POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale

Corso di Laurea Magistrale
in Ingegneria Gestionale

Tesi di Laurea Magistrale

Intelligenza artificiale applicata ai mercati finanziari: cluster analysis e reti neurali per individuare bolle speculative

Relatore
prof. Franco Varetto

Candidato
Riccardo De Giovanni

Anno accademico 2020-2021

Sommario

Introduzione

Nell'ultimo secolo il mercato azionario è stato più volte colpito da crolli importanti che hanno messo in ginocchio le principali economie mondiali provocando lunghe crisi. Questi eventi sono quasi sempre la conseguenza dello sviluppo e dello scoppio di bolle speculative, ovvero fasi di mercato in cui i prezzi aumentano considerevolmente senza una reale giustificazione nei fondamentali dei titoli sottostanti. La teoria dei mercati efficienti, il paradigma dominante fino all'inizio degli anni '90, non fornisce spiegazioni adatte a giustificare l'esistenza di questi fenomeni, ed è quindi necessario rivolgersi alla finanza comportamentale per cercare di capirne le cause e i sintomi. Tuttavia, anche con strumenti sempre più sofisticati, negli ultimi vent'anni il mercato ha subito due violenti crolli a causa dello scoppio di bolle speculative. Gli studi più recenti e innovativi, quindi, cercano di applicare l'intelligenza artificiale al mercato finanziario per riconoscere e gestire meglio queste anomalie. Lo scopo di questo elaborato è creare un modello in grado di individuare la presenza di bolle speculative all'interno del mercato azionario, utilizzando elementi di statistica multivariata, come la cluster analysis, e le reti neurali artificiali, dei modelli matematici che tentano di emulare il funzionamento del cervello umano per risolvere problemi molto complessi e dalla struttura non lineare.

Nella prima parte saranno presentate le principali teorie economiche riguardo le dinamiche del mercato azionario e i movimenti dei prezzi; nella seconda parte saranno approfondite le tecnologie utilizzate, concentrandosi soprattutto sulle reti neurali; nella terza parte verrà descritta la costruzione del modello con l'implementazione di una rete neurale ricorrente LSTM, e infine saranno commentati i risultati ottenuti inserendoli nel contesto attuale.

Capitolo 1

I mercati efficienti

1.1 Funzione dei mercati azionari e concetto di efficienza

Nei sistemi economici moderni i mercati finanziari svolgono un ruolo sempre più importante affiancandosi ai mercati dei capitali reali, rappresentandone la loro espressione finanziaria e consentendo, grazie alla loro natura flessibile e liquida, un disaccoppiamento nelle scelte di investimento. La finalità principale di un mercato dei capitali è allocare efficientemente le risorse trasferendole dalle unità in surplus alle unità in deficit.
Se nella teoria economica generale il concetto di efficienza è tendenzialmente associato all'ipotesi di mercato perfetto (concorrenza perfetta, assenza di costi di transazione), nei mercati finanziari l'efficienza viene declinata in più forme (Tobin, 1984):

- *Efficienza allocativa-funzionale*: riguarda la capacità del mercato di trasferire le risorse e le attività in modo da garantirne una valutazione coerente con i flussi di cassa attesi futuri (fundamental-valuation efficiency) e in modo da soddisfare le esigenze di tutti gli operatori
- *Efficienza tecnico-operativa*: riguarda la capacità del mercato di operare minimizzando tempistiche e costi di transazione
- *Efficienza informativa*: riguarda la capacità del mercato di elaborare e distribuire uniformemente le informazioni rilevanti per gli investitori. Gran parte dell'economia finanziaria, infatti, si è sviluppata attorno al concetto cardine di *informazione perfetta* (prontamente disponibile a tutti gli operatori in modo omogeneo), ed è proprio questo il tipo di efficienza dei mercati finanziari su cui sono nate e si sono sviluppate alcune tra le teorie finanziarie più importanti del XX secolo.

1.2 Background storico

A partire dall'ultimo quarto del XIX secolo iniziò a svilupparsi, in ambito economico, il periodo neoclassico, che ha visto tra i suoi massimi esponenti economisti del calibro di Léon Walras e Alfred Marshall. Lo sviluppo di tale periodo fu accompagnato e favorito da altre due dottrine: il positivismo di Comte e l'utilitarismo di Bentham e Mill. Il positivismo riteneva che fosse conoscibile solo ciò che era misurabile e quantificabile, perciò i metodi delle scienze sperimentali e delle scienze naturali si sarebbero dovuti applicare anche alle altre discipline. L'utilitarismo, riprendendo diversi concetti sviluppati già da Adam Smith durante il periodo classico, affermava che ogni individuo ha come obiettivo l'ottenimento di un benessere (o utilità) sempre maggiore in ogni decisione che prende. Queste due scuole di pensiero

contribuirono all'applicazione di un'enfasi sempre maggiore su rigore metodologico e matematica nelle scienze economiche. In risposta a questo cambio di paradigma, durante il periodo neoclassico, nacque la figura dell'*homo oeconomicus*, ovvero un individuo razionale in grado di compiere sempre la scelta, tra diverse combinazioni di beni e servizi, associata alla maggiore utilità possibile. Seppure solo qualche decennio dopo, nel 1921, John Maynard Keynes con il suo "*Trattato sulla probabilità*" mostrava evidenti deviazioni dal comportamento razionale da parte di numerosi attori economici, le influenze del periodo neoclassico hanno continuato a persistere nel corso di tutto il XX secolo e rappresentano i pilastri teorici da cui sono partite tantissime teorie economico finanziarie che hanno dominato il pensiero economico durante il 1900, tra cui la teoria dei mercati efficienti.

La paternità della teoria dei mercati efficienti si riconosce in larga misura a Eugene Fama e Paul Samuelson. Da un punto di vista storico è possibile dividere lo sviluppo della teoria in tre parti: gli anni '60, in cui Fama e Samuelson con tre celebri articoli nel 1965 misero le basi su cui ancora oggi poggia la teoria; gli anni '70, in cui la teoria ha ottenuto gran parte del suo consenso accademico e verifiche empiriche; gli anni '80, in cui sono nate le prime vere alternative all'efficienza dei mercati (Delcey, 2019).
Durante gli anni '60, il comportamento dei prezzi nei mercati finanziari era uno degli argomenti più discussi nel campo dell'economia finanziaria (Delcey, 2019) ed è in questo contesto che Paul Samuelson e Eugene Fama, nel 1965, sulla scia di Working, Kendall e Cowles, scrissero rispettivamente uno e due articoli in cui vennero per la prima volta presentate alcune delle idee alla base della teoria dei mercati efficienti.
Nel suo primo articolo del 1965, "*Behavior of Stock Market Prices*", Fama giunse al risultato che le variazioni di prezzi nei mercati finanziari sono indipendenti e fornì la prima definizione di "mercato efficiente": un mercato in cui, date le informazioni disponibili, il prezzo di un titolo rappresenta in ogni momento una buona stima del suo valore fondamentale, ovvero il valore attuale dei futuri flussi di cassa associati al titolo (Fama, 1965a). Nonostante l'indipendenza statistica delle variazioni dei prezzi non fosse empiricamente dimostrata, Fama sosteneva che i *sophisticated traders* (sostanzialmente analisti fondamentali e analisti tecnici esperti) avrebbero notato queste discrepanze, dovute principalmente alle limitate capacità di investitori poco esperti chiamati *noise traders*, e, tramite processi di arbitraggio, avrebbero riportato il prezzo al suo valore corretto molto velocemente, rendendo quindi non sfruttabili queste piccole anomalie.
Alla stessa conclusione di Fama riguardo la dinamica casuale dei prezzi azionari giunge anche Samuelson, sempre nel 1965, nell'articolo "*Proof That Properly Discounted Present Values of Assets Vibrate Randomly*" attraverso l'utilizzo del processo statistico martingala e il concetto di arbitraggio di Modigliani e Miller (Samuelson, 1965).
Più avanti, sempre nel 1965, Fama pubblica un altro articolo riguardo efficienza dei mercati. In "*Random Walk in Stock Market Price*" Fama sostituisce il concetto di sophisticated traders con la definizione più generale di *rational profit-maximizers*.
Per la prima volta in questo contesto viene posta l'enfasi sull'ottimizzazione del comportamento degli agenti che operano sui mercati: un mercato in cui sono presenti investitori che competono tra loro e che hanno un comportamento ottimizzante, e in cui le informazioni sono liberamente disponibili, è un mercato efficiente, e quindi il prezzo di un

titolo convergerà al suo valore fondamentale (Delcey, 2019; Fama, 1965b). Questa interpretazione fornita da Fama differisce, pur derivando dagli stessi assunti, da quella presentata da Samuelson. Secondo quest'ultimo, infatti, l'imprevedibilità dei movimenti azionari è unicamente dovuta alla competizione tra i diversi attori del mercato e non è legata al valore fondamentale del titolo. In ogni caso, a prescindere da questo punto di arrivo leggermente diverso, sia Fama sia Samuelson spiegano la dinamica casuale dei prezzi azionari come conseguenza della razionalità degli investitori. (Delcey, 2019)

1.3 EMH

Il primo riferimento all'Efficient Market Hypotesis (EMH), nome con cui viene riconosciuta la teoria di Fama sui mercati efficienti, risale al secondo articolo di Fama del 1965 "*Random Walk in Stock Market Price*". Alla fine degli anni '60 le principali conclusioni ottenute da studi sui mercati efficienti riguardavano l'andamento casuale dei prezzi e il fatto che essi riflettono in ogni momento il valore fondamentale del titolo sottostante.
Negli anni '70 la teoria raggiunge un nuovo livello di popolarità, grazie soprattutto al lavoro di Fama che nell'articolo "*Efficient Capital Markets: a Review of Theory and Empirical Work*" del 1970 presenta la definizione più famosa di mercato efficiente: un mercato finanziario è efficiente se i prezzi riflettono sempre le informazioni disponibili.
Questo passaggio rende intuitivo il collegamento tra casualità del movimento dei prezzi azionari e efficienza dei mercati: i prezzi di mercato rivelano le valutazioni degli investitori alle informazioni disponibili e si modificano solo in seguito a nuove informazioni, ma il processo di arrivo di nuove informazioni è per definizione casuale. Se le informazioni fossero in qualche modo prevedibili allora non sarebbero per definizione nuove e sarebbero pertanto già state scontate dal mercato.

Prima di presentare il modello di Fama è opportuno presentare il concetto degli *excess returns*. È possibile definire l'excess return come la differenza tra il rendimento ex-post, ovvero quello osservato sui mercati al tempo t+1, e il rendimento ex-ante, ovvero quello calcolato al tempo t sulla base delle informazioni disponibili.

$$\varepsilon_{t+1} = \frac{P_{t+1} - P_t}{P_t} - \frac{E(P_{t+1}|\Phi_t) - P_t}{P_t} = \frac{P_{t+1} - E(P_{t+1}|\Phi_t)}{P_t}$$

$$\varepsilon_{t+1} = r_{t+1} - E(r_{t+1}|\Phi_t) = \text{ex post} - \text{ex ante}$$

Il rendimento ex-post è osservabile sui mercati azionari al tempo t+1, mentre per il rendimento ex-ante bisogna utilizzare un modello che renda possibile farne una stima. L'alternativa più utilizzata in tal senso è il CAPM (Capital Asset Pricing Model), un'altra valida soluzione è rappresentata dall'APT (Arbitrage Pricing Theory).

Nel suo lavoro del 1970, Fama presenta tre modelli differenti per formalizzare il concetto di mercati efficienti:

- Fair Game Model: se il processo di formazione dei rendimenti attesi è avvenuto utilizzando pienamente le informazioni disponibili allora il valore atteso degli excess returns, nel lungo periodo, sarà pari a 0, ovvero la sequenza degli excess return è un "*fair game*" rispetto al set informativo Φ disponibile al tempo t (Fama, 1970).

  $$\mathrm{E}(\varepsilon_{t+1}|\Phi_t) = 0$$

  Se gli excess returns fossero diversi da 0 nel lungo periodo allora le informazioni disponibili non sarebbero pienamente incorporate nel prezzo e sarebbe quindi possibile sfruttarle per ottenere un rendimento sistematicamente maggiore di quello offerto dal mercato, conclusione che violerebbe le ipotesi valide in un contesto di mercato efficiente.

- Modello Martingala: una variabile statistica segue un processo dinamico Martingala rispetto ad un set Φ se il suo valore atteso futuro condizionale (quindi scontato al tasso di rendimento congruo per rischio) è uguale al suo valore corrente. Se il mercato è efficiente e quindi i prezzi attesi riflettono pienamente le informazioni disponibili allora:

  $$\mathrm{E}(\mathrm{P}_{t+1}|\Phi_t) = \mathrm{P}_t * (1 + \mathrm{E}(\mathrm{r}_{t+1}|\Phi_t))$$

  $$\mathrm{P}_t = \frac{\mathrm{E}(\mathrm{P}_{t+1}|\Phi_t)}{1 + \mathrm{E}(\mathrm{r}_{t+1}|\Phi_t)}$$

  Il prezzo atteso scontato segue un processo Martingala rispetto a Φ. Di conseguenza, in un contesto di mercato efficiente, il prezzo attuale è il miglior elemento, sulla base delle informazioni disponibili, per formulare una previsione per il prezzo di domani (Samuelson, 1965; Fama, 1970):

  $$E[Y_{t+1}|Y_0, \dots, Y_t] = Y_t$$

  Analizzare la serie storica dei prezzi non fornisce alcun vantaggio, in quanto tutte le informazioni da essa derivanti sono già state incorporate nel set informativo Φ.

- Modello Random Walk: una serie temporale segue un processo random walk se le sue variazioni successive sono indipendenti e identicamente distribuite: ciò che succede all'istante t+1 non dipende da ciò che è avvenuto negli istanti precedenti. Quindi, generalizzando il modello Fair Game e il modello Martingala, Fama afferma che la serie di cambiamenti nel prezzo dei titoli è priva di memoria e non segue un trend: la storia passata non può essere utilizzata per prevedere il futuro in nessuna maniera significativa.
  Con questo ultimo modello Fama non si limita più a descrivere il valore atteso dei prezzi e rendimenti in eccesso, ma fornisce una definizione più generale sulla forma della distribuzione statistica degli excess returns. Infatti, mentre i modelli Fair Game e Martingala affermano che il valore atteso dei rendimenti in eccesso è indipendente dalle informazioni disponibili, il modello Random Walk sostiene che l'intera distribuzione statistica è indipendente dal set informativo Φ:

$$f\left(r_{j,t+1}\middle|\Phi_t\right) = f(r_{j,t+1})$$

  La distribuzione condizionale e la distribuzione marginale di una variabile casuale che segue un processo random walk sono identiche.
  Fama ha da subito riconosciuto che il modello Random Walk non può essere una fedele rappresentazione dei mercati data l'impossibilità di individuare una serie storica dei prezzi completamente indipendente, ma in contesti pratici si accettano le ipotesi del modello nei casi in cui la correlazione della serie storica dei prezzi non superi una certa soglia.

L'efficienza del mercato può essere di tre tipi, a seconda della composizione del set informativo Φ pienamente incorporato nei prezzi:

- **Forma debole**: l'informazione è rappresentata dalla serie storica dei prezzi azionari. È un set informativo poco costoso e conoscibile da tutti gli investitori. Se un mercato è efficiente in forma debole non è possibile elaborare una strategia di investimento basata esclusivamente sulla storia dei corsi dei titoli che, a parità di livello di rischio, permetta di ottenere un rendimento in eccesso atteso superiore a quello del mercato. Negli anni '50 e '60 furono presentati diversi studi con l'obiettivo di dimostrare che i prezzi azionari si muovessero secondo un random walk, ma i primi lavori supportati da una rigorosa teoria economica risalgono al 1965 e sono opera di Benoit Mandelbrot con "*Forecasts of Future Prices, Unbiased Markets, and Martingale Models*" e Paul Samuelson con il già citato "*Proof That Properly Discounted Present Values of Assets Vibrate Randomly*" (Fama, 1970). Nella sua revisione della teoria dei mercati efficienti del 1970 Fama ha analizzato i trenta titoli appartenenti al Dow Jones dalla fine del 1957 al 1962 mostrando che, coerentemente con le ipotesi proposte, non sussistessero autocorrelazioni tra i rendimenti in eccesso dei prezzi a diversi intervalli di tempo (Fama, 1970). Le poche deviazioni da questo risultato

secondo Fama non sono abbastanza significative per rigettare le ipotesi di un mercato efficiente in forma debole. Tuttavia, sostiene che sia più efficace testare la non profittabilità di vari sistemi di trading operanti basandosi esclusivamente sulla storia dei prezzi. Vennero quindi paragonati le performance di un sistema *y% filter* e di una semplice strategia *buy and hold*, I risultati ottenuti sono concordi nell'affermare che non è possibile ottenere un rendimento in eccesso sistematicamente maggiore del mercato basandosi su un qualche sistema di trading. Gli spiragli di profittabilità, presenti soprattutto in un orizzonte temporale di breve termine se non addirittura intra-day, sono quasi sempre erosi dai costi di transazione (Fama, 1970).

- **Forma semi-forte**: il set informativo è rappresentato da tutte le informazioni pubbliche come ad esempio bilanci aziendali, notizie, dati macroeconomici e finanziari, prospettive settoriali etc. Queste informazioni sono teoricamente disponibili a tutti gli investitori, ma nella realtà dei fatti tali informazioni possono essere più o meno accessibili in termini di modalità e costo a seconda che si consideri un investitore istituzionale o un investitore individuale. Chi ha un miglior accesso ai dati pubblici è verosimilmente favorito, ma ciò non rappresenta una violazione delle ipotesi dell'EMH siccome l'ottenimento di determinate informazioni è un'attività costosa che andrà quindi in parte a erodere il rendimento ottenuto (Akintoye, 2019). Quindi, se un mercato è efficiente in forma semi-forte, non è possibile elaborare una strategia di investimento basata esclusivamente sulle informazioni pubblicamente disponibili e ottenere, a parità di livello di rischio, un rendimento in eccesso atteso superiore a quello del mercato.
  Le verifiche empiriche riguardo l'efficienza semi-forte del mercato sono molto ampie date le molte tipologie di dati pubblici che è possibile utilizzare per testare la reazione del mercato. Infatti, se il mercato è efficiente in forma semi-forte, la reazione dei prezzi a nuove informazioni dovrebbe essere istantanea. Eventuali overreactions e underreactions rappresentano delle deviazioni dal comportamento efficiente. Un esempio grafico di tale fenomeno è rappresentato in figura 1.1.

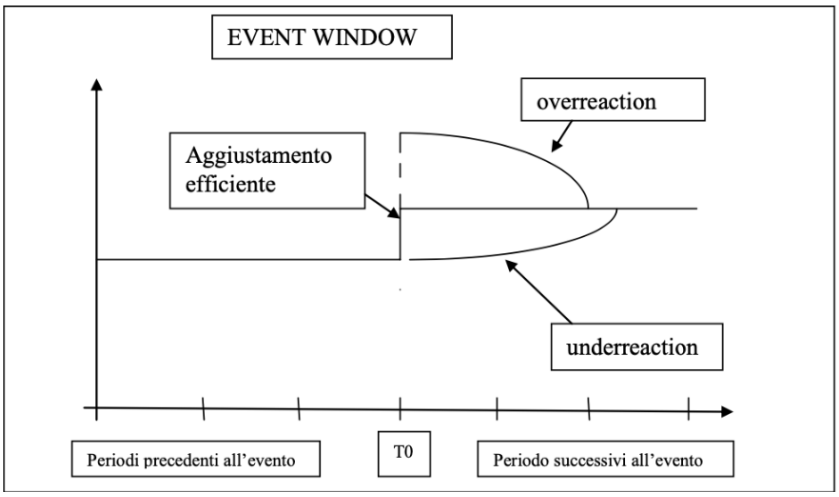


Figura 1.1: overreaction e underreaction del mercato

Gli eventi principalmente studiati sono stati gli stock splits, gli earnings announcements, le operazioni di merge and acquisition e le emissioni azionarie. Nel suo articolo del 1970 Fama analizza soprattutto la reazione dei mercati in seguito a degli stock splits prendendo in considerazione 940 titoli del N.Y.S.E. tra il 1927 e il 1959 e conclude che il mercato reagisce in modo significativamente efficiente (Fama, 1970). In generale, prendendo in considerazione anche gli altri eventi, si può affermare che il mercato risponde in maniera parzialmente efficiente in seguito alla pubblicazione di informazioni. Tuttavia, come nel caso dell'efficienza in forma debole, le opportunità di profitto, in questo caso più marcate, sono rese difficilmente sfruttabili dalla presenza dei costi di transazione e dalla breve durata delle anomalie (Malkiel, 2003).

- **Forma forte**: l'informazione è rappresentata da tutti i dati, anche quelli non disponibili pubblicamente. L'efficienza in forma forte del mercato ingloba quella in forma semi-forte e quella in forma debole. Oltre ad essere l'ipotesi più estrema, è anche la più difficile da verificare empiricamente a causa della poca mole di dati disponibili. Se il mercato fosse efficiente in forma forte sarebbe impossibile ottenere un rendimento in eccesso atteso sistematicamente maggiore del mercato utilizzando tutte le informazioni disponibili, anche quelle private, le cosiddette *insider information*. Fama stesso riconosce la non validità di tale assunto nella realtà dei fatti, in quanto numerosi studi ed episodi quotidiani mostrano che chi ha accesso ha informazioni monopolistiche è in grado di ottenere un rendimento maggiore del mercato senza assumersi un maggiore rischio (Sharpe, 1965).

1.4 Critiche ai mercati efficienti

Le critiche ai mercati efficienti hanno accompagnato la nascita e lo sviluppo della teoria e anche al giorno d'oggi ci sono intere scuole di pensiero molto scettiche nei confronti degli assunti proposti da Fama e Samuelson nell'EMH. Sino a quando Fama nel 1970 ha conferito un aspetto più rigoroso alla teoria, le critiche erano più generalmente riferite al contesto in cui la teoria dei mercati efficienti mirava a introdursi. Uno dei più critici, addirittura già negli anni '40, era Keynes, tra i primi a mettere in dubbio, già a suo tempo, la razionalità degli agenti economici nel prendere le decisioni di investimento. Secondo Keynes la distribuzione delle aspettative soggettive degli operatori, contagiata dai loro umori e sentimenti nei confronti del mercato, è il vero driver dei futuri prezzi azionari e non i fondamentali dell'impresa.
Nei decenni successivi le critiche all'EMH sono diventate sempre più frequenti, anche a causa di studi empirici che hanno rilevato delle anomalie sui mercati che, in qualche misura, sconfessavano le ipotesi dei mercati efficienti.

Le più evidenti e discusse tra queste anomali sono state:

- *Small-firm effect*: tra il 1951 e il 1979 le 50 più piccole imprese del N.Y.S.E. hanno ottenuto un rendimento del 20,65% rispetto ad un rendimento del 1,53% delle imprese di media-grande dimensione (Lustig and Leinbach, 1983). I rendimenti delle piccole imprese in quel periodo sono stati molto maggiori rispetto a quello attesi dal CAPM. Nel corso degli anni, in seguito alla scoperta di questa anomalia, meccanismi di arbitraggio hanno fatto scomparire questo effetto. Un'altra spiegazione di questo fenomeno è dovuta all'inadeguatezza del CAPM nello scontare correttamente certi rendimenti (Lustig and Leinbach, 1983).

- *Calendar effect*: ancora oggi, in certi intervalli di tempo, i rendimenti osservati a gennaio sono statisticamente maggiori di quelli osservati negli altri mesi dell'anno. Molte delle spiegazioni fornite sostengono che tale effetto è dovuto a vantaggi fiscali dovuti alla deducibilità delle perdite sui titoli, ma altri studi appoggiano la tesi di un'anomalia di mercato, dovuta anche a effetti psicologici, incoerente con le ipotesi dell'EMH (Jacobs and Levy, 1988; Thaler, 1987).

- *Book-to-market effect*: tra il 1963 e il 1990 le azioni con alto book-to-market ratio hanno ottenuto un rendimento mensile più alto dell'1,53% rispetto a quello ottenuto da azioni con basso book-to-market ratio (Fama and French, 1992). Fama, in studi successivi, ha affermato che questa non rappresenta un'anomalia: un alto book-to-market value, infatti, rappresenta una fonte di rischio per l'impresa, in quanto è spesso dovuto a scarse prestazioni, e il maggior rischio spiega il maggior rendimento garantito dal titolo. De Bondt e Thaler, invece, in uno studio del 1987 sostengono che la causa di questa anomalia sia un *overreaction* degli investitori ai fondamentali dell'impresa (Woo, Mai, McAleer, Wong, 2020).

- *Momentum effect*: le variazioni positive di un titolo tendono ad essere statisticamente seguite da variazioni dello stesso segno. In uno studio del 1993, Jegadeesh e Titman, hanno mostrato come, nell'arco temporale di sei mesi, un portfolio composto da *winner stocks* ottiene in media un rendimento maggiore del 9% rispetto ad un portfolio di *loser stocks*.

Una delle prime visioni alternative alla teoria dei mercati efficienti è sostanzialmente una variazione di essa. Nel 1980 Grossman e Stiglitz hanno sollevato un importante paradosso, ancora oggi non del tutto risolto, che prende il nome di paradosso di Grossman e Stiglitz. La loro idea è che, se il mercato è efficiente e quindi incorpora pienamente nei prezzi tutte le informazioni disponibili, e se la raccolta e di informazioni è un'attività costosa, allora nessun operatore ottiene un beneficio nel ricercare ed elaborare tali informazioni, ma si limiterebbero tutti ad osservare i livelli dei prezzi senza compiere alcuno sforzo attivando quindi dinamiche di free-riding. Ma se nessun investitore sostiene il costo per analizzare le informazioni allora quest'ultime non arriveranno mai sul mercato e quindi i prezzi non potranno incorporarle, conclusione non in linea con le ipotesi portanti dell'EMH (Grossman and Stiglitz, 1980). La soluzione più accreditata per risolvere questo paradosso prevede l'introduzione nel mercato di un rumore di fondo (noise) che non permette una perfetta trasparenza delle informazioni, e quindi chi sosterrà l'onere di elaborare tali informazioni otterrà un vantaggio competitivo, soprattutto nel breve termine (Black, 1986).

I maggiori sostenitori dell'EMH hanno continuato a far valere le idee della teoria nonostante il sopraggiungere di queste anomalie e delle forti critiche avanzate dagli anni '80 in poi. Le principali tesi avanzate in difesa dei mercati efficienti riguardano, in primis, il meccanismo di arbitraggio e i costi di transazione. È riconosciuta, infatti, la presenza di deviazioni dal comportamento ideale che un mercato efficiente dovrebbe avere, ma gli effetti prodotti sono deboli e soprattutto difficilmente sfruttabili sia a causa dei meccanismi di arbitraggio che riducono la durata temporale di questi puzzle, sia a causa dei costi di transazione che erodono quasi completamente le possibilità di guadagno di un qualsivoglia trading system (Malkiel, 2003). Il secondo elemento citato spesso in difesa dell'EMH riguarda proprio le performance dei fondi di investimento rispetto a quelle ottenute dal mercato. Se il mercato fosse veramente inefficiente allora gli investitori professionisti sul mercato dovrebbero essere in grado di sfruttare tali inefficienze e ottenere rendimenti mediamente maggiori del mercato. Burton Malkiel, nel suo articolo "*The Efficient Market Hypothesis and Its Critics*" del 2003 mostra, alla fine del 2001, il rendimento mediano degli equity funds con alta capitalizzazione in confronto a quello offerto dallo S&P 500 evidenziando come l'indice di mercato abbia in media ottenuto prestazioni migliori. Inoltre, sempre con dati che arrivano sino alla fine del 2001 e che considerano vari mercati, Malkiel mostra l'alta percentuale di fondi attivi che hanno ottenuto rendimenti minori rispetto a quelli del mercato. Questi dati sono rappresentati dalla figura 1.2 e dalla figura 1.3.

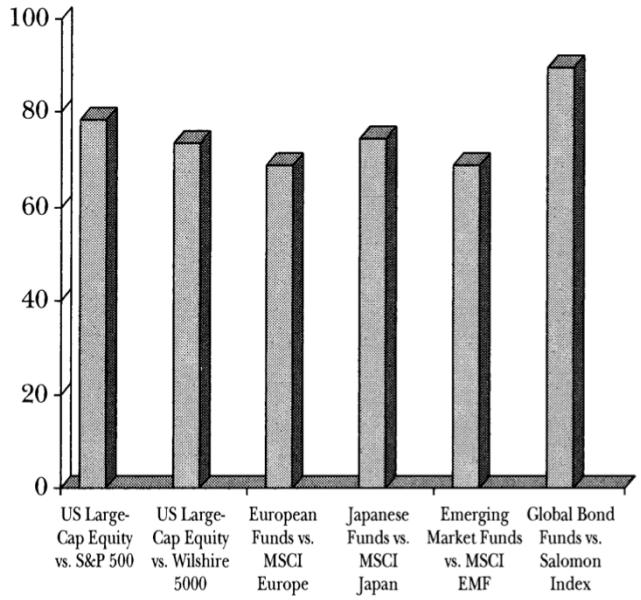
Median Total Returns Ending 12/31/2001

	10 years	15 years	20 years
Large Cap Equity Funds	10.98%	11.95%	13.42%
S&P 500 Index	12.94%	13.74%	15.24%

Source: Lipper Analytic Services, Wilshire Associates, Standard & Poor's and The Vanguard Group.

Figura 1.2: risultati dei fondi rispetto al mercato

Percentage of Various Actively Managed Funds Outperformed by Benchmark Index 10 Years to 12/31/01



Source: Lipper Analytic Services and Micropal.

Figura 1.3: percentuale di fondi battuti dal mercato

Dagli anni '80 in poi, l'EMH ha dovuto fronteggiare critiche più strutturali, rivolte alle ipotesi su cui si basa l'intero modello, e non solo legate ad anomalie osservate sui mercati. Il principale limite dell'EMH risiede nel fatto che non è in grado di prevedere e spiegare fenomeni estremi, come ad esempio le bolle speculative, in quanto in un contesto di mercato efficiente non sarebbe possibile il verificarsi di tali situazioni. Sono nate intere nuove scuole di pensiero riguardo la dinamica dei prezzi sul mercato azionario, tra cui la più importante è sicuramente la behavioural finance (finanza comportamentale), oggetto del prossimo capitolo.

Capitolo 2

Alternative ai mercati efficienti – La finanza comportamentale

2.1 L'eccessiva volatilità

La costante scoperta di nuove anomalie ed inefficienze di mercato iniziò a far traballare le fondamenta della teoria dei mercati efficienti. Per i sostenitori dell'EMH, infatti, era sempre più difficile trovare una spiegazione ai puzzle di mercato ed alle evidenze empiriche che mostravano distorsioni evidenti dagli assunti della teoria. Questo clima di incertezza attorno all'EMH fu esacerbato notevolmente duranti i primi anni '80, quando la robustezza del modello fu messa nuovamente alla prova dal problema dell'eccessiva volatilità (Hammond, 2015). Se le anomalie menzionate nel capitolo precedente possono essere considerate delle piccole deviazioni da quanto sostenuto dall'EMH, l'incapacità di spiegare gran parte della volatilità del mercato minacciava di dubitare dell'intera struttura della teoria (Shiller, 2003).

Secondo quanto proposto dall'EMH, il prezzo P_t al tempo t rappresenta, condizionatamente a tutte le informazioni disponibili, il valore fondamentale di un titolo, ovvero il valore atteso della somma dei flussi di cassa futuri generati. Tali flussi, non considerando i capital gains, si riducono ai dividendi futuri.
Segnando con P_t^* il valore atteso scontato dei dividendi futuri, possiamo scrivere che

$$P_t^* = P_t + U_t$$

dove U_t è l'errore della previsione.
Siccome ogni informazione disponibile al tempo t è incorporata nella formazione del prezzo P_t, U_t è una variabile non correlata con il set informativo Φ disponibile al tempo t e siccome il prezzo stesso P_t rappresenta un elemento del set informativo, P_t e U_t non sono correlate tra loro. Ne segue che P_t^* è la somma di due variabili non correlate e quindi la sua varianza può essere calcolata come somma delle varianze:

$$var(P_t^*) = var(P_t) + var(U_t)$$

Di conseguenza la varianza del valore atteso presente dei dividendi futuri non può essere maggiore della varianza della variabile P_t (Shiller, 2003).

Shiller (2003) ha analizzato i dati dello Standard & Poor's Composite Stock Price Index dal 1871 al 2002 confrontando il valore reale dei prezzi e il valore reale dei dividendi futuri, con tasso di sconto pari alla media geometrica del rendimento reale tra il 1871 e il 2002.
Nella figura 2.1 sono rappresentati su un grafico i risultati. È evidente che, mentre l'andamento dei dividendi reali (PDV, Constant Discount Rate) ha un trend piuttosto stabile con volatilità contenuta, l'andamento dei prezzi reali è molto più altalenante, con una forte volatilità in eccesso rispetto ai dividendi reali.
Questo risultato non può essere spiegato tramite la teoria dei mercati efficienti e rappresenta un'evidenza molto più solida e concreta rispetto alle diverse anomalie finanziarie scoperte.

Real Stock Prices and Present Values of Subsequent Real Dividends
(*annual data*)

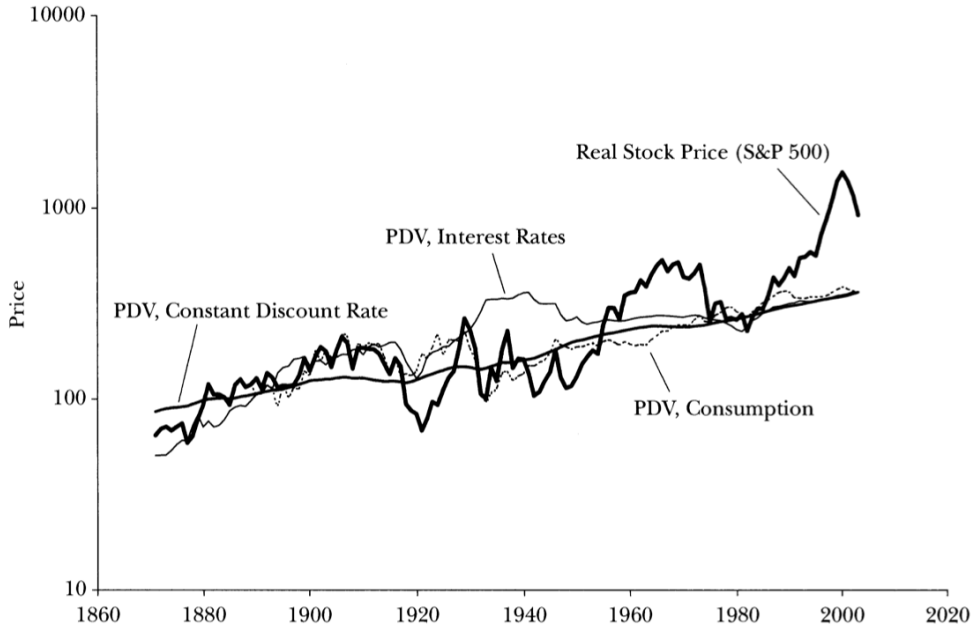


Figura 2.1: andamento reale dello S&P500 vs andamento dei dividendi reali futuri

Shiller (2003) sostiene che le evidenze riguardo l'eccessiva volatilità implicano il fatto che i prezzi non si modifichino per ragioni legate al valore fondamentale del titolo, almeno non interamente, ma che entrino in gioco dinamiche psicologiche degli investitori che portano la dinamica dei prezzi a deviare fortemente da quanto prevede l'EMH.

