore compreso fra 10.000 e 60.000 epoche sia sufficientemente sicuro per bloccare una rete la quale ormai con grande difficoltà può apprendere meglio di quanto abbia già fatto fino a quel momento. Naturalmente la scelta dipende anche dalla velocità con cui la rete raggiunge questi valori.

Un elemento per l'accettazione di una rete è quello della convergenza. Se anche gli errori risultano modesti ma l'oscillazione ha portato a un'elevata divergenza, è opportuno verificare l'adeguatezza dei parametri.

La figura 18 *Error(a)* mostra un processo di oscillazione sufficientemente contenuto senza una rilevante divergenza; meno positiva è invece l'evoluzione dell'apprendimento della rete mostrata dalla figura 18 *Error(b)*, con un'oscillazione che ha fatto perdere totalmente il sentiero del fenomeno studiato. La medesima considerazione deve essere fatta per quanto riguarda la distribuzione temporale degli errori commessi sul validation set.

Una volta verificata una corretta dinamica temporale dell'errore, almeno in termini grafici, è necessario misurarla quantitativamente. I programmi di apprendimento delle serie storiche tendono a utilizzare numerosi fra gli indicatori di errore sviluppati in ambito statistico [17], [18]. Fra questi spiccano:

- a. l'indice di determinazione (R^2);
- b. il Mean Absolute Error (MAE);

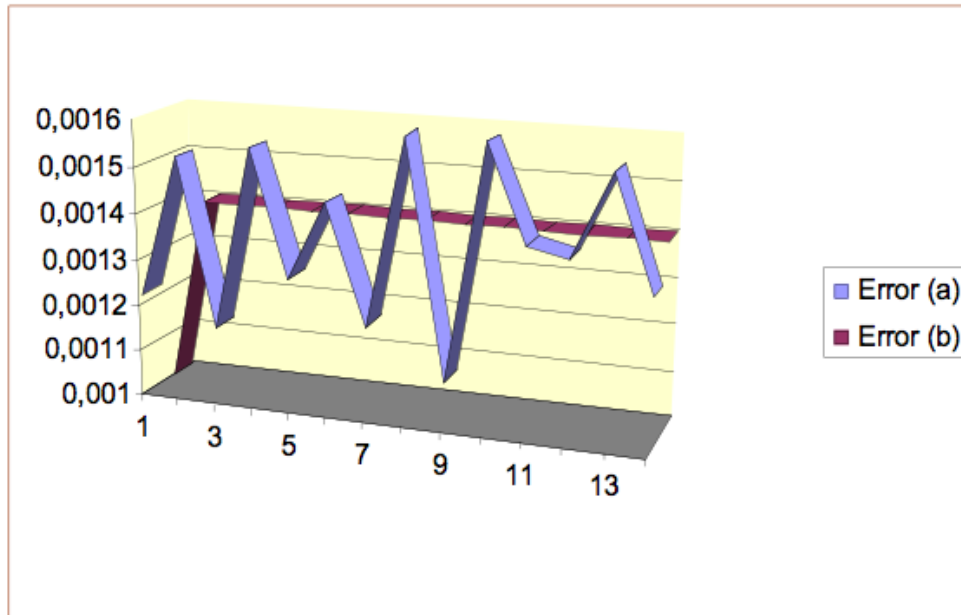


Figura 18: *Errori commessi dalla rete neurale sul training set.*

- c. il Mean Absolute Percentage Error (MAPE);
- d. il Mean Square Error (MSE).

Si tratta di indicatori che misurano, in vario modo, il differenziale fra l'output originario e quello stimato dalla rete; solo se gli input, i neuroni, le funzioni di attivazione e tutti i parametri precedentemente descritti fossero perfettamente in grado di individuare il fenomeno originale, gli scostamenti fra output reale e stimato sarebbero nulli, ottimizzando così i citati indicatori di errore.

L'esempio riportato nella tabella di figura 19 mostra come la rete sia stata in grado di approssimare il fenomeno (sul validation set) per l'82% circa e che si sia commesso un errore medio assoluto (MAE) pari a 0,446 e uno quadratico (MSE) di 0,383.

Un indicatore spesso utilizzato per ottimizzare la rete neurale è il coefficiente T di Theil, che presenta il pregio di variare entro un intervallo standardizzato $[0, 1]$, dove 0 si ottiene con modelli previsionali perfetti.

Il limite di questi indicatori di errore è quello di basarsi su un concetto di scostamento simmetrico rispetto al valore reale, mentre in finanza l'errore si misura solo in termini di perdita. Pertanto è opportuno stimare i pesi della rete sulla base dei profitti ottenuti [19]; in termini operativi, si posso-

Indicatori dell'errore	Valori
r squared	0,821
Mean squared error	0,383
Mean absolute error	0,446
Min. absolute error	0,000
Max. absolute error	3,343

Figura 19: *Indicatori di errore di una rete neurale.*

no adottare delle strategie di filtro per rimuovere *ex-post* il problema [20]. Anche in relazione alla validazione della rete occorre adattarsi alle finalità dell'analista.

2.10 La previsione della serie storica

Una volta che gli indicatori sono favorevoli alla scelta della rete neurale, è necessario verificare la sua bontà previsionale. È infatti possibile che un modello riesca a descrivere ottimamente il training e il test set ma poi risulti del tutto inadeguato per quanto riguarda la sua generalizzazione: cioè, nel caso finanziario, la previsione.

