

Rischio e valore nelle banche

A horizontal brushstroke in a vibrant yellow color, with a textured, painterly appearance, spanning across the width of the slide.

I modelli di *scoring*

Capitolo 10

Agenda



- Il rischio di credito: definizione, componenti e peculiarità
- Le tipologie di rischio di credito
- I modelli di scoring quantitativo: generalità
- L'analisi discriminante lineare
- I modelli di regressione
- I modelli di natura induttiva: reti neurali e algoritmi genetici
- Limiti e problemi dei modelli di scoring

Rischio di credito

Definizione

Deterioramento inatteso del merito di credito di una controparte nei confronti della quale vi è un'esposizione creditizia



- No approccio binomiale affidabile-non affidabile
- Il default è solo l'evento estremo
- Il rischio di credito è legato a diverse forme di esposizione (prestiti, bonds, derivati OTC)

Tre componenti rilevanti

- 1 Perdita attesa
- 2 Perdita inattesa (variabilità della perdita attesa)
- 3 Effetto diversificazione



Correlazione fra la perdita inattesa degli impieghi in portafoglio



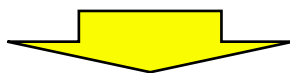
L'analisi tradizionale del rischio di credito si ferma alla prima componente



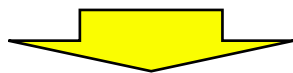
Perdita attesa → non diversificabile
Perdita inattesa → diversificabile → tanto minore
quanto minore è il grado di correlazione fra i
singoli impieghi

Le componenti del rischio di credito

Le variabili chiave



- Exposure at default (EAD)
- Probability of default (PD), o Expected Default Frequency (EDF)
- Recovery Rate (RR) = 1-Loss Given Default (LGD)



$$\text{Expected Loss (EL)} = \text{EAD} \times \text{PD} \times \text{LGD}$$

La stima della perdita attesa

Probabilità di insolvenza (PD o EDF)

1. Modelli statistici/automatici
2. Analisi di fido tradizionale
3. Tassi di insolvenza dai bonds
4. Approccio alla Merton (KMV)

Tasso di recupero (1-LGD)

1. Dati di bonds
2. Dati interni
3. Stime per classi di prodotto/garanzie

L'Adjusted Exposure

- La sua aleatorietà dipende dalla forma tecnica
 - ✓ nulla in caso di obbligazioni
 - ✓ elevata in caso di apertura di credito in c/c
- È stimata in base alla quota di fido utilizzata (DP – *drawn portion*); quella non utilizzata (UP – *undrawn portion*); lo stimato utilizzo in caso di default (UGD – *usage given default*)

$$AE = DP + UP \cdot UGD$$

La perdita inattesa



- È data dalla variabilità della perdita attorno alla media
- È rilevante per il concetto di diversificazione (effetto di portafoglio: variabilità della perdita di un portafoglio è tanto minore quanto più bassa è la correlazione tra posizioni)
- La perdita inattesa è coperta dal patrimonio (quella attesa da accantonamenti)