Il lavoro di Shiller (2003) è tra i più autorevoli riguardo il tema dell'eccessiva volatilità e riassume concetti e modelli che lui stesso e altri studiosi hanno iniziato a rimarcare dagli anni '80, quando le evidenze di eccessiva volatilità sui mercati non potevano più essere ignorate o inserite nel contesto dell'EMH.

In risposta alle problematiche relative alle anomalie ed inefficienze di mercato e all'eccessiva volatilità, gran parte del mondo accademico economico finanziario iniziò a cercare risposte al di fuori dall'EMH, concentrandosi su nuove teorie.

In questo contesto guadagnò molto spazio la finanza comportamentale, un campo di studio che lega psicologia cognitiva e finanza e che, seppur offrendo una prospettiva completamente diversa, ha rappresentato e rappresenta tutt'ora la più valida alternativa alla teoria dei mercati efficienti.

2.2 Finanza comportamentale

2.2.1 Background storico

Le prime contaminazioni di psicologia nell'ambito economico-finanziarie risalgono al periodo classico, quando Adam Smith, nel 1759, pubblicò "*The Theory of Moral Sentiments*" in cui descriveva i principi psicologici del comportamento individuale (Ashraf et al., 2005). Altri economisti come Walras, Menger, Jevons e Edgeworth, pur avvicinandosi ad un approccio di stampo neoclassico con l'utilizzo di scienze sperimentali in economia, ritenevano che il concetto di utilità avesse una forte connotazione psicologica (Lewin, 1996). Nonostante molti economisti del tempo sembravano accettare la complessa natura psicologica dell'uomo, il successo delle scienze naturali spostò l'attenzione verso una visione meccanicistica del mondo, aprendo le porte al positivismo ed all'utilitarismo che segnarono l'inizio del periodo neoclassico. Come specificato nel capitolo precedente, durante questo periodo il mondo economico-finanziario sposò la teoria dell'Homo Oeconomicus e della razionalità, estromettendo completamente qualsiasi ragionamento che ponesse l'accento su dinamiche psicologiche durante il processo di scelta dell'individuo.
Per avere le prime avvisaglie del ritorno del tema della psicologia all'interno dell'ambito economico-finanziario bisogna aspettare il 1912, anno in cui George Charles Selden scrisse "*Psicology of the Stock Market*" in cui tentò di identificare un legame di dipendenza tra la dinamica dei prezzi e il comportamento degli investitori.
Un importante punto di svolta avvenne tra gli anni '50 e gli anni '60 con l'avvento della psicologia cognitiva in contrasto al modello comportamentista. Pur avendo entrambe le teorie una forte impostazione naturalistica, la psicologia cognitiva, a differenza del comportamentismo che vede la mente umana come un black box incontrollabile a cui arrivano input e da cui fuoriescono ouptut, pensa che la mente umana sia un elaboratore che riceve informazioni dall'esterno, processa tali informazioni e le restituisce *semplificate* all'esterno. Il modo in cui la mente elabora le informazioni, secondo la psicologia cognitiva, è soggettivo e dipende da processi mentali come l'apprendimento, il ragionamento, la memoria, le emozioni e il linguaggio.
La rivoluzione cognitiva in psicologia, insieme alle prime evidenze di deviazioni dalla teoria dei mercati efficienti, contribuì a far rinascere un forte interesse nella psicologia e nelle sue applicazioni economiche (Illiashenko, 2017).
Inizialmente molti lavori di economia comportamentale non catturarono grande attenzione all'interno del mondo accademico a causa del fatto che molti esperimenti venivano condotti in situazioni astratte e ipotetiche, senza un concreto risvolto sul mondo reale (Heukelom, 2007). Con il passare degli anni gli esperimenti diventarono sempre più solidi e negli anni

’70 ci fu un cambio di passo decisivo grazie ai lavori di Amos Tversky e Daniel Kahneman, due psicologi cognitivi esperti dei processi di *decision-making* in condizioni di incertezza.

2.2.2 Prospect Theory

Negli anni ’60 Tversky stava lavorando a nuovi approcci alternativi alla Teoria dell’Utilità Attesa, mentre gli studi di Kahneman riguardavano l’errore della percezione umana in seguito a determinati stimoli esterni. La combinazione dei loro lavori li convinse a lavorare insieme al tema del decision-making nel mondo reale (Heukelom, 2007).
Nel 1979 Tversky e Kahneman pubblicarono “*Prospect Theory: A Study of Decision Making Under Risk*”, un lavoro che si considera abbia sancito uno dei punti chiave per la nascita dell’economia e della finanza comportamentale.
La Prospect Theory è la più valida alternativa alla Teoria dell’Utilità Attesa. Quest’ultima era stata, fino a quel momento, la teoria dominante per spiegare il processo di decision-making di un individuo ed è stata uno dei punti di forza del pensiero neoclassico in economia. La sua prima formulazione si deve a Daniel Bernoulli nel 1738 ed è stata poi ripresa e formalizzata da von Neuman e Morgenstern nel 1947. L’idea di fondo della teoria dell’utilità attesa ricalca concetti già presentati descrivendo l’Homo Oeconomicus: sostiene che ogni individuo compie delle scelte cercando di massimizzare la funzione obiettivo dell’utilità e in particolare l’utilità può essere calcolata come media pesata delle utilità in ogni stato del mondo possibile a seguito della decisione presa, utilizzando come pesi le probabilità del verificarsi dei singoli stati. L’individuo è perfettamente razionale nell’ordinare le proprie preferenze, nell’identificare gli stati del mondo e nell’associare ad ogni stato la probabilità di accadimento. Secondo questa teoria, dunque, l’utilità è un valore atteso.
Nella Prospect Theory, invece, Tversky e Kahneman (1979) dimostrano come, in molte situazioni, gli individui non seguano il ragionamento della teoria dell’utilità attesa, ma compiano scelte seguendo dinamiche *irrazionali*. Uno dei punti di forza della teoria di Tversky e Kahneman è il suo carattere puramente descrittivo: essi non mirano a proporre un modello che rappresenti come l’individuo *dovrebbe* prendere le decisioni, ma si limitano a osservare tramite evidenze empiriche come effettivamente le persone ragionano durante il processo di decision-making.
Cercando di cogliere gli aspetti principali della Prospect Theory, è possibile elencare tre punti che rappresentano le principali novità fornite dal lavoro di Tversky e Kahneman:

1. **Funzione di valore**: il concetto di utilità è sostituito dal concetto di *valore*. Seppur possa apparire un cambiamento esclusivamente semantico, in realtà la differenza è importante, soprattutto per i punti seguenti proposti dalla teoria. La funzione di utilità considera il benessere finale netto del soggetto, mentre la funzione di valore è riferita alla variazione di ricchezza, ragionando quindi in termini di guadagni e perdite partendo dalla posizione di partenza.

2. **Avversione alle perdite**: la funzione di valore è asimmetrica, con il dominio negativo relativo alle perdite formato da una curva convessa con andamento ripido, e il

dominio positivo formato da una curva concava con andamento più lento. Questa forma della funzione di valore è la conseguenza di quanto hanno notato Tversky e Kahneman (1979) con i loro esperimenti: gli individui hanno un'avversione per le perdite (*loss aversion*), ovvero modulano diversamente il rischio a seconda che ci sia una prospettiva di guadagno o una prospettiva di perdita.
Un esempio molto intuitivo è fornito dallo stesso Kahneman (2011) all'interno del suo famoso libro "*Thinking, Fast and Slow*":
Problema 1: oltre a quello che possiedi ti sono stati dati 1000 dollari. Ora ti si chiede di scegliere una di queste opzioni: 50% di probabilità di vincere 1000 dollari oppure ricevere sicuramente 500 dollari.
Problema 2: oltre a quello che possiedi ti sono stati dati 2000 dollari. Ora ti si chiede di scegliere una di queste opzioni: 50% di probabilità di perdere 1000 dollari oppure perdere sicuramente 500 dollari.
In entrambi i problemi gli stati finali di ricchezza sono identici: in tutti e due i casi si ha un'opzione che permette di avere sicuramente 1500 dollari in più rispetto alla ricchezza iniziale, e in entrambi i casi è presente un'opzione *rischiosa* che consentirebbe di diventare più ricco di 1000 o 2000 dollari. Secondo gli assunti della teoria dell'utilità attesa, in entrambi i problemi gli individui dovrebbero essere indifferenti tra le due alternative, in quanto il valore atteso tra i due scenari è uguale, e comunque i due problemi dovrebbero indurre nei soggetti preferenze analoghe. Tuttavia, negli esperimenti condotti da Tversky e Kahneman, i risultati sono stati diversi: nel primo problema la maggioranza dei soggetti ha preferito l'opzione più sicura ricevendo sicuramente 500 dollari, e nel secondo problema la maggioranza ha preferito l'opzione rischiosa accettando il 50% di probabilità di perdere 1000 dollari.

Sono stati condotti molti esperimenti simili dai due psicologi e i risultati hanno sempre mostrato queste dinamiche. Le conseguenze di questi esperimenti sono molto importanti. In primis si nota una forte deviazione dagli assiomi della teoria dell'utilità attesa, siccome in entrambi i problemi gli individui hanno mostrato una preferenza tra le alternative; e inoltre è evidente l'avversione alle perdite negli individui: quando sono confrontate direttamente o valutate l'una rispetto all'altra, le perdite appaiono molto più grandi dei guadagni (Kahneman, 2011). Nel problema 2, quindi, l'idea di perdere sicuramente 500 dollari ha più effetto rispetto allo scenario del problema 1 in cui si guadagnano sicuramente 500 dollari, e per questo la maggioranza degli individui sceglie l'alternativa probabilistica che potrebbe consentire nessuna perdita accettando il rischio di perdere ancora di più.

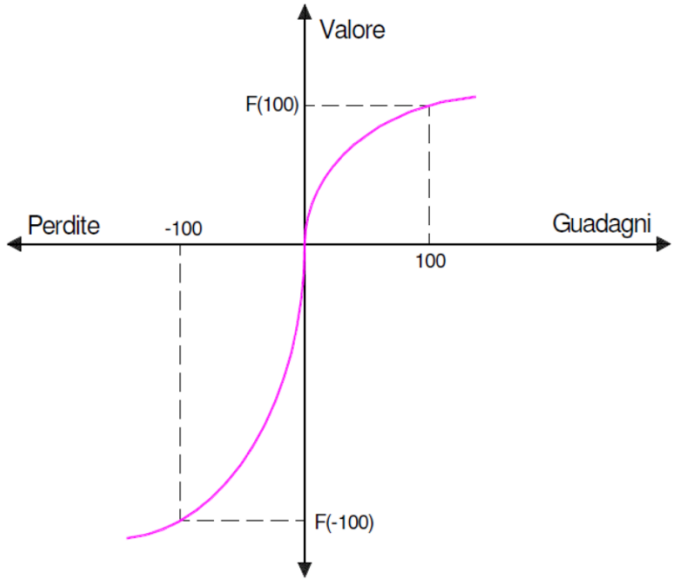


Figura 2.2: andamento asimmetrico della funzione di valore

La figura 2.2 mostra il valore psicologico dei guadagni e delle perdite e l'andamento asimmetrico della funzione dovuto appunto all'avversione alle perdite.

3. **Diminuzione di sensibilità**: osservando la figura 2 si può notare un altro aspetto importante della funzione di valore, ovvero l'andamento sigmoidale, che si concretizza in una variazione marginale del valore decrescente rispetto ai guadagni o alle perdite man mano che ci si allontana dal punto di riferimento iniziale della funzione (incontro tra gli assi). Ciò significa che la differenza in termini di valore soggettivo tra un guadagno di 100 dollari e uno di 200 dollari è maggiore della differenza soggettiva tra un guadagno di 1100 dollari ed un guadagno di 1200 dollari (Kahneman e Tversky, 1986).
   Quest'ultimo aspetto ha più rilevanza e applicazioni pratiche nel marketing rispetto alla finanza, in quanto permette di elaborare particolari strategie con cui approcciarsi ai consumatori.

La spiegazione del perché le persone siano così frequentemente portate a ragionare in modo non razionale, se non talvolta addirittura illogico, è fornita dalla teoria del processo duale, dalle euristiche e dai bias cognitivi. La componente psicologica della finanza comportamentale si basa quasi interamente su questi tre concetti che, come vedremo, sono strettamente collegati tra loro

2.2.3 Teoria del processo duale

La teoria del processo duale sostiene che ci sono due diversi modi attraverso cui possiamo formulare un pensiero, ciascuno dovuto all'azione di due processi cognitivi differenti.
Le fondamenta della teoria del processo duale sono opera dello psicologo americano William James, il quale sosteneva ci fossero due diversi tipi di pensiero: associativo e di ragionamento. Durante il XX secolo furono proposte moltissime versioni della teoria e con la rivoluzione cognitiva in campo psicologico si fece un ulteriore passo in avanti. Kahneman (2011) sceglie di utilizzare i termini coniati dagli psicologi Keith Stanovich e Richard West facendo riferimento a due sistemi mentali:

- *Sistema 1*: opera in fretta e automaticamente, con poco o nessuno sforzo e senza alcun senso di controllo volontario
- *Sistema 2*: indirizza l'attenzione verso le attività mentali impegnative che richiedono focalizzazione e concentrazione, come i calcoli e i ragionamenti complessi

Il sistema 1 genera modelli di idee sorprendentemente complessi, ma solo il sistema 2 è in grado di elaborare, più lentamente, pensieri in una serie ordinata di stadi (Kahneman, 2011). Questi due sistemi non lavorano separatamente, quanto piuttosto parallelamente, con il sistema intuitivo che genera impressioni, pensieri, e giudizi impulsivi che possono essere accettati, bloccati o corretti dal sistema 2 (Morewedge e Kahneman, 2010).
Illiashenko (2017) riassume gli aspetti fondamentali di questa dicotomia: il sistema 1 richiede molte risorse cognitive e per questo motivo il nostro cervello, in ottica di risparmio di energie, lavora maggiormente con il sistema 2. Lakoff e Johnson (1999) hanno stimato che le persone spendono circa il 95% del loro tempo sotto il controllo del sistema intuitivo e questo significa che molte delle decisioni che prendiamo sono opera del sistema 1 e non sono controllate dal sistema 2. Il sistema intuitivo è sempre coinvolto nel processo di decision-making, cioè il sistema limbico interferisce nei processi cognitivi (Lowenstein, 2008).
Inoltre, il sistema 1 non è adatto alla complessità del mondo di oggi (Hirshleifer, 2015) quindi si appoggia spesso su *shortcut* mentali, chiamati euristiche, che generano sistematici bias cognitivi nei processi decisionali.

2.2.4 Euristiche

Le euristiche sono delle scorciatoie mentali che utilizziamo a livello inconscio nel nostro processo decisionale. A causa della natura complessa di molti problemi che dobbiamo affrontare quotidianamente e a causa della difficoltà del nostro cervello nell'utilizzare in modo frequente il sistema 2, spesso riformuliamo il problema che dobbiamo affrontare in modo più semplice in modo che sia il sistema 1 a fornire la risposta, riducendo sensibilmente lo sforzo cognitivo necessario. Secondo Kahneman (2011) le euristiche aiutano a trovare risposte adeguate, anche se spesso imperfette, a quesiti difficili.

Le euristiche più comuni sono:

- **Euristica della rappresentatività**: è la tendenza a semplificare il nostro processo decisionale ricorrendo a stereotipi nel momento in cui elaboriamo le informazioni durante il processo di scelta. Attribuiamo la probabilità di accadimento ad un evento in base a quanto quell'evento è rappresentativo di una certa classe. Kahneman e Tversky (1974) nel loro articolo "*Judgment under Uncertainty: Heuristic and Biases*" svolsero un esperimento in cui chiesero ad un gruppo di persone di indovinare il lavoro svolto da un individuo sulla base di una descrizione di quella persona fornita da un conoscente. La descrizione fornita era: "*Steve è un ragazzo timido e introverso, disponibile, ma non molto interessato alle altre persone e a ciò che lo circonda. È mite, gli piace l'ordine e ama i dettagli*". I mestieri tra cui gli intervistati potevano scegliere erano: agricoltore, commercialista, pilota di aerei, fisico e bibliotecario. La maggioranza delle persone scelse l'alternativa *bibliotecario*, in quanto secondo loro Steve ben si adattava allo stereotipo di quel tipo di professione e non considerarono minimamente che professioni come il commercialista o l'agricoltore sono svolte da molte più persone. Molti intervistati, quindi, caddero in errore a causa di una distorsione probabilistica dovuta all'euristica della rappresentatività.

- **Euristica della disponibilità**: è la tendenza ad attribuire ad un evento una probabilità basata sulla facilità con cui ricordiamo il manifestarsi di tale evento. Questa euristica ci porta a sovrastimare o sottostimare le probabilità di accadimento di scenari alternativi che abbiamo di fronte. Shefrin e Statman (2000) nel loro articolo "*Behavioral Portfolio Theory*" hanno dimostrato che, a causa di questa euristica, spesso vengono preferiti titoli di grosse aziende, sovrastimandone il valore, rispetto a titoli di aziende più piccole. Questo accade perché, anche a causa di campagne di marketing, abbiamo molte più informazioni disponibili sulle big-company e nel momento della valutazione ci tornano facilmente in mente le loro buone prestazioni e associamo una probabilità di buon rendimento più alta di quanto in realtà non dovrebbe essere.

- **Euristica dell'ancoraggio**: è la tendenza ad *ancorarci* ad un'informazione ritenuta importante o ad un'ipotesi di partenza nel momento in cui effettuiamo una scelta. Shiller (2000) propone due tipologie di ancore: quantitative e morali, a seconda della tipologia di informazione alla quale ci si lega. Secondo Shefrin (2000), l'euristica dell'ancoraggio è la causa di analisi fallaci da parte degli analisti finanziari: nel momento in cui si trovano a valutare un'azienda si ancorano alle condizioni iniziali (ad esempio il prezzo e lo stato di salute dell'azienda) e quando vengono diffuse nuove informazioni tendono ad analizzarle basandosi comunque sulle probabilità iniziali che avevano formulato. Quindi, se ad esempio arrivano notizie finanziarie positive riguardo un'azienda che inizialmente aveva avuto performance negative (ipotesi a cui l'analista si è ancorato), la stima della possibilità che le performance future siano positive sarà al ribasso.

- **Euristica dell'affetto**: è la tendenza di un individuo nel coinvolgere le emozioni all'interno del processo decisionale alterando la valutazione di rischi e benefici. Finucane (2000) nel suo articolo "*The affect heuristic in judgments of risks and benefits*" ha dimostrato che questa euristica prevede una correlazione inversa tra rischi e benefici: un adolescente, ad esempio, giudica bassi i rischi di bere alcolici e allo stesso tempo ne giudica alti i benefici. Sempre nello stesso articolo Finucane sostiene che un'implicazione finanziaria di tale euristica sia il fatto che gli investitori attribuiscono un maggior valore ai titoli che hanno in portafoglio rispetto ad altri, anche più meritevoli, disponibili sul mercato. La componente affettiva che prende parte al processo di decision-making ne altera la sua efficacia.

2.2.5 Bias cognitivi

Le euristiche permettono al nostro cervello di risparmiare risorse. Il loro utilizzo ci consente di tenere il ritmo della mole e della complessità di problemi che dobbiamo risolvere quotidianamente, ma spesso sono la causa di errori di ragionamento e di valutazione, chiamati bias cognitivi, che rendono completamente fallace il nostro processo decisionale.
I bias cognitivi scoperti e analizzati sono tantissimi. In questo elaborato ne viene presentato un elenco dei più importanti, anche considerando i risvolti in campo finanziario:

- *Endowment effect*: consiste nella tendenza ad attribuire maggior valore ai beni in nostro possesso rispetto a quelli non in nostro possesso, ed è una stretta conseguenza dell'euristica dell'affetto. Kahneman, Knetsch e Thaler (1990) sostengono che questa anomalia sia dovuta all'incapacità degli individui di considerare il costo opportunità del bene che si possiede. Questo potrebbe portare a sovrastimare delle azioni in portafoglio a discapito di azioni più valide presenti sul mercato, creando così un portafoglio inefficiente. Kahneman, Knetsch e Thaler (1991) spiegano che la combinazione dell'avversione alle perdite e dell'endowment effect forma lo *status quo bias*, ovvero l'errore che genera una marcata preferenza per lo status quo nell'effettuare le proprie scelte.
- *Confirmation bias*: è la tendenza a ricercare attivamente informazioni che corroborano la tesi in cui si crede ignorando tutte le informazioni che invece non la confermano. Questo bias crea problemi soprattutto nella fase di raccolta di dati e prove per l'elaborazione di una strategia a lungo termine (Risen, Gilovich, 2007).
- *Self attribuition bias*: è la tendenza di un individuo ad attribuire i propri successi alle sue abilità e skills e i propri fallimenti a fattori fuori dal suo controllo, come ad esempio il caso e la sfortuna. Czaja e Roder (2020) hanno dimostrato che, soprattutto nei trader non professionisti, l'attribuzione dei successi esclusivamente alle proprie capacità causa delle conseguenti performance negative sui mercati. Una spiegazione plausibile è che questo bias aumenta l'*overconfidence* degli investitori diminuendo la loro capacità di imparare dai propri errori e causando l'elaborazione di strategie inefficienti.

- *Disposition effect*: è un bias collegato alle modalità con cui gli investitori gestiscono i guadagni e le perdite realizzate sui propri investimenti. Shefrin e Statman (1985) hanno dimostrato che è presente un'attitudine negli agenti di mercato a mantenere in portafoglio titoli in perdita, sperando in un recupero con il passare del tempo, e allo stesso tempo vendere troppo in fretta titoli in positivo. Questa tendenza è strettamente correlata con l'avversione alle perdite, e comporta la formazione di portafogli inefficienti e di distorsioni sui prezzi.
- *Gambler's fallacy*: è la tendenza a credere che eventi passati abbiano influenza sulla probabilità di accadimento di eventi futuri, quindi ad esempio un evento casuale ha più probabilità di verificarsi solo perché non si è verificato per un periodo di tempo. Secondo un rapporto del 2020 della Consob, nel merito di un esperimento effettuato, un quarto degli intervistati ha mostrato di aver commesso errori di investimento a causa di questo bias, come ad esempio credere, senza una concreta analisi dietro, che un titolo salga solo perché è da un po' che scende.

2.2.6 L'affermazione della finanza comportamentale

Il crescente successo della letteratura sulla finanza comportamentale ha spinto il mondo finanziario a tenere sempre più in considerazione queste prospettive. Sebbene la Prospect Theory di Kahneman e Tversky del 1979 fu un punto di svolta, essa e i risultati importanti ottenuti considerando il cognitivismo, le euristiche e i bias cognitivi faticarono a ottenere un riconoscimento formale da parte del mondo accademico economico finanziario, rimanendo un tema distaccato. Era infatti necessario far capire l'importanza dei temi psicologici in finanza riuscendo a spiegare anomalie ed inefficienze di mercato attraverso essi. Questo risultato, che secondo molti coincide con la vera e propria nascita della finanza comportamentale intesa come branca organica della finanza, si deve a Richard Thaler e a Werner De Bondt che nel 1985 con il loro articolo "*Does the Stock Market Overreact?*" ottennero consenso trasversale sia dal mondo della psicologia sia da quello finanziario (Illiashenko, 2017).
Nel loro lavoro De Bondt e Thaler (1985) dimostrarono la forte presenza di *overreaction* sui mercati, un bias causato dall'euristica della disponibilità e dall'eursitica dell'ancoraggio, che fa sì che gli investitori, nel valutare i titoli, diano troppo peso alle informazioni recenti e ne diano troppo poco a quelle passate. Questo causa andamenti esagerati, in positivo e in negativo, dei titoli che nell'arco di tempo successivo tendono poi a ottenere un rendimento opposto.
Le ipotesi di De Bondt e Thaler furono: movimenti estremi nei prezzi sono seguiti da movimenti in senso contrario; più è estremo il movimento iniziale e più lo sarà anche il successivo aggiustamento. Per verificare queste ipotesi costruirono un *winner portfolio*, formato da titoli che nel breve periodo avevano overperformato battendo il mercato, e un *loser portfolio*, con titoli che nel breve periodo avevano ottenuto performance al di sotto del livello di mercato.

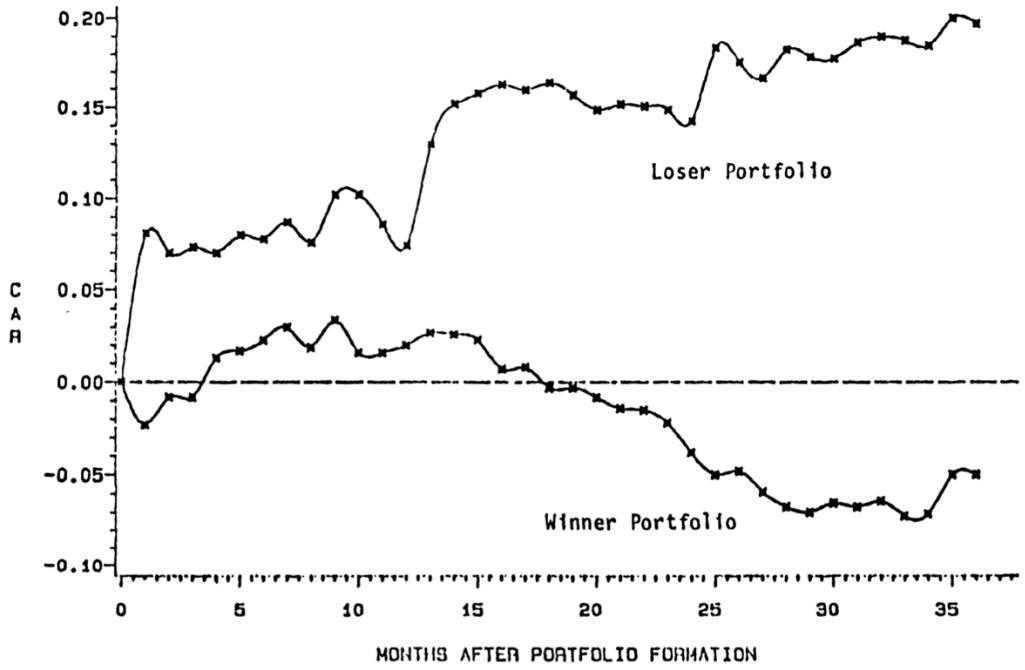


Figura 2.3: andamento del winner portfolio e del loser portfolio da 0 a 36 mesi dopo la loro formazione

La figura 2.3 mostra l'andamento dei due portafogli fino a 36 mesi dopo la loro formazione confermando in pieno le ipotesi formulate: entrambi hanno avuto un significativo aggiustamento nella direzione opposta dalla quale partivano.

Queste dinamiche di aggiustamento erano già state evidenziate in passato, ma l'importanza del lavoro di De Bondt e Thaler sta nel fatto che è stata fornita una spiegazione accettabile del fenomeno attraverso elementi di psicologia comportamentale (Illiashenko, 2017).

Negli anni '90 e nei primi anni del 2000 la finanza comportamentale ha raggiunto risultati sempre più importanti. Nel 1985 Shefrin e Statman scoprirono sui mercati finanziari il disposition effect citato precedentemente (Illiashenko, 2017), mentre nel 1995 Benartzi e Thaler, attraverso il concetto di miope avversione alle perdite, diedero una spiegazione all'equity premium puzzle. Nel 2000 Robert Shiller con il suo bestseller "Euforia irrazionale" predisse, fornendo spiegazioni attraverso concetti di finanza comportamentale, il crollo del mercato a causa dello scoppio della bolla delle dot-com.
Infine, la branca della finanza comportamentale guadagnò definitivamente riconoscibilità e rispetto dal mondo accademico in seguito all'assegnazione del Premio Nobel a Daniel Kahneman nel 2002 per "aver integrato la ricerca psicologica con la scienza economica, in particolare riguardo al giudizio umano e al processo decisionale in condizioni di incertezza" (Illiashenko, 2017).

Capitolo 3

Le bolle speculative

L'affermazione della finanza comportamentale e il progressivo superamento della teoria dei mercati efficienti hanno fornito gli strumenti per contestualizzare e comprendere meglio eventi anomali a cui periodicamente si assiste sui mercati finanziari.
Talvolta, l'inefficienza dei mercati e l'irrazionalità degli investitori, non provocano esclusivamente *market anomalies*, ma generano delle distorsioni strutturali con effetti potenzialmente molto pericolosi. Tra questi fenomeni il più famoso è sicuramente la bolla speculativa, ovvero un aumento insostenibile dei prezzi provocato dal comportamento di acquisto degli investitori e non da informazioni autentiche e fondamentali sul loro valore (Shiller, 2000). I mercati manifestano incredibili rialzi, anche per periodi di tempo piuttosto lunghi, senza che i fondamentali sottostanti giustifichino tale dinamica. Lo scoppio di una bolla speculativa genera un repentino crollo di mercato, spesso a ritmi più sostenuti di quelli osservati durante la crescita. Gli effetti sui mercati e, conseguentemente, sull'economia sono spesso dannosissimi.
La comprensione delle bolle speculative e di tutti gli aspetti che la caratterizzano è ancora oggi oggetto di studio all'interno del mondo accademico.
In questo capitolo saranno presentati i principali temi fino ad ora utilizzati nel settore finanziario per approcciarsi a questo fenomeno.

3.1 Cause

Capire il vero motivo per cui ha inizio una bolla speculativa è praticamente impossibile. Tutti i principali studiosi del tema, infatti, sono concordi nell'affermare che sarebbe riduttivo e superficiale trovare *una* causa scatenante. I principali esempi passati di bolle speculative da cui possiamo prendere esempio dimostrano che ci sono sempre più fattori, di diversa origine, che combinando tra loro creano una particolare situazione sui mercati da cui può formarsi una bolla. Se da un lato questo può far mancare una sorta di *certezza scientifica* sul problema in questione, dall'altro ne fa subito notare la sua ampia natura trasversale. In questo elaborato, prendendo spunto dall'impostazione utilizzata da Shiller (2000) nel suo "*Euforia irrazionale*", si presentano i fattori strutturali e culturali che possono far nascere o far accelerare lo sviluppo di una bolla, i meccanismi che ne possono amplificare l'effetto, e infine le dinamiche psicologiche che si nascondono dietro questo fenomeno finanziario.

3.1.1 Fattori strutturali e culturali

Tra i diversi fattori presentati da Shiller (2000), che ha pubblicato il suo lavoro poco prima dello scoppio della bolla delle dot-com, se ne riconoscono alcuni che prescindono dal periodo storico e dal contesto e sono adatti per rappresentare anche una situazione odierna:

- **I mass-media**: l'attenzione crescente dei media per i mercati finanziari e l'intensificazione di notizie economico-finanziarie generano un aumento nella domanda dei titoli, creando un meccanismo simile a quello che avviene con la pubblicità dei beni di consumo. I mezzi di comunicazione, nonostante si pongano come osservatori esterni e imparziali dell'andamento dei mercati, ne sono in realtà parte integrante (Shiller, 2000). Al giorno d'oggi ci sono servizi, via televisione o soprattutto su Internet, che forniscono un aggiornamento continuo sui mercati e su ogni notizia che possa ipoteticamente riguardarli. La competizione sempre più forte tra i diversi mass-media ha portato quest'ultimi ad una fruizione frenetica di notizie e dati, spesso presentati in modo *drastico* per attirare maggiormente l'attenzione alimentando, in certe situazioni, le manie speculative degli investitori. L'effetto dei mass-media si verifica anche quando l'oggetto delle notizie riportate non riguarda direttamente i mercati finanziari. Shiller (2000) propone un esempio utile per comprendere il concetto: il giorno dopo il terremoto del 1995 a Kobe in Giappone il Nikkei non registrò una caduta significativa, nonostante gli effetti di un tale evento avrebbero avuto sicuramente un forte impatto su molti settori dell'economia, ma circa una settimana più tardi il mercato giapponese iniziò a scendere e dieci giorni dopo aveva perso l'8% del suo valore, più del dovuto secondo stime fatte in seguito. Una spiegazione plausibile di questa dinamica riguarda il fatto che nei giorni successivi i media giapponesi hanno riportato incessantemente notizie sul terremoto generando un'ondata di attenzione verso quel tema e alimentando la preoccupazione e il pessimismo degli investitori che, presi dal panico, iniziarono a vendere in modo massiccio i titoli.[1]
- **Inflazione e illusione monetaria**: l'inflazione è uno dei temi economico-finanziari più chiacchierati e con più impatto sui mercati. È sufficiente pensare che l'obiettivo principale della Banca Centrale Europea sia proprio mantenere sotto controllo il livello di inflazione. Negli ultimi anni i mercati hanno sempre reagito in modo scorbutico a notizie riguardo l'indice dei prezzi, soprattutto in negativo al minimo accenno di aumento dell'inflazione, spaventati da un possibile rialzo dei tassi di interesse. Modigliani e Cohen (1979), sostengono che il mercato reagisce in modo inappropriato all'inflazione perché non ne capisce appieno l'effetto sui tassi di interesse (Shiller, 2000), attribuendo troppa importanza ai tassi di interesse nominali e troppo poca ai tassi di interesse reali. Queste distorsioni possono generare dei rialzi o dei ribassi di mercato non in linea con l'andamento dell'economia.
- **Livello delle negoziazioni**: scambiare titoli sul mercato è sempre più facile. Oltre ai più canonici servizi bancari, il numero di piattaforme di trading online sta crescendo di anno in anno creando una forte disintermediazione nel settore mobiliare e

[1] Quando il mercato subisce ingiustificati ribassi per lungo termine si parla di bolla negativa.

abbassando sensibilmente i costi di transazione. Questo genera un aumento costante delle contrattazioni e del livello di attenzione verso il mercato rischiando di ottenere gli effetti descritti nel primo punto: tra il 1982 e il 1999, poco prima dello scoppio della bolla delle dot-com, l'indice di rotazione dei titoli negoziati presso la borsa di New York è passato dal 42 all'88% (Shiller, 2000). Un altro aspetto da non sottovalutare legato alla crescente disponibilità di modalità per scambiare titoli è il fatto che un accesso così a portata di mano e semplice ai mercati rischia di portare sul mercato persone inesperte, i cosiddetti *naive investors* che distorcono l'efficienza del mercato.

