

Strumenti Statistici per l'Analisi di Dati Aziendali

Mariangela Guidolin

aa 2024-25

Informazioni generali

Docente: prof.ssa **Mariangela Guidolin**

email: **guidolin@stat.unipd.it**,

Ricevimento: su appuntamento, anche per via telematica

Programma del corso

- Parte generale
 - Introduzione generale sui problemi di Business Analytics.
 - Tecniche di presentazione dei risultati.
- Modelli di regressione
 - Modelli di regressione per dati ordinali
 - Modelli di regressione per dati gerarchici
 - Modelli di regressione nonlineare per serie storiche
 - Modelli di regressione non parametrica (Gradient Boosting)
- Relazioni fra osservazioni
 - Regole di associazione (market basket analysis)
 - Network Analysis: modelli per dati di rete
 - Cluster analysis: model-based clustering
- Analisi di case studies
- General Data Protection Regulation e Data Ethics

Svolgimento del corso e modalità d'esame

Svolgimento del corso

- Il corso si articola in 44 ore di lezione frontale e 20 ore di esercitazione in aula informatica.
- Sono previste conferenze didattiche su temi legati agli argomenti del corso.

Svolgimento del corso e modalità d'esame

Modalità d'esame

- ① Lavoro di gruppo (max 3 persone) su un caso di studio prescelto dal gruppo stesso.
La valutazione del lavoro di gruppo avverrà tramite una presentazione in aula (circa mezz'ora per gruppo).
- ② Prova scritta sui contenuti del programma del corso

Materiali di studio e testi di riferimento

Materiali di studio: Lucidi e letture forniti a lezione

Testi di riferimento:

- [Data analysis and data mining: An introduction](#). (Oxford University Press, 2012). Azzalini, A., Scarpa, B.
- [The elements of statistical learning](#) (Springer series in statistics, 2008). Hastie, T., Tibshirani, R., Friedman, J.
- [Data Analysis Using Regression and Multilevel/Hierarchical Models](#) (Cambridge University Press, 2007) Gelman, A., Hill, J.
- [Innovation diffusion models: Theory and Practice](#) (Wiley, 2023) Guidolin, M.
- [Analysis of ordinal categorical data](#) (Wiley, 2010) Agresti, A.
- [Storytelling with Data: A Data Visualization Guide for Business Professionals](#) (Wiley, 2015) Nussbaumer Knaflic, C.

Business Analytics

Business Analytics e Business Intelligence

Business Analytics: insieme di tecniche di analisi dei dati a supporto di decisioni aziendali.

Business Intelligence: tecniche di visualizzazione e reportistica per comprendere “cosa è successo e cosa sta succedendo”.

Dati, informazioni, conoscenza

Un ambiente di Business Intelligence deve offrire strumenti e metodologie atte a individuare decisioni **efficaci** e **tempestive**.

Vale un'importante distinzione:

- **Dati**: codifica strutturata di singole entità primarie (es: nella grande distribuzione dati si riferiscono a entità primarie quali clienti, punti vendita, articoli, scontrini)
- **Informazioni**: sono il risultato di operazioni di estrazione e elaborazione compiute a partire dai dati; hanno significato all'interno di uno specifico contesto.
- **Conoscenza**: informazioni diventano conoscenza se vengono utilizzate per prendere decisioni e sviluppare azioni conseguenti. Le informazioni, inserite in uno specifico contesto, sono arricchite da **esperienza** e competenze dei decision maker per affrontare problemi complessi.

Sistema di supporto alle decisioni

- Possiamo definire un **sistema** come un'entità costituita da un insieme di componenti legati tra loro da mutue relazioni e dotati di una finalità collettiva.
- Un sistema di supporto alle decisioni riceve **flussi in ingresso** e restituisce **flussi in uscita** mediante un processo di *trasformazione* regolato da condizioni interne ed esterne.
- In molti casi un sistema incorpora meccanismi di retroazione o feedback.
- Valgono due criteri essenziali: **efficacia** (azione giusta) ed **efficienza** (nel modo migliore).

Razionalità e problem solving

- Il processo decisionale si colloca entro una più ampia tematica indicata come **problem solving**, ovvero il percorso per colmare il divario tra le condizioni attuali *-as is-* e le condizioni ritenute migliori *-to be-*.
- Un orientamento razionale al processo decisionale implica di tenere conto di **criteri economici, tecnici, legali, etici, procedurali e politici**.

Processo decisionale

Il processo decisionale si compone di alcune fasi salienti:

- **Ricognizione**: individuazione e formulazione del problema.
- **Ideazione**: progettazione di azioni dirette a risolvere il problema analizzato.
- **Selezione**: valutazione delle alternative in base a criteri di prestazione ritenuti significativi.
- **Esecuzione**: trasformazione della decisione selezionata in azione.
- **Controllo**: verifica degli effetti prodotti rispetto agli intendimenti originari.

