

Modelli matematici e Data Mining

Introduzione

- I **modelli matematici** giocano un ruolo critico negli ambienti di business intelligence e **sistemi di supporto alle decisioni**.
- Essi rappresentano **un'astrazione selettiva** di un sistema reale e vengono utilizzati per capire in astratto come funziona un sistema reale.
- Nell'elaborazione di un modello *bisognerebbe rendere tutto il più semplice possibile, ma non troppo semplice.*

Struttura dei modelli

- **Iconici**: modelli che costituiscono una rappresentazione fisica della realtà (es. un plastico)
- **Analogici**: rappresenta fisicamente la realtà (es. la galleria del vento)
- **Simbolici**: costituiscono una rappresentazione astratta di un sistema concreto (es. modello matematico).

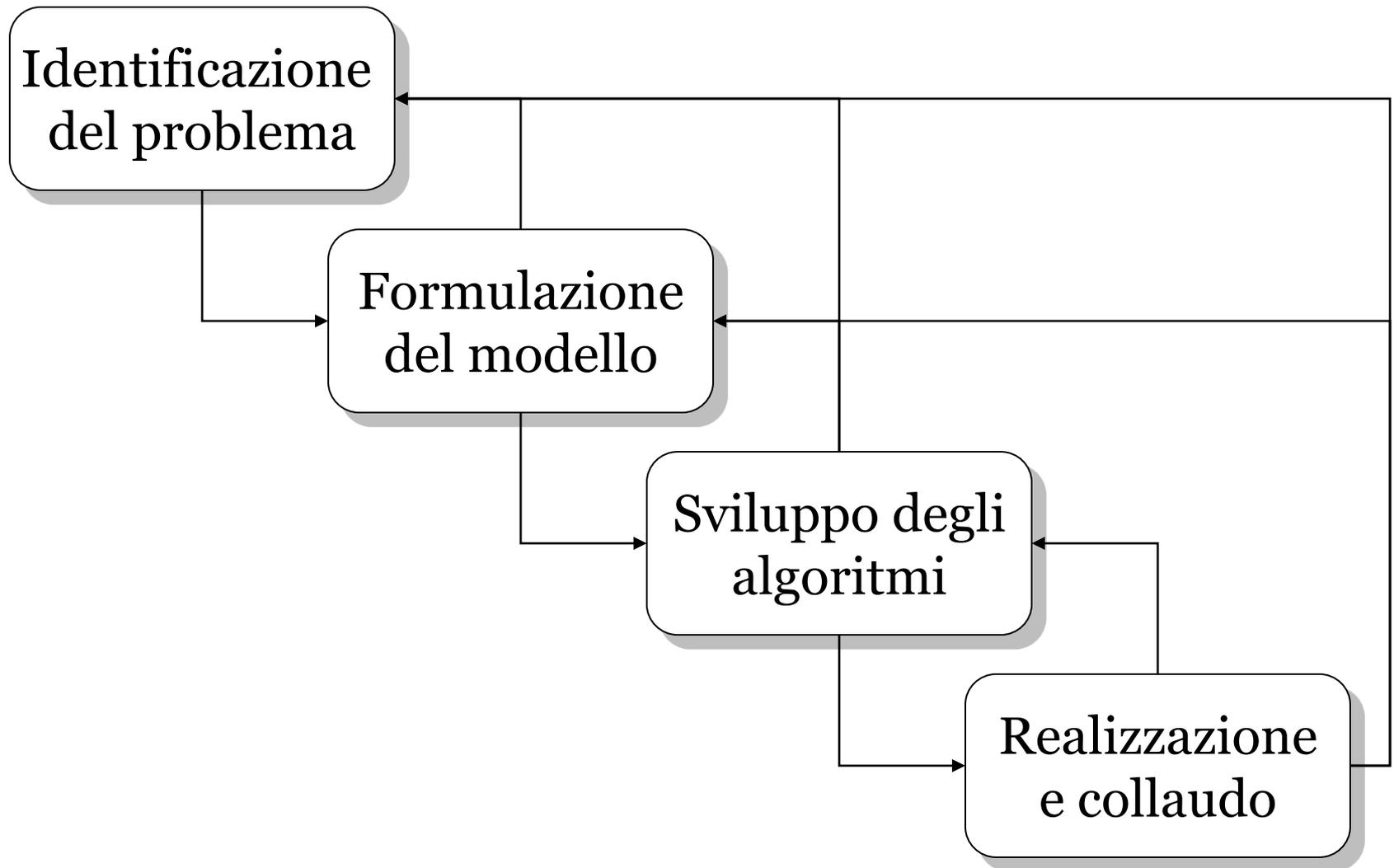
Modelli stocastici e deterministici

- Nei **modelli stocastici** alcune informazioni fornite in ingresso rappresentano eventi aleatori e sono caratterizzate da una distribuzione di probabilità assegnata oppure ignota
 - **Es. modelli della teoria delle code**
- Un modello si dice **deterministico** se tutti i dati in ingresso si suppongono noti a priori con certezza.