Le peculiarità del rischio di credito

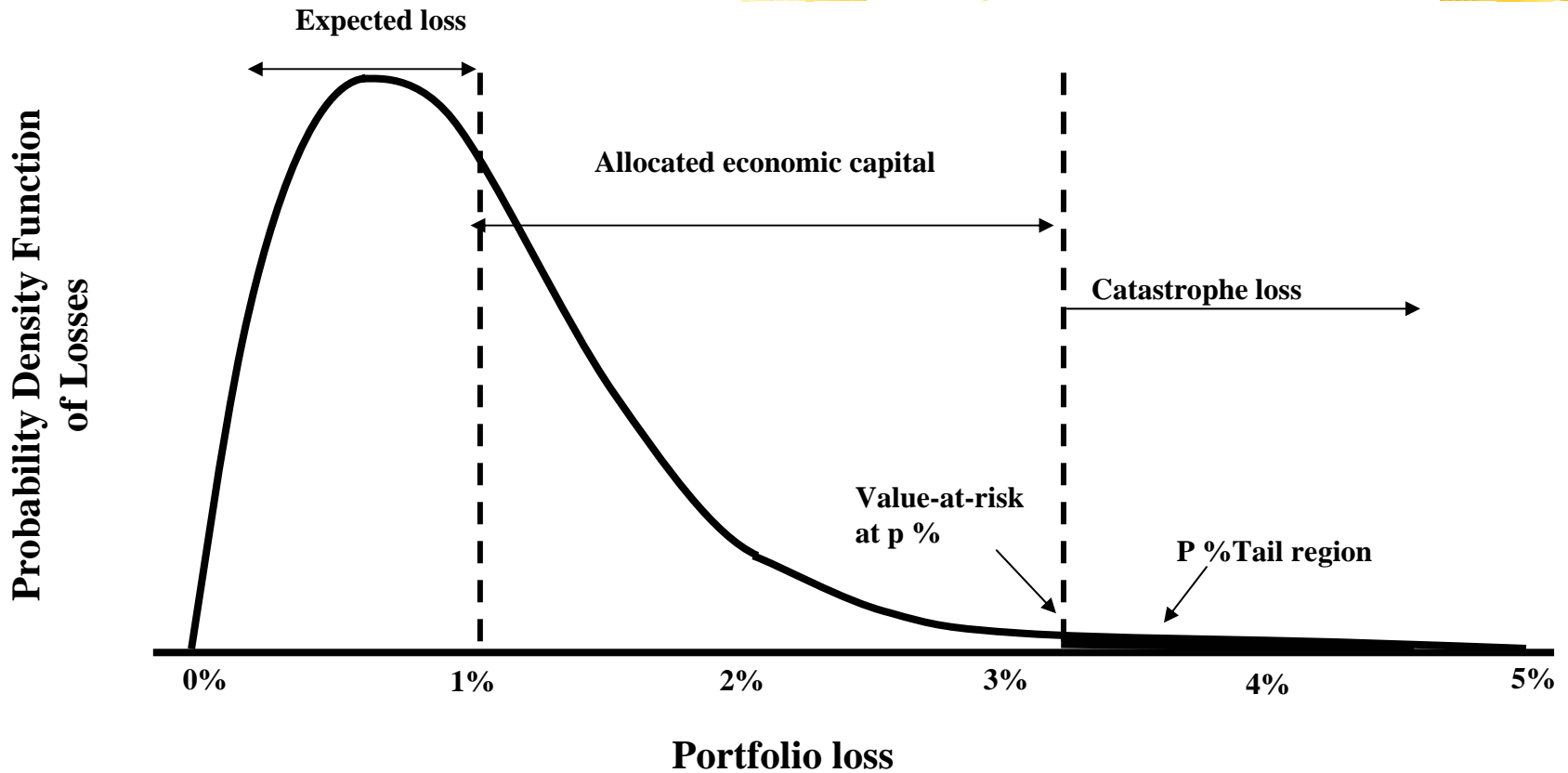


Il rischio di credito è più complesso da misurare rispetto al rischio di mercato

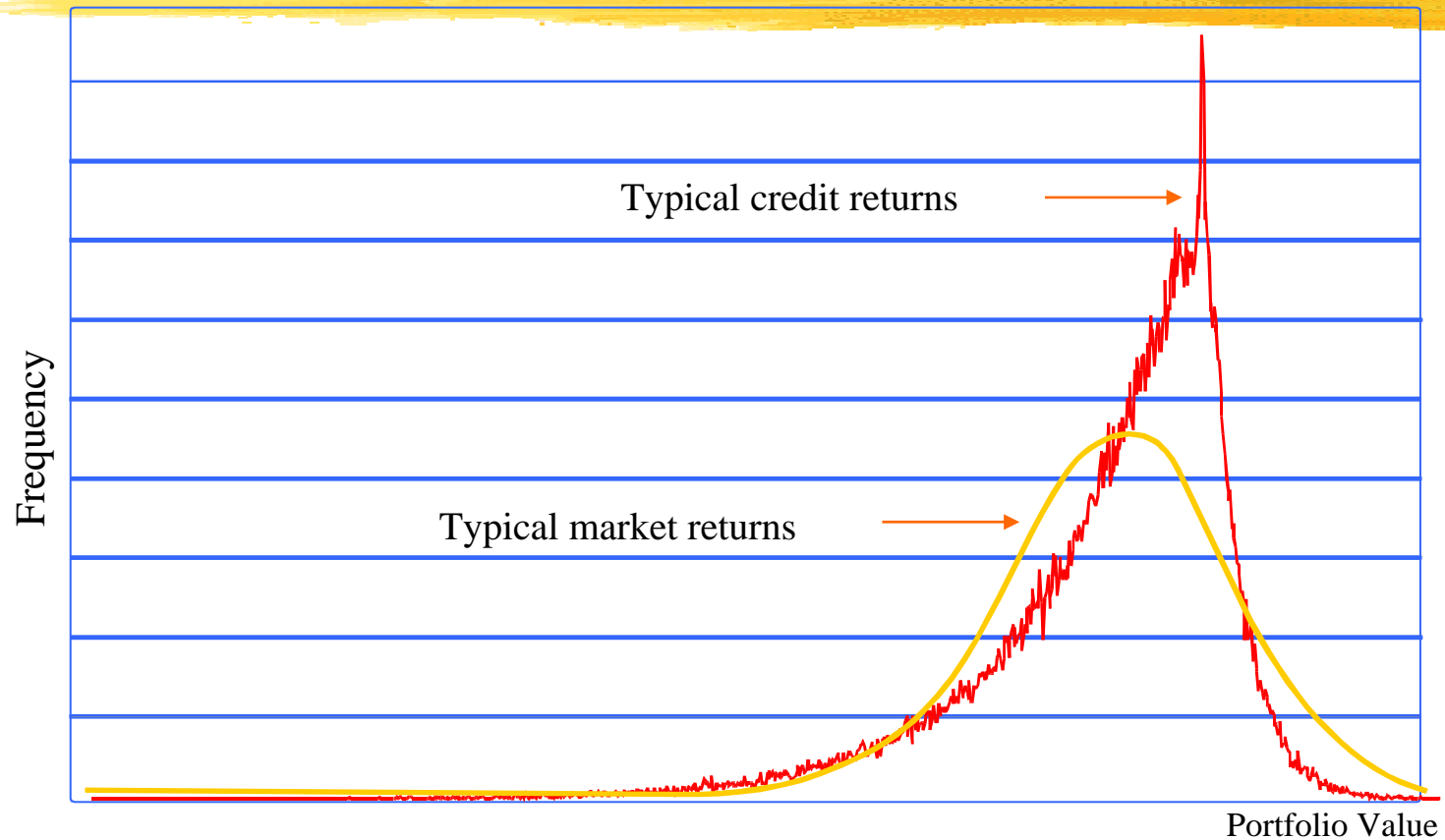


- Minore disponibilità di dati (default rates, recovery rates, correlations, etc.)
- Distribuzione non normale: asimmetrica e con coda spessa
- La media della distribuzione non è nulla

Le peculiarità del rischio di credito



Le peculiarità del rischio di credito



Confronto fra distribuzione dei rendimenti creditizi e rendimenti di mercato

Il rischio di credito: le tipologie



- Rischio insolvenza
- Rischio migrazione o di downgrading
- Rischio di recupero
- Rischio esposizione \Rightarrow discrezionalità fidi, derivati OTC
- Rischio spread \Rightarrow es. “flight to quality”
- Rischio concentrazione
- Rischio di pre-regolamento o rischio di sostituzione
- Rischio paese

La stima di PD



- ✓ Modelli contabili-analitici (judgmental)
- ✓ Modelli scoring (discriminant analysis, modelli logit/probit, reti neurali, algoritmi genetici)
- ✓ Frequenze di insolvenza da agenzie di rating
- ✓ Frequenze di insolvenza da rating interno
- ✓ Approccio a la Merton (KMV)

I metodi contabili-analitici



Stima della probabilità di insolvenza sulla base di
dati contabili/finanziari/settoriali/macro



Fattori rilevanti

Situazione economico-finanziaria

(leva finanziaria, redditività, cash-flow, liquidità)