- **Occasioni speculative**: Shiller (2000) sostiene che l'aumento delle agenzie da gioco e la maggiore frequenza del gioco d'azzardo hanno un impatto molto forte sulla nostra cultura e sugli atteggiamenti verso l'assunzione di rischio in altri campi come l'investimento azionario in quanto la scommessa induce a superare le naturali inibizioni nei confronti dei rischi. Le natura telematica dello scambio titoli e di molte tipologie di scommesse assottiglia l'apparente somiglianza tra le due attività. La scarsa considerazione del rischio crea distorsioni di mercato, come avvenuto, ad esempio, tra il 1929 e il 1933 quando sul mercato statunitense più che raddoppiò la volatilità e nello stesso periodo ci fu un boom del gioco d'azzardo (Grant, 1996), fattore che ha sicuramente contribuito al forte aumento dei prezzi.
- **Previsioni degli analisti**: negli esempi più famosi di bolle speculative le previsioni da parte degli analisti erano sempre più ottimiste con il passare del tempo. Questo può sicuramente essere dovuto in parte ad un errato giudizio della situazione dei mercati finanziari, ma spesso il principale motivo di questa dinamica è il conflitto di interessi tra gli analisti e le imprese che quest'ultimi devono analizzare e giudicare che potrebbero, in caso di rating negativo, escludere determinati analisti dagli incontri formativi (Shiller, 2000). Uno dei casi più famosi che ha riguardato questo conflitto di interessi è avvenuto durante la bolla immobiliare del 2008 quando le agenzie di rating continuavano ad assegnare un rating AAA ad aziende in pessimo stato di salute. Una ricerca della Zacks Investment Research ha mostrato che verso la fine del 1999, poco tempo prima dello scoppio della bolla delle dot-com, solo l'1% dei consigli dati ad un campione di 6000 clienti riguardava il vendere i titoli (Shiller, 2000). Lin e McNichols (1998) hanno invece analizzato la tendenza degli analisti, sempre a causa del conflitto di interesse, nel rivedere al ribasso le previsioni sugli utili di determinate aziende il giorno prima della pubblicazione del bilancio in modo da generare una sorpresa positiva sui mercati il giorno seguente.

3.1.2 Meccanismi amplificatori

I fattori descritti nel capitolo precedente creano delle distorsioni nei mercati azionari e il loro impatto può essere amplificato da alcuni meccanismi che, se innescati, aumentano il rischio di bolla speculativa. In questo capitolo si vedrà come queste dinamiche agiscono sulla fiducia degli investitori e il modo in cui si propagano seguendo una dinamica chiamata curva di retroazione.

Quando si è dentro una bolla speculativa, tendenzialmente il pubblico sembra avere una straordinaria fiducia nel mercato azionario. Shiller (2000), nel 1999, ha formulato un questionario formato da due domande e l'ha sottoposto a 147 intervistati, tra investitori istituzionali e investitori privati. Le due domande erano:

1. *Concorda con la seguente informazione? Il mercato azionario è la migliore forma di investimento per gli investitori di lungo periodo, i quali si possono permettere di acquistare e mantenere i titoli nel corso degli alti e bassi del mercato*.
2. *Quanto concorda con la seguente informazione? Se ci sarà un'altro crollo come quello del 19 ottobre 1987, il mercato tornerà certamente ai suoi livelli precedenti nel giro di un paio di anni.*

Alla prima domanda il 96% degli intervistati ha risposto di essere totalmente d'accordo o parzialmente d'accordo, mentre alla seconda domanda il 91% ha risposto *Accordo totale* o *Accordo parziale*. I risultati sono abbastanza emblematici, soprattutto considerando che poco tempo dopo il mercato crollò. È curioso notare come, nel secondo quesito posto da Shiller, la stragrande maggioranza degli intervistati fosse d'accordo nel credere in una rapida inversione di tendenza del mercato dopo un eventuale crollo: non avvertendo nessun pericolo di caduta gli investitori sono disponibili a comprare azioni anche se secondo indicatori solitamente utilizzati per la valutazione aziendale, come ad esempio il rapporto prezzo/utili, esse risultino decisamente sopravvalutate (Shiller, 2000). Ma il campanello d'allarme più significativo non è tanto la massiccia percentuale che si è trovata d'accordo con le domande poste da Shiller, quanto piuttosto il fatto che questa fiducia incondizionata cresca nel tempo. In un altro questionario proposto a diversi intervalli di tempo, infatti, Shiller fece a 145 intervistati la seguente domanda: "*Se domani il Dow Jones crollasse del 3%, cosa ritieni che succeda dopodomani all'indice?*". Nel 1989 il 35% rispose convinto che il mercato il giorno dopo salirebbe, nel 1996 tale percentuale salì al 46% e nel 1999 al 56%. Questa crescente convinzione nella forza del mercato, secondo Shiller (2000), è figlia del fatto che molti investitori stavano assistendo ad un'ininterrotta crescita del mercato dai minimi del 1982 e anche i più pessimisti, con il passare degli anni, rendendosi conto di avere ripetutamente torto, cambiarono idea. David Bell (1982), inoltre, analizzò il modo in cui il rammarico degli investitori sui mercati azionari fornisca una forte motivazione ad agire: il sentimento di invidia nel vedere altre persone guadagnare grosse somme senza un grosso sforzo abbassa l'autostima dell'individuo e lo convince che l'eventuale perdita che otterrebbe nel caso di cattivo investimento sarebbe comunque meglio della sensazione di rimanere fuori da quella che appare come una money machine senza freni. È possibile vedere un parallelismo con la situazione attuale, in cui la repentina crescita del mercato successivamente al crollo causato dal Covid-19, agisce su uno stato emotivo già molto vulnerabile di molti individui smorzando i loro freni inibitori e invitandoli a partecipare ad un rialzo che stenta a fermarsi. L'atteggiamento emozionale degli investitori quando decidono i loro investimenti è uno dei meccanismi che maggiormente amplifica i rialzi di mercato.

La natura auto-alimentante di queste dinamiche è alla base della teoria della curva di retroazione (*feedback loop theory*): gli aumenti iniziali dei prezzi, causati ad esempio da uno dei fattori presentati nel paragrafo precedente, provocano ulteriori aumenti dei prezzi a mano a mano che gli effetti degli aumenti dei prezzi iniziali agiscono retroattivamente su prezzi ancora più alti attraverso una maggiore domanda degli investitori (Shiller, 2000). Sostanzialmente gli investitori, stimolati dai prezzi in aumento, offrono prezzi ancora più alti invogliando a loro volta altri investitori a entrare sul mercato a causa di meccanismi emozionali spiegati precedentemente. Questo *loop* si rinnova di continuo amplificando l'effetto del fattore iniziale che aveva causato il rialzo.
In una delle versioni più riconosciute della feedback loop theory, la retroazione si verifica non a causa di un meccanismo emozionale, ma a causa di aspettative adattive, ovvero la retroazione avviene perché i passati aumenti di prezzi generano negli investitori aspettative di ulteriori aumenti (Barberis, Shleifer, Vishny, 1998). Secondo Barberis, Shleifer e Vishny (1998), inoltre, nel momento in cui qualche dinamica imprevedibile nella risposta degli investitori ai movimenti di prezzi provochi uno stop nell'aumento del mercato, ci si deve aspettare un crollo di quest'ultimo.

3.1.3 Fattori psicologici

Nell'approfondire i meccanismi amplificatori e la teoria della retroazione ci si è resi conto che per comprendere le dinamiche dietro la nascita e lo sviluppo di una bolla non è sufficiente ricercare uno o più fattori scatenanti, ma è necessario far riferimento a elementi apparentemente distanti dai mercati e, infine, rivolgersi alla psicologia comportamentale. I concetti presentati nel capitolo precedente, infatti, trovano nelle bolle speculative l'habitat di maggior espressione. In ogni caso passato di mania speculativa è difficile fornire una spiegazione dell'accaduto senza appoggiarsi a fattori psicologici.
Nonostante molti luoghi comuni descrivano gli investitori, durante un boom o un crollo, come individui totalmente irrazionali che si esaltano o terrorizzano, è noto da numerose ricerche che in situazioni di crisi gli investitori si affannano a fare la cosa giusta, ma hanno abilità limitate e istintivamente tendono a comportamenti inefficienti. Ma è importante sottolineare che questi modelli di comportamento non sono il risultato dell'estrema ignoranza dell'uomo, ma della sua intelligenza, riflettendone i limiti e la forza (Shiller, 2000).
In questo paragrafo saranno presentati gli elementi di psicologia che entrano maggiormente in gioco durante una bolla speculativa: il concetto di ancora, già menzionato nel capitolo precedente, l'overconfidence e i comportamenti gregari.

Ancore
Il concetto di ancora è legato all'euristica dell'ancoraggio presentata nel capitolo precedente: nel momento in cui si fornisce la probabilità di accadimento di un evento si tende a essere influenzati da un termine di paragone a cui ci si ancora. Le ancore possono essere:

- *Quantitative*: ci si lega ad un'informazione di natura quantitativa. Nei mercati finanziari un esempio di ancora quantitativa è il prezzo passato di un titolo:

nell'esprimere valutazioni sul livello dei prezzi azionari l'ancora sembra essere costituita dal prezzo più recente di cui ci si ricordi e nell'ambito di una decisione di investimento si considera più quello rispetto ad informazioni più utili per la valutazione aziendale. La tendenza degli investitori a servirsi di questa ancora rafforza la somiglianza dei prezzi giorno dopo giorno (Shiller, 2000).

- *Morali*: in situazioni di tensione finanziaria gli investitori faticano a basarsi su grandezze quantitative per prendere le loro decisioni, ma tendono ad ancorarsi ad una motivazione morale per operare la loro scelta di investimento. Secondo questa nozione gran parte del pensiero umano che sfocia in azione non è quantitativo, ma prende la forma di narrazione e giustificazione: sembra che la gente abbia necessità di trovare delle motivazioni per giustificare le loro decisioni e renderle accettabili e coerenti agli altri e a loro stessi (Shiller, 2000). Questa dinamica si lega molto bene al concetto di *fallacia narrativa* presentata da Nassim Taleb ne "Il cigno nero" (2007), ovvero la tendenza a costruire senza motivo storie intorno ai fatti per tentare di spiegarli. Secondo Taleb percepiamo continuamente narratività e causalità in ciò che succede per ridurre la dimensionalità del problema a causa della nostra grande fatica a comprendere ciò che ci sembra astratto. Un esempio intuitivo di questa fallacia è fornito da Taleb sempre ne "Il cigno nero" (2007): il giorno della cattura di Saddam Hussein, Bloomberg News diede la seguente notizia: "Titoli di stato in rialzo: la cattura di Saddam Hussein potrebbe non fermare il terrorismo" e mezz'ora dopo, in seguito ad un'apparente inversione della tendenza di mercato, diede la notizia: "Titoli di stato in ribasso: la cattura di Hussein aumenta l'attrattiva dei titoli ad alto rischio". La stessa notizia fu utilizzata per cercare di spiegare un evento finanziario (il rialzo dei titoli) ed il suo esatto opposto: fornire sempre una causa a ciò che succede aiuta a rendere gli argomenti più concreti.

  Durante i grandi rialzi di mercato i media spesso cavalcano l'entusiasmo fornendo continuamente storie che alimentano il contesto di euforia aiutando gli investitori a giustificare i continui acquisti di titoli e spingendo ancora più su la domanda. La stessa dinamica si evidenzia nel caso di crolli di mercato, o bolle negative, in cui le ancore morali e la fallacia narrativa aumentano il panico degli investitori.

Overconfidence

L'utilizzo di ancore incrementa ulteriormente l'eccessiva fiducia che le persone tendono ad avere in loro stessi e in ciò che credono. Fischof, Slovic e Lichtenstein (1977) hanno dimostrato, ponendo semplici domande fattuali a una serie di intervistati, che si tende spesso a sopravvalutare la probabilità di avere ragione: quando gli intervistati si dichiaravano certi di qualcosa, ad esempio, avevano ragione solo nell'80% dei casi. Una delle spiegazioni fornite per spiegare questa *overconfidence* si basa sul fatto che le persone, nel momento in cui valutano la correttezza delle proprie conclusioni, considerano solo l'ultima fase del loro ragionamento dimenticando altri elementi che potrebbero essere sbagliati (Pitz, 1975).

Se le persone fossero completamente razionali, infatti, la metà degli investitori sarebbe consapevole di essere al di sotto della media per quanto riguarda la capacità di negoziazione e non sarebbe quindi propensa a svolgere trattative speculative con la metà

più capace (Shiller, 2000). Barberis, Shleifer e Vishny (1998) hanno messo insieme i concetti di euristica della rappresentatività e di overconfidence sostenendo che gli investitori, quando vedono i prezzi seguire la stessa direzione, iniziano a pensare che la tendenza sia rappresentativa anche di altri dati economici che hanno osservato e rafforzano le loro opinioni. Questo, legato al principio di conservazione secondo cui cambiamo opinione lentamente, determina lo sviluppo di una retroazione speculativa. Shafir e Tversky (1992), inoltre, hanno dimostrato che, in molte dinamiche quotidiane, gli individui non sono in grado di ragionare consequenzialmente e non considerano quasi mai le ramificazioni logiche conseguenti a loro decisioni. Per questo motivo, gli effetti delle notizie sul mercato finanziario hanno più a che fare con il modo in cui ci si sente di fronte alle notizie che non con le reazioni logiche alle stesse (Shiller, 2000).

Effetto gregge ed epidemie finanziarie

In finanza l'effetto gregge indica una dinamica in cui il comportamento di un individuo si omologa a quello degli altri componenti del gruppo perché influenzato da essi. È considerato uno degli aspetti psicologici più importanti per la formazione di bolle speculative sui mercati. Se un pensiero non meccanicistico o irrazionale fosse comune ad un gran numero di persone, infatti, potrebbe essere fonte di euforia o depressione del mercato azionario (Shiller, 2000).

Le dinamiche che portano alla formazione di comportamenti gregari sono state oggetto di studi di psicologi e sociologi per molti anni, e lo sono tutt'ora. Tra le principali motivazioni individuate negli anni ci sono:

- *Influenza e pressione sociale*: secondo lo psicologo sociale Solomon Asch (1952), la pressione sociale influisce molto sul giudizio individuale. Asch organizzò un esperimento per dimostrare questo legame: inserì un soggetto in un gruppo di 7-9 persone che erano state raggruppate e istruite su come comportarsi all'interno del gruppo. Fu chiesto ai membri di rispondere a semplici domande riguardo la lunghezza di segmenti che venivano mostrati. Ciascun soggetto, prima di fornire la sua risposta, ascoltava quella di chi lo precedeva. I collaboratori, seguendo le istruzioni, sbagliarono apposta alcune risposte. In un terzo dei casi, nonostante il palese errore, il soggetto si adeguò e diede le stesse risposte errate fornite dal resto del gruppo. La pressione di sentirsi in difetto rispetto ad un gruppo di persone contribuisce sicuramente alla formazione di questo fenomeno, ma Deutsch e Gerard (1955) sostennero che il motivo principale che aveva portato il soggetto ad allinearsi agli errori era il fatto di aver pensato che le altre persone non potevano essersi sbagliate reagendo all'informazione che un gran numero di persone aveva espresso un giudizio diverso dal suo. Secondo Shiller (2000), questo comportamento è una questione di calcolo razionale dovuto al fatto che abbiamo imparato, nella vita di tutti i giorni, che quando un cospicuo numero di persone è d'accordo nel valutare qualcosa, gli appartenenti a quel gruppo hanno quasi sicuramente ragione. Altri esperimenti dello psicologo Stanley Milgram (1974) dimostrarono l'enorme potere dell'autorità sulla mente umana.

  In conclusione, questi esperimenti, insieme a molti altri che sono seguiti a quelli

proposti, testimoniano il fatto che le persone sono pronte a credere al punto di vista maggioritario e alle autorità anche quando sono in netta contraddizione con valutazioni fattuali (Shiller, 2000). I mercati azionari sono un terreno fertile per lo sviluppo di queste dinamiche generando movimenti dei prezzi del tutto ingiustificati, soprattutto quando si è già all'interno di una fase speculativa.

- *Cascata di informazioni*: le teorie sulla cascata di informazioni sostengono che anche soggetti individualmente razionali possono assumere un comportamento gregario quando considerano in modo eccessivo i giudizi altrui, anche se sono consapevoli del fatto che è in atto un comportamento gregario. L'applicazione sui mercati finanziari di questa teoria spiega che molti investitori hanno la convinzione che il livello del mercato azionario sia il risultato di un giudizio collettivo espresso da tutti gli investitori e scelgono così di accodarsi a tale giudizio senza effettuare un vero e proprio sforzo valutativo. L'effetto gregge si manifesta perché non ci si rende conto che anche molti altri soggetti hanno ragionato allo stesso modo, contribuendo a portare il mercato ad un livello che non è dovuto a giudizi di merito, ma a dinamiche psicologiche.
- *Trasmissione di informazioni e passaparola*: al giorno d'oggi le informazioni viaggiano ad una velocità estrema, spesso incontrollabile. Le modalità di trasmissione delle informazioni sono sempre di più grazie alla costante crescita dei social e della tecnologia. Le informazioni che tendono a fluire più rapidamente sono quelle inerenti ad aspetti utili alla società nella vita quotidiana: per questo motivo, nell'ambito dei mercati finanziari, le informazioni che viaggiano di più e più velocemente non riguardano statistica, economia o matematica finanziaria, quanto piuttosto opportunità imminenti di guadagno o perdita. In questo contesto, Pound e Shiller (1986) hanno dimostrato, in un questionario rivolto a investitori individuali, che solo il 6% di essi era stato attratto dai titoli che aveva acquistato grazie a notizie prese da periodici o giornali, ma la maggior parte degli intervistati indicò che aveva ottenuto fonti da comunicazioni interpersonali dirette. Questi meccanismi di *passaparola* rendono meno controllabile la trasmissione delle informazioni e il loro effetto sui mercati finanziari.

3.2 Teoria della riflessività

3.2.1 Struttura della teoria

Tra le più importanti e riconosciute teorie sulle bolle speculative è sicuramente presente la teoria della riflessività sviluppata da George Soros, secondo cui gli investitori non basano le loro decisioni sulla realtà, ma sulla loro percezione della realtà e le loro percezioni hanno il potere di influenzare i mercati finanziari cambiando a loro volta le percezioni di altri investitori.
La teoria della riflessività ha origine nelle scienze sociali e Soros negli anni ne ha fornito un'interpretazione nel campo economico-finanziario per spiegare il fenomeno delle bolle

speculative, riprendendo anche concetti della finanza comportamentale e della teoria di retroazione delle bolle.

Per comprendere le radici della teoria della riflessività è necessario introdurre brevemente l'aspetto epistemologico del pensiero di Karl Popper, filosofo di cui Soros è stato allievo.
Secondo Popper, la più importante caratteristica della scienza è la sua falsificabilità. Tutto ciò che è scienza non ha a che fare con la verità, ma con semplici congetture che, una volta confutate, contribuiranno al vero progresso scientifico. Non è possibile ottenere una conoscenza certa siccome non si può mai escludere l'errore. Il processo di congetture e confutazioni permette il raggiungimento di teorie sempre più verosimili che forniscono una maggiore comprensione del mondo. Tutto ciò che non è confutabile, secondo Popper, non è scienza, ma metafisica.

Questo impianto logico popperiano ha influenzato profondamente Soros, la cui teoria della riflessività si appoggia sulla falsificabilità di Popper. Soros, infatti, si distacca sin da subito dalla visione positivista secondo cui gli agenti economici sono razionali nel comprendere la realtà attorno a loro.
Secondo Soros, ciascun essere umano si relaziona davanti ad ogni situazione attraverso due funzioni: la funzione cognitiva, con cui si cerca di comprendere e interpretare la realtà; e quella manipolativa, attraverso cui gli agenti cercano di intervenire ed interagire con la situazione. La teoria della riflessività prevede la concomitanza di queste due funzioni che, agendo l'una sull'altra, innescano un meccanismo circolare che può portare ad un progressivo allontanamento dalla realtà. Tale meccanismo è innescato da un evento che, modificando lo stato della realtà, influenza la funzione cognitiva e quindi le opinioni degli agenti. Di conseguenza, la modifica delle opinioni degli agenti influenza la funzione manipolativa e quindi le loro scelte, che a loro volta influiranno sulle opinioni di altri agenti, e così via. Questo meccanismo può generare un feedback positivo, provocando un progressivo allontanamento dalla realtà, oppure un feedback negativo, che permette di riavvicinarsi parzialmente alla realtà.

3.2.2 Applicazione ai mercati finanziari

Secondo Soros (1987), le aspettative e le opinioni che abbiamo riguardo gli eventi futuri non attendono che quegli eventi si realizzino, ma possono cambiare in qualsiasi momento modificandone l'esito. Nei mercati finanziari accade esattamente questo: l'essenza dell'investimento sta nel cercare di prevedere il futuro, ma il prezzo che gli investitori sono disposti a pagare oggi per un'azione può influenzare in diversi modi le sorti dell'azienda interessata. Di conseguenza, il modificarsi delle aspettative presenti influisce sul futuro che viene previsto. Nei mercati finanziari speculativi, molto spesso si acquista un titolo non perché si è realmente interessati al suo utilizzo (ad esempio partecipando alle assemblee dei soci), ma perché si crede che il prezzo salirà e che si potrà quindi realizzare un profitto. Se questo ragionamento viene applicato da molti investitori contemporaneamente allora il prezzo di un titolo salirà in modo importante non grazie ad una reale crescita dell'azienda,

ma grazie all'aspetto auto-avverante di questo meccanismo. Il valore speculativo dei titoli supera il loro valore intrinseco, e questo può essere la causa dell'alternanza di momenti di euforia e panico sui mercati (Intropido, 2014).
La profonda irrazionalità che caratterizza alcuni aspetti degli agenti economici, dovuta a euristiche e bias, può portare alla nascita di tendenze che drogano il mercato provocando l'innesco per meccanismi riflessivi. Secondo la teoria della riflessività è l'insieme di questi aspetti che provoca la genesi, lo sviluppo e lo scoppio di una bolla speculativa (French, 1992): la funzione cognitiva determina la percezione di appetibilità di un titolo azionario, prevedendo che le quotazioni saliranno. La funzione manipolativa, guidata dall'interpretazione cognitiva della realtà, interagisce attivando un comportamento attivo di acquisto del titolo aumentandone la domanda. La riflessività di questo meccanismo fornisce un'illusione auto-avverante dell'ipotesi iniziale rinforzandone le basi. Questo comportamento è, inoltre, fortemente amplificato da comportamenti di gregge e dal cosiddetto *bandwagon effect*, conseguenza della feedback loop theory presentata precedentemente. Si verifica, nel tempo, la nota spirale domanda-prezzo in cui la prima traina il secondo, indipendentemente dall'effettiva capacità di quest'ultimo di riflettere le caratteristiche ed il valore reale del titolo di riferimento (Castellani, 2014).
Ad un certo punto, parte dei soggetti economici inizia a rendersi conto dell'ingiustificabilità di una tendenza così marcata e cambia opinione iniziando ad invertire il verso di azione della teoria della riflessività. Nell'ultima fase della bolla i diffidenti aumentano di numero e la tendenza continua ad aumentare ad un ritmo più basso e più volatile, richiedendo oltretutto sempre maggior indebitamento e leva finanziaria a causa degli alti costi di ingresso sul mercato.
Lo stesso circolo vizioso che aveva contribuito all'espansione della bolla ne causa, man mano che i pessimisti aumentano di numero, il suo scoppio, che quasi sempre denota un tasso di decrescita maggiore rispetto al tasso di crescita osservato durante lo sviluppo della bolla.

3.3 Bolle speculative nella storia

Le prime bolle speculative hanno un'origine molto antica, precedente allo sviluppo finanziario delle società moderne.
La prima famosa bolla risale al XVII secolo quando, tra il 1634 e il 1637, si sviluppò in Olanda, la cosiddetta *tulipmania* (febbre dei tulipani). In quel periodo un virus non letale colpì alcuni tulipani olandesi attribuendo loro un particolare aspetto. Questa peculiarità aumentò esponenzialmente l'interesse per questa specie di tulipani, che iniziò ad essere scambiata pesantemente in borsa. Venivano acquistati tulipani solo con lo scopo di rivenderli dopo poco tempo, con la sicurezza che il prezzo sarebbe salito. Nel picco della bolla il valore di mercato di un bulbo superò quello di uno stipendio medio annuale di una famiglia olandese. Nel 1637, in seguito ad un'asta deserta, iniziò a diffondersi il panico e i prezzi dei tulipani crollarono vertiginosamente causando una profonda crisi nella società olandese.

Negli anni ’20 del 1700, fu il turno della bolla riguardante la Compagnia del Mississipi e, qualche anno dopo, la bolla dei Mari del Sud con la Sowth Sea Company. In entrambi i casi, lo scarso sviluppo dei mercati finanziari e una forte asimmetria informativa crearono un effetto distorsivo su questi titoli che salirono vertiginosamente per poi crollare. Successivamente ci furono i primi interventi regolatori sui mercati finanziari che servirono a diminuire la frequenza e l’impatto di questi eventi. Tuttavia, i mercati, ciclicamente, ripresentano situazioni estremamente anomale in cui trovano spazio bolle speculative.
Solo nell’ultimo secolo il mercato statunitense (e collateralmente anche altri mercati) hanno assistito alla nascita, sviluppo e scoppio di bolle speculative che hanno avuto conseguenze significative negli anni successivi per l’economia del Paese.
Le espansioni più marcate del mercato azionario sono state spesso associate alla percezione per cui il futuro è più roseo e meno incerto del passato. Shiller (2000) utilizza l’espressione *nuova era* per descrivere questi periodi di forte entusiasmo sui mercati azionari.
Tra le bolle speculative più famose a partire dal 1900 è sicuramente presente la bolla di fine anni Venti. Durante quel periodo ci fu una rapida crescita economica e si diffusero alcune innovazioni tecnologiche che contribuirono a creare entusiasmo e rialzi sul mercato azionario che, da circa metà del 1928, iniziarono a sembrare del tutto ingiustificati dai fondamentali dell’economia reale. Il 24 ottobre del 1929 (giovedì nero di Wall Street) la bolla scoppiò e 13 milioni di azioni furono vendute senza limite di prezzo. Il mercatò crollo violentemente e ne seguì un periodo di crisi, denominato Grande Depressione, che mise in ginocchio gli Stati Uniti. Nella figura 3.1 è mostrato l’andamento del Dow Jones alla fine degli anni ’20 ed è cerchiato in rosso il periodo caratterizzato dalla bolla speculativa.

Figura 3.1: andamento del Dow Jones a fine degli anni '20

Un altro caso estremamente famoso è rappresentato dalla bolla delle dot-com, il cui scoppio è stato previsto da Shiller nel suo “*Euforia irrazionale*”. Da circa metà degli anni ’90, le forti innovazioni tecnologiche che seguirono lo sviluppo di internet, diedero il via ad un nuovo ciclo economico chiamato *new economy*, in cui assunsero un ruolo centrale le cosiddette

dot-com companies, ovvero aziende del settore tecnologico e informatico, la cui crescita sembrava inarrestabile. Negli anni successivi si alimentò una forte euforia sui mercati riguardo questi titoli nonostante i tradizionali indicatori di redditività non solo non giustificavano tali comportamenti, ma anzi spesso segnalavano addirittura trend opposti. Nella prima metà del 2000, molti bilanci delle aziende tecnologiche mostrarono risultati estremamente deludenti, invertendo col passare dei mesi la tendenza sul mercato azionario. L'indice Nasdaq 100 perse il 70% in circa un anno, come mostra la figura 3.2.

Figura 3.2: andamento del Nasdaq durante la bolla delle dot-com

Un ultimo caso, più recente e di cui alcune economie risentono ancora oggi, è la bolla immobiliare che trovò nei mercati finanziari il veicolo perfetto per espandersi enormemente. Le procedure di cartolarizzazione comportarono il passaggio del modello di business delle banche da *originate and hold* a *originate and distribute*. I mercati iniziarono ad essere inondati di nuovi derivati composti da pacchetti di mutui sempre più scadenti, i cosiddetti mutui subprime. Con il passare degli anni gli investitori e le banche si staccarono completamente da una valutazione attenta dei fondamentali degli strumenti scambiati e contribuirono ad un'incredibile distorsione del mercato. Dalla seconda metà del 2007 e per tutto il 2008 molti dei mutui su cui erano costruiti i derivati scambiati divennero insolventi e crollò di conseguenza tutto il castello finanziario costruito alle spalle. La crisi che ne seguì fu devastante per molti paesi, e in molti casi dovettero intervenire le rispettive banche centrali con aiuti sostanziosi.

Capitolo 4

Tecnologie utilizzate

4.1 Cluster analysis

La cluster analysis è un insieme di tecniche statistiche atte ad individuare gruppi di unità tra loro simili rispetto ad un insieme di caratteristiche prese in considerazione (Fabbris, 1983). L'obiettivo della cluster analysis è quindi quello di analizzare un insieme di elementi sulla base di determinati attributi e dividerli in gruppi il più possibile omogenei.

Esistono diverse tecniche e algoritmi di clustering. Una prima classificazione più generale riguarda l'approccio iniziale ai dati:

- *Algoritmi di clusterizzazione agglomerativi*: iniziano inserendo ogni oggetto dell'insieme in un proprio cluster per poi raggrupparli iterativamente fino al raggiungimento del numero di clusters desiderato.
- *Algoritmi di clusterizzazione divisivi*: iniziano inserendo tutti gli oggetti dell'insieme in un unico cluster per poi separarlo iterativamente in cluster più piccoli fino al raggiungimento del numero di clusters desiderato.

Ciascun algoritmo utilizza una sua personale metrica geometrica per capire quanto due oggetti siano *simili* tra loro. Ciascun elemento è visto dall'elaboratore come un vettore di valori rappresentato dalle sue caratteristiche e per ogni coppia di vettori viene calcolata la *distanza* intesa come grado di somiglianza.

Le principali metriche di distanza utilizzate sono:

- *Distanza euclidea*:

$$d(x,y) = \left(\sum_{i=1}^{n} (x_i - y_i)^2\right)^{1/2}$$

- *Distanza di Manhattan*:

$$L_1(P_1,P_2) = |x_1 - x_2| + |y_1 - y_2|$$

Sono presenti numerosi algoritmi di clustering, ciascuno con le sue peculiarità e i suoi punti di forza. La scelta tra i vari algoritmi varia in base alla capacità computazionale del sistema hardware, al grado di linearità dei dati e al numero di gruppi che si desidera ottenere alla fine dell'analisi. In questo elaborato viene presentato l'algoritmo K-means, in quanto ben rappresentativo di come una tecnica di clustering lavori sui dati. Un maggior approfondimento su altri algoritmi esula dalla finalità di questo elaborato, in quanto le differenze implementative a livello di codice non sono ampie.

Il K-means è un algoritmo iterativo che prevede la specificazione a priori del numero di clusters desiderato e che si appoggia al concetto di centroide. Il centroide è un punto dello spazio che ha la funzione di *rappresentare* un cluster. A livello matematico corrisponde al punto medio degli elementi del cluster in questione, come mostra la Figura 4.1.

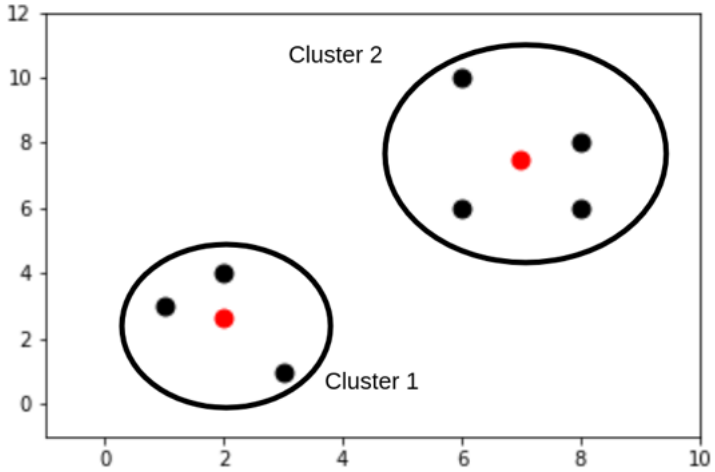


Figura 4.1: centroidi dei clusters

L'algoritmo K-means può essere descritto dai seguenti passi:

1. Scegliere il numero K di clusters da creare
2. Selezionare casualmente K centroidi all'interno dello spazio dimensionale
3. Calcolare la distanza tra ogni centroide e tutti gli oggetti attraverso la metrica di distanza utilizzata
4. Assegnare ogni elemento al cluster rappresentato dal centroide più vicino
5. Ricalcolare i centroidi di ogni cluster calcolando la media degli esempi
6. Ripetere dal punto 2 finché nessun elemento cambia cluster di appartenenza

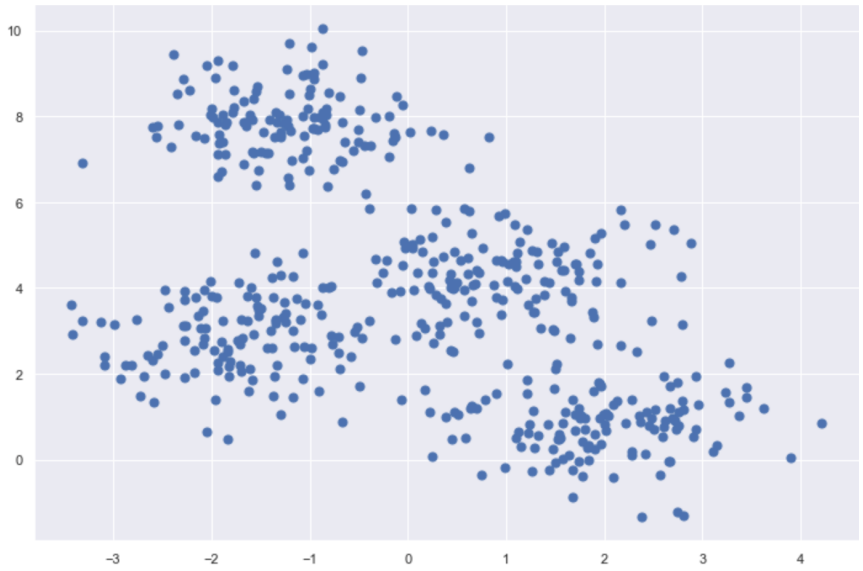


Figura 4.2: dati pre clustering

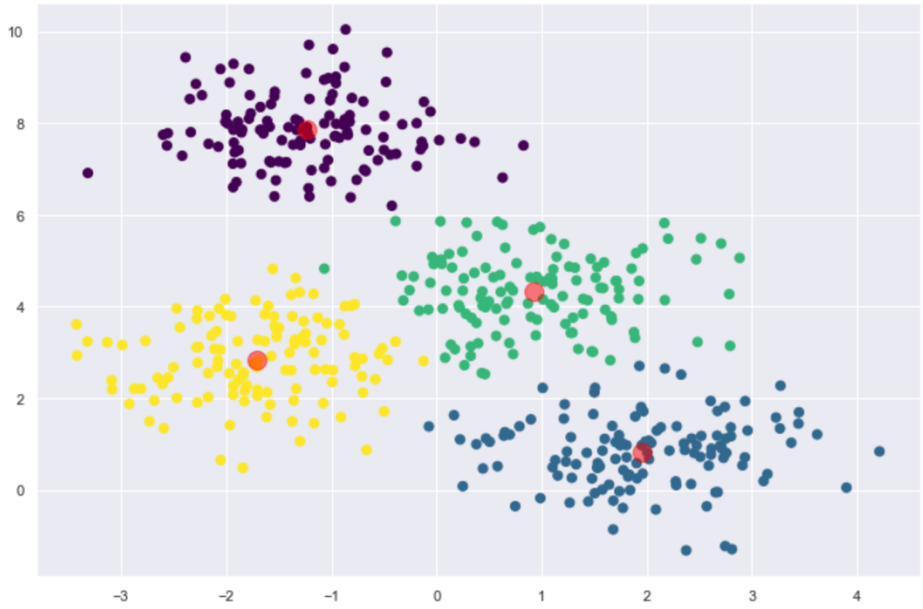


Figura 4.3: dati post clustering

Nella Figura 4.2 sono rappresentati su un piano cartesiano un insieme di punti da sottoporre alla Cluster Analysis utilizzando come features per l'analisi il valore sull'asse x e il valore sull'asse y (che rappresentano le due caratteristiche di interesse degli elementi sulla base di cui si vuole dividerli in gruppi omogenei). La Figura 4.3 mostra il risultato ottenuto dopo aver eseguito un clustering K-means sui dati, avendo scelto di dividerli in 4 clusters. All'interno di ogni gruppo è presente un punto rosso che rappresenta il centroide.
In questo esempio il vettore di caratteristiche degli elementi su cui effettuare il clustering è formato da due features per rendere graficamente rappresentabile il risultato, ma in contesti reali vengono considerate più caratteristiche e il meccanismo di calcolo della distanza e di raggruppamento non cambia. La scelta ottimale del numero di clusters da creare dipende dal problema che si deve affrontare: più caratteristiche dei dati si considerano per effettuare

la cluster analysis più clusters sarà necessario creare per cogliere tutte le differenze tra i gruppi e raggiungere un grado accettabile di omogeneità all'interno di ogni cluster.