L'analista dovrà pertanto testare sul generalisation set la rete neurale con le medesime tecniche già descritte. In primo luogo, si dovranno misurare sulla serie storica mai osservata dalla rete gli indicatori di errore già descritti. Qualora questi dovessero risultare significativamente peggiori e, comunque, non accettabili dall'analista sulla base degli obiettivi originari, la rete dovrà essere ulteriormente testata.

La previsione finanziaria viene poi spesso condizionata dalla capacità del modello di individuare per tempo le inversioni cicliche del fenomeno.

Inoltre, è possibile utilizzare l'output originale e quello stimato dalla rete per misurare quantitativamente il ritardo con cui la rete apprende l'inver-

sione del ciclo. Numerosi sono gli strumenti utili per analizzare la capacità previsionale della rete neurale: a ogni inversione del ciclo la serie dei valori del fenomeno deve registrare una variazione del segno della variazione. Se, dunque, si passa da un trend crescente a uno decrescente, i segni delle variazioni passeranno da “+” a “-”.

Un sistema efficace per verificare con quale capacità la rete segnala questa inversione, è quello che si basa sul calcolo dell’intervallo di confidenza dell’output della rete. La misurazione dell’intervallo di confidenza può essere costruito seguendo questo schema:

1. calcolo dello scarto quadratico medio della serie di output della rete neurale;
2. definizione del limite di confidenza dell’analista: se questo é pari al 95%, in ipotesi di distribuzione normale, occorrerà calcolare 1,96 volte lo scarto quadratico medio;
3. misurazione dei valori estremi dell’intervallo: il valore massimo è dato dall’output cui si aggiunge $n\sigma$ base all’intervallo scelto al punto repunto2, mentre il valore minimo è dato dall’output cui si sottrae $n\sigma$;
4. si verifica in quali casi il segno del valore minimo e di quello massimo coincidono: fissato l’intervallo di confidenza, si può affermare che la probabilità di prevedere correttamente il trend della serie coincide con quella percentuale;
5. si controlla quante volte si commette realmente un errore di segnalazione, valutando in tal modo l’affidabilità dello strumento previsionale;
6. l’analista deciderà quale posizione di mercato prendere solo nei casi in cui valore minimo e massimo dell’intervallo hanno il medesimo segno: questo segno indicherà naturalmente la posizione coerente da prendere sul mercato [21];
7. nel caso in cui due o più osservazioni consecutive presentassero lo stesso segno, l’operatore avrebbe un’indicazione di mantenimento della posizione esistente.

3 Applicazioni delle Reti Neurali in ambito finanziario

Gli sviluppi applicativi delle reti neurali vedono impegnate in prima linea diverse componenti del mondo finanziario: banche, società di intermediazione

mobiliare, fondi comuni di investimento, semplici investitori privati. Lo sforzo intellettuale di chi opera sulle reti è sostenuto in questo caso anche da risorse finanziarie non indifferenti; si stima che, dopo le applicazioni militari, proprio le applicazioni finanziarie siano il campo in cui la ricerca sulle reti neurali ha ricevuto il maggior supporto finanziario. Del resto, lo stimolo a investire nel campo è molto forte: l'obiettivo è infatti quello di arrivare a un sistema superiore (rispetto al resto del mercato) di selezione, di classificazione o di previsione. Illustrare e classificare i risultati delle ricerche sulle reti neurali in ambito finanziario non è semplice per una serie articolata di motivi:

- a. anzitutto in molti casi i risultati delle ricerche si traducono in sistemi proprietari delle grandi istituzioni finanziarie; ciò del resto è coerente con l'obiettivo della ricerca, vale a dire trovare un modello superiore che può restare tale solo in quanto non condiviso dalla concorrenza;
- b. è difficile trovare più ricerche indipendenti condotte con la medesima finalità su un medesimo set di dati. Ogni ricerca e ogni sviluppo di rete trascinano inevitabilmente con sé un contesto empirico di riferimento; e dato che la rete non spiega il perché dei fenomeni, ma si limita ad analizzarli trovando regolarità empiriche, la riproponibilità di un modello fuori dal suo contesto di riferimento è sempre problematica;
- c. anche a prescindere dall'eterogeneità dei dati e delle finalità, è straordinariamente ricca la varietà dei modelli utilizzati in termini di tipologia della rete, numero di variabili di input e di output, funzioni di attivazione, tipologia di apprendimento e così via.

Paradossalmente, ciò che ostacola lo sforzo del ricercatore in questo campo non è tanto la difficoltà a reperire informazioni quanto piuttosto l'abbondanza delle stesse. E come sempre succede quando il campo di ricerca è nuovo, i contributi importanti sono mischiati agli irrilevanti; non c'è ancora stata cioè, se non in via marginale, quella sedimentazione dei risultati che consente di semplificare molto l'attività di ricerca. In parte ciò dipende, come accennato, dalla natura empirica dell'analisi, per cui il concetto stesso di superiorità di una rete su un'altra è molte volte discutibile e contingente; in parte dipende, invece, dalla frammentazione delle conoscenze specialistiche sulle reti. Una buona parte dell'attività di ricerca e sperimentazione e dell'offerta commerciale di reti proviene, quando non da singoli individui, da numerose istituzioni di piccole dimensioni, quasi artigianali, più che da grandi centri di ricerca.