Tipi di decisioni

In base alla loro **natura** le decisioni vengono classificate come strutturate, semi-strutturate e non strutturate.

- **Strutturate**: riconducibili a una procedura di scelta ben definita e replicabile.
- **Non strutturate**: quando il processo decisionale viene affrontato per la prima volta o se si presenta molto di rado.
- **Semi-strutturate**: se alcune fasi sono strutturate e altre non lo sono.

Tipi di decisioni

In base alla loro **portata** le decisioni vengono classificate come strategiche, tattiche e operative.

- **Strategiche**: riguardano un'intera organizzazione per un lungo periodo di tempo. Determinano gli obiettivi e le politiche generali di una impresa.
- **Tattiche**: riguardano una singola funzione aziendale e hanno un orizzonte temporale di medio termine
- **Operative**: riguardano singole specifiche unità e hanno un impatto limitato sul futuro.

Concetti generali di Business Analytics

- **Classificazione:** la più semplice forma di analisi: il destinatario di una offerta può rispondere oppure no; una persona che chiede un prestito può restituirlo in tempo, in ritardo, oppure dichiarare bancarotta; una transazione di una carta di credito può essere normale o fraudolenta; un taxi può essere in servizio oppure no.
- **Regressione:** in questo caso si tenta di prevedere il valore di una variabile numerica (quanto è stato acquistato), piuttosto che una classe (acquisto sì o no).

Concetti generali di Business Analytics

- **Regole di associazione:** Grandi raccolte di dati sulle transazioni dei clienti danno luogo naturalmente ad associazioni fra prodotti comprati, secondo la regola “che cosa va con cosa”. Le regole associative sono utilizzate per trovare delle strutture di relazione e possono essere impiegate in molti modi: posizionamento del prodotto, offerte promozionali; in un ospedale possono essere utilizzate per riconoscere “quale sintomo è seguito da quale altro sintomo”.

Concetti generali di Business Analytics

- I metodi che verranno presentati possono essere indicati con il termine generale **predictive analytics**.
- Vale una fondamentale distinzione tra metodi impiegati: **metodi supervisionati** e **non supervisionati**.
- Molti metodi classici come la regressione lineare e la regressione logistica, e metodi più recenti come GAM, boosting e support vector machines sono metodi di apprendimento supervisionato, nel senso che per ogni $x_i, i = 1, \dots, n$ c'è una risposta associata y_i .
Diversamente, l'apprendimento non supervisionato descrive la situazione in cui per ogni vettore di x_i non c'è la corrispondente variabile risposta y_i .

Tipici step di un processo di Business Analytics

- **Comprendere lo scopo del progetto:** come verranno impiegati i risultati? a chi interessano i risultati? si tratta di una analisi one-shot o di una procedura continuativa?
- **Ottenere il dataset da utilizzare:** campionamento; integrazione di diverse fonti di dati;
- **Esplorare e pulire i dati:** si tratta di verificare che ci siano delle condizioni ragionevoli: come gestire i dati mancanti? i valori sono in un range ragionevole? sono presenti degli outliers?
- **Ridurre la dimensione dei dati, se necessario:** eliminazione di variabili non necessarie, trasformazione o creazione di nuove variabili.
- **Determinare il tipo di analisi necessaria:** la domanda generale al punto 1 deve essere tradotta in un problema statistico.

Tipici step di un processo di Business Analytics

- **Dividere i dati:** stima e verifica.
- **Scegliere il metodo di analisi più adatto:** regressione, classificazione, clustering, regole associative ...
- **Iterare il processo più volte:** questo step prevede di utilizzare più varianti dello modello.
- **Interpretare i risultati dei modelli:** questo step richiede di fare una scelta sul miglior modello.

Come scegliere il modello migliore

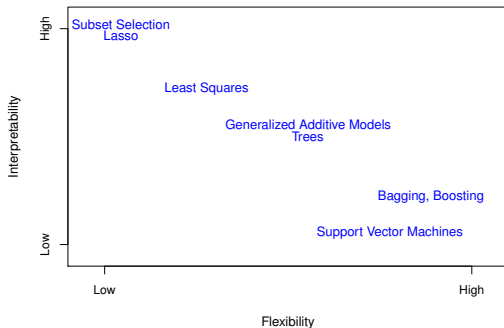
Tipicamente, siamo interessati a prevedere e spiegare

- Valore numerico (regressione)
- Appartenenza a una classe (classificazione)
- probabilità di appartenenza a una classe (ranking)

Ricordiamo che accuratezza previsiva e adattamento (goodness-of-fit) sono due concetti distinti.

Flessibilità e Interpretabilità dei modelli

Perchè scegliere un metodo più restrittivo invece di uno più flessibile?



Flessibilità e interpretabilità: trade off

(fonte: G. James, D. Witten, T. Hastie and R. Tibshirani, 2022).