Settore produttivo, Area geografica, Classe
dimensionale, Qualità del management

I modelli di scoring



- Sono modelli multivariati
 - ✓ input: indici economico-finanziari dell'affidato
 - ✓ output: score che rappresenta la probabilità di insolvenza
- Possono essere impiegati per
 - ✓ prevedere default
 - ✓ classificare gli impieghi in categorie di rischio

Modelli di scoring: categorie



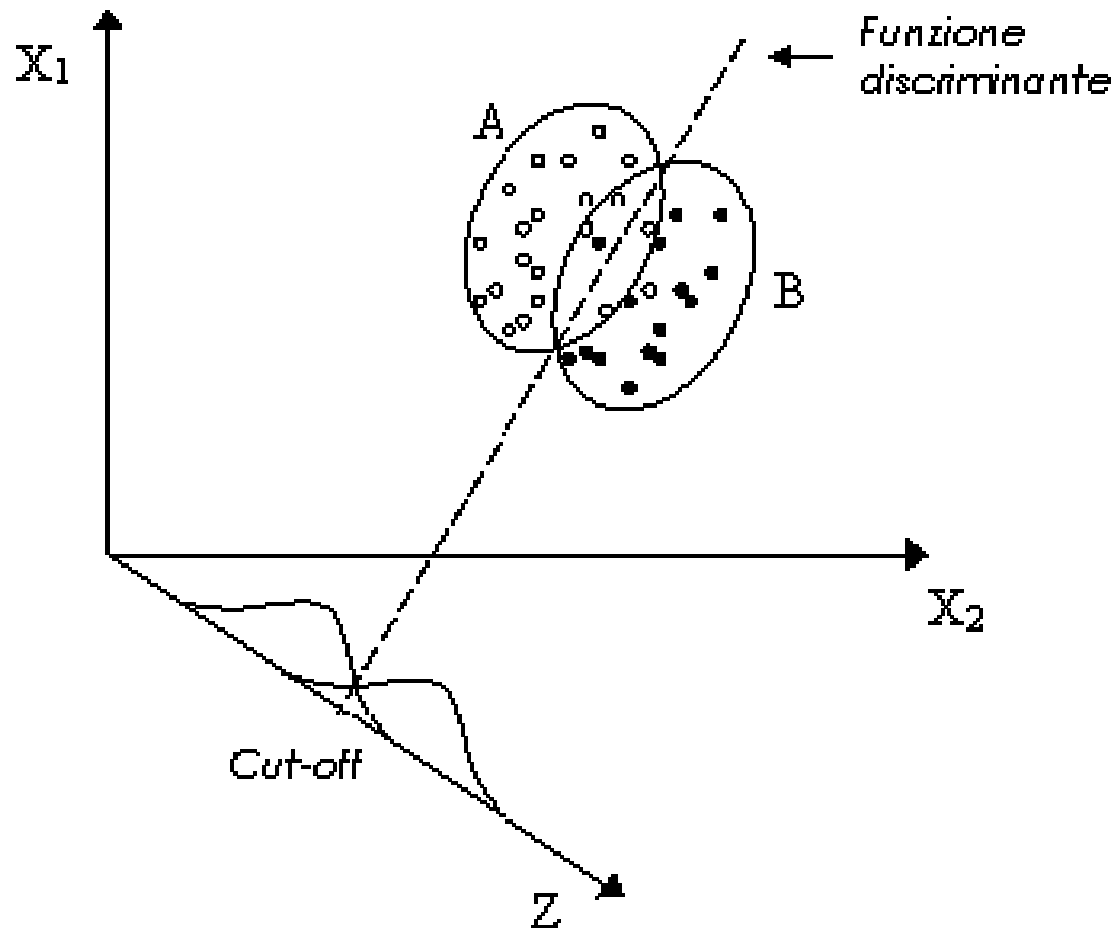
- Analisi discriminante lineare (Altman)
- Regressione lineare (linear probabilistic model)
- Logit e probit
- Modelli di natura euristica (reti neurali e algoritmi genetici)