Tralasciando i casi di informazione privata, la possibilità di ottenere profitti sfruttando solo la serie storica dei prezzi oggetto di previsione (ed eventualmente la serie storica di altre variabili) sembrerebbe contrastare con l'i-

potesi di efficienza dei mercati [22]. Se il prezzo di un'attività davvero seguisse una logica di *random walk* ne conseguirebbe l'impossibilità di trarre profitti da uno studio della serie storica dei prezzi e di altre informazioni pubblicamente disponibili.

Proprio il concetto di disponibilità pubblica delle informazioni può essere utilizzato per aggirare il problema. Se esiste un modello superiore per descrivere la struttura dei prezzi, e quel modello non è pubblicamente disponibile, allora è possibile per chi lo detiene ottenere profitti dall'attività di previsione. La tecnologia superiore (in questo caso il modello neurale) sarebbe lo strumento tramite il quale l'efficienza dei mercati viene perseguita.

Un secondo aspetto di interesse può essere la valutazione comparata delle reti neurali e dei modelli statistici di previsione. In fin dei conti, sia le reti neurali sia i modelli statistici elaborano un certo numero di dati di input per giungere a previsioni riguardo ad alcune variabili. Per quale motivo allora non dovrebbe essere possibile ottenere con modelli statistici tradizionali i risultati delle reti neurali? In altre parole, in quali contesti le reti neurali risultano superiori rispetto ai modelli statistici?

In linea di principio non è difficile associare differenti tipologie di reti a modelli statistici consolidati quali la regressione semplice e multipla, lineare e non lineare, analisi discriminante, analisi delle componenti principali e la cluster analysis. In parecchi casi alle reti è richiesto di svolgere compiti che possono essere condotti anche attraverso analisi statistica tradizionale. Resta quindi da chiedersi se le ostentazioni di novità che accompagnano le presentazioni di molti modelli di rete siano giustificate⁵. Da questo punto di vista occorrerebbe un certo grado di cautela. E vero che i modelli previsionali neurali battono in genere senza difficoltà la previsione random, ma ciò vale in larga misura anche per modelli statistici tradizionali. Dove le reti neurali mostrano vantaggi significativi è nel caso di forme funzionali fortemente complesse e altamente non lineari.

Sotto certe ipotesi è dimostrato che alcune reti possono fungere da "approssimatori universali di funzione". Ciò implica che possiamo trattare con le reti forme funzionali di dipendenza che sarebbero gestibili con grande difficoltà a livello statistico tradizionale. Un'ultima osservazione preliminare riguarda il presunto aspetto di "intelligenza" della rete, di solito confortato da affermazioni secondo cui "*la rete neurale simula il cervello umano*" oppure "*la rete neurale impara*". C'è molta mistificazione in queste affermazioni; in ambito commerciale spesso ciò si traduce in abuso della buona fede della

⁵È con sarcasmo che Sarle [23] osserva: "*neural network researchers routinely reinvent methods that have been known in the statistical or mathematical literature for decades or centuries*".

controparte chiamata ad adottare la rete. Anche ammettendo che l'ispirazione del principio delle reti sia il funzionamento del cervello umano, resta da sottolineare il differente ordine di complessità dei due sistemi. È difficile che una rete neurale abbia più di qualche decina di nodi (l'equivalente in rete del neurone nel cervello); un cervello umano ha circa cento miliardi di neuroni. È difficile fare analogie su queste basi. Riguardo poi all'aspetto dell'apprendimento, è evidente che non è la rete in sé ad apprendere, ma il costruttore della rete a creare un meccanismo di adattamento dei pesi. Se diciamo che la rete neurale è intelligente, allora dobbiamo riservare lo stesso trattamento a una regressione tradizionale, in cui la scelta dei regressori minimizza l'errore.

3.1 I principali ambiti applicativi

Premesso che le applicazioni delle reti neurali sono talmente numerose da rendere arduo ogni tentativo di classificazione, nondimeno si può tentare di schematizzare gli ambiti applicativi in campo finanziario⁶ in tre grandi categorie⁷: previsione di serie temporali; classificazione; approssimazione di funzioni.

3.2 La previsione di serie temporali

È in questo ambito che senza dubbio ricade la maggioranza delle applicazioni delle reti neurali in campo finanziario. Lo scopo di gran parte di queste applicazioni è quello di realizzare profitti speculativi tramite trading, prevalentemente di breve o brevissimo periodo, di attività finanziarie quotate: principalmente azioni, tassi di cambio, futures, commodities. Normalmente la serie storica viene divisa in due parti: il passato (osservazioni *in-sample*) e il futuro (osservazioni *out-of-sample*). L'allenamento della rete avviene sulla prima serie di dati, mentre la seconda ha lo scopo di verificarne la validità. La rete può essere allenata a fornire previsioni per diversi orizzonti temporali, sfruttando le sue stesse previsioni a breve come input per le previsioni a lungo. Inoltre, una rete può essere costruita per fornire previsioni riguardo a più di una variabile; il che consente di cogliere eventuali fenomeni di correlazione tra le variabili stesse. I problemi tipici di un utilizzo previsionale delle reti neurali possono essere così sintetizzati:

- a) *scelta delle variabili di input*: dato che la rete non è in grado di fornire alcun contributo esplicativo, la stessa si ritrova a trattare acriticamente

⁶La precisazione è opportuna perché, al di fuori dell'ambito strettamente finanziario, le applicazioni si moltiplicano: dalla meccanica alla robotica, dalla biologia al controllo dei processi.