Sono meritevoli di citazione altri due algoritmi di clustering molto utilizzati: l'Agglomerative clustering, che ha il vantaggio di non dover specificare in anticipo il numero di clusters; e il DBSCAN, adatto ad individuare e clusterizzare dati con relazioni non lineari, e molto efficace nel rilevare gli outliers. Per questo motivo il DBSCAN è molto utilizzato nel campo dell'anomaly detection.

4.2 Deep learning e reti neurali

4.2.1 Introduzione all'intelligenza artificiale

La nascita dell'intelligenza artificiale risale al 1956, anno in cui, durante il seminario presso il Dartmouth College nel New Hampshire, sono state programmate le prime basi della materia partendo dai contributi sviluppati negli anni precedenti da numerosi accademici. In aggiunta alla più tradizionale ricerca informatica utilizzata per costruire i pilastri della materia, è importante sottolineare che altre discipline come la filosofia, la matematica e la psicologia hanno contribuito allo sviluppo dell'intelligenza artificiale modellandone i contorni. L'obiettivo dell'intelligenza artificiale è creare sistemi hardware e software che siano in grado di fornire prestazioni apparentemente esclusive dell'intelligenza umana, sia nei risultati ottenuti, sia nei meccanismi utilizzati per arrivare a tale risultato: l'inferenza, la deduzione, la generalizzazione, la particolarizzazione, la valutazione di ipotesi e l'apprendimento.

Data la complessità nel definire un concetto universale di *intelligenza*, negli anni si è cercato di identificare diverse caratteristiche associabili all'intelligenza umana in modo che computer dotati di intelligenza artificiale cercassero di riprodurre una o più di queste caratteristiche (*Narrow artificial intelligence*). Dotare un'unica macchina di tutte queste facoltà (*Artificial General Intelligence*) è infatti ancora oggi poco meno che un'utopia.
A partire dalla caratteristica dell'intelligenza naturale tentata di replicare dall'elaboratore si sono create diverse branche dell'intelligenza artificiale:

Caratteristiche intelligenza naturale	Branca intelligenza artificiale
Percezione	Computer vision, robotica
Esplorazione	Robotica
Comunicazione	Natural Language Processing
Ragionamento / Problem solving	Pianificazione automatica
Apprendimento	Machine Learning

Al giorno d'oggi le applicazioni dell'intelligenza artificiale sono innumerevoli sia come numero assoluto sia considerando lo spettro dei settori interessati, e secondo molti studiosi le potenzialità della materia sono ancora moltissime. Nella quotidianità siamo spesso immersi in sistemi di intelligenza artificiale senza neanche rendercene conto: assistenti virtuali come Siri e Alexa, quasi tutti i servizi di Google, i sistemi di raccomandazione utilizzati da Amazon, Netflix e Youtube e praticamente tutti i videogiochi, solo per fare qualche esempio.

4.2.2 Machine Learning

Il machine learning è una branca dell'intelligenza artificiale che si pone l'obiettivo di fornire all'elaboratore la capacità di apprendere direttamente dall'esperienza, senza dover scrivere un apposito algoritmo per ogni ipotetico compito da svolgere.
La programmazione classica, basata appunto sul concetto di algoritmo, prevede che i computer e i programmi vengano guidati nei loro compiti da un insieme di istruzioni sequenziali e logiche nel risolvere i diversi problemi per cui sono pensati. Nel corso del tempo ci si è però accorti che anche i più sofisticati algoritmi non riuscivano a gestire la crescente complessità dei problemi con cui si ha a che fare quotidianamente. Pensare di scrivere una serie di istruzioni logiche e sequenziali per insegnare ad un'automobile a guidarsi da sola, ad esempio, è impossibile a causa dell'incredibile numero di variabili che entrano in gioco.
Il machine learning ha rappresentato la soluzione a questi problemi, grazie ad un approccio opposto a quello della programmazione classica: anziché scrivere un algoritmo a partire dai dati di input per ottenere l'output desiderato, l'idea è utilizzare i dati di input e output insieme sin dall'inizio, lasciare che la macchina comprenda le relazioni tra i dati e far sì che sia direttamente lei a scrivere l'algoritmo. La figura 4.4 schematizza questa differenza di approccio. Proprio per questo motivo le due fasi principali del machine learning sono l'apprendimento e la predizione.

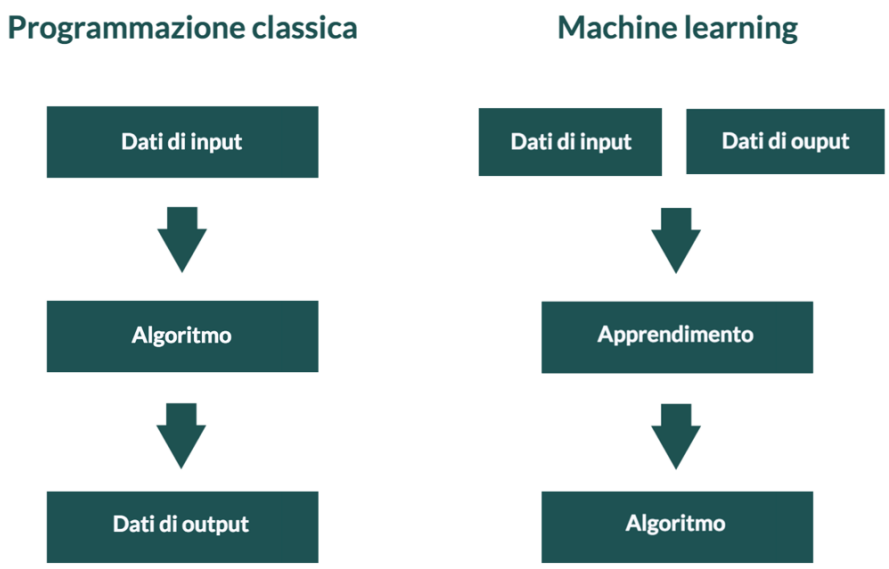


Figura 4.4: approccio classico vs machine learning
Fonte: https://blog.profession.ai

Un classico esempio per comprendere più nel dettaglio l'approccio radicalmente diverso del machine learning riguarda il riconoscimento di immagini. Se si volesse ottenere un programma in grado di riconoscere che animale è rappresentato in una foto attraverso istruzioni logiche e sequenziali e con i costrutti della programmazione classica non si otterrebbe un risultato soddisfacente a causa del numero e della complessità di casistiche e condizioni da verificare; cercando di risolvere il problema con il machine learning, invece, durante la fase di apprendimento la macchina studia le coppie di input (le singole foto) e di output (un'etichetta che indica l'animale rappresentato) imparando a riconoscere le diverse strutture e caratteristiche dei dati associate ad ogni animale (ad esempio i valori dei pixel) creando delle rappresentazioni interne. Una volta finita questa prima fase, durante la predizione la macchina utilizzerà le relazioni imparate per classificare automaticamente nuove immagini non etichettate. L'idea di fondo è lasciare che sia la macchina a programmarsi da sola accumulando esperienza dai dati che analizza trasferendo al computer la gestione della complessità del problema.

Si identificano tre grosse categorie all'interno del machine learning, a seconda dei dati a disposizione e dell'obiettivo prefissato:

- **Apprendimento supervisionato**: vengono presentati al computer degli input di esempio ed i relativi output desiderati, con lo scopo di apprendere una regola generale in grado di mappare gli input negli output. La macchina, durante la fase di addestramento, impara a riconoscere le relazioni e le dinamiche dei dati di input e ad associarli al corretto output.
- **Apprendimento non supervisionato**: al computer vengono forniti solo dei dati in input, senza alcun output atteso, con lo scopo di apprendere una qualche struttura nei dati d'ingresso. L'apprendimento non supervisionato può essere utilizzato per scoprire dei pattern nascosti nei dati oppure per estrapolare le principali

caratteristiche (features) di un insieme di dati e utilizzare quanto appreso per un altro task di machine learning.

- **Apprendimento per rinforzo**: il computer interagisce con un ambiente dinamico nel quale deve raggiungere un certo obiettivo (ad esempio, guidare un'automobile o affrontare un avversario in un gioco). Man mano che il computer esplora il dominio del problema gli vengono forniti dei feedback in termini di ricompense o punizioni a seconda del successo conseguente a azioni intraprese, in modo da indirizzarlo verso la soluzione migliore tramite un meccanismo di rinforzo.

Proprio per le peculiarità del machine learning, i dati spesso rivestono un ruolo fondamentale, talvolta anche più importante di quello della tecnologia in sé. I dati, infatti, sono lo strumento attraverso cui algoritmi di machine learning vengono addestrati. Più dati si hanno a disposizione maggiori sono le potenzialità di tali algoritmi ed è proprio per questo che con la diffusione dei big data il machine learning ha ottenuto uno sbalzo in avanti notevole.

4.2.3 Reti neurali

4.2.3.1 Basi di reti neurali artificiali

I modelli più classici di machine learning, a partire dalle regressioni lineari e logistiche e arrivando a random forest e alberi decisionali, hanno sin dai primi anni mostrato evidenti limiti nel rilevare relazioni non lineari nei dati. Questo problema ha rappresentato per decenni uno scoglio importante siccome nelle più disparate applicazioni pratiche il grado di linearità dei dati è spesso basso. In questo contesto si sono fatte strada le reti neurali artificiali, che grazie alla loro capacità di astrazione hanno permesso al mondo del machine learning di compiere un importante passo in avanti.
Le reti neurali artificiali tentano di riprodurre il modello di apprendimento neuronale che avviene all'interno del cervello umano.
All'interno del nostro cervello sono presenti cellule chiamate neuroni, responsabili della nostra capacità di comprendere l'ambiente esterno e di fornire risposte ai segnali che si presentano. All'interno dei neuroni troviamo:

- *Somi neuronali*: i corpi dei neuroni che ricevono e processano le informazioni
- *Neurotrasmettitori*: composti biologici sintetizzati nei somi che si occupano della modulazione degli impulsi nervosi
- *Assoni*: elemento del neurone che permette di comunicare in uscita con altri neuroni
- *Dendriti*: rappresentano la via di comunicazione in ingresso per ogni neurone. Ogni neurone ne possiede molti, si parla infatti di albero dendritico
- *Sinapsi*: i siti in cui avviene il passaggio di informazione tra i neuroni. Il numero di sinapsi può incrementare o diminuire a seconda degli stimoli che riceve la rete. Più sono numerosi, maggiori connessioni sinaptiche vengono create, e viceversa, rafforzando o smorzando determinati segnali.

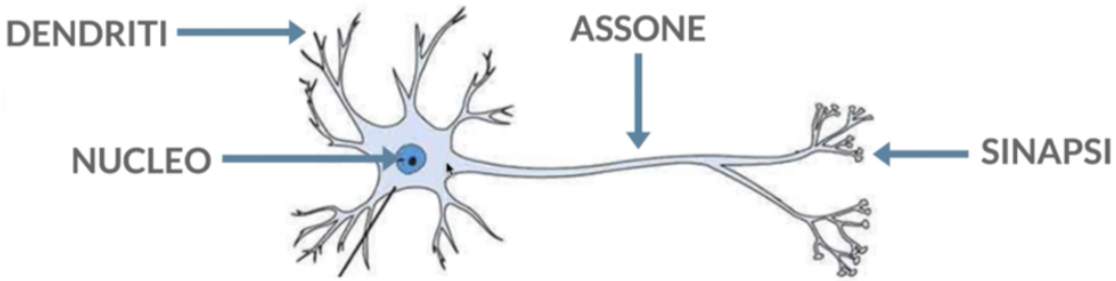


Figura 4.5: struttura di un neurone

La figura 4.5 mostra una rappresentazione semplificata della struttura di un neurone. Ogni neurone riceve un'informazione tramite un impulso dai dendriti, tale informazione viene processata all'interno dei somi. I neurotrasmettitori modulano l'impulso di risposta e lo trasmettono tramite l'assone. Le sinapsi hanno una funzione eccitatoria che tende a trasmettere l'impulso nervoso ad altri neuroni, oppure una funzione inibitoria che ne smorza l'effetto. Se il potenziale elettrico del segnale in uscita supera una soglia chiamata *potenziale d'azione* allora il successivo neurone in ingresso riceve il segnale e lo elaborerà a sua volta. Il propagarsi di questo procedimento è alla base delle reti neurali biologiche.

Il primo modello teorico di rete neurale artificiale risale al 1943 ed è opera di McCulloch e Pitts i quali crearono una macchina in grado di processare esclusivamente funzioni booleane elementari. Nel 1958 Rosenblatt propone la prima rete neurale chiamata Perceptron, che possiede solo uno strato di nodi in input e uno strato di nodi in output. I pesi sinaptici, che rappresentano la forza di connessione tra due nodi, sono dinamici e si modificano durante la fase di apprendimento della macchina, che lavora esclusivamente feedforward propagandosi in avanti. Nel 1974 Werbos migliora il modello aggiungendo uno strato nascosto (*hidden layer*) e nel 1986 Rumelhart, Hinton e Williams elaborano un algoritmo di retropropagazione, l'Error Back-Propagation, che permette alla rete neurale di perfezionarsi in stadi successivi e soprattutto di lavorare in modo bidirezionale.

La Figura 4.6 fornisce una rappresentazione grafica semplificata di ciò che avviene all'interno di un neurone e del meccanismo che si vuole ricreare all'interno di un calcolatore.

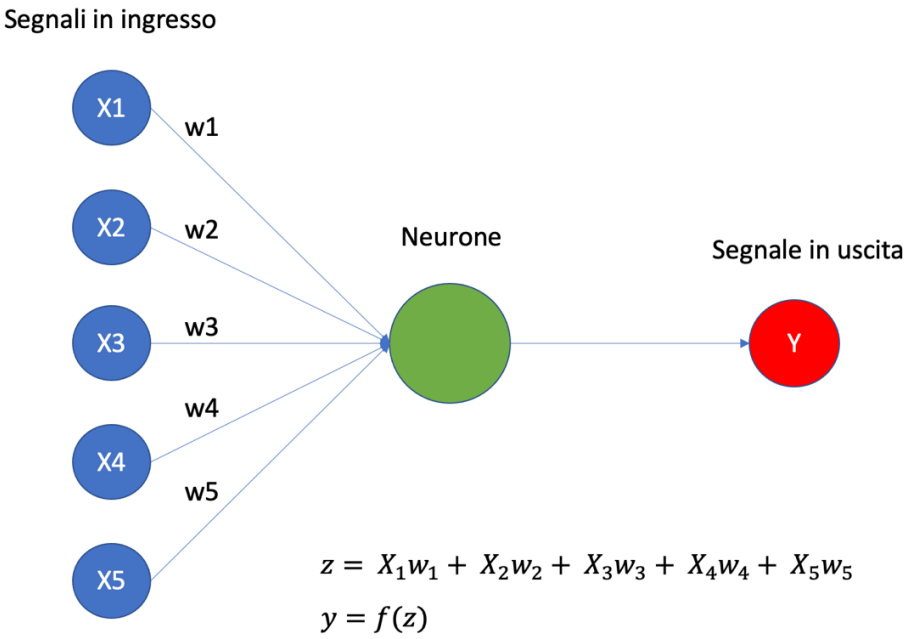


Figura 4.6: input e output in un neurone

Il neurone riceve diversi segnali in ingresso da diversi neuroni, ciascun peso w_i rappresenta la forza della connessione sinaptica. Z rappresenta il potenziale d'azione ricevuto dal neurone e dipende dai segnali in ingresso e dai pesi della connessione; y è il segnale di uscita ed è una funzione del potenziale d'azione.

Estendendo questo modello base si può ottenere una rappresentazione più generale riguardo una rete neurale, in cui sono presenti più strati di neuroni che interagiscono tra loro.

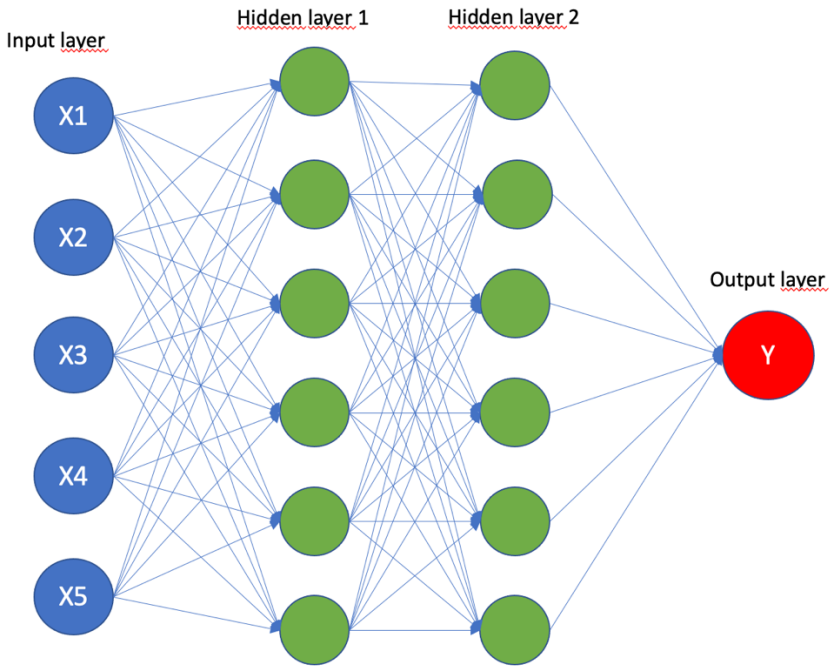


Figura 4.7: rete neurale artificiale

In ogni strato sono presenti più neuroni e all'interno di ognuno di esso avviene il procedimento di ricezione e trasmissione del segnale descritto dalla Figura 4.6.
La Figura 4.7 permette di descrivere alcune delle caratteristiche di una rete neurale artificiale:

- **Features extraction**: partendo dai dati di input presenti nell'input layer la rete neurale artificiale, negli strati nascosti, applica un'astrazione sempre maggiore dei dati creando nuove features partendo da combinazioni di quelle di cui già dispone. Questo procedimento permette di rilevare anche relazioni non lineari tra i dati. Durante la fase di addestramento la rete ottimizzerà tutti i pesi delle connessioni. Nella maggioranza dei casi non si conoscono le features intermedie, ma sono elaborazioni che rimangono interne alla rete.
- **Struttura delle connessioni**: una rete neurale in cui ogni neurone dello strato i_{-1} è collegato ad ogni neurone dello strato i è detta densa. Una rete neurale in cui è presente almeno uno strato nascosto è detta profonda.
- **Funzioni di attivazione**: come precisato nella Figura 6, il segnale in ingresso di un neurone non è esattamente il potenziale d'azione z rappresentato dai segnali dei neuroni dello strato precedente, ma è presente una funzione f(z) che filtra tale segnale. Questa funzione è chiamata funzione di attivazione, ed è utilizzata per rendere più efficiente l'apprendimento della rete. A seconda dei casi vengono utilizzate diverse funzioni di attivazione:
  - *Rectified Linear Unit (ReLU)*: è la funzione di attivazione utilizzata tra strati nascosti. Ritorna il valore massimo tra 0 e il potenziale d'azione z. $y = max(0, z)$
  - *Sigmoide*: è una delle funzioni di attivazioni utilizzate prima dello strato di output nel caso in cui il problema richieda la classificazione dell'output in categorie indipendenti tra loro. Il valore di ogni nodo di output fornisce un valore tra 0 e 1 che rappresenta la probabilità che l'elemento appartenga a quella determinata classe, come rappresentato in figura 4.8. Un elemento può appartenere a più classi.

$$y = \frac{1}{1 + e^{-z}}$$

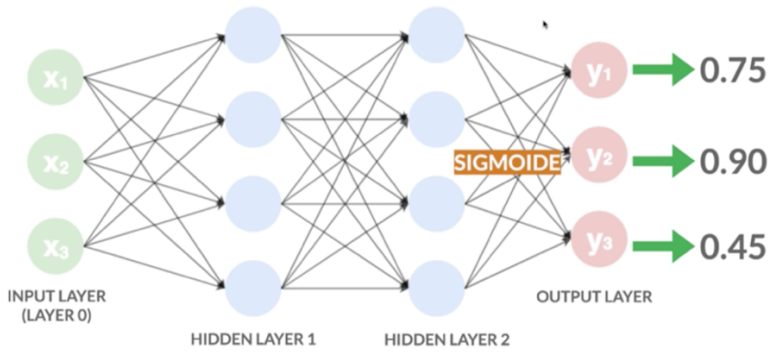


Figura 4.8: funzione di attivazione sigmoide
Fonte: https://www.profession.ai

- Softmax: è una delle funzioni di attivazione utilizzate prima dello strato di output nel caso in cui il problema richieda la classificazione dell'output in categorie non indipendenti tra loro. Un elemento può appartenere solo ad una classe quindi la somma dei valori dei nodi di output deve essere 1. La figura 4.9 mostra un esempio dell'utilizzo della funzione di attivazione softmax

$$\sigma(\mathbf{z})_j = \frac{e^z}{\sum_{k=1}^{K} e^{z_k}}, \quad \text{per } j = 1, \ldots, K$$

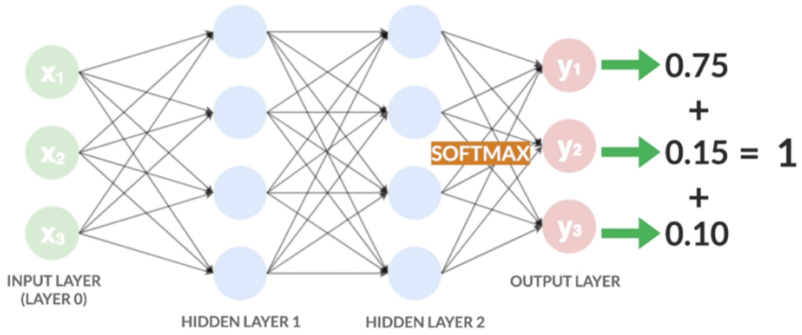


Figura 4.9: funzione di attivazione softmax
Fonte: https://www.profession.ai

La Figura 4.10 fornisce un esempio semplice e intuitivo del modello presentato: partendo da una serie di dati riguardanti uno studente si vuole predire se sarà bocciato. Lo strato di input è formato dalle features associate ad ogni studente, all'interno dello strato nascosto la rete applica la *features extraction* combinando le features a disposizione creando un'astrazione sempre maggiore fino a ottenere una previsione. Ai fini di questo esempio le features dello strato nascosto hanno un senso logico, ma, come già specificato, nel reale funzionamento di una rete neurale artificiale questo passaggio avviene internamente alla macchina e quindi non sappiamo nulla riguardo l'astrazione compiuta.

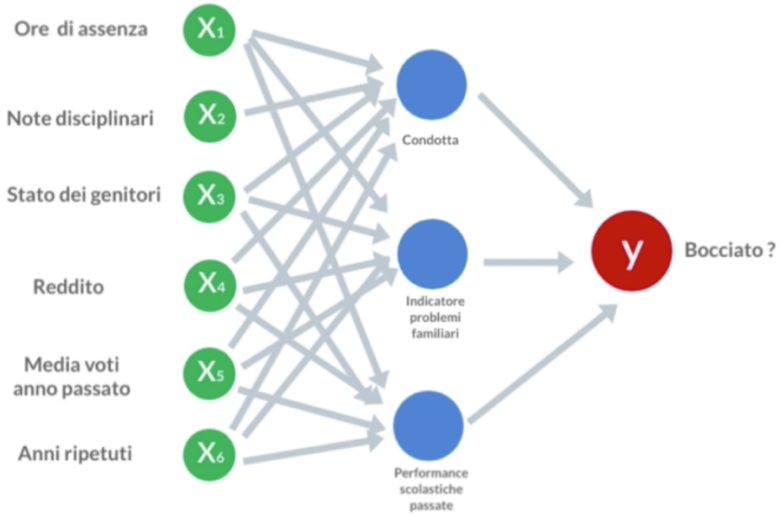


Figura 4.10: esempio rete neurale e features extraction
Fonte: https://www.profession.ai

4.2.3.2 Addestramento di una rete neurale artificiale

Fino ad ora è stata presentata la logica strutturale di una rete neurale senza specificare la sua dinamica di apprendimento e predizione. È quindi importante adesso focalizzarsi sul meccanismo di addestramento della rete neurale.

L'obiettivo che la rete tenta di raggiungere è quello di ottenere i valori ottimi di tutti i pesi w_i, ovvero le connessioni tra i neuroni dei vari strati. Il database a disposizione viene diviso in due set diversi: il set di addestramento, su cui appunto la rete si addestra e cerca di trovare la combinazione ottima dei pesi; e il set di test, su cui la rete verifica la sua accuratezza.

L'algoritmo di addestramento più comune è il Gradient Descent, che avviene in due step:

1. **Forward propagation**: i pesi vengono inizializzati in modo casuale e si propagano in avanti tra i diversi strati attraverso le loro combinazioni lineari e le funzioni di attivazione, come descritto precedentemente. Una volta giunti al nodo finale si ottiene un valore di output. Questo valore è confrontato con l'output corretto che la rete avrebbe dovuto fornire. Tale confronto genera un valore della *funzione di costo*, che indica l'errore commesso dalla rete nel fornire l'output.
2. **Backward propagation**: è il cuore dell'apprendimento della rete neurale. A partire dal valore della funzione di costo ottenuto si calcolano le derivate parziali della funzione di costo rispetto ai pesi dell'ultimo strato per capire quanto ciascun peso ha contribuito all'errore. Il gradiente ottenuto viene propagato all'indietro tramite una semplice concatenazione di derivate parziali che vengono moltiplicate tra loro (*chain rule*) in modo da ottenere la derivata parziale di ciascun peso rispetto alla funzione di costo. Questo permette di capire quanto ogni peso di ogni strato ha contribuito all'errore del modello e di quanto sia necessario modificarlo. Una volta calcolato l'intero gradiente si applica un update dei pesi.

Questi due passi vengono ripetuti iterativamente, per il numero di epoche prestabilito, e ad ogni iterazione la rete sarà in grado di fornire un output più accurato.

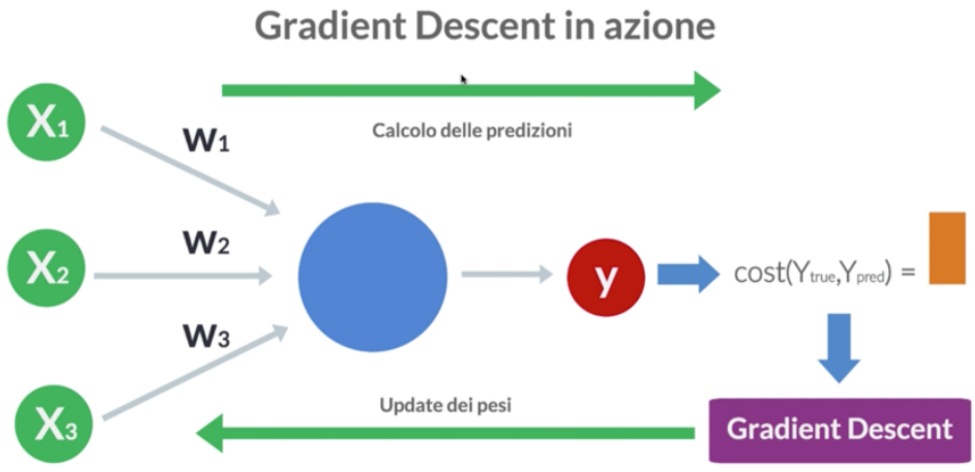


Figura 4.11: funzionamento Gradient Descent
Fonte: https://www.profession.ai

La Figura 4.11 mostra una rappresentazione grafica del procedimento di addestramento, con il calcolo delle predizioni partendo dai pesi w_i (forward propagation) e l'update dei pesi partendo dall'errore (backward propagation).

Una versione semplificata del codice del Gradient Descent potrebbe essere questa:

```
w = rand()
for epoch in epochs:
      dw = cost_derivative()
      w = w – alpha*dw
return w
```

Le epoche rappresentano il numero di iterazioni dell'algoritmo e sono scelte a priori, mentre alpha è chiamato *learning rate* e rappresenta la velocità con cui i pesi si aggiornano ad ogni epoca: è un iperparametro e deve essere specificato al momento della costruzione della rete neurale. Un learning rate troppo piccolo rallenta eccessivamente il processo di addestramento non riuscendo a ottenere in tempi utili una combinazione ottimale dei pesi, mentre un learning rate troppo grande fa oscillare i pesi tra valori sempre lontani dall'ottimo. il suo valore consigliato è compreso tra 10^{-4} e 1.

Ci sono 3 principali versioni del Gradient Descent:

- **Full Batch Gradient Descent**: esegue il Gradient Descent su tutti gli esempi del set di addestramento ad ogni step. Quindi l'update dei pesi avviene una sola volta per ogni epoca considerando tutti gli elementi del set. Ha dei forti limiti in quanto è inefficiente per dataset grandi dovendo caricare l'intero dataset in memoria. È inoltre soggetto al problema dei minimi locali: la poca dinamicità del processo di addestramento rischia di bloccarsi in un minimo locale confondendolo per il punto di ottimo della funzione di costo.
- **Stochastic Gradient Descent**: esegue il Gradient Descent utilizzando un esempio del set di addestramento per volta. Un'epoca si conclude quando il Gradient Descent è stato applicato a tutti gli elementi del set. Questa versione del Gradient Descent permette di superare i limiti del Full Batch grazie al poco peso che ha in memoria (dovendo caricare solo un esempio per volta) e alla sua dinamicità che evita il problema dei minimi locali. Aggiornando i pesi del modello dopo ogni esempio analizzato si ottiene una funzione di costo oscillante e non più in costante diminuzione come nel caso del Full Batch. Questo aspetto è uno dei pochi svantaggi dello Stochastic Gradient Descent: l'andamento oscillante della funzione di costo, infatti, può comportare il rischio di saltare il punto di minimo globale della funzione.
- **Mini Batch Gradient Descent**: è la versione più utilizzata, e consiste in una via di mezzo tra il Full Batch Gradient Descent e lo Stochastic Gradient Descent. Esegue il Gradient Descent su un numero prestabilito di esempi del set di addestramento. La dimensione consigliata del batch va da 32 a 512.

4.2.3.3 Problematiche durante l'addestramento di una rete neurale artificiale

Durante l'addestramento di una rete neurale artificiale possono presentarsi diverse problematiche che necessitano particolare attenzione nella definizione dell'architettura della rete. In questo elaborato vengono menzionate la scomparsa del gradiente e l'overfitting.

Scomparsa del gradiente

La scomparsa del gradiente è un problema molto comune nelle reti neurali troppo profonde che causa difficoltà nell'addestramento. Consiste in una riduzione eccessiva del gradiente durante la backward propagation a cui segue un update dei pesi molto piccolo che quindi non consente un effettivo miglioramento nel calcolo dell'output. Questo fenomeno è causato principalmente da due aspetti. In primis il fatto che alcune funzioni di attivazione, come la sigmoide, comprimono un input di dimensioni molto grandi tra 0 e 1: una grande variazione nell'input della sigmoide causa una piccola variazione nel suo output, e di conseguenza la derivata parziale avrà un valore basso; e il fatto che durante la backward propagation le derivate parziali vengono moltiplicate tra loro e quindi, nel caso di derivate parziali minori di 1, si ottengono numeri sempre più prossimi allo 0.

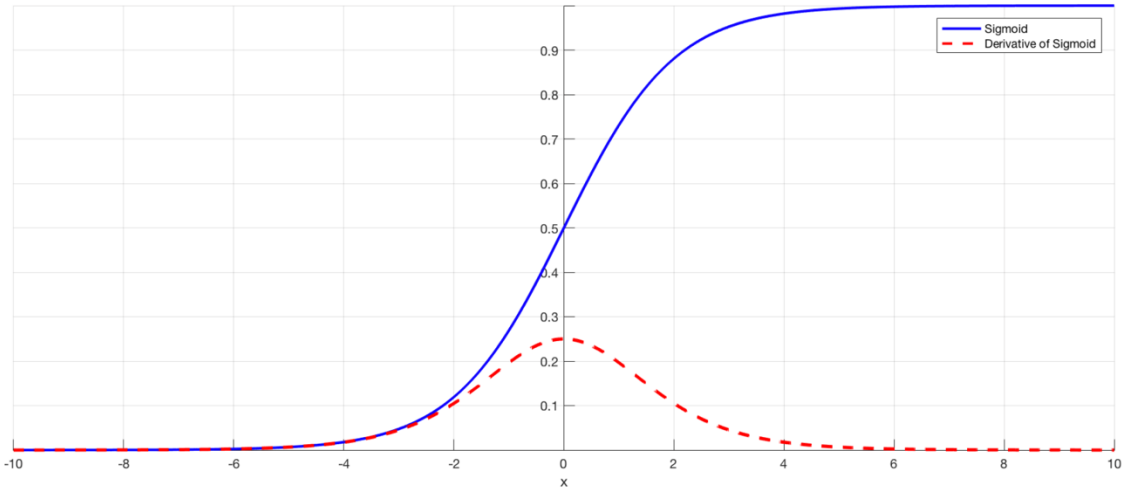


Figura 4.12: andamento della funzioene sigmoide e della sua derivata

La Figura 4.12 mostra l'andamento della funzione sigmoide e della sua derivata: quando l'input della funzione sigmoide (asse x) cresce in valore assoluto, la derivata si avvicina sempre di più a 0. Questo è uno dei motivi per cui, negli strati intermedi di una rete neurale, è consigliato utilizzare la ReLU, che non crea questa compressione.