Analisi discriminante lineare



- Obiettivo: identificare le variabili che consentono di discriminare tra imprese sane e imprese anomale (fallite; oggetto di ristrutturazione finanziaria; con debito in sofferenza)
- Procedura: utilizzo di informazioni tratte da un campione di imprese (alcune sane e alcune anomale) per derivare una funzione discriminante che separa le sane dalle anomale

Imprese sane e imprese anomale



La funzione discriminante

- Combinazione lineare di variabili indipendenti che consentono di discriminare fra sane e insolventi
 - ✓ massimizza la varianza **fra** i due gruppi e minimizza la varianza **in** ciascun gruppo; massimizza il rapporto tra varianza intergruppi e varianza infragruppo
- I coefficienti di discriminazione (ovvero quelli delle singole variabili) sono tali da produrre score che:
 - ✓ massimizzino distanza tra *score* dei 2 gruppi
 - ✓ minimizzino la varianza degli *score* all'interno di ciascun gruppo

La funzione discriminante (segue)

$$Z_i = \sum_{j=1}^n \alpha_j X_{i,j} \quad [1]$$

dove

Z_i = valore discriminante dell'impresa i

α_j = coefficiente di discriminazione della variabile j -esima (X_j)

A seconda che il risultato Z_i si collochi sopra o sotto una data soglia (risultante dall'analisi empirica su un campione di imprese) l'impresa è classificata potenzialmente sana (affidabile) oppure potenzialmente anomala (non affidabile).

Il calcolo dei coefficienti di discriminazione

- Il valore α dei coefficienti di discriminazione è dato da

$$\alpha = (\bar{X}_1 - \bar{X}_2)' S^{-1} \quad [2]$$

dove \bar{X}_1 (\bar{X}_2) è il vettore delle medie delle imprese del gruppo 1, sane (gruppo 2, insolventi)

Il calcolo dei coefficienti di discriminazione (segue)

1. Si calcolano le medie delle variabili X_i per il gruppo 1 e per il gruppo 2
2. Si calcola la matrice varianze-covarianze tra le variabili X_i
3. In base alla formula [2] si calcolano i coefficienti α
4. Si calcola il *cut-off point* come semisomma della media degli score delle imprese sane e di quella delle imprese insolventi

Lo Z-score di Altman

- Il modello di Altman:

$$Z_i = 1,2X_{i,1} + 1,4X_{i,2} + 3,3X_{i,3} + 0,6X_{i,4} + 1,0X_{i,5}$$

dove

X_1 = capitale circolante / totale attivo

X_2 = utili non distribuiti / totale attivo

X_3 = EBIT / totale attivo

X_4 = valore di mercato patrimonio / valore contabile debiti I.t.

X_5 = fatturato / totale attivo

Lo Z-score di Altman (segue)

- Il modello di Altman:
 - ✓ tanto maggiore è il valore Z tanto minore è la probabilità di insolvenza
 - ✓ valore soglia 1,81 ($Z > 1,81$: impresa sana; $Z < 1,81$: elevato rischio di default)
 - ✓ due valori soglia (intervallo di confidenza):
 - $Z >$ soglia sup.: impresa sana
 - $Z <$ soglia inf.: impresa insolvente
 - soglia inf. $< Z <$ soglia sup.: *grey area*

Analisi discriminante: attenzione alle ipotesi

- Le variabili discriminanti sono caratterizzate da distribuzione normale multivariata
 - ✓ ipotesi violata per imprese in default
 - problemi soprattutto per efficacia previsionale
 - ✓ multicollinearità tra variabili discriminanti
- Le matrici varianze-covarianze per le variabili discriminanti sono uguali per i due gruppi di imprese

La selezione delle variabili discriminanti

- Metodo simultaneo (diretto)
 - ✓ in base ad "a priori" teorici (inclusione delle variabili che ci si attende siano rilevanti)
- Metodo *stepwise*
 - ✓ si parte da un elevato numero di variabili e
 - si eliminano via via quelle con minor potere esplicativo (*backward elimination*) oppure
 - si inseriscono progressivamente quelle con maggior potere esplicativo (*forward selection*) oppure
 - si inserisce una variabile alla volta che poi viene eliminata se perde potere discriminante a seguito dell'inserimento di altre (*stepwise selection*)