⁷Per una classificazione analoga, vedi [24]

tutte le variabili che le vengono date in pasto, con il rischio di utilizzare variabili non significative se durante il periodo di training queste hanno casualmente un qualche potere esplicativo;

- b) *livello ottimo di apprendimento*: è il grande trade off del processo di training, poiché troppo poco training non consente alla rete di cogliere le relazioni tra le variabili, mentre troppo training (*overfitting* o *over-learning*) rende la rete incapace di generalizzare: una volta tolta dal contesto dei dati di training la rete è incapace di prevedere. Va rilevato come, durante il procedimento di training, l'errore di previsione in sample continui a diminuire; ciò rende particolarmente difficile la scelta di un momento in cui arrestare il processo di apprendimento;
- c) *scelta dell'orizzonte temporale di riferimento per la previsione*. Anche in questo caso ci si trova di fronte a un sostanziale trade off: orizzonti temporali di previsione molto brevi accrescono il numero di previsioni corrette (almeno nel segno) ma ciascuna di queste ha un minore profitto unitario; al contrario poche indicazioni di trading di periodo più lungo saranno mediamente meno corrette, ma le corrette comporteranno un profitto medio più elevato. Il problema è aggravato dai costi di transazione, che avvantaggiano strategie di trading su orizzonti temporali più protratti;
- d) *instabilità delle relazioni tra variabili*. Le relazioni tra variabili cambiano nel tempo. Input che sono significativi oggi potrebbero non esserlo più in futuro. Il problema può già manifestarsi nel passaggio dai dati di training ai dati di verifica;
- e) *scarsa lunghezza delle serie storiche*. Il problema è tipico delle variabili macroeconomiche fondamentali (per esempio PIL) il cui data base storico potrebbe non essere sufficiente ad allenare adeguatamente la rete;
- f) *passaggio da output della rete a decisione operativa*. In molti casi le indicazioni della rete riguardano il valore dell'attività che si vuole prevedere. Ciò non si traduce automaticamente in una decisione di acquisto o di vendita. Il passaggio al segnale di acquisto o vendita può essere demandato alla rete neurale o può essere il frutto di regole stabilite sulla base di risultanze empiriche o del confronto dei risultati di più reti.

3.3 La classificazione e la discriminazione

In un compito di classificazione la rete neurale ha il compito di assegnare gli input a un certo numero di categorie predefinite. Ogni unità di output corrisponde a una di queste categorie. In un modello destinato alla discriminazione alla rete è anche demandato il compito di creare le classi in cui suddividere i dati di input. Applicazioni tipiche di queste finalità in campo economico riguardano il rischio di credito: ad esempio suddivisione dei debitori in classi di rating, classificazione delle aziende tra destinate a fallire o meno, decisioni di affidamento.

3.4 L'approssimazione di funzioni

Una prima obiezione a questa classificazione potrebbe essere quella che in fin dei conti anche le due categorie precedenti rientrano nella approssimazione di funzioni, per quanto complesse. Ciò è senz'altro vero, ma quanto si vuole qui evidenziare è la circostanza che le reti possono essere applicate in tutte le funzioni avanzate di *pricing* e di *risk management* in cui manca una forma funzionale precisa per la valutazione degli strumenti. È questo il caso ad esempio delle opzioni di tipo americano o di molte opzioni esotiche (asian, barrier) e di portafogli di opzioni. Se adeguatamente allenate con un buon set di dati, alcune reti sembrano essere particolarmente promettenti a questo fine. Da questo punto di vista le reti neurali si presentano come concorrenti delle metodologie Monte Carlo, da cui peraltro sono completamente distinte da un punto di vista concettuale. Rispetto al caso delle previsioni di variabili finanziarie, inoltre, in questa fattispecie di applicazioni si può fare totale affidamento sulla stabilità della funzione di pricing che, pur ignota, non ha motivo di cambiare nel tempo.

A titolo di esempio, può essere interessante riportare le aree di ricerca attuali del Neuroforecasting Club di Londra: fondato dalla London Business School e dallo University College di Londra, ha il compito di sviluppare le metodologie basate sulle reti neurali e di valutarne le applicazioni pratiche nei mercati dei capitali. È finanziato al 50% dal Department of Trade and Industry e per il restante 50% da grandi istituzioni finanziarie come Barclays de Zoed Wedd, Citicorp, Société Generale e altre istituzioni. I progetti di ricerca attuali sono:

- asset allocation tattica;
- sensibilità dei prezzi future al volume e all'*open interest*;
- negoziazione di valuta tattica infragionaliera; modelli fattoriali per l'investimento azionario;

- modelli per la valutazione e la negoziazione di futures su indice;
- previsione delle volatilità per applicazioni di option pricing;
- metodologie e apprendimento avanzati.

Come è possibile osservare, si tratta di tematiche certamente avanzate, tendenzialmente più complesse della semplice “previsione”; in particolare, in molti casi viene affrontato il passaggio dalla previsione del valore alla decisione di negoziazione.