In altri casi può esistere anche il problema opposto: l'esplosione del gradiente. Quando in una rete neurale molto profonda le derivate parziali sono troppo grandi, durante la backward propagation il gradiente diventerà un numero troppo grande (NaN) non più gestibile dal calcolatore.

Per limitare questi problemi, oltre ad utilizzare altre funzioni di attivazione come la ReLU, si effettua un'inizializzazione intelligente dei pesi w_i, cercando di mantenerne la varianza costante mentre si calcola l'output della rete. Tra le tecniche più utilizzate di inizializzazione c'è la Glorot normal Initializer e la Glorot uniform Initializer.

$$w^{[l]} = rand[-\sqrt{\frac{2}{n^{[l-1]} + n^{[l]}}}, \sqrt{\frac{2}{n^{[l-1]} + n^{[l]}}})$$

Overfitting

L'overfitting è probabilmente il problema più ostico che si può incontrare durante l'addestramento di una rete neurale artificiale. Per comprenderlo a pieno è necessario prima introdurre i concetti di bias, varianza e del cosiddetto *bias-variance trade-off*.

Il bias indica quanto le predizioni della rete sono distanti dai valori corretti nel caso si costruisse il modello più volte utilizzando parti diverse del dataset. Misura l'errore sistematico, che non dipende dalla casualità dei dati del dataset utilizzati dal modello, ma dal modello in sé. Una rete con basso bias ottiene risultati accurati a prescindere da quale parte del set di addestramento viene utilizzata.

La varianza misura la differenza nelle predizioni ottenute utilizzando diverse parti del dataset; indica quanto la rete è sensibile alla casualità dei dati di addestramento. Una rete con una bassa varianza ottiene risultati simili utilizzando diverse parti del set di addestramento.

Una buona rete neurale deve trovare un equilibrio tra bias e varianza, come mostrato nella Figura 4.13:

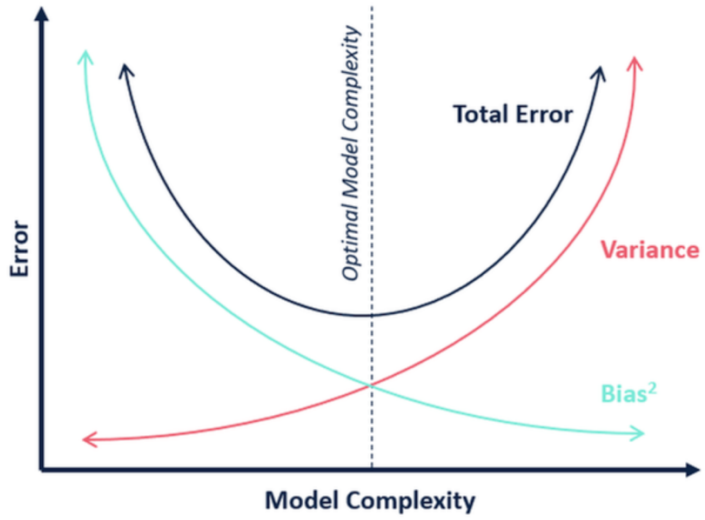


Figura 4.13: bias-variance trade-off

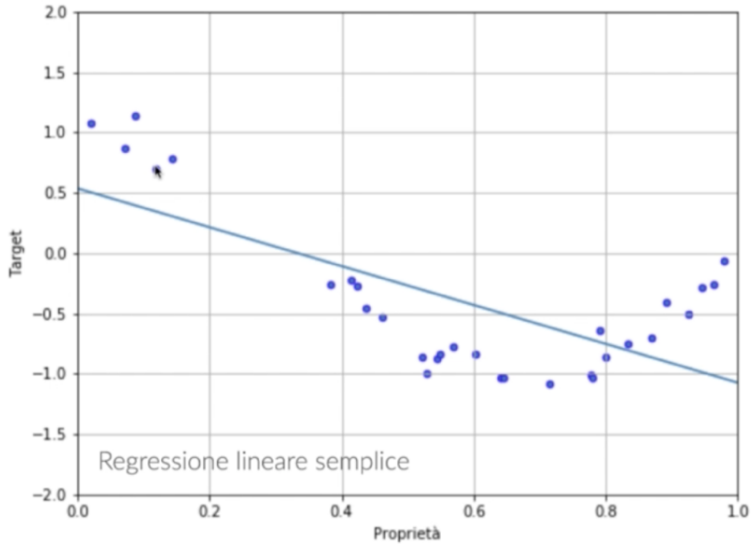


Figura 4.14: underfitting
Fonte: https://www.profession.ai

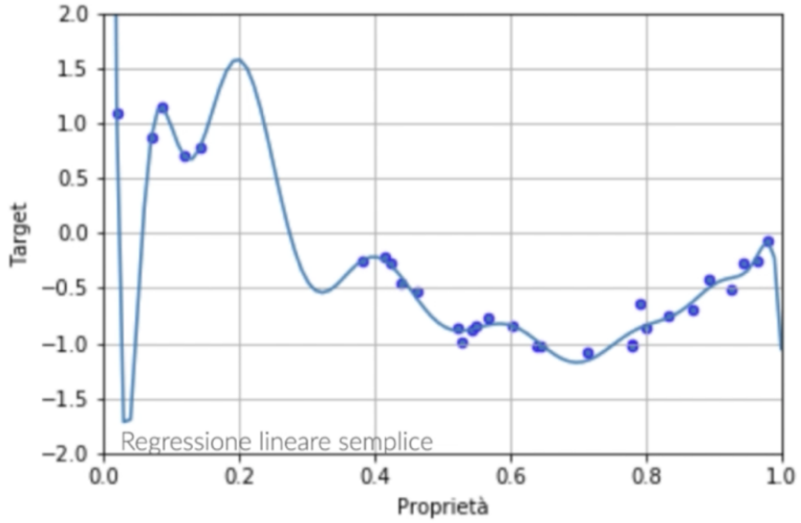


Figura 4.15: overfitting
Fonte: https://www.profession.ai

Nella Figura 4.14 la varianza è bassa e il bias alto. Il modello è troppo semplice e non è in grado di catturare le complessità del problema e di fornire predizioni sufficientemente accurate. È un caso di *underfitting*.
Nella Figura 4.15, al contrario, la varianza è alta e il bias è basso. Quindi si ottengono delle predizioni molto accurate ma solo utilizzando quella parte di dati del set di addestramento: nel momento in cui si forniscono altri dati alla rete su cui addestrarsi i risultati sarebbero molto più scarsi. Il modello è eccessivamente complesso: è un caso di *overfitting*.
L'overfitting, quindi, è un problema di cui soffre la rete quando la sua architettura è eccessivamente complessa. In questo caso la rete non ha *imparato dai dati*, ma li ha sostanzialmente memorizzati e non è quindi in grado di astrarre le dinamiche relazionali tra quei dati ed applicarle ad altri dataset: non è in grado di generalizzare quanto imparato su

nuovi dati. Nella pratica si riconosce un problema di overfitting quando l'errore sul set di test è notevolmente maggiore di quello sul set di addestramento.

La soluzione più utilizzata per combattere l'overfitting, nel caso in cui non si voglia ridurre il numero di proprietà dei dati o cambiare l'architettura della rete, è applicare una tecnica di regolarizzazione.
All'interno della funzione di costo, la regolarizzazione associa una penalità ai pesi w_i più alti permettendo così di diminuirli maggiormente durante la fase di update dei pesi. Smorzando l'effetto dei coefficienti più alti si rende meno complesso il modello, diminuendo la varianza e aumentando il bias.

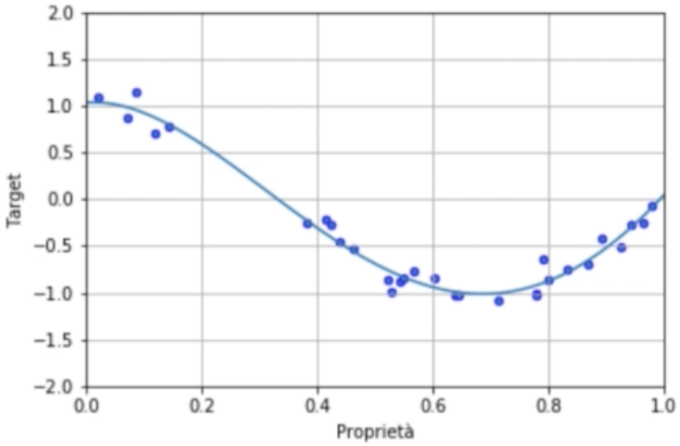


Figura 4.16. Effetto della regolarizzazione sul modello con overfitting

La Figura 4.16 mostra il modello della Figura 4.15 dopo la regolarizzazione. Si nota come sia stata ridotta la complessità del modello smorzando l'effetto dei pesi più alti.

Una tra le tecniche di regolarizzazione più utilizzate è la regolarizzazione L2 (Figura 4.17), che aggiunge un secondo termine alla funzione di costo costringendo il Gradient Descent a cercare il minimo della funzione per valori di pesi più piccoli. λ è detto parametro della regolarizzazione, è un iperparametro della rete neurale e rappresenta il peso della regolarizzazione sul modello. Se λ è troppo grande allora quasi tutti i pesi saranno troppo piccoli e si ottiene l'effetto opposto con un modello troppo semplice rientrando nel caso di underfitting.

REGOLARIZZAZIONE L2
(weight decay)

$$cost(W, b) + \lambda \sum_{j=1}^{M} W_j^2$$

Figura 4.17: regolarizzazione L2

Un'altra tecnica di regolarizzazione, specifica per le reti neurali, è il dropout. Il dropout è una tecnica che consiste nel rimuovere un determinato numero di nodi in uno strato della rete ad ogni iterazione della fase di addestramento. Questi nodi vengono selezionati casualmente ad ogni iterazione e non parteciperanno né alla forward propagation né alla backward propagation. Il risultato è che ad ogni iterazione viene addestrato un modello leggermente diverso per poi ottenerne uno complessivo mettendo insieme tutti quelli trattati, per questo motivo il dropout può essere considerato una forma di apprendimento ensemble. Il dropout funziona perché, a causa di un meccanismo chiamato *co-adaptations*, durante la fase di addestramento alcuni nodi tendono a farsi carico degli errori commessi da altri nodi. I coefficienti dei nodi che hanno *trasferito* gli errori tendono ad azzerarsi mentre i coefficienti che si sono fatti carico degli errori tendono a crescere. Questo porta a overfitting perché il modello si lega troppo ai coefficienti alti. Eliminando ad ogni iterazione dei nodi il modello ottimizzerà i coefficienti dei nodi rimasti e sarà meno probabile che si verifichi questo problema.

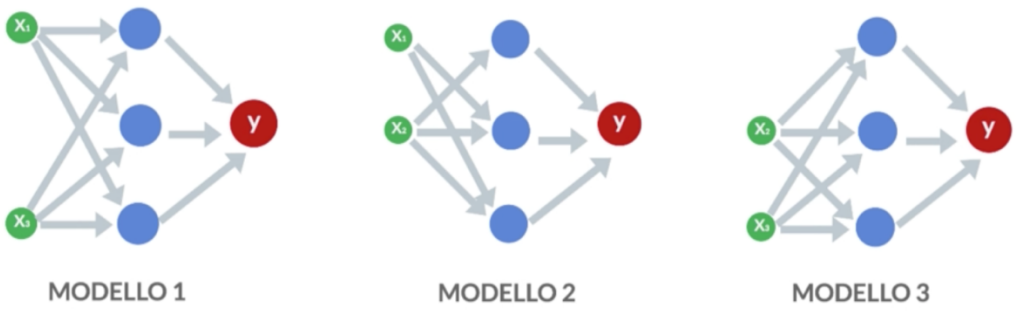
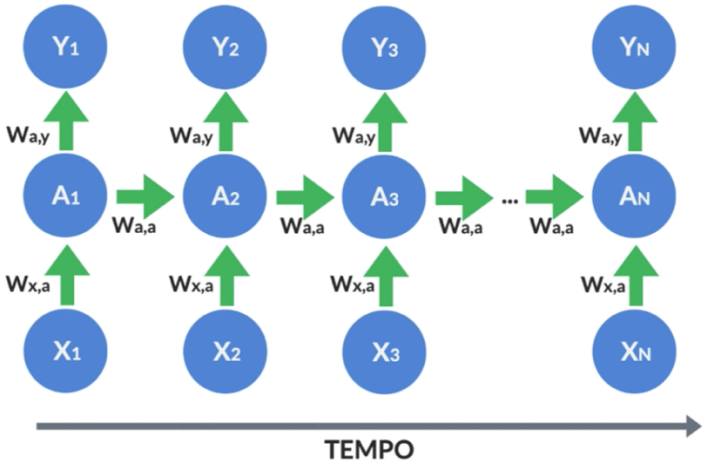


Figura 4.18: dropout
Fonte: https://www.profession.ai

La Figura 4.18 mostra una rappresentazione intuitiva di ciò che avviene con il dropout.

4.2.3.4 Reti neurali ricorrenti

Nel corso degli anni sono state sviluppate numerose tipologie di reti neurali artificiali per trovare soluzioni sempre più specifiche a determinati problemi. In questo elaborato vengono presentate le reti neurali ricorrenti, le più adatte a gestire dati sequenziali, come ad esempio le serie temporali. Questa particolare architettura permette di passare l'output di un'esecuzione a quella successiva, creando un loop che connette uno strato nascosto tra diverse esecuzioni, creando quindi una memoria all'interno della rete.



$$a_1 = \tanh(W_{x,a}x_1 + b_a)$$
$$y_1 = \text{softmax}(W_{a,y}a_1 + b_y)$$
$$a_2 = \tanh(W_{a,a}a_1 + W_{x,a}x_2 + b_a)$$
$$y_2 = \text{softmax}(W_{a,y}a_2 + b_y)$$
$$a_3 = \tanh(W_{a,a}a_2 + W_{x,a}x_3 + b_a)$$
$$y_3 = \text{softmax}(W_{a,y}a_3 + b_y)$$
$$\dots$$
$$a_N = \tanh(W_{a,a}a_{N-1} + W_{x,a}x_N + b_a)$$
$$y_N = \text{softmax}(W_{a,y}a_N + b_y)$$

Figura 4.19: rete neurale ricorrente
Fonte: https://www.profession.ai

La Figura 4.19 mostra una rappresentazione esemplificativa di una rete neurale ricorrente dipendente dalla variabile tempo, la funzione di attivazione di ogni strato di output è la softmax mentre la funzione di attivazione di ogni strato nascosto è la tangente iperbolica che, in modo simile alla ReLU, limita il problema di scomparsa o esplosione del gradiente.

X_1 è il vettore composto dai nodi dello strato di input, A_1 è il vettore composto dai nodi dello strato nascosto, Y_1 è il vettore di output. $W_{x,a}$ e $W_{a,y}$ sono i vettori dei pesi delle connessioni tra i diversi strati. Come si nota, l'output di uno strato nascosto non viene passato solo al corrispondente vettore di output, ma anche allo strato nascosto dell'esecuzione successiva, e questo è ciò che permette alla rete di tenere conto della relazione sequenziale dei dati creando una linea temporale tra le esecuzioni. Allo strato nascosto A_2, oltre all'informazione dallo strato X_2, arriva anche quella dallo strato nascosto dell'esecuzione precedente A_1 tramite il vettore di pesi $W_{a,a}$, in modo che il vettore di output Y_2 consideri anche le informazioni elaborate nell'esecuzione precedente.
Un esempio intuitivo per capire l'efficacia di questo tipo di architettura è quello del *machine translation*: immaginando di avere una frase da tradurre, in cui ogni parola rappresenta un'esecuzione della rete, è evidente che per effettuare una traduzione più precisa che mantenga il senso intero della frase non si possa semplicemente tradurre parola per parola, ma sia necessario mantenere anche l'informazione relativa alle parole precedenti. Perciò la memoria che si crea all'interno di una rete neurale ricorrente permette di considerare il contesto della frase per tradurre una determinata parola.
Il numero di esecuzioni sarà pari al numero di dati in una sequenza, e la relazione tra input e output permette di creare diversi tipi di reti neurali:

- Relazione uno a uno: un dato in input e un dato in output; tipologia adatta alla classificazione di immagini
- Relazione uno a molti: un dato in input e una serie di dati in output; tipologia adatta alla descrizione di immagini
- Relazione molti a molti: come nella Figura 4.19, una serie di dati in input e una serie di dati in output. L'output di ogni esecuzione non si basa solo sull'input che riceve,

ma anche sulle informazioni processate fino a quel punto. Tipologia adatta al machine translation

- Relazione molti a uno: in input una sequenza di dati e in output una sola informazione; tipologia adatta alla sentiment analysis o alla classificazione di serie temporali di dati, in cui ogni elemento della serie rappresenta un'esecuzione della rete neurale ricorrente (figura 4.20).

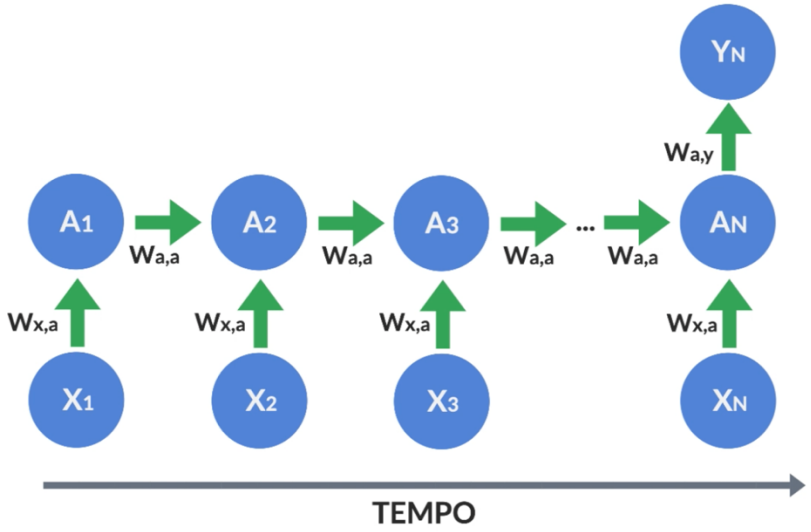


Figura 4.20: rete neurale ricorrente molti a uno
Fonte: https://www.profession.ai

L'addestramento di una rete neurale ricorrente avviene in modo simile ad una normale rete neurale: la propagazione in avanti permette di calcolare i pesi e la funzione di costo, mentre la Back Propagation Through Time (BPTT) consente la propagazione all'indietro del gradiente non solo attraverso i diversi strati, ma anche attraverso le esecuzioni sequenziali della rete, come mostrato in Figura 4.21.

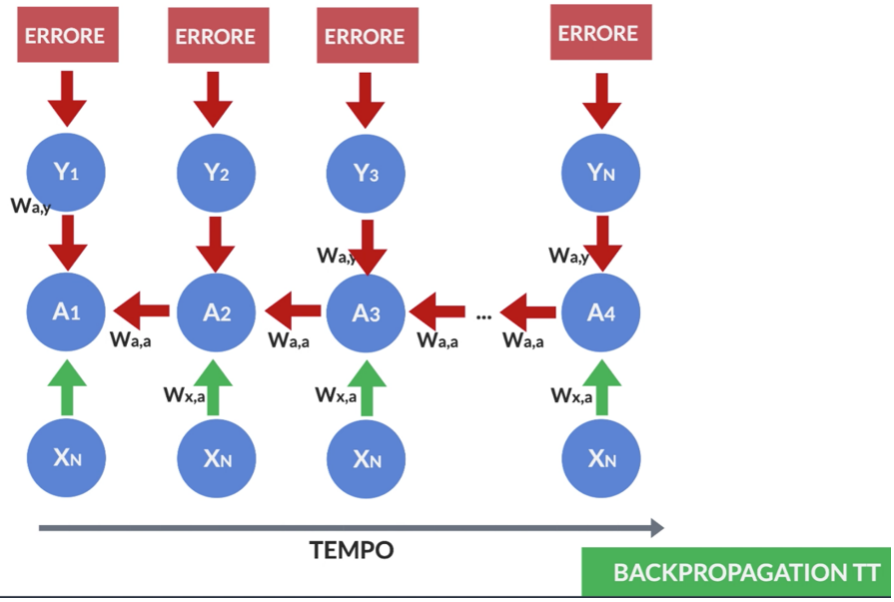


Figura 4.21: back Propagation Through Time
Fonte: https://www.profession.ai

4.2.3.5 LSTM

Il funzionamento della Back Propagation Through Time rende ancora più frequente il problema della scomparsa del gradiente che, oltre a rallentare notevolmente l'addestramento della rete, nel caso delle reti neurali ricorrenti provoca la perdita di dipendenza dai primi dati della sequenza in quanto l'aggiornamento dei pesi dei nodi dei primi strati sarà pressoché nullo. Per questo motivo, i modelli più classici di reti neurali ricorrenti, chiamati Vanilla Recurrent Neural Network (Vanilla RNN), sono adatti solo a brevi sequenze di dati.
La soluzione a questo problema è stata fornita nel 1997, quando per la prima volta è stata teorizzata la Long Short Term Memory, abbreviata LSTM, un tipo particolare di rete neurale ricorrente costruita appositamente per imparare le dipendenze a lungo termine dei dati, oltre che quelle a breve termine. L'architettura di questo tipo di rete è molto complicata e una spiegazione dettagliata esula dalle finalità di questo elaborato. Tuttavia, ne viene fornita una descrizione generale per comprendere i concetti e le idee chiave dietro al suo funzionamento.

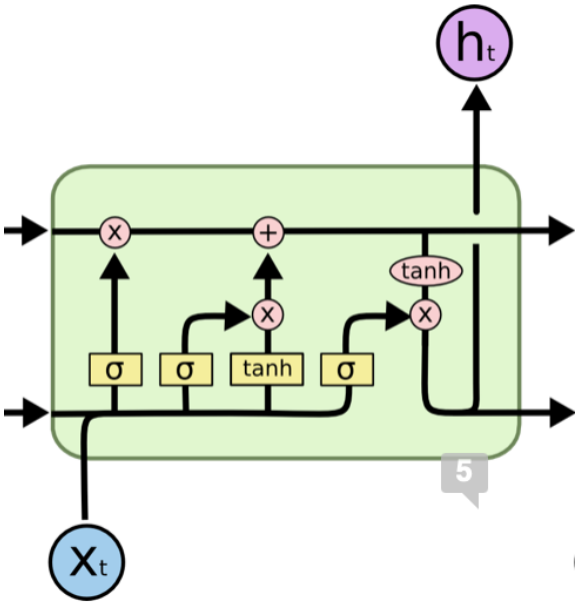


Figura 4.22: struttura di uno strato nascosto in una LSTM

La Figura 4.22 mostra la struttura di uno strato nascosto di una LSTM. A differenza di una Vanilla RNN, in cui in ogni strato nascosto è presente un solo layer con la funzione di attivazione tanh, in una LSTM sono presenti un core centrale chiamato Cell State e diversi *gates* che interagiscono tra loro e gestiscono le nuove e vecchie informazioni:

- **Cell state**: è l'idea chiave dietro alle LSTM. Funge da canale di comunicazione prioritario che permette alle informazioni di fluire da un'esecuzione ad un'altra. Tali informazioni vengono modificate esclusivamente da alcune operazioni lineari che permettono di rimuovere o aggiungere nuove informazioni al Cell State. È la memoria

della rete che trasporta le informazioni lungo le diverse esecuzioni della rete.

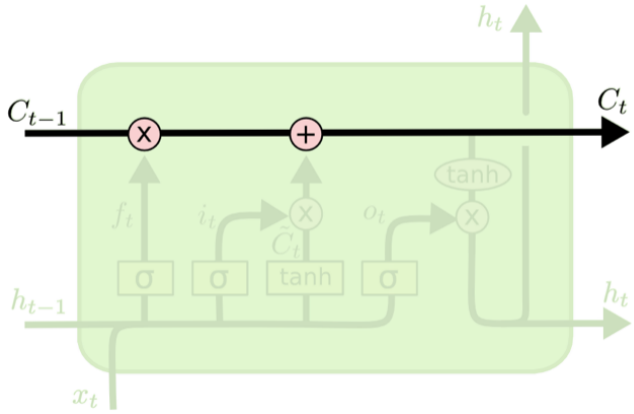
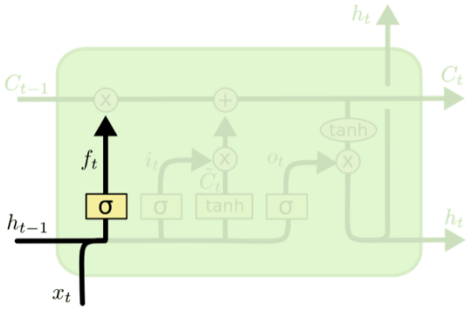


Figura 4.23: cell State

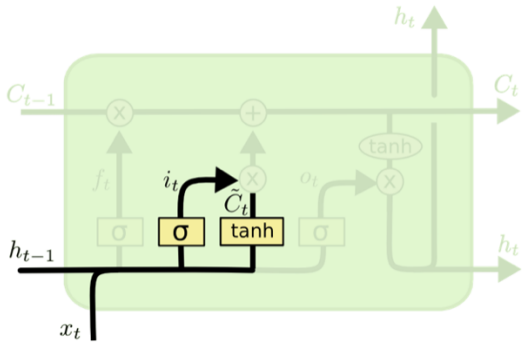
- **Forget gate**: all'interno di questo gate viene stabilito se l'informazione all'interno del Cell State deve essere modificata o meno. La funzione di attivazione sigmoide restituisce un valore tra 0 e 1 che indica il *grado di sostituibilità* dell'informazione.



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Figura 4.24: forget gate

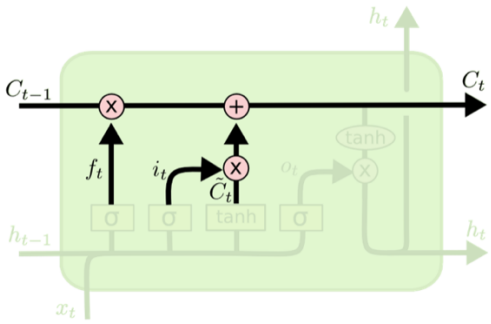
- **Input gate e *new candidate***: nel caso in cui il forget gate abbia stabilito che va aggiunta o sostituita un'informazione, in questa fase si stabilisce l'informazione da rimuovere e quella nuova in ingresso, chiamata appunto *new candidate*



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Figura 4.25: input gate e new candidate

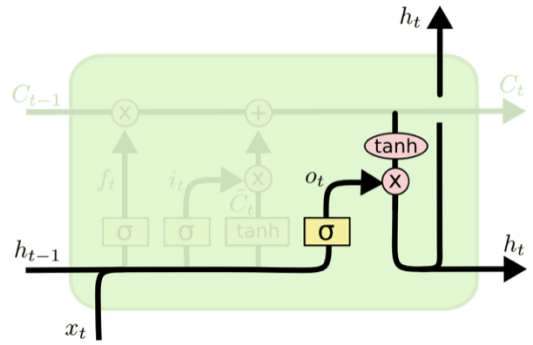
- **Update del Cell State**: viene modificato il Cell State in seguito alle elaborazioni dei gate precedenti.



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Figura 4.26: update del Cell State

- **Output gate**: in questo gate viene fornito l'output allo strato successivo della LSTM. Tale output sarà una versione filtrata del Cell State: una funzione sigmoide stabilisce la parte di Cell State da considerare e la funzione tanh comprime l'output finale tra -1 e 1 prima di passarlo allo strato successivo.



$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Figura 4.27: output gate

Capitolo 5

Raccolta dati e cluster analysis

Nei capitoli precedenti è stato presentato il contesto finanziario e tecnologico in cui questo elaborato mira ad inserirsi, analizzando la letteratura più importante riguardo le dinamiche dei prezzi e bolle speculative, e presentando gli strumenti che verranno utilizzati per creare il modello.
In questo capitolo e in quello successivo sarà invece sviluppata la parte sperimentale della tesi, cioè la vera e propria creazione da zero del modello basato su una rete neurale artificiale.

Questo capitolo riguarderà la raccolta dei dati, la loro elaborazione, e lo svolgimento della cluster analysis, mentre il capitolo seguente tratterà la creazione, l'addestramento e il testing della rete neurale.

5.1 Raccolta dei dati

L'obiettivo finale del modello che si intende costruire è individuare, attraverso una rete neurale, delle bolle speculative all'interno del mercato azionario. Essendo le bolle speculative un fenomeno finanziario di medio-lungo termine è necessario considerare un periodo storico molto lungo sui mercati in modo da fornire alla rete neurale dati a sufficienza per essere addestrata efficacemente.
Si è scelto quindi di utilizzare i dati degli ultimi 93 anni, dal 1928 al 2021, dello Standard & Poor 500, l'indice azionario statunitense formato da un paniere con le 500 aziende statunitensi a maggiore capitalizzazione. Tale indice consente fornisce una prospettiva più generale sul mercato statunitense rispetto a quella offerta dal Nasdaq 100, indice incentrato sui titoli tecnologici, e dal Dow Jones, formato da titoli di natura prettamente industriale.
È possibile ottenere i dati dello S&P 500 direttamente da Yahoo Finance, oppure da linea di codice su Python sfruttando la libreria DataReader di Pandas.

```
import pandas as pd
import datetime as dt
import pandas_datareader.data as web

start = dt.datetime(1928,1,3)
end = dt.datetime(2021,5,12)

df = web.DataReader("^GSPC","yahoo",start,end)

df.to_excel(excel_writer="datiS&P500.xlsx")
```

Si ottengono, direttamente in un file Excel, i dati storici dell'indice S&P 500 dal 3/01/1928 al 12/5/2021.

Date	High	Low	Open	Close	Volume	Adj Close
21/03/85	180,22	178,89	179,08	179,35	95930000,00	179,35
22/03/85	179,92	178,86	179,35	179,04	99250000,00	179,04
25/03/85	179,04	177,85	179,04	177,97	74040000,00	177,97
26/03/85	178,86	177,88	177,97	178,43	89930000,00	178,43
27/03/85	179,80	178,43	178,43	179,54	101000000,00	179,54
28/03/85	180,60	179,43	179,54	179,54	99780000,00	179,54
29/03/85	180,66	179,54	179,54	180,66	101400000,00	180,66
01/04/85	181,27	180,43	180,66	181,27	89900000,00	181,27
02/04/85	181,86	180,28	181,27	180,53	101700000,00	180,53
03/04/85	180,53	178,64	180,53	179,11	95480000,00	179,11
04/04/85	179,13	178,29	179,11	179,03	86910000,00	179,03
08/04/85	179,46	177,86	179,03	178,03	79960000,00	178,03

Nella tabella sono mostrati, come esempio, 12 giorni dello storico.
Per ciascun giorno sono segnati:

- *High*: prezzo massimo giornaliero
- *Low*: prezzo minimo giornaliero
- *Open*: prezzo di apertura giornaliero
- *Close*: prezzo di chiusura giornaliero
- *Volume*: ammontare giornaliero di operazioni di compravendita dell'indice
- *Adj Close*: prezzo di chiusura giornaliero rettificato considerando anche frazionamenti azionari e dividendi

Ai fini del modello che si vuole costruire, considerando l'orizzonte di lungo periodo dell'analisi, si è optato per considerare esclusivamente il prezzo di chiusura e i volumi scambiati.

5.2 Stima dei volumi mancanti

Un primo problema da affrontare riguarda i volumi: dal 1928 al 1950, infatti, non sono disponibili i dati sui volumi scambiati dello S&P 500. Questo comporterebbe il fatto di non poter considerare le stesse metriche per tutti gli elementi del database viziando sin da subito l'analisi. È necessario, perciò, effettuare una retropolazione storica per stimare i volumi dello S&P 500 prima del 1950. Tale procedimento avviene in due fasi:

1. Fino al 3/02/1930 ci si appoggia all'indice Dow Jones e ai suoi volumi di scambio, considerati un buon proxy per quelli dello S&P 500. Si effettua quindi una regressione lineare tra i volumi dello S&P 500 e quelli del Dow Jones sui dati dal 1950 al 1952:

   $$Volume\ S\&P500 = \beta\ Volume\ Dow\ Jones + \alpha$$

Ottenendo α e β della regressione si procede stimando i volumi mancanti dello S&P 500 dal 1950 fino al 1930.

2. Dal 3/02/1930 al 3/01/1928 non sono disponibili neanche i volumi del Dow Jones. Si effettua quindi un'ulteriore approssimazione, stimando una regressione tra i volumi e i prezzi dello S&P 500 dal 1930 al 1932 e applicando i risultati ai due anni mancanti, in modo da ottenere una prosecuzione approssimata della regressione del punto 1:

   $$Volume\ S\&P500 = \beta\ Prezzo\ S\&P500 + \alpha$$

   Questo approccio per ottenere i volumi mancanti è sicuramente un'approssimazione, ma il trade-off è positivo: per la bontà del modello finale, infatti, è molto importante avere i dati completi sui volumi di scambio sino al 1928, in modo da poter considerare anche le dinamiche azionarie di fine anni '20, momento in cui, a causa dello scoppio di una bolla speculativa, il mercato azionario entrò in forte crisi.

