



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

High Dimensional Data Analysis

2021-2-F9101Q016

Obiettivi formativi

Questo è un corso avanzato di statistica che ha come oggetto principale l'analisi di high-dimensional data. 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 high-dimensional data.

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. Nearest neighbors. 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.

Nel periodo di **emergenza Covid-19** le lezioni si svolgeranno da remoto asincrono con eventi in videoconferenza sincrona.

Modalità di verifica dell'apprendimento

Presentazione di un lavoro di gruppo su un progetto concordato con il docente e una prova individuale scritta. Il voto finale sarà una media ponderata tra: prova scritta (50%), progetto di lavoro (30%), presentazione e discussione orale del progetto (20%).

Ciascun progetto può riguardare un articolo o un capitolo di libro su un argomento specifico trattato nel corso. La relazione deve comprendere una descrizione della metodologia utilizzata, una sua discussione critica e l'implementazione del metodo descritto attraverso R, dopo aver scelto opportunamente un set di dati. Il lavoro di gruppo può prevedere al massimo tre studenti per gruppo.

Tipologia di prova:

-prova individuale scritta: domande aperte e project work

- prova individuale orale: colloquio di discussione sullo scritto, sul project work e su argomenti trattati a lezione

Nel periodo di **emergenza Covid-19** gli esami individuali orali saranno solo telematici. Verranno svolti utilizzando la piattaforma WebEx e nella pagina e-learning dell'insegnamento verrà riportato un link pubblico per l'accesso all'esame di possibili spettatori virtuali.

Testi di riferimento

- Materiale didattico fornito dal docente
- _____
- 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