3.5 Un esempio pratico di rete neurale

In questa parte viene sviluppata, in maniera più tecnica, un’applicazione di reti neurali alla previsione dei tassi di cambio. L’enfasi è posta sulle problematiche metodologiche e sulla valutazione dei risultati.

In questo esempio, i dati riguardano i tassi di cambio giornalieri della sterlina e del marco confrontati con il dollaro statunitense nel decennio 1987–1997; l’analisi stata effettuata con l’ambiente di calcolo GridMathematica sul cluster computazionale Apple disponibile nel laboratorio per l’analisi quantitativa dei dati del Dipartimento di Scienze Economiche, Matematiche e Statistiche.

Sono disponibili circa 2900 giorni di tassi di cambio, con quattro tassi per giorno rappresentanti i valori di apertura, massimo, minimo e di chiusura. Il grafico del valore massimo del marco rispetto al giorno è rappresentato in figura 20.

Si supponga di voler prevedere il valore di apertura utilizzando i tassi di cambio del giorno precedente. Perciò, il valore di apertura è definito come il risultato y del processo e gli altri tre tassi sono gli input u .

Il modello predittivo può essere descritto dalla seguente equazione:

$$\hat{y} = g(\theta, x(t)) \quad (8)$$

Qui $\hat{y}(t)$ è la previsione dell’output $y(t)$, la funzione g è il modello di rete neurale, θ rappresenta i parametri del modello, e $x(t)$ è il regressore del modello, dato dalla seguente equazione:

$$x(t) = [y(t-1)u_1(t-1)u_2(t-1)u_3(t-1)]^T \quad (9)$$

Per un migliore test del predittore il data set è diviso in dati di training (ue e ye) e di validazione (uv e yv). Il secondo data set è utilizzato per validare e confrontare i diversi predittori.

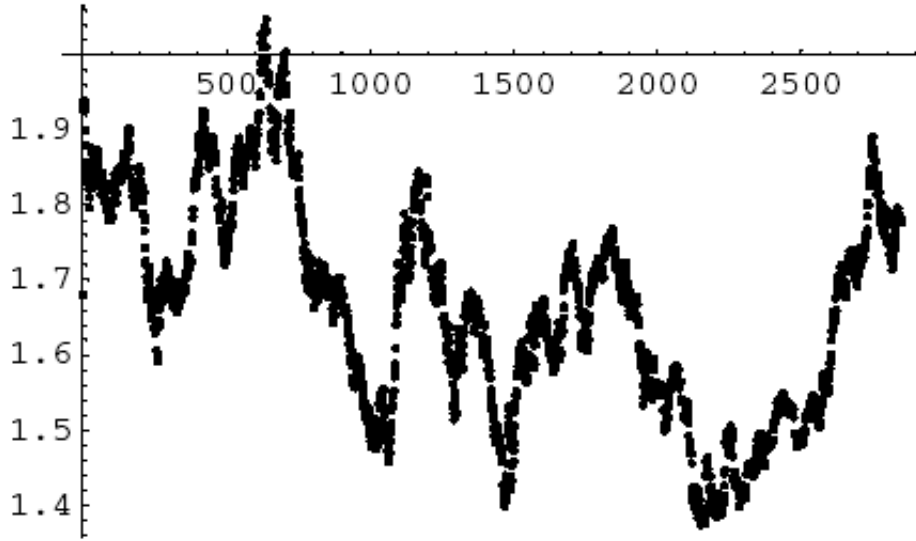


Figura 20: *Tassi di cambio del marco rispetto al dollaro.*

NeuralARX è il modello neurale utilizzato per questa applicazione. Tale rete viene inizializzata e stimata con il comando `NeuralARXFit` di Mathematica. Inizialmente viene stimato un modello di predizione lineare. Per trovare tale predizione, viene scelto `FeedForwardNet` senza neuroni nascosti. Per ottenere il regressore nella forma (9) occorre scegliere gli indici del regressore come segue: $n_a = 1$, $n_b = 1, 1, 1$ e $n_k = 1, 1, 1$.

Stimiamo un modello lineare per la previsione dei tassi. Il modello lineare dell'equazione (8) può essere espresso come segue:

$$\hat{y}(t) = a_1 y(t-1) + b_1 u_1(t-1) + b_2 u_2(t-1) + b_3 u_3(t-1) + b_4 \quad (10)$$

Se si osservano i parametri del modello “addestrato”, si vede che b_3 è vicino a 1 ed il resto dei parametri sono vicini a 0. Ciò significa che il tasso di cambio di apertura è per lo più correlato al tasso di chiusura del giorno precedente. Questa sembra una caratteristica abbastanza naturale, ed è quindi possibile avere una qualche fiducia nel modello.

Prima dell'addestramento di un modello nonlineare, occorre valutare la previsione single-step sui dati di validazione. Poiché il mercato cambio nel tempo, la previsione viene valutata solo sui 100 giorni successivi ai dati di stima. Il grafico mostra il tasso previsto insieme al tasso reale. Le due curve sono così vicine che risulta abbastanza difficile discernerele. Viene anche fornito l'errore quadratico medio (RMSE), utilizzabile come misura della qualità della previsione.