5.3 Creazione trimestri

Una bolla speculativa è un fenomeno che può essere associato a degli orizzonti temporali medio-lunghi. Non avrebbe alcun significato finanziario sostenere che in un dato giorno, o in una data settimana, il mercato registra la presenza di una bolla speculativa. Ai fini della costruzione del modello è quindi necessario scegliere una cadenza temporale sufficientemente ampia in base alla quale partizionare i dati di borsa.
Perciò si è scelto di considerare come time bucket un trimestre. Partendo dai dati di mercato giornalieri, quindi, bisogna ottenere delle osservazioni trimestrali su cui calcolare successivamente diverse metriche. Dovendo necessariamente fornire al modello elementi della stessa dimensione, si è scelto di considerare come lunghezza fissa di un trimestre 65 giorni di day trading: alcuni elementi del database rimodellato copriranno qualche giorno in più di tre mesi di borsa e altri qualche giorno in meno, ma considerando l'ordine di grandezza del problema in questione sono differenze trascurabili. Inoltre, al fine di fornire alla rete neurale più dati possibili, i trimestri sono stati calcolati considerando una finestra mobile di 1 giorno:

Trimestre i-esimo	Data inizio	Data fine
Trimestre 1	03/01/28	04/04/28
Trimestre 2	04/01/28	05/04/28
Trimestre 3	05/01/28	09/04/28
Trimestre 4	06/01/28	10/04/28
Trimestre 5	09/01/28	11/04/28
…	…	…
…	…	…

A questo punto bisogna calcolare diverse misure e statistiche per ciascun trimestre in modo da ottenere elementi con elevato contenuto informativo riguardo ciò che è avvenuto sui mercati finanziari in quel determinato periodo.
Siccome i dati coprono più di 90 anni tempo, è prima necessario applicare qualche trasformazione sui dati grezzi in modo da calcolare, successivamente, metriche confrontabili tra loro anche quando appartenenti a trimestri temporalmente molto lontani, in cui ovviamente il valore assoluto dei prezzi e dei volumi è molto diverso.
I prezzi e i volumi di ogni trimestre, quindi, sono stati normalizzati fissando a 100 il prezzo e il volume del primo giorno e calcolando i dati dei seguenti 64 giorni di conseguenza. Su Excel la procedura è piuttosto semplice: per ogni elemento del trimestre si effettua la seguente operazione:

$$prezzo = (prezzo \,/\, prezzo\ giorno\ 1) * 100$$
$$volume = (volume \,/\, volume\ giorno\ 1) * 100$$

Prezzi grezzi	Prezzi normalizzati	Volumi normalizzati
17,76	100,00	100,00
17,72	99,77	99,73
17,55	98,82	98,60
17,66	99,44	99,33
17,50	98,54	98,27
17,37	97,80	97,40
17,35	97,69	97,27
17,47	98,37	98,07
17,58	98,99	98,80
17,29	97,35	96,87
…	…	…
…	…	…

[2]

A questo punto, per ciascun trimestre, sui prezzi e volumi normalizzati, è stato calcolato:

- **Log-rendimento trimestrale**: $ln \frac{prezzo\ giorno\ 65}{prezzo\ giorno\ 1}$
- **Log-rendimento medio del trimestre** : $\frac{\sum ln \frac{prezzo\ giorno\ i+1}{prezzo\ giorno\ i}}{65}$

[2] Tale operazione non è stata esplicitamente calcolata nel file Excel, in quanto avendo utilizzato trimestri a finestra mobile di 1 giorno la visualizzazione grafica del risultato sarebbe stata eccessivamente pesante. È stata comunque utilizzata la versione normalizzata dei prezzi e dei volumi per i calcoli successivi condensando la normalizzazione all'interno di ogni formula di Excel da applicare.

- **β volumi**: β della regressione che stima l'andamento trimestrale dei volumi: $volume_i = \beta\ giorno_i + \alpha$
  L'andamento dei volumi è un indicatore molto importante riguardo i trend attuali e futuri del mercato. I volumi in crescita rafforzano il trend dei prezzi che si manifesta, mentre i volumi in decrescita sono un segnale di una possibile inversione della tendenza a cui si sta assistendo.

- **Deviazione standard trimestrale**: $deviazione\ standard(prezzo1, \ldots, prezzo65)$

- **β trimestre**: β della regressione che stima l'andamento trimestrale dei prezzi: $prezzo_i = \beta\ giorno_i + \alpha + \varepsilon_i$
  Rispetto al log-rendimento è una metrica più robusta in quanto è calcolata su tutti i 65 giorni del trimestre e non solo sul primo e sull'ultimo giorno.

- **Volatilità residui trimestre**: deviazione standard degli errori ε_i della regressione sui prezzi. Questo indicatore permette di valutare la volatilità dei prezzi attorno al trend trimestrale. Una bassa volatilità attorno al trend indica poche incertezze sul mercato nel seguire la tendenza che si sta manifestando, mentre un'alta volatilità attorno al trend è sintomo di instabilità.

Tutte queste metriche sono state calcolate ugualmente non solo per il trimestre di riferimento, ma anche relativamente al trimestre precedente[3] e a due trimestri precedenti, in modo da fornire a ciascun elemento del database anche delle misure backward looking, molto utili in fase di classificazione. Sono stati inoltre calcolati anche il log-rendimento e il beta corrispondente del semestre precedente, per dare un'indicazione della tendenza di lungo periodo che ha preceduto il trimestre in questione. Si ottiene quindi la seguente rappresentazione su Excel:

Data inizio	Data fine	Beta volumi	Log rendimento t	Log rendimento medio t	Beta t	Volatilità residui t	…
19/07/28	19/10/28	0,288987876	0,145310062	0,002203167	0,247227088	1,484891763	…
20/07/28	22/10/28	0,286560492	0,136351838	0,002130093	0,245237521	1,488039842	…
23/07/28	23/10/28	0,284072808	0,138032223	0,002196161	0,243301393	1,482658296	…
24/07/28	24/10/28	0,284035358	0,141423148	0,002151582	0,243205253	1,486365971	…
25/07/28	25/10/28	0,279876455	0,132032828	0,002119669	0,239874691	1,484700971	…

Data inizio	Data fine	…	Beta volumi t-1	Log rendimento t-1	Beta t-1	Volatilità residui t-1	Beta volumi t-2	…
19/07/28	19/10/28	…	-0,108699015	-0,025422887	-0,093417732	2,378783884	0,254846536	…
20/07/28	22/10/28	…	-0,111528385	-0,027527092	-0,09584934	2,359686697	0,262676602	…
23/07/28	23/10/28	…	-0,113144301	-0,031042248	-0,097327217	2,328891422	0,268028234	…
24/07/28	24/10/28	…	-0,115661292	-0,011440546	-0,099249894	2,359776948	0,267601521	…
25/07/28	25/10/28	…	-0,120216334	-0,00313805	-0,10298937	2,349783867	0,269107422	…

[3] Come trimestre precedente si intende il primo trimestre precedente a quello di riferimento e interamente non sovrapposto. Ad esempio, se il trimestre in questione va dal giorno 100 al giorno 164, il trimestre precedente comprenderà il periodo che va dal giorno 35 al giorno 99.

Data inizio	Data fine	…	Log rendimento t-2	Beta t-2	Volatilità residui t-2	Log rendimento t-1,2	Beta t-1,2
19/07/28	19/10/28	…	0,121311252	0,214398472	2,444259468	0,095888365	0,10874784
20/07/28	22/10/28	…	0,126352508	0,22100988	2,433560836	0,093206055	0,106215431
23/07/28	23/10/28	…	0,113783799	0,22541409	2,41570346	0,097625018	0,104035529
24/07/28	24/10/28	…	0,096982624	0,225348978	2,386453456	0,09541488	0,100954185
25/07/28	25/10/28	…	0,096453717	0,226860188	2,347374715	0,088107302	0,098157248

5.4 Necessità di ottenere una classificazione a priori

Una volta calcolate tutte le metriche e le statistiche desiderate si è ottenuto un database in cui ogni riga rappresenta un trimestre e ogni colonna uno specifico indicatore. Tutte le informazioni relative a un trimestre servono per poter effettuare una classificazione logica e intelligente in modo che la rete neurale lavori con degli elementi etichettati. Per le finalità di questo elaborato, infatti, è necessario fornire alla rete neurale un problema di apprendimento supervisionato in cui nel database su cui sarà addestrata siano presenti variabili di input (ovvero i trimestri di borsa) e variabili di ouptut (tipologia di trimestre) cosicché la rete impari a mappare una funzione tra i dati in input e quelli in output generalizzando quanto appreso dall'esperienza attraverso i passaggi descritti nel Capitolo 4.
Uno dei principali problemi di questo approccio, infatti, è l'indisponibilità a priori di un database classificato: non esiste un database che classifichi dei trimestri di borsa etichettandoli in bolla speculativa, crollo di mercato, crescita stabile, ecc. Questo perché, come specificato nei primi capitoli, non si è ancora raggiunto un punto di incontro tra le diverse scuole di pensiero economiche nel definire la natura finanziaria di un trimestre di borsa e tantomeno non c'è unanimità riguardo tutti i momenti storici caratterizzati da una bolla. Anche i grandi eventi speculativi a cui si è assistito nell'ultimo secolo sono ancora oggi argomento di dibattito tra gli studiosi: alcuni economisti della scuola dei mercati efficienti, ad esempio, sostengono che il crollo di inizio anni 2000 non fu causato dallo scoppio della bolla delle dot-com, ma semplicemente dall'aggiustamento rapido ed efficiente del mercato a nuove informazioni (Malkiel, 2003).
In questo elaborato, tuttavia, si sposa la teoria della finanza comportamentale e della critica ai mercati efficienti in merito alla formazione di distorsioni di mercato e bolle speculative.

In ogni caso rimane la necessità, per proseguire nella costruzione del modello, di classificare a priori in qualche modo i trimestri in modo da fornire degli elementi etichettati alla rete neurale, che altrimenti non riuscirebbe ad imparare le dinamiche relazionali tra variabili in input e in output, fallendo quindi nell'individuare eventuali bolle speculative.
Un primo approccio potrebbe essere quello di effettuare una classificazione *a spanne*, etichettando come bolle speculative i trimestri appartenenti a periodi storici ormai da quasi tutti ritenuti fortemente speculativi, come ad esempio i mesi che anticiparono i crolli del 1929, del 1987, del 2001 e del 2008, per citare i più famosi. Questa modalità, però, è stata ritenuta troppo approssimativa e non sufficiente: l'obiettivo che si vuole raggiungere è quello di

creare una rete neurale in grado di riconoscere diverse tipologie di trimestri con lo scopo principale di individuare possibili bolle speculative, ma che non si limiti a stabilire se un trimestre sia o meno una bolla.
Si vuole ottenere perciò un modello che sia in grado di coprire in modo capillare tutto il mercato e le diverse fasi che può attraversare: crescita, decrescita, bolla speculativa, crollo, stagnazione, ecc. A tal fine non è sufficiente una classificazione semplicistica e binaria basata sulla conoscenza generale riguardo i periodi più speculativi della storia. Inoltre, ci sono stati molti altri periodi speculativi, di minor durata e di minor impatto, che hanno comunque recato molti danni al mercato e agli investitori e di cui si vuole tenere conto.
Oltre ai trimestri di borsa la cui natura fortemente speculativa è ormai generalmente riconosciuta, quindi, per effettuare una classificazione completa anche di tutti gli altri trimestri presenti nel database non ci si può esclusivamente affidare a metodi qualitativi perché il risultato sarebbe troppo grossolano. Inoltre, considerando la dimensione dei dati a disposizione, non è sostenibile neanche analizzare ogni trimestre singolarmente: è necessario accelerare in qualche modo il procedimento di tassonomia dei trimestri di borsa.
Di conseguenza si è deciso di effettuare una cluster analysis, tecnica di statistica multivariata presentata nel Capitolo 4, che permette di dividere gli elementi di analisi in gruppi il più possibile omogenei tra loro.
A posteriori della cluster analysis sarà possibile interpretare a livello economico-finanziario ogni gruppo e ottenere quindi una mappatura completa dei trimestri.

5.5 Preparazione della cluster analysis

Prima di effettuare la cluster analysis si procede ad una fase di pulizia e preparazione del dataset, sempre su Python.

```
# Importazione delle librerie pandas, numpy e matplotlib
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt


# Importazione del database dei trimestri
nameDB = "S&P500.xlsx"
dataset = pd.read_excel(nameDB)


# Eliminazione dei trimestri per cui non sono disponibili tutte le misurazioni:
# per i primi trimestri non sono disponibili le metriche riguardo i trimestri
# precedenti e il semestre precedente
dataset = dataset[dataset.Indice > 128 ]
```

```
# Esclusione trimestri con volatilità residui anomala
q1 = df["Volatilità residui trimestre"].quantile(0.97)
q2 = df["Volatilità residui t-1"].quantile(0.97)

# Creazione del database outliers che sarà analizzato a parte
outliers = df[(df["Volatilità residui trimestre"] >=q1) | (df["Volatilità resid
ui t-1"] >= q2)]
df = df[(df["Volatilità residui trimestre"] < q1) &
(df["Volatilità residui t-1"] < q2)]
```

Gli outliers selezionati sono trimestri in cui la volatilità dei residui del trimestre in questione, o di quello precedente, hanno un valore estremamente alto. Per evitare di viziare la cluster analysis con trimestri che hanno avuto un comportamento particolarmente anomalo e in cui i vari beta dei prezzi perdono buona parte del loro significato, si è preferito estrometterli e analizzarli successivamente a parte.
A questo punto, per scegliere in modo adeguato le metriche su cui effettuare la cluster analysis, si vuole ottenere una matrice di correlazione tra tutte le variabili calcolate per ogni singolo trimestre: l'obiettivo è effettuare una clusterizzazione su più metriche possibili, in modo da considerare il maggior contenuto informativo, senza però includere variabili tra loro fortemente correlate, in quanto causerebbero solo maggior rumore e ridondanza nella cluster analysis senza aggiungere informazioni importanti. Anche questo passaggio avviene con Python.

```
# Importazione della libreria seaborn per la data visualization
import seaborn as sn


# Definizione funzione per creare la matrice di correlazione
def checkCorrelation(db):
    cm = db.corr() # Effettua la correlazione tra ogni coppia di indicatori
    mask = np.array(cm)
    mask[np.tril_indices_from(mask)] = False
    fig = plt.gcf()
    fig.set_size_inches(25,10)
    sn.heatmap(data=cm, mask=mask, square=True, annot=True, cbar=True)
```

La figura 5.1 mostra la matrice di correlazione che si ottiene:

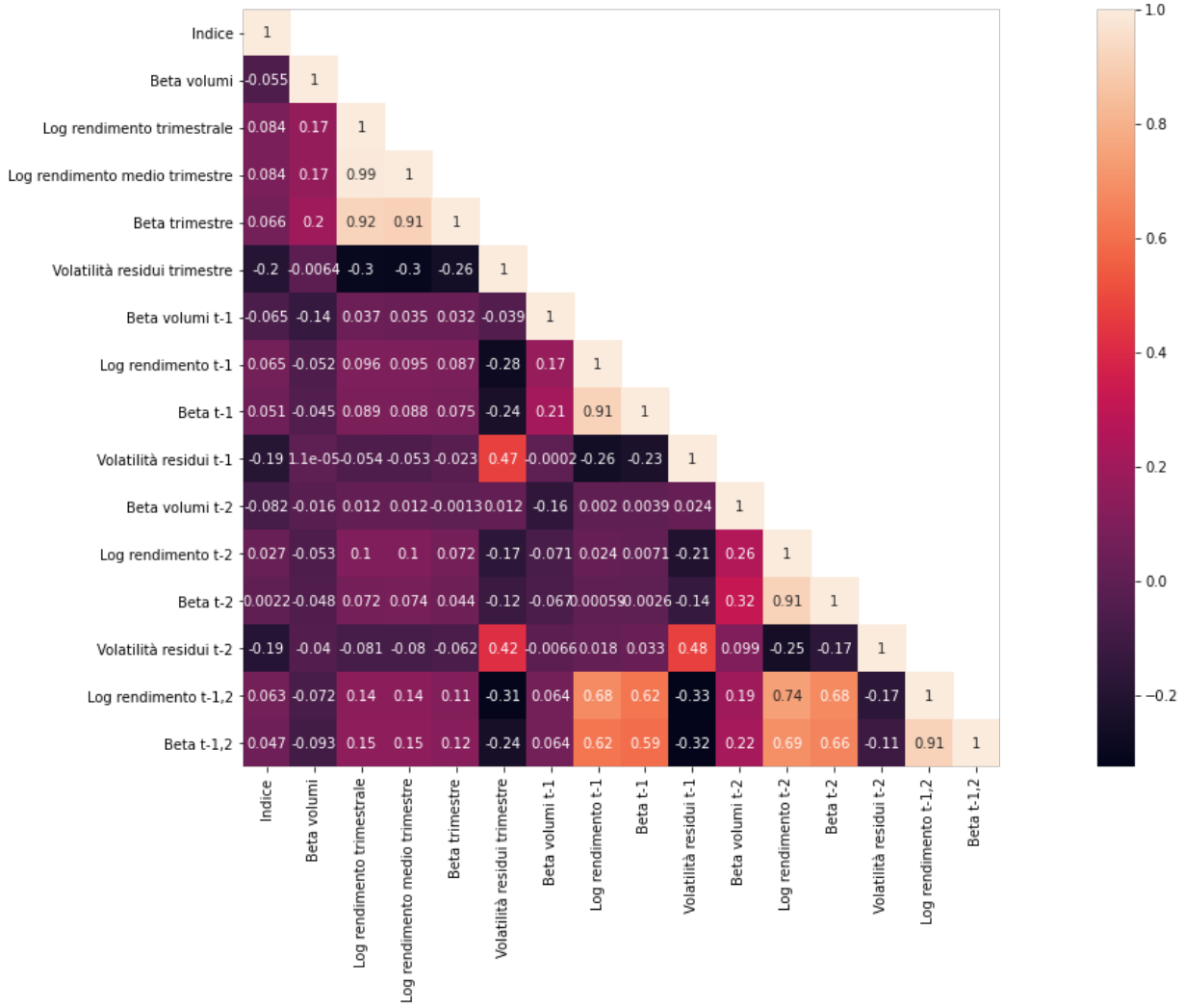


Figura 5.1: matrice di correlazione

5.6 Svolgimento della cluster analysis

Analizzando la matrice di correlazione, si sceglie di effettuare la cluster analysis sulle seguenti metriche:

- **β volumi**
- **β trimestre**
- **Volatilità residui trimestre**
- **β volumi t-1**
- **β t-1[4]**
- **Volatilità residui t-1**
- **β t-2[5]**

[4] Beta trimestre precedente
[5] Beta del secondo trimestre precedente

Si è preferito non inserire gli indicatori *β volume t-2* e *volatilità residui t-2* nelle variabili da considerare per la cluster analysis per non appesantirla eccessivamente. Inoltre, per quanto importanti, sono misure che hanno più valore informativo se riferite al trimestre attuale o, al limite, al trimestre precedente.

Per effettuare la cluster analysis si utilizza la libreria sci-kit learn di python:

```
from sklearn.cluster import KMeans

# trasportiamo i valori del database in un array numpy X per poter effettuare
# la cluster analysis
X = df.values

# Dopo diverse prove si è optato per la creazione di 14 clusters, giudicato un
# punto di equilibrio tra complessità del modello e omogeneità dei gruppi che
# si ottiene
kmeans = KMeans(n_clusters=14)

kmeans.fit(X[:,[3,6,7,8,10,11,14]])

KMeans(n_clusters=14)

y = kmeans.predict(X[:,[3,6,7,8,10,11,14]])
```

Il vettore y ottenuto dal clustering contiene, per ogni elemento di X, un label tra 0 e 13 che indica il cluster di appartenenza.

A questo punto si creano dei database separati contenenti gli elementi di ogni cluster:

```
# Creazione dei dataset separati per cluster
cluster1 = pd.DataFrame(X[y==0], columns = df.columns)
cluster2 = pd.DataFrame(X[y==1], columns = df.columns)
cluster3 = pd.DataFrame(X[y==2], columns = df.columns)
cluster4 = pd.DataFrame(X[y==3], columns = df.columns)
cluster5 = pd.DataFrame(X[y==4], columns = df.columns)
cluster6 = pd.DataFrame(X[y==5], columns = df.columns)
cluster7 = pd.DataFrame(X[y==6], columns = df.columns)
cluster8 = pd.DataFrame(X[y==7], columns = df.columns)
cluster9 = pd.DataFrame(X[y==8], columns = df.columns)
cluster10 = pd.DataFrame(X[y==9], columns = df.columns)
cluster11 = pd.DataFrame(X[y==10], columns = df.columns)
cluster12 = pd.DataFrame(X[y==11], columns = df.columns)
cluster13 = pd.DataFrame(X[y==12], columns = df.columns)
cluster14 = pd.DataFrame(X[y==13], columns = df.columns)
```

L'ultimo step di questa fase consiste nell'utilizzare la funzione describe dell'oggetto DataFrame di Pandas per ottenere delle tabelle riassuntive sulle metriche di ogni cluster per comprendere le caratteristiche economico-finanziarie di ciascun gruppo in modo da poterne fornire un'interpretazione adeguata:

```
statistiche_cl1 = cluster1.describe()
statistiche_cl2 = cluster2.describe()
statistiche_cl3 = cluster3.describe()
statistiche_cl4 = cluster4.describe()
statistiche_cl5 = cluster5.describe()
statistiche_cl6 = cluster6.describe()
statistiche_cl7 = cluster7.describe()
statistiche_cl8 = cluster8.describe()
statistiche_cl9 = cluster9.describe()
statistiche_cl10 = cluster10.describe()
statistiche_cl11 = cluster11.describe()
statistiche_cl12 = cluster12.describe()
statistiche_cl13 = cluster13.describe()
statistiche_cl14 = cluster14.describe()
```

Un esempio della tabella riassuntiva di ogni cluster che si ottiene con la funzione `describe`:

	Beta volumi	Beta t	Volatilità residui t	Beta volumi t-1	Beta t-1	Volatilità residui t-1	Beta t-2
mean	0,0812	0,0779	0,7359	0,1420	0,0465	1,2167	0,0396
std	0,4317	0,0693	0,1187	0,4264	0,0674	0,5004	0,0774
min	-0,8084	-0,0514	0,3388	-0,8084	-0,1563	0,3995	-0,2477
25%	-0,1010	0,0304	0,6551	-0,1158	0,0072	0,8687	-0,0072
50%	0,0701	0,0801	0,7562	0,0353	0,0497	1,1603	0,0376
75%	0,2761	0,1307	0,8335	0,3276	0,0980	1,4358	0,0940
max	0,8342	0,3077	0,9096	2,4668	0,2359	3,3539	0,2857

5.7 Creazione dei gruppi

La cluster analysis svolta ha permesso di fare una prima importante divisione in clusters lavorando su una grande mole di dati (più di 20.000 esempi). Senza una procedura simile sarebbe stato eccessivamente oneroso applicare una classificazione ai dati, vista la loro complessità e la loro numerosità. Fermarsi ai risultati ottenuti dalla cluster analysis, però, sarebbe estremamente limitativo. Per quanto indispensabile, infatti, il risultato offerto dalla cluster analysis non è sufficientemente preciso: i trimestri sono divisi in gruppi abbastanza omogenei, ma al loro interno sono presenti elementi ancora troppo diversi tra loro e sarebbe come minimo approssimativoclassificarli allo stesso modo.

Per questo motivo, partendo dai risultati ottenuti con la cluster analysis, è stata effettuata una verifica critica di ciascun gruppo escludendo gli elementi più estremi in modo da aumentarne l'omogeneità. Gli elementi eliminati dai clusters sono stati aggiunti al database di outliers creato precedentemente.

Il database di outliers contiene circa 2000 elementi, tra quelli selezionati inizialmente e quelli aggiunti successivamente alla divisione in gruppi. Questi elementi sono stati analizzati manualmente e, considerando il periodo storico e le principali metriche calcolate, sono stati inseriti nel gruppo che si è reputato essere il più rappresentativo, tra quelli ottenuti con la cluster analysis. Questo approccio non è sicuramente il più efficiente, ma permette una maggiore precisione del risultato finale e soprattutto permette di considerare tutti i trimestri senza escluderne nessuno.
In seguito a questa ulteriore fase elaborativa si ottengono i 14 gruppi definitivi. Di seguito, per ciascun gruppo, è presentata una tabella riassuntiva con le principali metriche calcolate, un'etichetta con l'interpretazione economico-finziaria corrispondente al gruppo, e un grafico che mostra l'andamento dei prezzi rappresentativo di ogni cluster su un orizzonte temporale di 9 mesi (il trimestre attuale e i due trimestri precedenti). Ogni punto del grafico è stato ottenuto dalla media dei prezzi degli elementi del cluster in quel giorno.

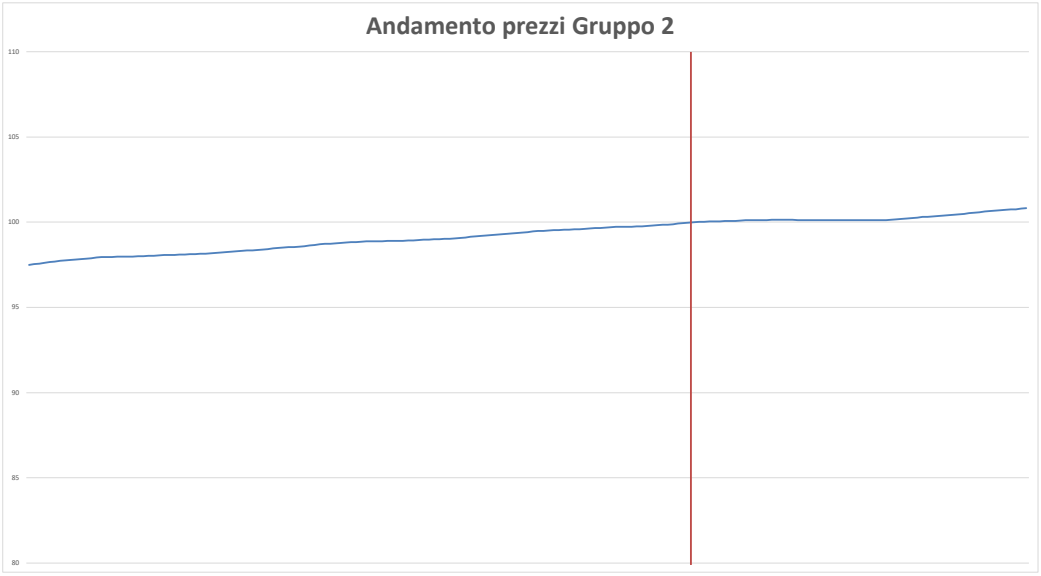
Gruppo 1

Gruppo 1 - Numero elementi: 3051						
	Beta volumi	**Beta t**	**Volatilità residui t**	**Beta t-1**	**Beta t-2**	**Beta t-1,2**
mean	-0,097239	0,018340	1,716713	0,106927	0,127126	0,130242
std	0,390398	0,052329	0,606005	0,083925	0,091324	0,046939
25%	-0,303429	-0,017563	1,308372	0,048464	0,076432	0,090536
50%	-0,092805	0,017633	1,588204	0,105455	0,125198	0,119623
75%	0,105065	0,057515	2,046569	0,158688	0,180131	0,163341

Interpretazione: trimestre medio con tendenza di lungo periodo in crescita

Gruppo 2

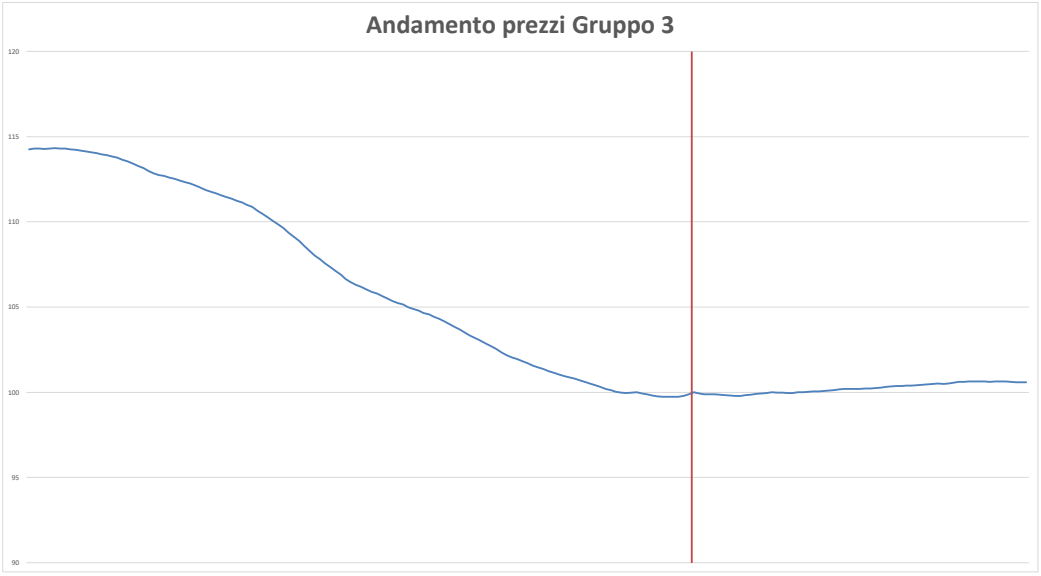
Gruppo 2 - Numero elementi: 4160						
	Beta volumi	Beta t	Volatilità residui t	Beta t-1	Beta t-2	Beta t-1,2
mean	0,034894	0,010755	1,528891	0,022699	0,021267	0,020351
std	0,404493	0,054106	0,630006	0,085703	0,084427	0,032121
25%	-0,197594	-0,022406	1,099739	-0,025876	-0,024067	-0,003032
50%	0,008352	0,013405	1,379175	0,023238	0,024096	0,026073
75%	0,252536	0,045294	1,817269	0,072806	0,073637	0,046605



Interpretazione: trimestre medio stabile

Gruppo 3

Gruppo 3 - Numero elementi: 938						
	Beta volumi	Beta t	Volatilità residui t	Beta t-1	Beta t-2	Beta t-1,2
mean	0,010266	0,014815	2,178297	-0,093260	-0,109339	-0,118326
std	0,610290	0,068193	0,712864	0,151953	0,145330	0,062471
25%	-0,355082	-0,034035	1,659631	-0,188728	-0,150075	-0,154421
50%	-0,073832	0,012496	2,105590	-0,082699	-0,097185	-0,097360
75%	0,309176	0,059114	2,424673	-0,002796	-0,033438	-0,068601



Interpretazione: trimestre medio in stabilizzazione post-decrescita

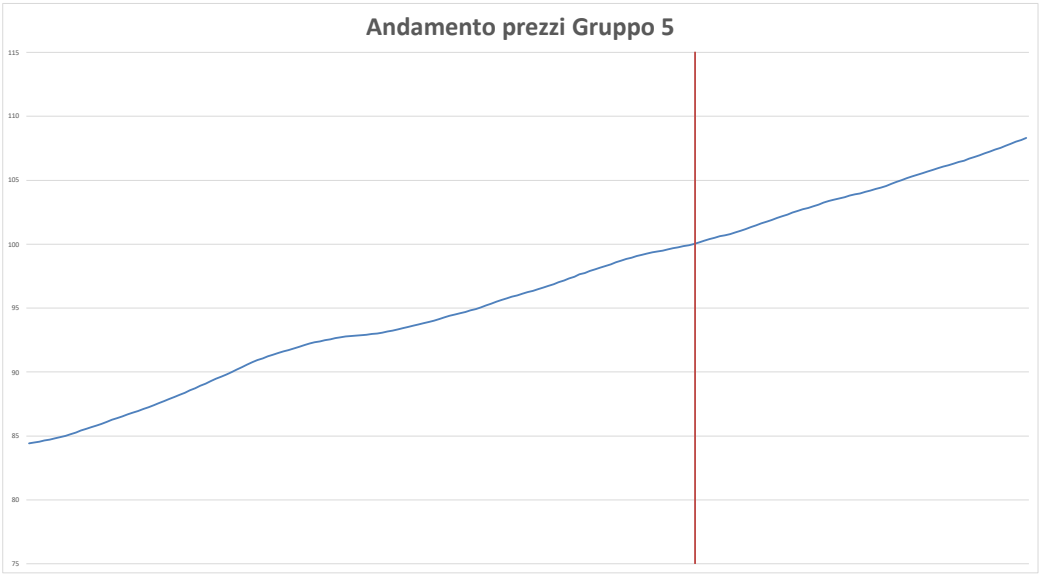
Gruppo 4

Gruppo 4 - Numero elementi: 762						
	Beta volumi	**Beta t**	**Volatilità residui t**	**Beta t-1**	**Beta t-2**	**Beta t-1,2**
mean	0,325128	0,131024	1,227593	0,137951	0,207508	0,167114
std	0,328952	0,078229	0,433138	0,080246	0,083802	0,067614
25%	0,194782	0,074957	0,528371	0,093921	0,141675	0,126013
50%	0,316693	0,131274	1,235921	0,131971	0,199797	0,159546
75%	0,463920	0,182331	1,894325	0,185243	0,251843	0,221666

Interpretazione: bolla in fase di accumulo

Gruppo 5

Gruppo 5 - Numero elementi: 966						
	Beta volumi	Beta t	Volatilità residui t	Beta t-1	Beta t-2	Beta t-1,2
mean	-0,084185	0,127655	2,275781	0,132838	0,178341	0,147401
std	0,337618	0,051946	0,416204	0,068957	0,077038	0,052331
25%	-0,402741	0,091986	1,742951	0,097896	0,122952	0,119231
50%	-0,087512	0,125936	2,239031	0,129786	0,152845	0,140131
75%	0,004523	0,159098	2,874853	0,167536	0,217700	0,166611



Interpretazione: bolla vicino allo scoppio

Gruppo 6

Gruppo 6 - Numero elementi: 1113						
	Beta volumi	Beta t	Volatilità residui t	Beta t-1	Beta t-2	Beta t-1,2
mean	-0,093814	0,149913	2,218777	0,207011	-0,093581	0,045932
std	0,557429	0,080384	0,518916	0,094657	0,134211	0,090178
25%	-0,157837	0,093091	1,828406	0,130788	-0,165983	-0,006508
50%	-0,089463	0,144901	2,113591	0,190756	-0,072849	0,059296
75%	0,004892	0,205465	2,460918	0,250444	0,015412	0,098817

Interpretazione: mercato in crescita instabile nel medio periodo

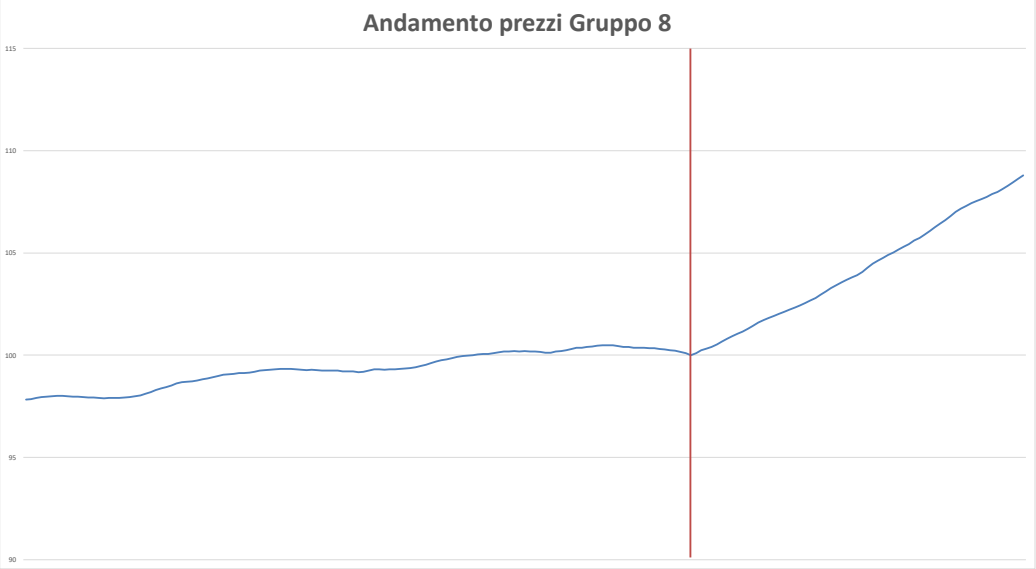
Gruppo 7

Gruppo 7 - Numero elementi: 1081						
	Beta volumi	**Beta t**	**Volatilità residui t**	**Beta t-1**	**Beta t-2**	**Beta t-1,2**
mean	0,398437	0,131666	1,127859	0,152247	-0,028589	0,062164
std	0,583512	0,056422	0,284386	0,053270	0,087581	0,059754
25%	0,185139	0,098660	0,938846	0,110009	-0,081233	0,024498
50%	0,318943	0,129711	1,130122	0,136502	-0,011414	0,063905
75%	0,540867	0,161336	1,337500	0,185033	0,040147	0,102151

