UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

High Dimensional Data Analysis

2122-2-F9101Q016

Obiettivi formativi

Questo è un corso avanzato di statistica che ha come oggetto principale l'analisi di dati ad alta dimensionalità. L'obietto 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. Sezione 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, algebra lineare, programmazione.

Metodi didattici

Tutte le lezioni si svolgono in laboratorio, integrando aspetti di carattere teorico con quelli computazionali attraverso l'uso di R.

Modalità di verifica dell'apprendimento

Prova individuale orale su argomenti trattati a lezione. Viene valutata la completezza, la correttezza delle risposte e la proprietà di linguaggio.

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
- Hastie, Tibshirani and Wainwright (2015) Statistical Learning with Sparsity: The Lasso and Generalizations . CRC Press

Periodo di erogazione dell’insegnamento

Primo Semestre

Lingua di insegnamento

Italiano