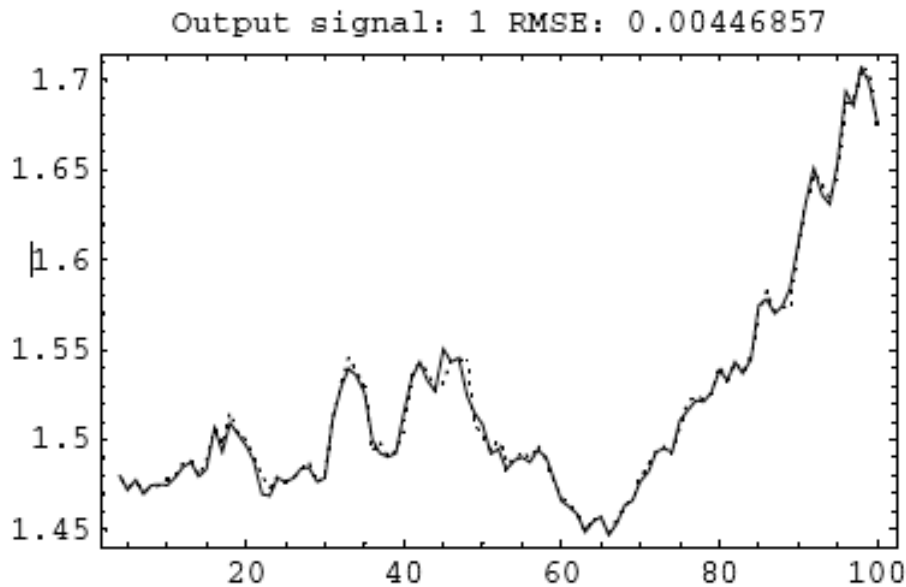


Figura 21: *Valutazione della previsione one-step sui dati di validazione.*

È ora possibile provare modelli nonlineari per vedere se sono in grado di prevedere meglio di quello lineare. Utilizziamo una rete *Feed-Forward* con due neuroni nascosti. Il regressore scelto è lo stesso del modello lineare. Un modello lineare è incluso in parallelo con la rete. Poiché i dati di validazione sono inclusi nel successivo addestramento, la stima è ottenuta con la “stopped search”.

Addestrando una rete FF con due neuroni sui dati dei tassi di cambio si ottiene: Il miglioramento della rete neurale durante l’addestramento è molto piccolo poiché l’inizializzazione della rete fa uso dei minimi quadrati per adattare i parametri lineari.

Calcolando la previsione one-step sullo stesso intervallo di dati utilizzato per il modello lineare si vede (figura 23) che l’errore è leggermente inferiore per il modello nonlineare rispetto a quello lineare: perciò, la previsione è leggermente migliore. È tipico che il miglioramento di dati economici sia piccolo; altrimenti, sarebbe troppo facile fare denaro.

Consideriamo ora una rete *Radial Basis Function* (RBF) con due neuroni, mantenendo gli stessi argomenti utilizzati per la rete FF. Inizializzando ed addestrando una rete RBF sui dati disponibili otteniamo il grafico di figura 24. Le prestazioni della rete RBF sui dati di validazione peggiorano durante l’addestramento, restituendo la rete RBF inizializzata.

