UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Statistical Learning

2627-1-F9103Q023

Obiettivi formativi

Il corso si pone come obiettivo l'acquisizione delle principali tecniche di statistical learning (SL) e la loro implementazione nell'ambiente di programmazione R. Durante il corso verrà data particolare enfasi alla algorithmic modeling culture, prestando anche attenzione alla stima dell'incertezza nelle previsioni.

Il corso contribuisce al raggiungimento degli obiettivi formativi nell'area di apprendimento del CdS: "Artifical Intelligence".

Contenuti sintetici

Gli argomenti principali sono:

- Metodi basati sugli alberi e aspetti computazioni.
- Moving beyond linearity.
- Stima dell'incertezza.

Programma esteso

Metodi basati sugli alberi.

- Alberi decisionali: classificazione e regressione.
- Bagging.
- Random forests.
- Boosting e alberi additivi.

- Ensemble learning.
  Focus: algoritmo gradient boosting.

Moving beyond linearity.

- Regression splines and smoothing splines.
- Generalized additive models.
- Multivariate adaptive regression splines.

Uncertainty estimation.

- Conformal prediction.

Prerequisiti

Si consiglia la conoscenza degli argomenti trattati nell'insegnamento di “Advanced Foundetions of Statistics for AI”.

Metodi didattici

Le lezioni si svolgono sia in aula che in laboratorio, integrando aspetti di carattere teorico con quelli pratico-applicativi di analisi dei dati e di programmazione in R.

Le 54 ore di didattica saranno così suddivise:

- 32 ore di lezione;
- 24 ore di attività di laboratorio.

Modalità di verifica dell'apprendimento

L'esame si compone di un esame scritto.

(31 punti su 31) Prova scritta a domande aperte e chiuse, in cui vengono valutati gli aspetti teorici del corso. Saranno inoltre presenti domande pratiche relative a R.

Lo studente potrà svolgere un progetto, non obbligatorio, dove potrà raggiungere un massimo di 2 punti.

Testi di riferimento

T. Hastie, R. Tibshirani, J. Friedman (2017) The Elements of Statistical Learning. Springer.
D. Efron, T. Hastie (2016) Computer-Age Statistical Inference: Algorithms, Evidence, and Data Science. Cambridge University Press.
M. Kuhn, L. Johnson (2019) Feature Engineering and Selection: A Practical Approach for Predictive Models.

Chapman & Hall/CRC Data Science.

Altro materiale verrà suggerito durante il corso.

Periodo di erogazione dell'insegnamento

Secondo semestre.

Lingua di insegnamento

Le lezioni saranno svolte in inglese. Il materiale e i libri di testo sono in Inglese.

Sustainable Development Goals

ISTRUZIONE DI QUALITÁ