Il problema degli errori

- Del primo tipo (costo pari a C_I)
 - ✓ impresa insolvente classificata sana
 - perdita pari a interessi e capitale perduti per default
- Del secondo tipo (costo pari a C_{II})
 - ✓ impresa sana classificata insolvente
 - perdita pari a mancato guadagno per commissioni e margine di interesse
- Il *cut-off point* può tenere conto del diverso livello dei costi

$$P_{I^\circ} \cdot C_{I^\circ} = P_{II^\circ} \cdot C_{II^\circ} \quad \text{Se } C_{I^\circ} > C_{II^\circ} \text{ allora } P_{I^\circ} < P_{II^\circ}$$

 Cut-off point più elevato

Modelli di regressione: probabilità lineare

- Le variabili che determinano le probabilità di insolvenza e il loro peso sono identificate in base a regressione lineare
- Fasi
 - ✓ selezione del campione: imprese sane e imprese insolventi (variabile dipendente, Z)
 - attenzione ai problemi di dimensione del campione
 - ✓ selezione delle variabili indipendenti
 - indici economico-finanziari
 - ✓ stima dei coefficienti del modello
 - ✓ stima della probabilità di insolvenza

$$Z_i = \sum_{j=1}^n \beta_j X_{i,j} + \varepsilon_i$$

Probabilità lineare: problemi



- La probabilità determinata in base al modello potrebbe essere fuori dall'intervallo 0-1
- La varianza dei residui della regressione dipende dalla variabile dipendente: eteroschedasticità
- Soluzione: modelli logit e probit

Modelli logit e probit

- Logit: regressione logistica
 - ✓ mediante trasformazione esponenziale (funzione logistica cumulata) il risultato della regressione è compreso tra 0 e 1
 - ✓ si inputa il risultato Z derivante dalla regressione lineare nella formula

$$F(Z_i) = \frac{1}{1 + e^{-Z_i}} = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

- la trasformazione assume valori compresi tra 0 e 0,5 se $Z_i < 0$

Modelli logit e probit



- Probit: utilizza distribuzione normale standardizzata anziché logistica
 - ✓ mediante la trasformazione il risultato è tra 0 e 1
 - la distribuzione logistica è caratterizzata da code più spesse rispetto alla distribuzione normale
 - nella prassi i modelli possono essere utilizzati alternativamente senza differenze di rilievo

Modelli di natura induttiva



- I modelli di scoring strutturali tendono a identificare le relazioni fondamentali che spiegano l'equilibrio economico-finanziario di un'impresa
 - ✓ le variabili rilevanti sono accettate se
 - statisticamente significative
 - economicamente fondate (segno atteso)
- I modelli di natura induttiva applicano regole empiriche, senza indagare "cosa sta dietro" economicamente alle relazioni tra fenomeni e probabilità di insolvenza

Modelli strutturali: rischi di manipolazione

- Generazione del modello strutturale da parte della banca:
 1. raccolta input
 2. connessione input tramite funzione del modello
 3. produzione output
- Apprendimento e manipolazione da parte delle imprese
 4. tentativo delle imprese analizzate di ricostruire il modello
 5. adozione da parte delle imprese di politiche di bilancio per manipolare i risultati del modello

Esempio di metodo induttivo

1. azienda $\phi \in \bar{X}$ con $R_1^\phi < k$, è fallita
2. azienda $\vartheta \in \bar{X}$ con $R_1^\vartheta > k$, è sana
3. azienda $\zeta \in \bar{X}$ con $R_1^\zeta < k$, è fallita
4. azienda $\psi \in \bar{X}$ con $R_1^\psi < k$, è fallita
5. azienda $\theta \in \bar{X}$ con $R_1^\theta > k$, è sana

...

