



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

High Dimensional Data Analysis

2425-2-FDS01Q022

Obiettivi formativi

Questo è un corso avanzato di statistica che ha come oggetto principale l'analisi di dati ad alta dimensionalità. L'obiettivo del corso è quello di presentare le moderne tecniche di analisi dei dati e la teoria statistica sottostante, coniugando armoniosamente aspetti teorici, pratici e computazionali.

Contenuti sintetici

Il corso riguarda metodi di regressione e classificazione che possono essere impiegati nel caso di dati ad alta dimensionalità

Programma esteso

1. Regressione lineare, bias/variance trade-off
2. Regressione penalizzata, ridge regression e lasso
3. Selezione del modello, metodi di validazione incrociata
4. Regressione nonparametrica. *k-nearest neighbors* (k-NN). Kernel smoothing. Regression splines, Smoothing splines, Local regression

Prerequisiti

Sono necessarie conoscenze di probabilità ed inferenza, modellizzazione statistica, algebra lineare e

programmazione in R.

Tali prerequisiti devono essere presi molto seriamente. Se non si possiede una solida comprensione dell'algebra lineare, della probabilità e dell'inferenza, oltre a una notevole esperienza di programmazione con una certa attenzione alle strutture dati, consiglio vivamente di non seguire questo corso. Tuttavia, questi prerequisiti non sono formali; piuttosto, sono imposti dal fatto che non si può comprendere il corso senza di essi.

Metodi didattici

Tutte le lezioni integreranno aspetti di carattere teorico con quelli computazionali attraverso l'uso del software R.

- 12 di lezioni in remoto con modalità sincrona;
- 30 ore di lezioni in presenza con modalità erogativa.

Modalità di verifica dell'apprendimento

Studenti frequentanti: esame scritto.

Studenti non frequentanti: esame scritto.

Durante l'esame sarà valutata la correttezza e la chiarezza delle risposte. L'esame mira a valutare le competenze descritte negli obiettivi formativi.

L'esame scritto consta di 2 domande a risposta aperta.

Testi di riferimento

- Materiale didattico fornito dal docente
- Azzalini, Scarpa (2012) Data analysis and data mining, an introduction . New York: Oxford University Press
- Gareth, Witten, Hastie, Tibshirani (2014) An Introduction to Statistical Learning, with Applications in R . Springer
- Hastie, Tibshirani, Friedman (2009) The Elements of Statistical Learning. Data Mining, Inference and Prediction . Springer

Periodo di erogazione dell'insegnamento

Primo Semestre

Lingua di insegnamento

Inglese

Sustainable Development Goals

ISTRUZIONE DI QUALITÀ