Interpretazione: mercato in crescita nel medio periodo

Gruppo 8

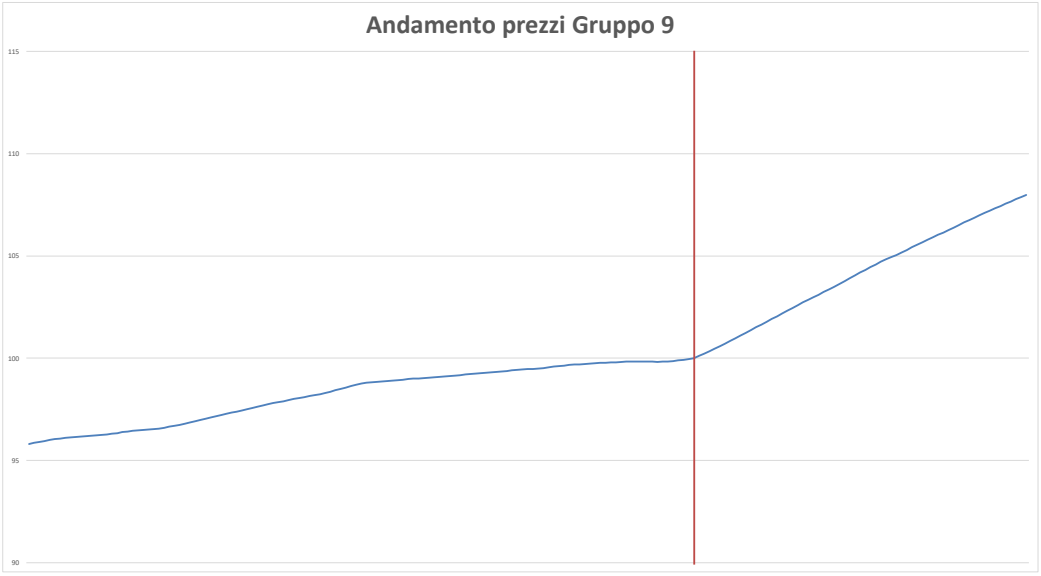
Gruppo 8 - Numero elementi: 895						
	Beta volumi	Beta t	Volatilità residui t	Beta t-1	Beta t-2	Beta t-1,2
mean	0,281562	0,139556	2,262686	0,020583	0,034761	0,026942
std	0,579968	0,079963	0,599942	0,040838	0,090295	0,071057
25%	-0,069156	0,085026	1,828027	-0,009009	-0,020885	-0,016309
50%	0,162133	0,134660	2,101201	0,021087	0,033887	0,031152
75%	0,442398	0,185937	2,468577	0,056136	0,107281	0,075512



Interpretazione: boom di mercato

Gruppo 9

Gruppo 9 - Numero elementi: 1979						
	Beta volumi	Beta t	Volatilità residui t	Beta t-1	Beta t-2	Beta t-1,2
mean	0,317976	0,126976	1,096169	0,020612	0,050323	0,039169
std	0,752935	0,057869	0,304640	0,041777	0,071740	0,047104
25%	-0,017710	0,094044	0,864100	-0,014588	0,006331	0,013364
50%	0,283237	0,122758	1,118664	0,022666	0,055368	0,040972
75%	0,683948	0,151212	1,341312	0,057008	0,099808	0,067928



Interpretazione: mercato in rialzo stabile

Gruppo 10

Gruppo 10 - Numero elementi: 1223						
	Beta volumi	**Beta t**	**Volatilità residui t**	**Beta t-1**	**Beta t-2**	**Beta t-1,2**
mean	0,382038	0,158415	2,008056	-0,124500	-0,101023	-0,089200
std	0,934042	0,091848	0,857482	0,102914	0,119428	0,069177
25%	-0,203835	0,094377	1,424522	-0,163997	-0,127034	-0,131712
50%	0,244807	0,141351	1,933552	-0,113184	-0,067495	-0,064852
75%	0,799131	0,209273	2,488639	-0,077401	-0,025835	-0,040446

Interpretazione: ripresa del mercato

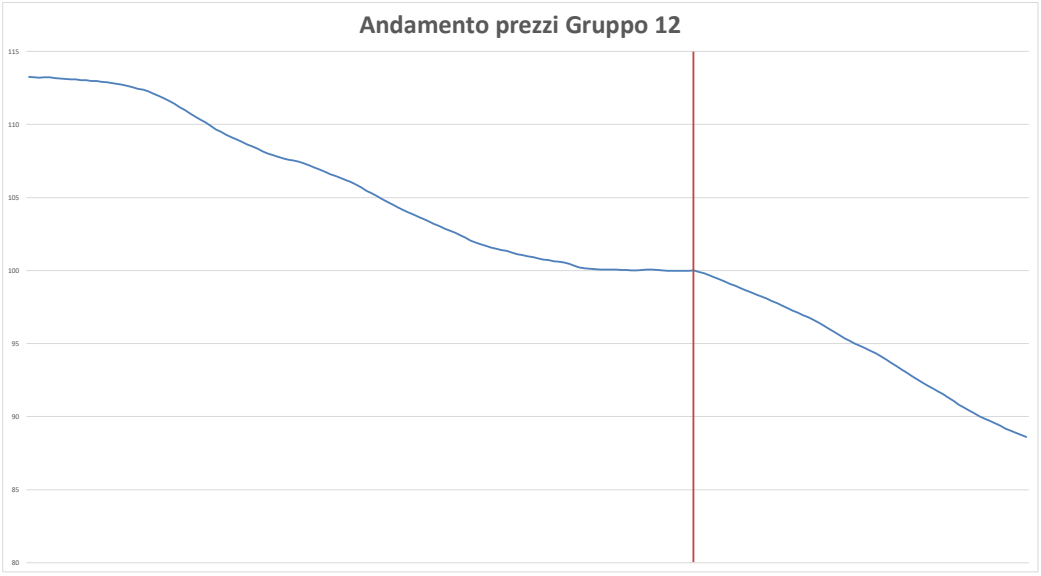
Gruppo 11

Gruppo 11 - Numero elementi: 781						
	Beta volumi	Beta t	Volatilità residui t	Beta t-1	Beta t-2	Beta t-1,2
mean	0,280968	0,171624	2,624681	-0,141901	0,128123	-0,006144
std	0,556275	0,098224	0,731163	0,073334	0,096953	0,068915
25%	-0,080752	0,104230	1,710818	-0,181023	0,073518	-0,036993
50%	0,154035	0,148943	2,551673	-0,122660	0,118108	-0,011706
75%	0,506136	0,241707	2,973945	-0,087160	0,168018	0,036060

Interpretazione: mercato fortemente instabile

Gruppo 12

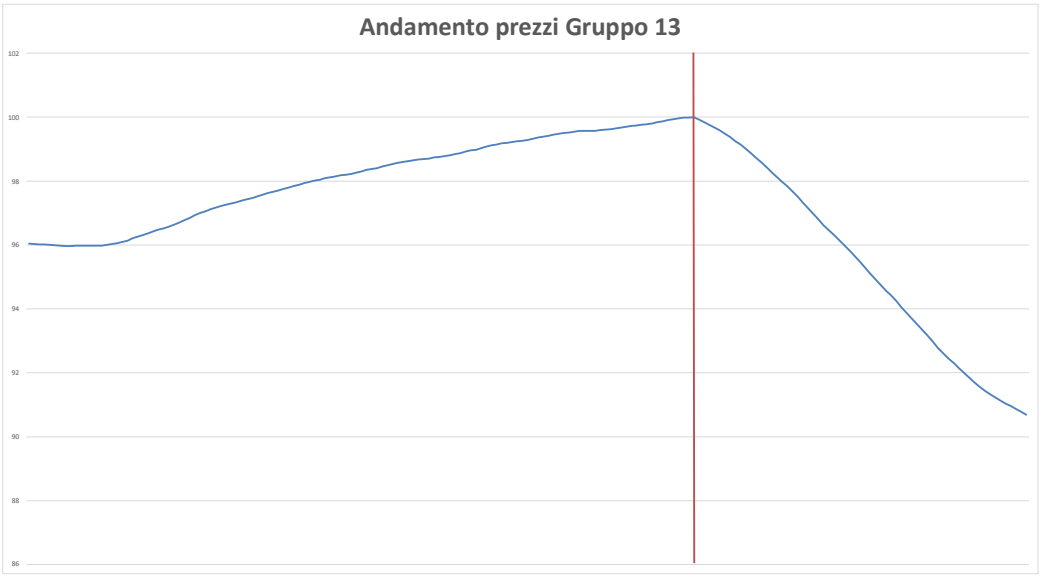
Gruppo 12 - Numero elementi: 1017						
	Beta volumi	Beta t	Volatilità residui t	Beta t-1	Beta t-2	Beta t-1,2
mean	-0,000809	-0,187601	2,562062	-0,070318	-0,111107	-0,109925
std	0,806577	0,135467	0,920501	0,164036	0,165482	0,053661
25%	-0,389966	-0,233180	1,798345	-0,133824	-0,184250	-0,133628
50%	-0,067819	-0,153881	2,448890	-0,066613	-0,087196	-0,092766
75%	0,237712	-0,102438	3,206003	0,025985	-0,012040	-0,070861



Interpretazione: mercato in depressione

Gruppo 13

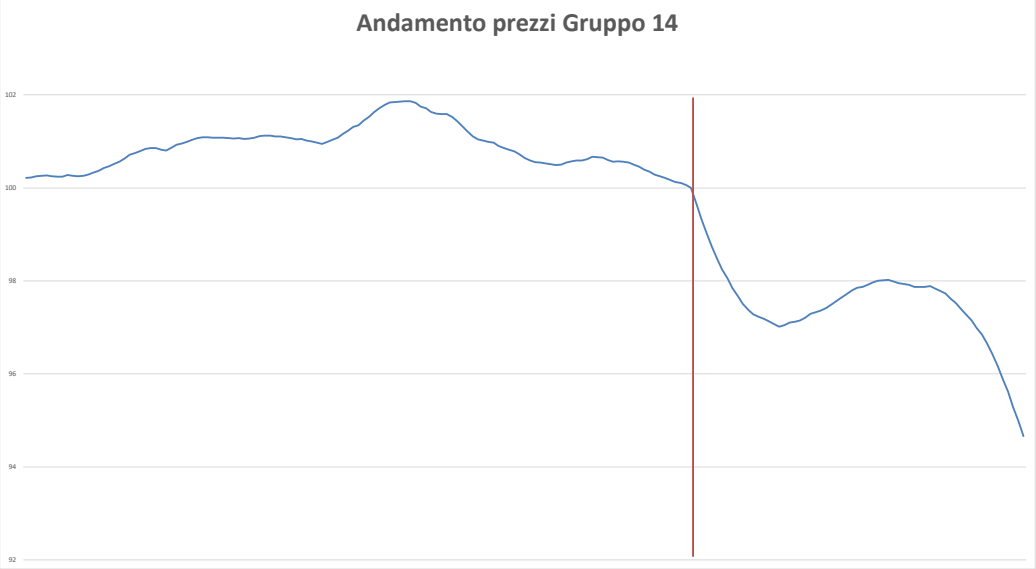
Gruppo 13 - Numero elementi: 2742						
	Beta volumi	Beta t	Volatilità residui t	Beta t-1	Beta t-2	Beta t-1,2
mean	0,154732	-0,159285	2,419262	0,037411	0,055061	0,047365
std	0,584902	0,113152	1,247248	0,132472	0,135370	0,092078
25%	0,002378	-0,198461	1,584938	-0,043398	-0,029760	0,000340
50%	0,139046	-0,127195	2,057670	0,036696	0,061410	0,045112
75%	0,274194	-0,094969	2,816645	0,130720	0,141482	0,094352



Interpretazione: crollo del mercato

Gruppo 14

Gruppo 14 - Numero elementi: 661						
	Beta volumi	Beta t	Volatilità residui t	Beta t-1	Beta t-2	Beta t-1,2
mean	-0,088544	-0,030161	3,794725	-0,021308	0,020113	0,002420
std	0,596461	0,037892	0,626476	0,102840	0,101786	0,043561
25%	-0,300946	-0,059685	1,963628	-0,092466	-0,053065	-0,027088
50%	-0,044869	-0,037346	3,536021	-0,016850	0,029451	-0,005707
75%	0,252567	-0,011530	4,118943	0,052009	0,100548	0,030628



Interpretazione: mercato in calo con tendenza di medio periodo instabile

Si può notare che alcuni grafici sembrano somigliarsi molto, nonostante rappresentino gruppi diversi. Queste somiglianze, tuttavia, sono accettabili: in primis sono dovute al fatto che, essendo ogni punto di ogni grafico la media di un insieme di valori, molti trend irregolari sono smorzati, quindi la rappresentazione grafica dell'insieme dei prezzi dei trimestri del cluster non coglie bene le differenze nella volatilità attorno al trend; e inoltre bisogna considerare che l'andamento dei prezzi è solo uno degli aspetti considerati per la divisione in gruppi, perciò è plausibile che trimestri in cui l'andamento dei prezzi è stato simile siano alla fine in gruppi diversi.

Seppur, per completezza di analisi, il modello che si sta costruendo ambisce a classificare e riconoscere tutte le tipologie di trimestri che si possono osservare sul mercato, lo scopo principale rimane quello di riuscire a identificare delle bolle speculative. A tal proposito, in seguito alla formazione dei 14 gruppi, è utile menzionare ulteriormente quelli su cui si desidera che la rete neurale ottenga risultati soddisfacenti in quanto possibili indicatori della presenza di una bolla speculativa: il *Gruppo 4*, il *Gruppo 5* e il *Gruppo 6*.

Capitolo 6

Costruzione della rete neurale e analisi dei risultati

6.1 Preparazione dei dati per la rete neurale

Terminato il raggruppamento dei trimestri si è ottenuto un database classificato dei trimestri azionari dello S&P 500 dal 1928 al 2021.
Ogni riga rappresenta un trimestre con le sue rispettive metriche e con un'etichetta che indica a quale dei 14 gruppi appartiene:

Data inizio	Data fine	Beta volumi	Beta trimestre	…	…	Gruppo
06/07/28	05/10/28	0,2769	0,2377	…	…	9
09/07/28	08/10/28	0,2776	0,2385	…		11
10/07/28	09/10/28	0,2807	0,2411			11
11/07/28	10/10/28	0,2929	0,2505			11

Questo passaggio è fondamentale perché ci consente di impostare un problema di apprendimento supervisionato in cui la rete neurale verrà addestrata su dati in input e dati in output.
Il label di ogni trimestre rappresenta la variabile in output che si vuole che la rete neurale, una volta completato l'addestramento, sia in grado di fornire in autonomia.
Per quanto riguarda le variabili in input, è necessario applicare ulteriori trasformazioni ai dati. Passare alla rete neurale come dato in input il trimestre *monodimensionale* su una riga con le sue rispettive metriche sarebbe estremamente riduttivo: in tal caso, infatti, la rete non farebbe altro che imparare ciò che si è già ottenuto con la cluster analysis, ovvero dividere in gruppi i trimestri sulla base delle metriche fornite. Non si sfrutterebbero le potenzialità delle reti neurali e, soprattutto, non si otterrebbe nulla di più dal modello.
L'obiettivo è creare una rete neurale che sia in grado, con dati grezzi, di assegnare ogni trimestre al gruppo corretto replicando in modo molto più snello tutte le operazioni svolte in fase di raggruppamento (cluster analysis e analisi manuale) e che sia in grado di astrarre ulteriormente le caratteristiche dei dati per trovare pattern nuovi.
A tal fine si vuole costruire una rete *comoda*, che non ha bisogno di complicate elaborazioni prima di essere utilizzata, ma che produca dei risultati con i dati più semplici a disposizione. Per questo motivo si è deciso di fornire alla rete solo i dati sui prezzi e sui volumi giornalieri per la durata di tre trimestri (quello su cui si vuole ottenere una previsione e i due precedenti) e sfruttare le sue capacità di elaborazione di dati sequenziali.

Il dato in input della rete neurale sarà quindi una matrice:

Data	Prezzo	Volume
03/01/28	17,76	2489823
04/01/28	17,72	2483190
05/01/28	17,55	2455002
06/01/28	17,66	2473241
09/01/28	17,50	2446711
10/01/28	17,37	2425155
…	…	…
06/07/28	19,39	2760098
09/07/28	19,46	2771705
10/07/28	19,43	2766731
11/07/28	18,95	2687141
…	…	…
02/10/28	21,26	3070169
03/10/28	21,19	3058563
04/10/28	21,26	3070169
05/10/28	21,22	3063537

La tabella mostra la matrice fornita alla rete neurale come dato in input per il primo trimestre. Il primo trimestre parte il 06/07/1928 e finisce il 05/10/1928. Partendo dai dati di due trimestri prima, la matrice contiene i dati giornalieri su prezzi e volumi dal 03/10/1928 al 05/10/1928[6].

Per ogni trimestre si vuole quindi ottenere una matrice di 193 righe (numero di trading days in 3 trimestri) e due colonne (una per i prezzi e una per i volumi). Partendo da un database bidimensionale, bisogna ottenerne uno tridimensionale formato da un insieme di matrici. La rete neurale riceverà come dato in input una matrice e avrà come dato in output un label che indica a quale gruppo appartiene la matrice (cioè il trimestre azionario).

Queste operazioni di trasformazione dei dati vengono svolte su Python.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

nameDB = "S&P_NN.xlsx"
dataset = pd.read_excel(nameDB)

dataset = dataset.drop(["Indice", "Data inizio", "Data fine"], axis = 1)

X = dataset.values

# definizione della funzione che divide i dati grezzi in periodi di 3
# trimestri con finestra mobile 1 giorno. Si otterrà quindi una struttura
# dati tridimensionale
```

[6] La colonna “Data” non è presente nella matrice fornita alla rete neurale. È inserita in questa tabella per facilitare la comprensione dell’esempio.

```
def reshapeDB(quarters, days, features, dataset):
    db = np.zeros((quarters, days, features))
    for i in range(0, quarters):
        for j in range(0, days):
            db[i,j] = dataset[j+i]
    return db

# definizione del numero di elementi totali, numero di giorni per ogni
# elemento e numero di features per ogni giorno (ovvero 2: prezzo e volume)
n_quarters = X.shape[0]
n_days = 193
n_features = 2

# applicazione della funzione
array3D = reshapeDB(n_quarters, n_days, n_features, X)
```

Si è rimodellato il database iniziale ottenendo un array3D di dimensione (23258, 193, 2), quindi 23258 matrici (una per ogni trimestre) di 193 righe e 2 colonne.

6.2 Pre-processing sui dati

A questo punto si effettuano delle operazioni di pre-processing sui dati. È buona norma, infatti, far lavorare la rete neurale su dati adeguatamente trasformati. Sui dati in input, quindi sulle matrici, si applica una normalizzazione fissando a 100 i dati del primo giorno del trimestre (ovvero la riga 129 di ogni matrice) in analisi e trasformando i dati degli altri giorni di conseguenza (sia i giorni successivi del trimestre sia i giorni dei due trimestri precedenti). È importante avere dati sulla stessa scala perché altrimenti l'algoritmo di apprendimento potrebbe dare più importanza alle colonne con range di valori più alti.

Successivamente si standardizzano i dati. Si sceglie la standardizzazione in quanto mantiene le informazioni riguardo gli outliers di ogni matrice. Inoltre, spesso, avere i dati in forma di distribuzione normale agevola la fase di addestramento.

È possibile effettuare queste operazioni di pre-processing su Python.

```
# definizione della funzione di normalizzazione
def db_normalization(db):
    for i in range(0, db.shape[0]):
        firstPrice = db[i, 128, 0]
        firstVolume = db[i, 128, 1]
        for j in range(0, db.shape[1]):
            db[i,j,0] = db[i,j,0] / firstPrice
            db[i,j,1] = db[i,j,1] / firstVolume

array3D = db_normalization(array3D)
```

```
# applicazione della funzione di standardizzazione in cui ogni prezzo e volume
# si divide per la media dei 9 mesi e si divide per la deviazione standard
def db_standardization(db):
    for i in range(0, db.shape[0]):
        mean = np.mean(db[i], axis = 0)
        std_dev = np.mean(db[i], axis = 0)
        for j in range(0, db.shape[1]):
            db[i,j,0] = (db[i,j,0] - mean[0]) / std_dev[0]
            db[i,j,1] = (db[i,j,1] - mean[1]) / std_dev[1]

array3D = db_normalization(array3D)
```

A questo punto si dividono i dati, sia quelli in input sia i label di output, in due set: il training set, su cui la rete si addestrerà, e il test set, su cui la rete testerà quanto è riuscita ad apprendere.

```
nameDB = "LabeledDB.xlsx"
labeledDB = pd.read_excel(nameDB)

y = labeledDB["Gruppo"].values

from sklearn.model_selection import train_test_split

# utilizzo della funzione train_test_split di sci-kit learn, si sceglie
# di utilizzare il 70% dei dati per il training set e il 30% per il test set
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.3)
```

A questo punto si effettua una trasformazione anche sui dati in output. La variabile di output è un label che può assumere un valore tra 1 e 14, a seconda del gruppo a cui appartiene il trimestre di riferimento. La rete neurale dovrà avere 14 nodi di output, quindi si devono creare 14 output differenti, trasformando l'output da un numero intero ad un array.

```
from keras.utils import to_categorical

# utilizzo della funzione to_categorical di keras per trasformare l'output
# da numero intero a vettore
y_train_cat = to_categorical(y_train)
y_test_cat = to_categorical(y_test)
```

Esempio output pre-trasformazione: 5

Esempio output post-trasformazione:

0	0	0	0	1	0	0	0	0	0	0	0	0	0

6.3 Implementazione della rete neurale

La tipologia di rete neurale più adatta a lavorare su una sequenza temporale di dati è una rete neurale ricorrente. In questo modello, in particolare, si sceglie di utilizzare l'architettura Long Short Term Memory, abbreviata LSTM. Come approfondito nel Capitolo 4, le LSTM sono tra le più innovative e complesse tipologie di reti neurali ricorrenti. Il loro grande successo è dovuto alla capacità di limitare fortemente il problema di scomparsa del gradiente. Inoltre, hanno dimostrato di essere molto efficaci nell'apprendere dinamiche temporali di breve e lungo periodo tra i dati su cui vengono addestrate, risultando molto utili in problemi in cui la sequenza di dati è molto lunga, come in questo caso.

Configurare l'architettura di una rete neurale, tuttavia, è un'operazione molto più complessa ed articolata rispetto a scegliere semplicemente quale tipo di rete utilizzare. Nel capitolo 4 si è parlato degli iperparametri di una rete neurale. Gli iperparametri sono un insieme di parametri che determinano il modo in cui viene addestrata la rete neurale e la sua struttura. Sono presenti molti iperparametri che, se impostati in modo intelligente, permettono di migliorare le prestazioni della rete. La difficoltà sta nel fatto che non c'è una vera e propria teoria che spieghi quali siano gli iperparametri ottimali a seconda del problema che si sta affrontando.
A tal fine, per trovare una configurazione della rete efficiente e funzionale al problema che si intende svolgere, anziché provare in modo randomico diverse alternative cambiando di volta in volta gli iperparametri, è stata effettuata una *grid search* sull'architettura di base scelta per trovare, per ciascun iperparametro desiderato, il valore che garantisse le migliori prestazioni.

La *grid search* è una tecnica di tuning che serve a ottimizzare il valore degli iperparametri. L'architettura di base scelta, sulla quale effettuare l'ottimizzazione dei parametri, consiste in:

- Layer LSTM
- Layer Dropout
- Layer LSTM
- Layer Dropout
- Layer LSTM
- Layer Dropout
- Layer Dense con 100 nodi
- Layer Dense con 14 nodi di output

Gli iperparametri che saranno ottimizzati su questa architettura sono:

- Numero di nodi negli strati LSTM: si stabilirà, tra le alternative fornite, il numero di nodi di ciascun layer LSTM che permette di ottenere le migliori prestazioni
- Numero di nodi nel primo strato Dense: allo stesso modo si ottimizzerà il numero di nodi presente nel primo strato denso della rete. Il secondo strato denso invece, essendo l'ultimo, avrà 14 nodi di output siccome ogni trimestre può appartenere ad uno dei 14 gruppi.

- Dropout_rate: dopo aver stabilito il miglior numero di nodi nei layer LSTM si ottimizzerà il dropout_rate, ovvero la percentuale di nodi da non considerare nell'addestramento ad ogni esecuzione. Come spiegato nel Capitolo 4, il dropout è una tecnica che permette di limitare il problema dell'overfitting
- Optimizer: sarà scelto l'algoritmo di addestramento che ottimizza le prestazioni della rete neurale. Ce ne sono diversi e a seconda del problema da affrontare può cambiare la miglior soluzione. Le alternative che saranno provate durante la grid search sono lo Stochastic Gradient Descent (SGD), l'RMSProp, l'Adam e l'Adagrad.
- Learning rate: una volta selezionato il miglior algoritmo di ottimizzazione, sarà ottimizzato il learning rate, ovvero la velocità con cui la rete aggiorna i propri pesi.
- Modalità di inizializzazione dei pesi: se i pesi dei nodi della rete vengono inizializzati casualmente, c'è un maggior rischio che si presenti il problema di esplosione o scomparsa del gradiente. Per questo motivo sono presenti delle *inizializzazioni intelligenti* da impostare sui pesi. Sarà scelta l'inizializzazione che ottimizza i risultati della rete

Sono presenti anche altri iperparametri, ma si è ritenuto che per il livello di complessità di questo modello fosse sufficiente ottimizzare quelli presentati.
Ogni *grid search* sarà effettuata su 100 epoche e utilizzando un *batch size* di 1024 e proverà ogni combinazione possibile con i valori forniti per gli iperparametri restituendo quella che ha ottenuto i risultati migliori.
Questa fase non rappresenta il vero e proprio addestramento della rete neurale finale, ma serve solo a stabilire quali sono i migliori parametri con cui configurare la rete neurale. Per questo motivo non è necessario svolgere la *grid search* su tutto il training set: è sufficiente selezionarne un sottoinsieme in modo che i risultati ottenuti, seppur meno prestanti di quelli garantiti da un addestramento sull'intero training set, consentano di individuare i migliori parametri.
La rete neurale finale ottimizzata, invece, lavorerà sull'intero training set, per un numero di epoche anche maggiore.

6.3.1 Tuning del numero di nodi nei layer LSTM e Dense

```
# importazione delle librerie necessarie
from keras.models import Sequential
from keras.layers import Dense, LSTM, Dropout
from sklearn.model_selection import GridSearchCV
from keras.wrappers.scikit_learn import KerasClassifier
from keras.constraints import maxnorm
```

```
# funzione che crea il modello su cui effettuare la grid search:
# sono passati come parametri della funzione dei valori di default da
# impostare nei layer LSTM. Durante la grid search avverrà proprio
# l'ottimizzazione di questi valori
def create_model(LSTM1=100, LSTM2=64, LSTM3=32, Dense1=50):

    model = Sequential() # è il metodo che crea l'istanza di una rete neurale

    # creazione della rete aggiungendo i layer desiderati
    model.add(LSTM(LSTM1, input_shape=(X_train.shape[1], X_train.shape[2]),
    return_sequences=True, dropout=0.3, recurrent_dropout=0.2))
    model.add(LSTM(LSTM2, dropout=0.3, recurrent_dropout=0.2,
    return_sequences = True))
    model.add(LSTM(LSTM3, dropout=0.3, recurrent_dropout=0.2))
    model.add(Dense(Dense1, activation="relu"))
    model.add(Dense(14, activation="softmax"))

    # si compila il modello. Come funzione di costo è stata scelta la
    # categorical_crossentropy che è adatta nei casi di classificazione
    # multiclasse
    model.compile(loss='categorical_crossentropy', optimizer='rmsprop',
    metrics = ['accuracy'])

    return model


# Per ogni iperparametro da ottimizzare è specificata una lista di valori
# tra cui trovare l'ottimo
LSTM1_config = [64, 100, 128]
LSTM2_config = [64,100,128]
LSTM3_config = [32, 64, 100, 128]
Dense1_config = [32, 100, 150]


# specificazione del numero di epoche e batch size con cui avviare la ricerca
model = KerasClassifier(build_fn=create_model, epochs=100, batch_size=1024)


# creazione di un dizionario con gli iperparametri da ottimizzare e le loro
# specifiche possibili configurazioni
param_grid = dict(LSTM1=LSTM1_config, LSTM2=LSTM2_config, LSTM3=LSTM3_config,
Dense1 = Dense1_config)
```

```
# definizione della grid search. La grid search lavora con il metodo
# statistico della k-fold cross-validation, ovvero valuta la bontà dei
# parametri su k campioni del set che le viene fornito. Quindi cv = 3
# significa che il set su cui sarà svolta la grid search sarà diviso in
# 3 campioni e ogni configurazione sarà valutata su ognuno dei campioni
grid = GridSearchCV(estimator=model, param_grid=param_grid, cv=3)


# si sceglie di effettuare la grid search su un set di 3072 elementi
X_grid = X_train[:3072]
y_grid = y_train_cat[:3072]


# esecuzione della grid search
grid_result = grid.fit(X_grid, y_grid)
```

Il metodo grid_result.best_params fornisce alla fine della *grid search* l'elenco dei migliori parametri e dei rispettivi valori.
Per quanto riguarda questa prima fase di tuning si ottiene:

- Miglior valore LSTM1: 128
- Miglior valore LSTM2: 128
- Miglior valore LSTM3: 100
- Miglior valore Dense1: 100

Analizzando le prestazioni delle varie configurazioni, si nota che anche la combinazione [LSTM1 = 128, LSTM2 = 128, LSTM3 = 64, Dense1 = 100] e la combinazione [LSTM1 = 128, LSTM2 = 128, LSTM3 = 64, Dense1 = 0] hanno risultati soddisfacenti, e sono quindi architetture che è possibile prendere in considerazione per la rete neurale finale.

6.3.2 Tuning del Dropout e dell'algoritmo di ottimizzazione

Una volta ottenuta la combinazione ottimale dei nodi nei layer LSTM e Dense, si procede alla seconda fase di tuning, ovvero quella relativa al *dropout_rate* e all'algoritmo di ottimizzazione. Il procedimento è analogo a quello svolto per la prima ottimizzazione.

```
# creazione del modello. Questa volta come parametri della funzione
# sono passati il dropout_rate e l'optimizer, ovvero l'algoritmo
# ottimizzazione. Per gli iperparametri già ottimizzati si inseriscono
# i valori migliori
def create_model(dropout_rate = 0.3, optimizer ='SGD'):
    model = Sequential()
    model.add(LSTM(128, input_shape=(X_train.shape[1], X_train.shape[2]),
    return_sequences=True, dropout= dropout_rate, recurrent_dropout=0.2))
    model.add(LSTM(128, dropout= dropout_rate, recurrent_dropout=0.2,
    return_sequences = True))
```

```
    model.add(LSTM(100, dropout= dropout_rate, recurrent_dropout=0.2))
    model.add(Dense(100, activation="relu"))
    model.add(Dense(14, activation="softmax"))

    # Compilazione del modello
    model.compile(loss='categorical_crossentropy', optimizer= optimizer,
    metrics=['accuracy'])

    return model


dropout_rate_config = [0.1, 0.2, 0.3, 0.4]
optimizer_config = ['SGD', 'RMSprop', 'Adagrad', 'Adam']

model = KerasClassifier(build_fn=create_model, epochs=100, batch_size=1024)

param_grid = dict(dropout_rate = dropout_rate_config,
optimizer = optimizer_config)

grid = GridSearchCV(estimator=model, param_grid=param_grid, cv=3)

X_grid = X_train[:3072]
y_grid = y_train_cat[:3072]

grid_result = grid.fit(X_grid, y_grid)
```

A fine elaborazione, richiamando il metodo grid_result.best_params, si ottiene la combinazione migliore degli iperparametri ottimizzati:

- Miglior valore dropout_rate: 0.2
- Miglior valore optimizer: Adam

Il tuning sull'algoritmo di ottimizzazione ha confermato ciò che si evince dalla letteratura. L'algoritmo Adam, infatti, è uno dei più performanti per molte tipologie di problemi. L'Adam mette insieme le caratteristiche dell'RMSProp, un altro algoritmo adatto a limitare la riduzione eccessiva del learning rate, e il Nesterov Momentum, un parametro che, aggiunto al learning rate, permette di velocizzare la fase di apprendimento.

6.3.3 Tuning del learning rate e della modalità di inizializzazione dei pesi

L'ultima fase di tuning mira ad ottimizzare il learning rate e l'inizializzazione dei pesi. All'interno del modello vengono inseriti i valori corretti degli iperparametri già ottimizzati, per il resto il procedimento non cambia.

```
# Importazione degli ottimizzatori di keras
from keras import optimizers

def create_model(learning_rate = 0.1, init_mode ='uniform'):
    model = Sequential()
    model.add(LSTM(128, input_shape=(X_train.shape[1], X_train.shape[2]),
    return_sequences=True, dropout=0.2, recurrent_dropout=0.2,
    kernel_initializer=init_mode))
    model.add(LSTM(128, dropout= 0.2, recurrent_dropout=0.2,
    return_sequences = True, kernel_initializer=init_mode))
    model.add(LSTM(100, dropout= 0.2, recurrent_dropout=0.2,
    kernel_initializer=init_mode))
    model.add(Dense(100, activation="relu", kernel_initializer=init_mode))
    model.add(Dense(14, activation="softmax", kernel_initializer=init_mode))

    optimizer = optimizers.Adam(learning_rate=learning_rate)

    # Compilazione del modello
    model.compile(loss='categorical_crossentropy', optimizer= optimizer,
    metrics=['accuracy'])

    return model

learning_rate_config = [0.0001, 0.001, 0.01, 0.1]
init_mode_config = ['uniform', 'glorot_normal', 'glorot_uniform']

model = KerasClassifier(build_fn=create_model, epochs=100, batch_size=1024)

param_grid = dict(learning_rate = learning_rate_config,
init_mode = init_mode_config)

grid = GridSearchCV(estimator=model, param_grid=param_grid, cv=3)

X_grid = X_train[:3072]
y_grid = y_train_cat[:3072]

grid_result = grid.fit(X_grid, y_grid)
```

I valori ottimizzati dei parametri sono:

- Miglior valore learning rate: 0.001
- Miglior valore init mode: glorot_uniform

6.3.4 Costruzione della rete neurale finale

Terminata l'ottimizzazione degli iperparametri si ha a disposizione la miglior configurazione di essi, tra le alternative analizzate. È possibile quindi costruire la rete neurale finale, che sarà addestrata sull'intero set di addestramento e sarà verificata sul set di test.