n - 1. azienda $\varphi \in \bar{X}$ con $R_1^\varphi < k$, è fallita

quindi è probabile che

n. se $R_1^\alpha > k$, azienda $\alpha \in \bar{X}$ è sana

se $R_1^\alpha < k$, azienda $\alpha \in \bar{X}$ è destinata a fallire

Reti neurali



- Metodo induttivo; black-box
- Genera risultati in base ad una serie di relazioni non lineari con le variabili di input
- Riproduce il meccanismo di apprendimento tipico della mente umana
- Si compone di un numero elevato di elementi ("neuroni") collegati tra di loro tramite "sinapsi"

Reti neurali

- Ogni neurone assume come input un vettore n-dimensionale $x = [1, x_1, x_2, \dots, x_n]^T$ e applica a tale vettore di input un vettore di ponderazione $w = [w_0, w_1, w_2, \dots, w_n]^T$ per ottenere la somma ponderata $w^T x$
- Le ponderazioni della rete sono apprese dalla medesima mediante meccanismi iterativi che modificano i pesi con l'obiettivo di associare correttamente input (variabili economico-finanziarie e di altro tipo) e output (*default* oppure no)

Algoritmi genetici



- Alternativa alle reti neurali
- Applicazione del principio "*survival of the fittest*":
 - ✓ il processo evolutivo porta ad un miglioramento della specie perché gli individui più adatti ad un dato ambiente hanno le maggiori probabilità di trasferire alle generazioni future il loro patrimonio genetico
 - ✓ la ricombinazione (*cross-over*) fra individui con elevata fitness all'ambiente porta a una progenie con caratteristiche migliori (*superfit*) di quelle di ognuno degli individui che la origina
- Gli algoritmi genetici riproducono tale processo evolutivo

Algoritmi genetici

- Definito un problema da risolvere (es. prevedere l'insolvenza in base a combinazione di indici di bilancio), l'algoritmo:
 1. genera casualmente una popolazione di n individui, ciascuno dei quali rappresenta una possibile soluzione, data dagli n indicatori I e dagli $n+1$ coefficienti α della funzione

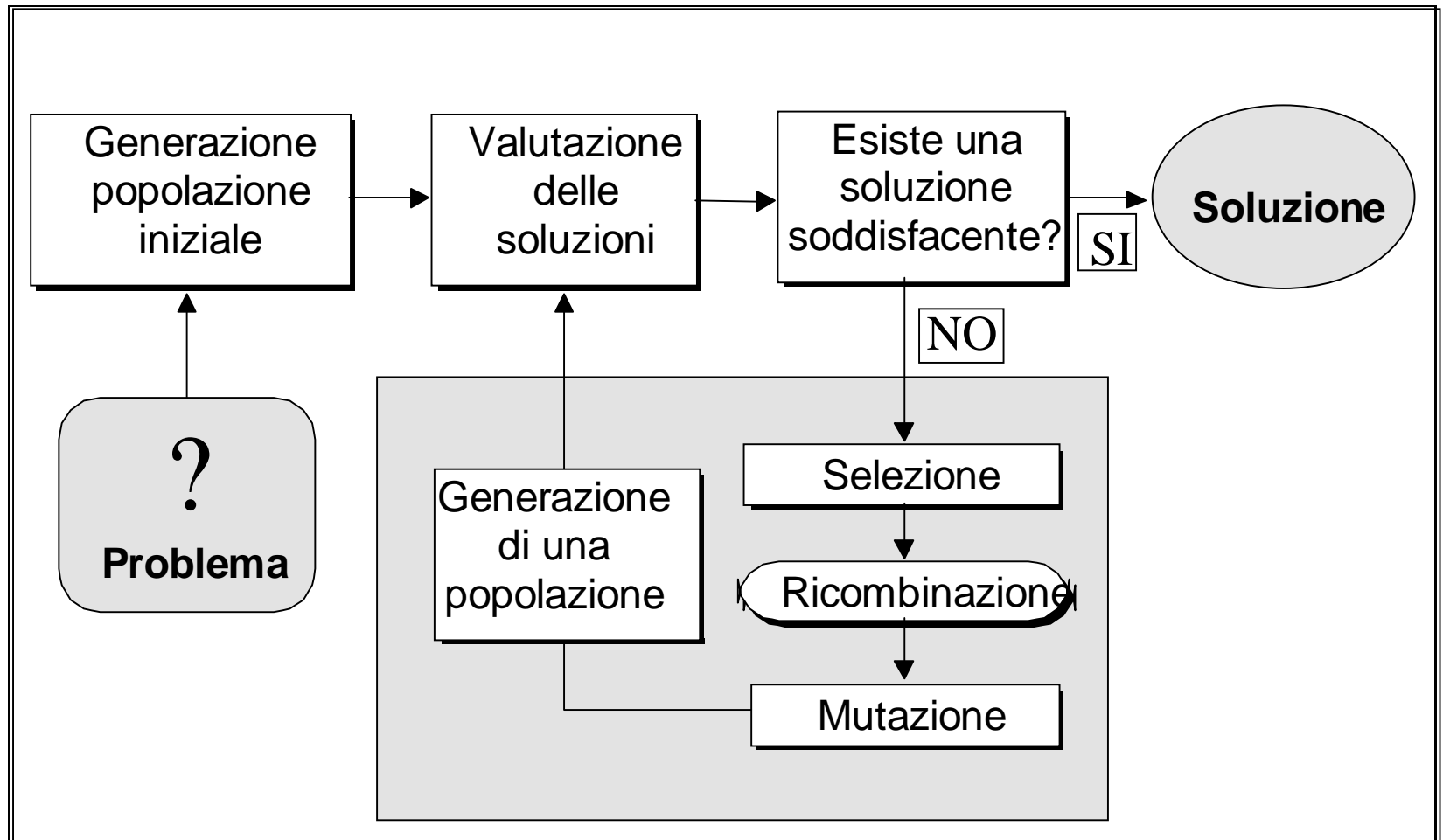
$$Z = \alpha_0 + \alpha_1 I_1 + \alpha_2 I_2 + \dots + \alpha_n I_n$$