La valutazione della previsione one-step con la rete RBF è rappresenta-

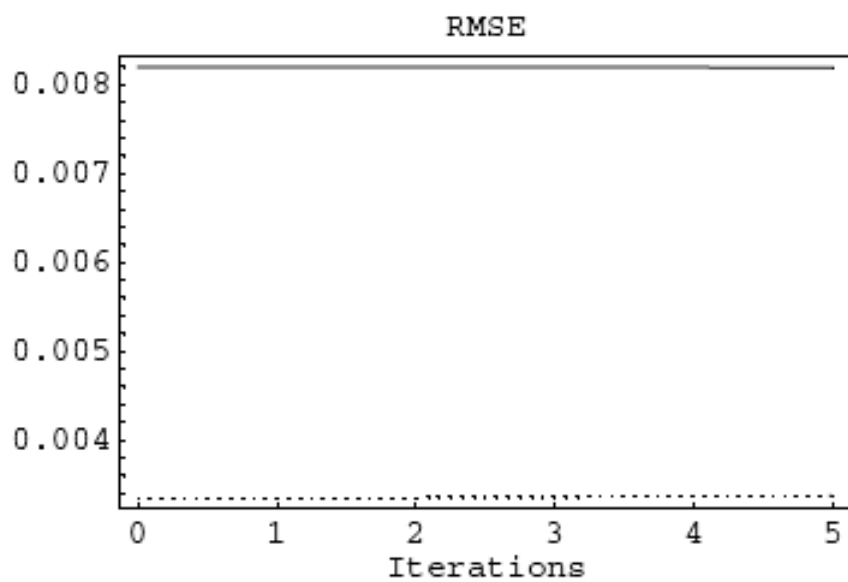


Figura 22: *Training di una rete FF con due neuroni nascosti.*

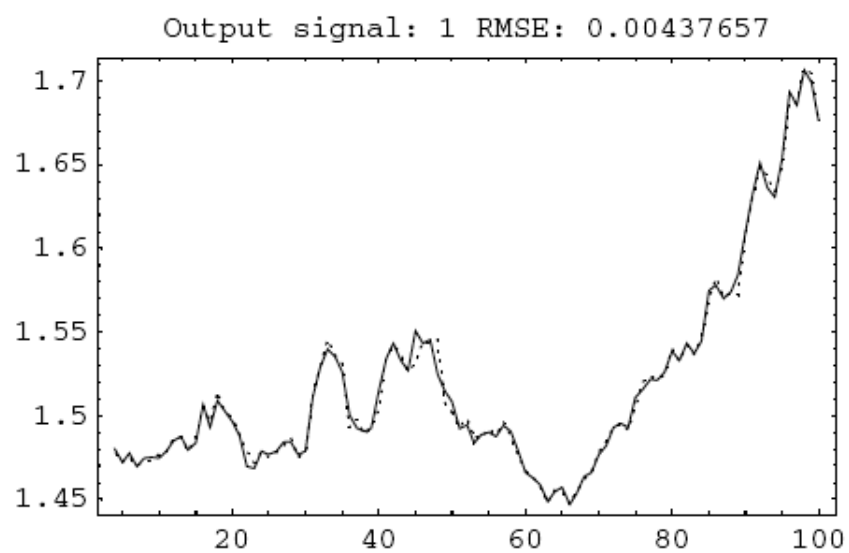


Figura 23: *Previsione one-step.*

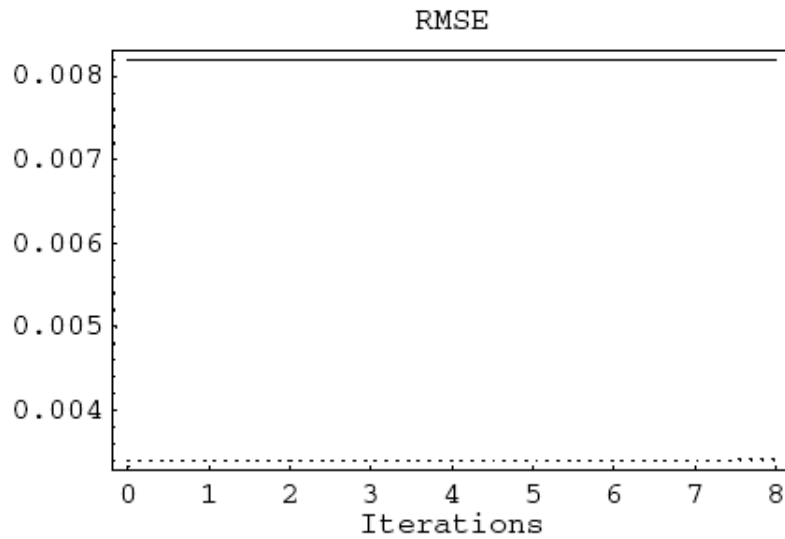


Figura 24: *Inizializzazione e addestramento della rete RBF.*

ta nel grafico di figura 25. La rete RBF è leggermente migliore della rete lineare, ma non altrettanto buona quanto la rete FF. Se si ripete la valutazione dell'esempio, il risultato potrebbe cambiare leggermente a seguito della casualità nell'inizializzazione.

Si potrebbe ripetere l'esempio modificando diverse opzioni, come ad esempio:

- cambiare il numero di neuroni;
- cambiare il regressore per contenere un maggior numero di risultati precedenti;
- escludere alcuni dei segnali di input dal regressore.

È anche possibile cambiare l'intervallo di dati utilizzato nell'addestramento e nella validazione, come pure provare a prevedere la sterlina invece del marco.

L'esempio illustra che è possibile prevedere il tasso di cambio di apertura facendo uso dei tassi del giorno precedente. la relazione è naturalmente nonlineare, dato che i modelli nonlineari basati su reti FF e RBF operano leggermente meglio del modello lineare.

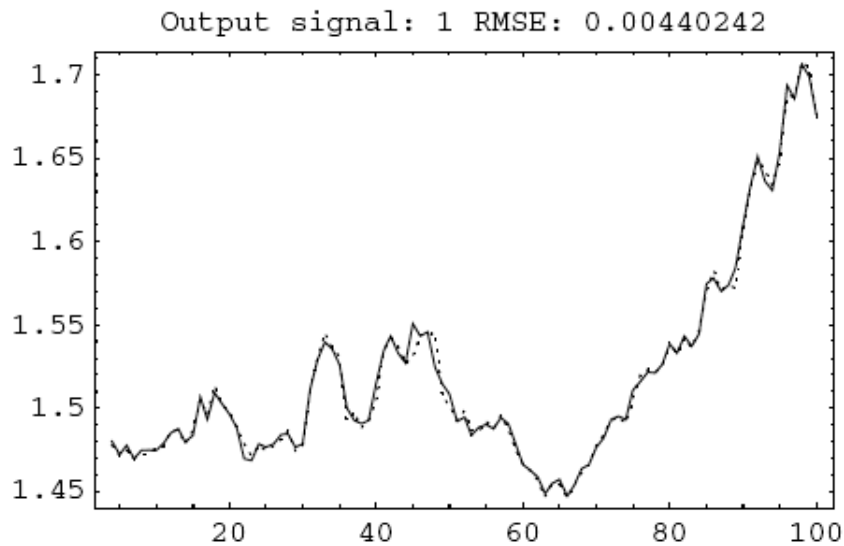


Figura 25: *Valutazione della previsione one-step con la rete RBF.*

4 Conclusioni

Le Reti Neurali sono attualmente in forte espansione e vengono utilizzate quale principale metodologia di modellizzazione per le previsioni finanziarie e per l'analisi di mercato. In questo lavoro viene passata in rassegna la teoria alla base delle reti neurali, con uno sguardo critico allo “stato dell'arte” delle applicazioni nella finanza.