Modelli statici e dinamici

- I **modelli statici** considerano un sistema ed il relativo processo decisionale in un'**unica fase temporale**.
 - Es. alcuni modelli di ottimizzazione
- I **modelli dinamici** considerano un sistema attraverso **varie fasi temporali**, di solito di ampiezza predefinita.
 - Es. serie temporali

Fasi di sviluppo



Identificazione del problema

- La prima fase consiste nell'identificare correttamente il problema da affrontare.
- E' necessario analizzare i sintomi e formulare le prime ipotesi.
 - Es. Il livello di giacenza a fronte di indici di rotazione elevati, è sintomo di un'inefficace pianificazione?

Formulazione del modello

- Orizzonte temporale
 - Es. Qual è il periodo di tempo considerato?
- Valutazione delle decisioni
 - Indicatori di prestazione: costi e ricavi, livello di efficacia, qualità dei prodotti, flessibilità, affidabilità,...
- Variabili di decisione
 - Es. In un piano di produzione vanno prese in considerazione i volumi produttivi dei prodotti/ processi/ periodi.

Formulazione del modello

- Parametri numerici
 - Es. capacità produttiva, capacità di assorbimento,...
- Relazioni matematiche
 - Legami deterministici o probabilistici.

Sviluppo degli algoritmi

- La **risoluzione** di un modello richiede l'identificazione l'**algoritmo** risolutivo e degli strumenti **software** che lo incorporano.
- Non è possibile prescindere dalla conoscenza dei metodi risolutivi esistenti e delle loro caratteristiche.

Realizzazione e collaudo

- E' necessario verificare:
 - Plausibilità e verosimiglianza delle conclusioni raggiunte;
 - Consistenza dei risultati per valori estremi;
 - Stabilità dei risultati.

Classi di modelli

- Modelli predittivi
 - **Esplicativi**: identificano la forma funzionale di una relazione.
 - Es. Modelli di regressione, classificazione,
 - **Serie storiche**: identificano eventuali regolarità.
- Modelli di apprendimento
 - **Modelli di interpretazione**: identificano regolarità e le esprimono tramite regole criteri.
 - **Apprendimento** supervisionato e non supervisionato
- Modelli di ottimizzazione
 - **Pianificazione** logistica, produttiva, finanziaria

Classi di modelli

- Modelli per la gestione dei progetti
 - PERT: project evaluation and review techniques.
- Modelli di analisi del rischio
 - Teoria bayesiana delle decisioni.
- Modelli di teoria delle code

Data Mining

- Il complesso delle attività volte all'analisi di basi di dati di grandi dimensioni per ricavarne conoscenze utilizzabili nel corso dei processi decisionali.
- Il data mining indica il processo di esplorazione ed analisi di un insieme di dati per individuarne regolarità, estrarne conoscenza e ricavare regole ricorrenti.

Interpretazione e predizione

- L'**interpretazione** è volta ad identificare schemi di regolarità presenti nei dati e a esprimerli attraverso regole e criteri che risultino comprensibili agli esperti.
- La **predizione** è volta a prevedere il valore che una variabile casuale assumerà in futuro, oppure a stimare la probabilità di un certo evento.
- Le due attività **non sono mutuamente esclusive**: un modello sviluppato per la predizione può risultare efficace anche per l'interpretazione.

Applicazioni di data mining

- Marketing relazionale
 - Identificazione di segmenti di clienti,
 - Stima del tasso di risposta,
 - Comprensione dei comportamenti d'acquisto,
 - Identificazione delle combinazioni di acquisto.
- Identificazione di frodi
 - Telefoniche assicurative, bancaria,...
- Valutazione del rischio
 - Prestiti, mutui,...

Applicazioni di data mining

- Text mining
 - Spam, classificazione di notizie d'agenzia, ...
- Web mining
 - Caching, personalizzazione, pagine dinamiche, ...
- Diagnostica medica
 - Predizione delle prognosi
 - E-healthcare
- Pubblica amministrazione
 - Predizione dei flussi del traffico
 - Analisi dei fattori di rischio nei fenomeni criminali
 - Analisi dei gruppi nelle aree urbane

Rappresentazione dei dati di input

- **Categorici**
 - Conteggi (associazione di una carta ad un conto)
 - Nominali (città di residenza)
 - Ordinali (livello di scolarità)

- **Numerici**
 - Discreti
 - Continui

Processo di data mining

Predizione
Interpretazione

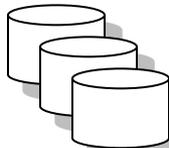
Sviluppo modelli

Selezione degli attributi

Analisi esplorativa

Data mart

Raccolta dati e
integrazione



Definizione

Metodologie di analisi

- Apprendimento supervisionato
 - Caratterizzare e discriminare
 - Classificare
 - Stimare
 - Serie storiche

- Apprendimento non supervisionato
 - Correlare
 - Raggruppare
 - Trovare casi anomali

Sommario

- Abbiamo visto:
 - Struttura dei modelli matematici
 - Fasi di sviluppo dei modelli
 - Classi principali di modelli
 - Definizione, modelli, applicazioni di data mining
 - Rappresentazione dei dati
 - Metodologie di analisi

Nella prossima lezione

- Preparazione dei dati
 - Validazione
 - Trasformazione
 - Riduzione