```
model = Sequential()

model.add(LSTM(128, input_shape=(X_train.shape[1], X_train.shape[2]),
return_sequences=True, dropout=0.2, recurrent_dropout=0.2,
kernel_initializer="glorot_uniform"))

model.add(LSTM(128, dropout=0.2, recurrent_dropout=0.2,
return_sequences = True, kernel_initializer="glorot_uniform"))

model.add(LSTM(100, dropout=0.2, recurrent_dropout=0.2,
kernel_initializer="glorot_uniform"))

model.add(Dense(100, activation="relu", kernel_initializer="glorot_uniform"))

model.add(Dense(14, activation="softmax", kernel_initializer="glorot_uniform"))

optimizer = optimizers.Adam(learning_rate=0.001)

# Compilazione del modello
model.compile(loss='categorical_crossentropy', optimizer=optimizer,
metrics=['accuracy'])

# Addestramento della rete. Si sceglie un batch_size di 1024 e 200 epoche
model.fit(X_train, y_train_cat, batch_size=1024, validation_split=0.2,
epochs=200)
```

La performance della rete neurale sui dati del training set è un'accuracy del **86,13%.**

6.4 Valutazione dei risultati

Per valutare i risultati della rete neurale si utilizza la matrice di confusione. La matrice di confusione è una tabella che restituisce l'accuratezza di un modello. Ogni colonna della matrice contiene i valori predetti, e ogni riga rappresenta i valori reali. Oltre a fornire un'indicazione sull'accuratezza generale, la matrice di confusione è uno strumento molto utile perché permette di stabilire le prestazioni del modello su ogni specifica classe.

Nella tabella sottostante è rappresentato un esempio di matrice di confusione, in cui sono presenti due classi: classe negativa e classe positiva.

		Valori predetti	
		Classe negativa	*Classe positiva*
Valori reali	*Classe negativa*	Veri negativi	Falsi positivi
	Classe positiva	Falsi negativi	Veri positivi

$$\text{accuracy} = \frac{(\text{veri negativi} + \text{veri positivi})}{(\text{veri negativi} + \text{falsi positivi} + \text{falsi negativi} + \text{veri positivi})}$$

Nel caso di problema multiclasse non cambia nulla: i valori correttamente predetti sono sulla diagonale principale della matrice, e le altre celle rappresentano i valori erroneamente predetti classe per classe.

È possibile creare la matrice di confusione della rete neurale appena addestrata attraverso Python.

```
# si crea un vettore y_pred che conterrà le previsioni fornite dalla rete
# neurale su ciascun esempio del set di test
y_pred = model.predict_classes(X_test)

# si crea la matrice di confusion in cui y_test contiene i valori reali e
# y_pred contiene i valori predetti
cm = confusion_matrix(y_test, y_pred)

# esportazione della matrice di confusione su excel
df = pd.DataFrame(cm)
df.to_excel(excel_writer="Confusion_matrix.xlsx")
```

Si ottiene la seguente matrice di confusione:

		Valori predetti													
	Gruppi	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Valori reali	1	773	19	20	27	9	11	6	4	16	0	0	0	11	8
	2	11	1123	11	0	1	9	13	6	15	3	9	0	16	20
	3	0	3	258	0	0	5	0	1	0	9	5	12	2	3
	4	6	1	0	164	20	13	9	2	8	0	1	0	0	0
	5	5	5	0	16	253	5	2	5	3	0	0	0	0	0
	6	3	3	1	12	14	265	9	6	5	2	0	0	0	0
	7	2	9	0	0	14	12	258	18	6	1	0	0	0	0
	8	2	23	3	6	0	3	10	214	12	5	4	0	0	0
	9	6	39	1	0	2	0	6	14	520	8	14	0	0	0
	10	0	9	31	0	0	3	0	4	4	304	8	0	1	0
	11	2	5	1	0	0	0	0	1	4	0	204	2	0	0
	12	0	0	6	0	0	0	0	0	7	8	5	269	6	4
	13	15	18	0	0	0	0	0	0	0	0	0	21	766	18
	14	3	22	4	0	0	0	0	0	0	0	0	11	19	137

Accuracy = 85,91%
Bubble accuracy = 81,38%

Oltre all'*accuracy* del modello, è stata calcolata la metrica *bubble accuracy*, che misura l'accuratezza del modello nel riconoscere i trimestri appartenenti ai gruppi 4, 5 e 6, ovvero quelli indicati come sintomatici di bolle speculative. Siccome l'obiettivo principale del modello è individuare questo tipo di fenomeno, è importante verificare la sua precisione anche in questo senso.

Per completezza, sono presentati anche i risultati ottenuti con altre due architetture della rete neurale. In fase di tuning del numero di nodi dei layer, infatti, altre due configurazioni avevano restituito discreti risultati:

Prima architettura alternativa:
[LSTM1 = 128, LSTM2 = 128, LSTM3 = 64, Dense1 = 100, Dense2 = 14]

		Valori predetti													
	Gruppi	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Valori reali	1	824	10	0	6	21	6	3	4	11	0	0	0	14	5
	2	31	1047	3	1	4	16	18	6	20	1	19	0	33	38
	3	0	12	215	0	0	7	0	1	0	32	2	14	0	15
	4	8	1	5	108	73	12	6	4	4	0	3	0	0	0
	5	4	0	0	9	258	9	8	2	4	0	0	0	0	0
	6	3	6	4	9	7	257	26	7	1	0	0	0	0	0
	7	10	13	0	0	25	32	217	15	7	1	0	0	0	0
	8	21	23	1	9	0	15	3	164	31	7	8	0	0	0
	9	21	40	1	4	28	4	22	3	468	6	13	0	0	0
	10	0	9	22	0	0	7	2	6	6	294	18	0	0	0
	11	3	9	2	0	0	0	0	4	11	6	183	1	0	0
	12	0	0	32	0	0	0	0	0	0	0	0	251	12	10
	13	25	21	2	0	0	0	0	0	0	0	0	10	772	8
	14	7	35	5	0	0	0	0	0	0	0	0	1	16	132

Accuracy = 80,95%
Bubble accuracy = 74,34%

Seconda architettura alternativa:
[LSTM1 = 128, LSTM2 = 128, LSTM3 = 64, Dense1 = 0, Dense2 = 15]

		Valori predetti													
	Gruppi	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Valori reali	1	733	23	0	10	50	22	12	4	12	0	0	0	32	6
	2	16	954	11	1	7	25	22	11	33	3	20	0	71	63
	3	0	4	219	0	0	5	0	2	0	24	11	13	7	13
	4	6	3	0	132	46	15	10	6	3	0	3	0	0	0
	5	7	4	0	7	262	0	4	0	10	0	0	0	0	0
	6	5	8	2	4	8	246	32	3	7	5	0	0	0	0
	7	8	16	1	0	21	33	230	0	8	3	0	0	0	0
	8	17	29	5	11	4	16	3	143	31	14	9	0	0	0
	9	14	43	1	7	25	2	49	7	435	16	11	0	0	0
	10	2	6	22	0	0	2	3	10	9	292	16	0	2	0
	11	1	8	1	0	1	0	0	7	13	7	179	1	1	0
	12	0	0	25	0	0	0	0	0	0	0	0	248	28	4
	13	40	14	0	0	0	0	0	0	0	0	0	16	757	11
	14	3	20	4	0	0	0	1	0	0	0	0	4	47	117

Accuracy = 77,16%
Bubble accuracy = 76,37%

I risultati sono piuttosto buoni. Il primo modello, con tutti i parametri ottimizzati dalla fase di tuning, ha ottenuto delle buone performance, mostrando un'accuracy generale superiore all'85% e un'accuracy relativa ai trimestri critici superiore all'80%.
Inoltre, considerando l'accuracy di 86,13% sui dati del training set, si può affermare che non è presente overfitting.

Le due architetture alternative, seppur con prestazioni inferiori rispetto al modello principale, hanno ottenuto comunque discreti risultati, con un'accuracy relativa ai trimestri critici vicina all'80%.

6.5 Errori di modello

Analizzando gli errori del modello si intende capire nello specifico quali sono le tipologie di errori maggiormente commesse dalla rete e la loro gravità.
Per distinguere i diversi tipi di errori, è possibile effettuare una macro-divisione dei gruppi in base alle tendenze generali di mercato che rappresentano:

- Gruppo 1, Gruppo 2, Gruppo 3: il mercato, nel trimestre di riferimento, mostra un trend tendenzialmente laterale [7]
- Gruppo 4, Gruppo 5, Gruppo 6, Gruppo 7, Gruppo 8, Gruppo 9: il mercato mostra un trend in rialzo
- Gruppo 10, Gruppo 11: il mercato segnala una ripresa in seguito ad un periodo di depressione o di instabilità
- Gruppo 12, Gruppo 13, Gruppo 14: il mercato mostra un trend in calo

Una classificazione errata della rete neurale è reputata più grave ai fini della bontà del modello se classifica un trimestre come appartenente ad un gruppo che non è nella stessa macro-classe di quello corretto. Errori di classificazione riguardanti gruppi non eccessivamente dissimili sono comunque da tenere in considerazione perché rappresentano una debolezza del modello, ma comunque meno gravi.
Per ciascun modello sono state calcolate tre misure:

$$\%\ errori\ gravi\ = \frac{numero\ errori\ gravi}{numero\ totale\ previsioni} * 100$$

$$incidenza\ errori\ gravi\ = \frac{numero\ errori\ gravi}{numero\ totale\ errori} * 100$$

$$\text{incidenza errori gravi } 4{,}5{,}6\ = \frac{numero\ errori\ gravi\ gruppi\ 4,5,6}{numero\ totale\ errori\ gruppi\ 4,5,6} * 100$$

Applicando queste formule ai modelli si ottengono i seguenti risultati:

- **Modello finale**:
  % errori gravi = 8,11%
  incidenza errori gravi = 57,59%
  incidenza errori gravi gruppi 4, 5, 6 = 17,30%
- **Prima architettura alternativa**:
  % errori gravi = 10,99%

[7] Un trend si dice laterale quando il prezzo si muove all'interno di un intervallo senza instaurare una tendenza stabile al rialzo o al ribasso.

incidenza errori gravi = 57,73%
incidenza errori gravi gruppi 4, 5, 6 = 15,81%

- **Seconda architettura alternativa**:
  % errori gravi = 14,12%
  incidenza errori gravi = 61,82%
  incidenza errori gravi gruppi 4, 5, 6 = 21,72%

In tutte e tre le architetture, in generale, il maggior numero di errori è dovuto al fatto che i trimestri tendono ad essere classificati nei gruppi 1, 2 e 3 più del dovuto. Ciò è probabilmente causato dal fatto che i primi 3 gruppi sono tra quelli con maggior numerosità (soprattutto il gruppo 1 e il gruppo 2, i più numerosi in assoluto) e quindi la rete neurale *impara* meglio i pattern relativi a quei gruppi e tende ad assegnarli anche ad altri tipi di trimestri.

6.6 Applicazione del modello al contesto attuale

Per concludere, l'intenzione è applicare il modello finale all'ultimo trimestre dello S&P 500, dall'8/04/21 al 9/07/21, e vedere la previsione fornita dalla rete neurale sulla situazione attuale di borsa.
Il procedimento segue esattamente i passi visti in questo capitolo per quanto riguarda l'elaborazione dei dati e il loro utilizzo all'interno della rete neurale. Per ottenere una previsione su quest'ultimo trimestre è necessario passare al modello gli ultimi 9 mesi dei dati su prezzi e volumi dello S&P 500.
Rilanciando la cluster analysis inserendo anche l'ultimo trimestre il modello lo classifica all'interno del Gruppo 6, quindi non uno dei due gruppi relativi a bolle speculative, ma comunque uno dei gruppi *critici* relativo a trend di crescita instabile nel medio lungo periodo. Ciò che è più interessante fare, però, è osservare la predizione della rete neurale sulla natura di questo trimestre e trarre delle conclusioni.
L'output della rete neurale è un vettore di 14 celle, in cui nella cella i-esima è inserita la probabilità del trimestre di appartenere all'i-esimo gruppo.
I tre gruppi di appartenenza più probabili secondo la rete neurale sono:

- Gruppo 6 (mercato in crescita instabile): 62,43%[8]
- Gruppo 4 (bolla in accumulo): 30,67%
- Gruppo 7 (mercato in crescita stabile nel medio periodo): 5,12%

È interessante notare che tutti e tre i gruppi più probabili sono gruppi che mostrano comunque tendenza al rialzo del mercato azionario. Il modello è stato quindi innanzitutto in grado di individuare correttamente questo trend. Inoltre, secondo il risultato ottenuto, c'è più del 93% di probabilità di trovarsi in un momento *critico* sul mercato. Secondo la rete c'è addirittura oltre il 30% di probabilità di trovarsi all'interno di una bolla.

[8] La rete neurale in questo caso ha confermato la classificazione ottenuta con la cluster analysis

6.7 Punti di forza, limiti e possibili miglioramenti del modello

Il modello ha ottenuto risultati piuttosto soddisfacenti. Sono stati combinati elementi di machine learning, statistica multivariata e deep learning, e il modello finale ha mostrato buone capacità nel riconoscere le tipologie di trimestri sul mercato finanziario. È sicuramente un approccio innovativo per individuare trend di mercato senza utilizzare framework tipici dell'analisi tecnica.
La rete neurale creata permette di giungere a risultati più che accettabili lavorando con pochi dati grezzi, considerando che per ogni trimestre sono stati passati alla rete solo i prezzi e i volumi. Questo aspetto conferisce al modello una facilità d'uso da non sottovalutare: per ottenere una previsione della rete sulla natura di un periodo di borsa è sufficiente fornirle esclusivamente i prezzi e i volumi scambiati in quel periodo, senza compiere ulteriori elaborazioni time-consuming.
Un altro punto di forza del modello è rappresentato dalla sua versatilità: la rete è stata costruita per valutare elementi di cadenza trimestrale e sui dati dello S&P 500, ma la struttura del modello, dalla fase di raccolta e analisi dei dati all'implementazione della rete neurale, permette di cambiare facilmente l'orizzonte temporale dell'analisi e anche la fonte dei dati. Ad oggi, pur con una crescente omogeneizzazione dei mercati finanziari più sviluppati, permangono delle differenze di paese in paese, e quindi l'elasticità del modello permette di renderlo adatto a diversi contesti.
Un possibile limite del modello è legato al fatto che sembra, in parte, verificarsi da solo, in quanto la rete neurale, riconoscendo le tipologie di trimestri, sostanzialmente riconferma i risultati ottenuti con la cluster analysis, che, per quanto approfondita, rappresenta comunque una procedura automatica basata su determinate metriche e assunzioni. Si potrebbe quindi pensare che non sia necessario l'utilizzo di una rete neurale, ma che sia sufficiente svolgere la cluster analysis per ottenere le informazioni desiderate. Questa osservazione non terrebbe però conto di alcuni aspetti importanti. Innanzitutto, c'è da sottolineare il fatto che i risultati della rete neurale sono stati ottenuti utilizzando molti meno dati rispetto a quelli necessari per svolgere la cluster analysis: se ci si limitasse all'uso di quest'ultima il modello perderebbe le sue doti di versatilità e elasticità, e ogni suo utilizzo su nuovi dati richiederebbe numerose elaborazioni. Inoltre, aspetto ancora più importante, la classificazione dei trimestri non è stata effettuata interamente con la clusyer analysis: molti elementi, soprattutto quelli più estremi, sono stati analizzati una seconda volta più nel dettaglio e inseriti manualmente nel gruppo ritenuto più corretto. Questo significa che, soprattutto per i trimestri caratterizzati da comportamenti non standard, come ad esempio le bolle speculative, la cui individuazione è il vero obiettivo di questo modello, la rete neurale non fornisce una previsione basata su una classificazione automatica effettuata precedentemente, ma imparerà a riconoscere dinamiche relazionali tra i dati in input e i dati in output e le astrarrà ad un livello sempre maggiore per apprendere cosa caratterizza un trimestre con bolla speculativa e cosa lo differenzia rispetto, ad esempio, ad un normale trimestre di crescita del mercato. Considerato che l'obiettivo iniziale del modello, questo aspetto del modello è molto importante ed è ciò che rende significativo l'utilizzo di una rete neurale. Quanto appena affermato è confermato dal fatto che, in tutte le architetture finali proposte, l'accuratezza nel riconoscere i trimestri con bolle speculative è minore rispetto

all'accuratezza generale del modello: per la rete neurale è più facile riconoscere gli elementi classificati automaticamente dalla cluster analysis sulla base di certe metriche piuttosto che riconoscere i trimestri analizzati maggiormente a livello qualitativo.
La scelta di considerare tutti i tipi di trimestri, e non solo quelli con potenziali bolle speculative, conferisce al modello la capacità di riuscire a fornire una prospettiva completa sul mercato azionario, ma i risultati su cui si pone la maggiore attenzione riguardano i trimestri con bolle.

Un limite del modello, legato strettamente alla natura delle reti neurali, riguarda la ricerca dell'architettura ottimale della rete neurale. Non ci sono regole o teorie che permettono di conoscere la corretta impostazione di tutti gli iperparametri, e perciò non si può avere la certezza di aver costruito la miglior architettura possibile. L'ottimizzazione iterativa degli iperparametri tramite grid search è sicuramente un approccio efficiente per trovare una buona configurazione, ma ci sono comunque altri iperparametri che non sono stati considerati (come, ad esempio, il numero di epoche e il batch size) e, per quelli di cui si è tenuto conto, si sono proposti solo alcuni valori. Un altro aspetto importante che potrebbe rappresentare un limite del modello è il fatto che, prima di iniziare le grid search sugli iperparametri, è stata scelta un'architettura di base formata da 3 layer LSTM e 2 layer Dense. Non è detto che un'altra scelta non avrebbe garantito dei risultati migliori. Si sarebbero potuti inserire più strati, o anche provare architetture differenti come le reti neurali convoluzionali. La scelta delle LSTM, tuttavita, come spiegato in fase di costruzione del modello, è dovuta alla natura sequenziale dei dati da analizzare.
Un ulteriore problema delle reti neurali che si ripercuote sul modello è il cosiddetto *black box problem*: la rete impara, con l'addestramento, a mappare una funzione tra input e output. Anche se ottiene notevoli risultati, non sappiamo cosa effettivamente la rete ha imparato. Non conosciamo tutte le configurazioni di pesi e bias, non capiamo quali features la rete ha ritenuto importanti e quali irrilevanti. Nel nostro caso, concentrandoci sui trimestri con bolle speculative, nonostante il modello ottenga risultati interessanti, non sappiamo quali pattern la rete ha imparato per identificare delle bolle speculative. Beneficiamo solo del risultato finale. Per questo motivo, è importante sottolineare che, anche con risultati notevoli, non si può avere un approccio serio a questo tema utilizzando come unico strumento la rete neurale, che sia questa o una versione migliore. È fondamentale utilizzare i risultati ottenuti come supporto ad un'analisi organica e completa del problema.

I principali miglioramenti del modello da applicare in una futura versione di esso possono essere divisi in tre categorie:

1. *Classificazione dei trimestri*: si è visto che la rete neurale lavora ad un problema di apprendimento supervisionato grazie ad una classificazione dei trimestri, in parte automatica e in parte manuale, svolta nella fase precedente. In questo elaborato è stata proposta la tecnica della cluster analysis unita ad una valutazione più dettagliata di un insieme di elementi esclusi dall'analisi. Il miglioramento di questa fase di classificazione dei trimestri consentirebbe alla rete neurale di fornire previsioni sempre più coerenti e in linea con quanto si osserva sui mercati. È possibile effettuare

una cluster analysis più dettagliata e approfondita, o anche basata su metriche diverse; oppure aumentare il numero di elementi analizzati manualmente anche più a livello qualitativo.

2. *Architettura della rete neruale*: nella fase di tuning, attraverso la *grid search* sono stati ottimizzati alcuni degli iperparametri della rete. Si è però spiegato che non si può avere la certezza di aver ottenuto la miglior architettura possibile per il problema da affrontare. Alcuni iperparametri, infatti, non sono stati considerati nella fase di tuning e, inoltre, tra quelli considerati, sono stati proposti solo alcuni valori tra cui trovare l'ottimo. Perciò allargare la fase di ottimizzazione degli iperparametri potrebbe consentire di ottenere una versione più performante della rete neurale con conseguenti migliori risultati.
   Sarebbe inoltre interessante costruire la rete neurale in modo da non dover passare, per ogni trimestre, anche i due precedenti, mantenendo comunque l'informazione su cosa è successo nei 6 mesi prima. Si potrebbe provare a costruire una struttura iterativa composta da tre reti neurali: nella prima verrebbero passati in input semplicemente i dati grezzi su prezzi e volumi di ciascun trimestre e si otterrebbe come output la previsione della rete su ciascun trimestre; la seconda rete neurale prenderebbe in input, per ciascun trimestre, i dati grezzi di prezzi e volumi e la previsione riguardo al trimestre precedente fornita dalla prima rete neurale; e la terza rete, allo stesso modo, avrebbe in input i dati su prezzi e volumi e previsioni sui due trimestri precedenti fornite dalle prime due reti. Ciò permetterebbe di mantenere un'informazione di lungo periodo senza dover inserire per ogni elemento 9 mesi di dati. È chiaramente solo una proposta che andrebbe poi analizzata nel dettaglio, ma l'idea alla base è sicuramente interessante.

3. *Database utilizzati*: come spiegato nel paragrafo sugli errori di modello, uno dei principali fattori che potrebbero causare errori della rete è la non perfetta omogeneità nella numerosità dei diversi gruppi. Le reti neurali soffrono il problema delle classi non bilanciate durante la fase di addestramento, quindi un'idea potrebbe essere utilizzare anche altri database per ottenere più trimestri nei gruppi che mancano. Ad esempio, se si vuole aumentare la numerosità dei gruppi contenenti bolle speculative si potrebbero prendere i dati del Nasdaq 100 durante la bolla delle dot-com per ottenere trimestri che sicuramente ben rappresentano quelle dinamiche.

Capitolo 7

Conclusione

Il corretto utilizzo di un modello simile non prevede la sua completa sostituzione a ogni qualsivoglia valutazione critica per prendere decisioni riguardo strategie di investimento di medio lungo periodo. L'idea è che uno strumento del genere possa offrire un supporto importante nel comprendere le dinamiche di mercato. Eventuali risultati ottenuti dal modello vanno criticamente analizzati e inseriti nel contesto attuale per svolgere un processo di decision-making più efficiente possibile.
In questo momento sui mercati è presente grande ottimismo: dopo il crollo causato dal Covid, i mercati stanno attraversando un'incredibile fase di rialzo che sembra inarrestabile. Alcuni investitori molto noti hanno iniziato a manifestare preoccupazione per la situazione, esprimendo dubbi sulla possibile presenza di una bolla speculativa. Due spunti interessanti che potrebbero confermare queste preoccupazioni sono offerti da due indici. Il primo è lo SKEW index, un indice che misura la volatilità di mercato attesa dagli investitori, senza specificare però la direzione: l'indice segue l'interesse degli investitori nei confronti delle opzioni out-of-the-money, ovvero le opzioni il cui prezzo di esercizio, se l'opzione scadesse subito, non garantirebbe guadagni. Se cresce l'interesse degli investitori nei confronti delle opzioni out-of-the-money significa che si aspettano importanti variazioni nei prezzi azionari nel breve e medio periodo. L'indice SKEW ha toccato i massimi storici verso fine giugno e tutt'ora si mantiene su livelli molto alti. Tale situazione sembra inoltre coerente con il risultato della rete neurale riguardo una forte instabilità sui mercati. Il secondo indice che conferma un importante ottimismo sui mercati è l'*equity put/call ratio*, un indice che rapporta l'interesse per le opzioni put su azioni e l'interesse per le opzioni call su azioni. L'indice è intorno al valore 0,5. Siccome un'opzione call equivale ad una posizione rialzista sul mercato (e viceversa un'opzione put rappresenta una posizione ribassista), l'indice mostra che al momento per ogni 100 persone ribassiste ce ne sono circa 200 rialziste.
Sembrano esserci dei campanelli d'allarme da tenere in considerazione, anche in merito ad analogie con vecchie bolle speculative. Michael Burry, il famoso investitore che aveva previsto il crollo del mercato nel 2008, ha sostenuto verso metà giugno che, secondo lui, siamo all'interno di un'enorme bolla speculativa ("*Greatest Speculative Bubble of All Time in All Things*"). Ray Dalio, investitore tra i più noti e gestore del fondo Bridgewater, a febbraio 2021 ha pubblicato un report in cui manifestava il rischio di presenza di bolla speculativa, soprattutto all'interno del settore delle start-up tecnologiche.
È utile considerare pareri autorevoli anche e soprattutto quando vanno contro il generale sentiment di mercato. Il cigno nero del Covid ha ricordato una volta di più che i mercati finanziari manterranno sempre una forte componente di imprevidibilità, ed è perciò importante utilizzare gli strumenti sempre più sofisticati che si hanno a disposizione per avere un approccio di lungo periodo equilibrato ed efficiente sui mercati finanziari.

Bibliografia

Akintoye, I. (2008). *Efficient Market Hypothesis and Behavioural Finance: A Review of Literature*. European Journal of Social Sciences – Volume 7, Number 2 (2008).

Asch, S. (1952). *Social Psychology*. Englewood Cliffs, N.J.: Prentice Hall, 1952), pp. 450-501.

Ashraf, N., Loewenstein, G., Camerer, C. (2005). *Adam Smith, Behavioral Economist*. ournal of Economic Perspectives 19(3):131-145.

Barberis, N., Shleifer, A., Vishny, R. (1998). *A Model of Investor Sentiment*. Journal of Financial Economics 49 (3): 307-343.

Bell, D. (1982). *Regret in Decision Making Under Uncertainty*. Operations Research, 30(5): 961-981.

Benartzi, S., Thaler, R. (1995). *Myopic loss aversion and the equity premium puzzle*. Quarterly Journal of Economics, 110, 73–92.

Black, F. (1986). *Noise*. The Journal of Finance, Vol. 41, No. 3, pp. 529-543.

Castellani, M. (2014). *Riflessività e dinamiche socio-cognitive nell'evoluzione dei mercati finanziari: alcune proposte interpretative*. Articolo in Sistemi Intelligenti, 2014.

Czaja, D., Röder, F. (2020). *Self-attribution bias and overconfidence among nonprofessional traders*. The Quarterly Review of Economics and Finance, Elsevier, vol. 78(C), pages 186-198.

De Bondt, W., Thaler, R. (1985). *Does the Stock Market Overreact?* The Journal of Finance, 40(3), 793-805.

De Bondt, W., Thaler, R. (1987). *Further Evidence on Investor Overreaction and Stock Market Seasonality*. The Journal of Finance, 42(3), 557-581.

Delcey, T. (2019). *Samuelson vs Fama on the Efficient Market Hypothesis: The Point of View of Ex- pertise*. Œconomia - History/Methodology/Philosophy, NecPlus/Association Œconomia, 2019, Varia, 9 (1), pp.37-58.

Deutsch, M., Gerard, H. (1955). *A study of normative and informational social influences upon individual judgment*. Journal of Abnormal and Social Psychology, 51 (1955), pp. 629-636.

Fama, E. (1965a). *The Behavior of Stock-Market Prices*. The Journal of Business, 38(1), 34-105.

Fama, E. (1965b). *Random Walks in Stock Market Prices*. Selected Papers of the Graduate School of Business, University of Chicago, reprinted in the Financial Analysts Journal (September - October 1965), The Analysts Journal, London (1966), The Institutional Investor (1968)

Fama, E. (1970). *Efficient Capital Markets: A Review of Theory and Empirical Work*. The Journal of Finance 25 (2): 383-417.

Fama, E., French, K. (1992). *The cross-section of expected stock returns*. Journal of Finance 47, 427– 465.

Finucane, M. (2000). *The Affect Heuristic in Judgments of Risks and Benefits*. January 2000. Journal of Behavioral Decision Making 13(1):1-17.

Fischhoff, B., Slovic, P., Lichtenstein, S. (1977). *Knowing with Certainty: The Appropriateness of Extreme Confidence*. Journal of Experimental Psychology: Human Perception and Performance 1977, Vol. 3, No. 4, 552-56.

French, E. (1992). *Early speculative bubbles and increases in the supply of money*. University of Nevada, Las Vegas. UNLV Retrospective Theses and Dissertations.

Grant, J. (1996). *The trouble whit prosperity: a contrarian tale of boom, bust and speculation*. New York: John Wiley and Sons, 1996.

Grossman, S., Stiglitz, J. (1980). *On the Impossibility of Informationally Efficient Markets*. The American Economic Review, 70(3), 393-408.

Hammond, R. *Behavioral finance: Its history and its future*. (2015). Selected Honors Theses. https://firescholars.seu.edu/honors/30.

Heukelom, F. (2007). *Kahneman and Tversky and the origin of behavioral economics*. Tinbergen institute discussion Paper No. 07-003/1. Available at SSRN: https://ssrn.com/abstract=956887.

Hirshleifer, D. (2015). *Behavioral finance*. Annual review of Financial Economics, 7, pp. 133-159.

Illiashenko, P. (2017). *Behavioral finance: history and foundations*.

Intropido, M. (2014). *La riflessività dei mercati finanziari*.

Jacobs, B., Levy, K. (1988). *Calendar Anomalies: Abnormal Returns at Calendar Turning Points*. Financial Analysts Journal, Vol. 44, No. 6, pp. 28-39, November/December 1988.

Jegadeesh, N., Titman, S. (1993). *Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency*. The Journal of Finance, 48(1), 65-91.

Kahneman, D. (2011). *Pensieri lenti e veloci*. 1° ed. Milano: Mondadori.

Kahneman, D., Tversky, A. (1974). *Judgement under Uncertainty: Heuristics and Biases*. Science, 185 (4157), pp. 1124-1131.

Kahneman, D., Tversky, A. (1979). *Prospect theory: an analysis of decision under risk*. Econometrica: Journal of the econometric society, 47(2), pp. 263-292.

Kahneman, D., Tversky, A. (1986). *Rational Choice and the Framing of Decisions*. The Journal of Business, 59(4), S251-S278.

Kahneman, D., Knetsch, J., Thaler, R. (1991). *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. Journal of Economic Perspectives*. Vol. 5, No. 1, Winter 1991, pp. 193-206.

Lakoff, G., Johnson, M. (1999). *Philosophy in the flesh: the embodied mind and its challenge to western thought*. Basic books, New York.

Lewin, S. (1996). *Economics and psychology: lessons for our own day from the early twentieth century*. Journal of economic literature, 34(3), pp. 1293-1323.

Lin, H., McNichols, F. (1998). *Underwriting relationships, analysts' earnings forecasts and investment recommendations*. Journal of Accounting and Economics, 1998, vol. 25, issue 1, 101-127.

Loewenstein G., Rick, S., Cohen, J. (2008). Neuroeconomics. Annual review of Psychology, 59, pp. 647-672.

Lustig I., Leinbach, P. (1983). *The small-firm effect*. Financial Analysts Journal, May 1983 Volume 39 Issue 3.

Malkiel, B. (2003). *The Efficient Market Hypothesis and Its Critics*. The Journal of Economic Perspectives, Winter, 2003, Vol. 17, No. 1 (Winter, 2003), pp. 59-82

Mandelbrot, Benoit. (1965). *Forecasts of Future Prices, Unbiased Markets, and "Martingale" Models*. The Journal of Business 39 (1): 242-242.

Milgram, S. (1974). *Obedience to Authority*. New York: Harper and Row, 1974, pp. 13-54.

Modigliani, F., Cohn,R. (1979). *Inflation, Rational Valuation, and the Market*. Financial Analysts' Journal, 35 (1979), pp. 22-44.

Morewedge, K., Kahneman, D. (2010). *Associative processes in intuitive judgment*. Trends in cognitive sciences, 14(10), pp. 435-440.

Pitz, G. (1975). *Subjective probability distributions for imperfectly known quantities*. Knowledge and Cognition, pp 29-41.

Risen, J., Gilovich, T. (2007). *Informal Logical Fallacies*. In R. J. Sternberg, H. L. Roediger III, & D. F. Halpern (Eds.), Critical thinking in psychology (pp. 110–130). Cambridge University Press.

Samuleson, P. (1973*). Proof That Properly Discounted Present Values of Assets Vibrate Randomly*. Bell Journal of Economics 4 (2): 369-74.

Sharpe, W. (1966). *Mutual fund Performance*. Journal of Business, 39 (Special Supplement, January 1966), 119-38.

Sharpe, W. (1965). *Risk Aversion in the Stock Market*. Journal of Finance, 20 (September 1965), 416-22.

Shefrin, H., Statman, M. (2000). *Behavioral portfolio theory*. Journal of financial and quantitative analysis, 35(02), pp. 127-151.

Shiller, R. (2000). Irrational Exuberance. Princeton University Press, 2000.

Shiller, R. (2003). *From Efficient Markets Theory to Behavioral Finance*. The Journal of Economic Perspectives, Winter, 2003, Vol. 17, No. 1 (Winter, 2003), pp. 83-104.

Shiller, R., Pound, J. (1986). *Survey Evidence on Diffusion of Investment Among Institutional Investors*. National Bureau of Economic Research, paper 1851.

Soros, G. (1987). *L'alchimia della finanza. La logica, le tendenze e i segreti del mercato*. Ponte alle Grazie collana Saggi.

Taleb, N. (2007). *Il cigno nero*. Il Saggiatore S.r.l., Milano.

Thaler, R. (1987). *The January Effect*. Journal of Economic Perspectives 1(1):197-201.

Tobin, J. (1984). *On the efficiency of the financial system.*

Treynor, J. (1965). *How to Rate Management of Investment Funds*. Harvard Business Review, 43 (January-February 1965), 63-75.

Tversky, A., Shafir, E. (1992). *Choice under Conflict: The Dynamics of Deferred Decision*. Psychological Science, Vol. 3, issue 6, pp. 358-361.

Woo, K., Mai, C., McAleer, M., Wong, W. (2020). *Review on Efficiency and Anomalies in Stock Markets*. Economies, MDPI, Open Access Journal, vol. 8(1), pages 1-51, March.

Sitografia

https://keras.io/api/

https://numpy.org/doc/stable/reference/routines

https://pandas.pydata.org/docs/reference/index

https://blog.profession.ai

https://www.treccani.it/enciclopedia/intelligenza-artificiale

https://www.igorvitale.org/la-prospect-theory-di-tversky-e-kahneman/

https://colah.github.io/posts/2015-08-Understanding-LSTMs/

https://www.algofj.com/teoria-riflessivita-da-popper-a-soros

https://machinelearningmastery.com