Ricerche sperimentali mostrano che le reti neurali possono risultare più efficaci dei modelli convenzionali in molti casi, o perlomeno possono rappresentare uno strumento complementare. L'implementazione di una rete neurale è un'arte. Prestazioni occasionalmente funzionanti non implicano necessariamente un modello efficace: vi è probabilmente sempre un altro modello più efficace per altri set di dati, e ciò consiglia l'approccio della segmentazione in subset di dati e nel relativo training per la miglior verifica dell'ottimalità del modello ipotizzato.

Riferimenti bibliografici

- [1] E. Gately: *Neural Networks for Financial Forecasting*, Series Editor: P.J. Kaufman, Wiley Trader's Advantage, 1995.
- [2] A. Colombo, L. Mussio: *The Statistical Behaviour of Correlated Data*, Int. Archives of Photogrammetry and Remote Sensing, vol. XXXII, part 6W1, Padua, 1997, pp. 112–115.
- [3] L. Galati, A. Tumietto et al.: *La previsione nei mercati finanziari*, Bancaria Editrice, Curatore: G. Gabbi, 1999, ISBN 88-449-0089-0.
- [4] J.J. Murphy: *The Visual Investor*, Wiley, 1996, ISBN 0471144479.
- [5] M.P. Niemira, P.A. Klein: *Forecasting Financial and Economic Cycles*, Addison Wesley, 1995.
- [6] M. Dahlquist, L.E.O. Svensson: *Estimating the Term Structure of Interest Rates for Monetary Policy Analysis*, The Scandinavian Journal of Economics, Vol. 98, No. 2 (Jun., 1996), pp. 163–183(21).
- [7] P. Jorion, L.J. Sweeney: *Mean reversion in real exchange rates: evidence and implications*, J. Int. Money Finance 15, 1996, pp. 535-550.
- [8] E. Azoff: *Neural Network Time Series Forecasting of Financial Markets*, John Wiley & Sons, 1994.
- [9] G.P. Zhang, B.E. Patuwo, M.Y. Hu: *Forecasting with artificial neural networks: The state of the art*, International Journal of Forecasting, Vol. 1, No. 14, 1998, pp. 35–62.
- [10] M. Shanker, M. Hu, M. Hung: *Effect of data standardization on neural network training*, Omega International Journal of Management Science, Vol. 24, No. 4, 1996, pp. 385-397.
- [11] A. Cecchini, F. Viola: *Approaching a generalised urban automata with help! on line (AUGH!)*, in Workshop su reti Neurali e Automi Cellulari Urbani, Dicembre 1995, Milano.
- [12] G.J. Deboeck: *Neural network in finance: A tutorial*, Milano: Università Cattolica del Sacro Cuore, 1995.
- [13] H. Bandy: *Developing a neural network system*, AAIL Working Paper, 1997.

- [14] K. Hornik, M. Stinchcombe, H. White: *Multilayer Feedforward Networks are Universal Approximators*, Neural Networks, Vol. 2, pp. 359–366, 1989.
- [15] B. Zhang, H. Mühlenbein: *Synthesis of Sigma-Pi Neural Networks by the Breeder Genetic Programming*, Proceedings of IEEE International Conference on Evolutionary Computation (ICEC-94), World Congress on Computational Intelligence, 1994.
- [16] P. Viola: *Alignment of 2- and 3-dimensional objects based on maximization of mutual information*, PhD Thesis, Massachusetts Inst of Tech, Cambridge, 1997.
- [17] B. Colombo, A. Cortese, L. Fabbri: *L'errore statistico nella produzione di microdati e macrodati*, (a cura di) La Produzione di Statistiche Ufficiali, pp. 247–259, Padova, Cleup, 1994.
- [18] A. Sironi, M. Marsella: *La misurazione e la gestione dei rischi di mercato*, Il Mulino, Bologna, 1997.
- [19] G. Leitch, J.E. Tanner: *Economic Forecast Evaluation: Profits Versus the Conventional Error Measures*, The American Economic Review, Vol. 81, No. 3 (Jun., 1991), pp. 580–590(11).
- [20] R. Bramante, R. Colombo, G. Gabbi: *Are Neural Network and Econometric Forecasts Good for Trading? Stochastic Variance Model as Filter Rule*, in A.P.N. Refenes, A.N. Burgess, J.E. Moody (eds), Decision Technologies for Computational Finance, Kluwer Academic Publishers, Dordrecht, 1998.
- [21] T.B. Rubino Jr., D.P. Nimey II: *Option Strategies And Neural Networks*, Technical Analysis of STOCKS & COMMODITIES (Sep., 1996).
- [22] K. Swingler: *Financial Prediction, Some Pointers, Pitfalls, and Common Errors*, Center of Cognitive and Computational Neuroscience, Stirling Univ., Stirling, FK9 4LA, *ftp source*, 1994.
- [23] W.S. Sarle: *Neural network and statistical models*, in Proceedings of the 19th Annual SAS Users Group International Conference, Cary, NC, April 1994.
- [24] A. Episcopos: *Predicting Returns on Canadian Exchange Rates with Artificial Neural Networks and EGARCH-M Models*, NNCM-95 Third International Conference On NEURAL NETWORKS IN THE CAPITAL MARKETS, Poster No. 3, London, 1995.