2. in base ad una funzione di valutazione (f) calcola la bontà (*fitness*) degli n individui (es. capacità di ogni soluzione di classificare correttamente le imprese sane e insolventi in base a un campione storico)

Algoritmi genetici (segue)

3. si applica l'operatore genetico di selezione per individuare i soggetti destinati a sopravvivere (alle soluzioni migliori è assegnata più alta probabilità di sopravvivenza)
4. si applica l'operatore ricombinazione per riprodurre gli individui sopravvissuti
5. si applica l'operatore mutazione che consente di registrare un'improvvisa modifica di una o più soluzioni
6. le soluzioni figlie della i -ma generazione costituiscono una nuova popolazione
7. il processo continua sinché il miglioramento da una generazione all'altra è nullo (il modello classifica correttamente)

Struttura di un algoritmo genetico



Algoritmi genetici



- Il processo di miglioramento del fitness si arresta quando le soluzioni raggiungono l'ottimo e le loro caratteristiche sono trasmesse alle generazioni future
 - ✓ a questo punto gli individui di una popolazione assumono caratteristiche sempre più simili
- L'evoluzione non è casuale: valutazione, selezione e ricombinazione realizzano una ricerca adattiva attraverso la quale le soluzioni hanno "memoria"
- L'esplorazione avviene attraverso lo sfruttamento delle informazioni già acquisite mediante la competizione tra soluzioni

Algoritmi genetici



- Non richiedono di conoscere lo spazio delle soluzioni possibili (spazio delle soluzioni ampio e rumoroso - *noisy*)
- Richiedono una funzione di valutazione della fitness delle diverse soluzioni
- Non garantiscono soluzione ottimale, ma soluzione buona e rapida

Limiti e problemi dei modelli di *scoring*

- La definizione di *default* per la classificazione di imprese sane e imprese insolventi influenza il risultato
- L'importanza relativa (peso) delle variabili indipendenti non è stabile nel tempo
- Sono trascurati fattori qualitativi importanti
- Quando manca l'omogeneità settoriale delle imprese del campione di stima
 - ✓ indici economico-finanziari hanno medie molto diverse tra settori
 - ✓ stesso indice ha diversa importanza in diversi settori
- E' necessario un ampio numero di osservazioni, ma l'insolvenza è evento raro e quindi campioni sbilanciati