



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Statistical Learning M

2526-2-F8204B015

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: "Statistica".

Contenuti sintetici

Gli argomenti principali sono:

- Metodi basati sugli alberi e aspetti computazionali.
- Deep Learning per dati non strutturati.
- 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.

Deep learning

- Neural networks.
- Convolutional neural networks.
- Recurrent neural networks.

Uncertainty estimation.

- Conformal prediction.

Prerequisiti

Si consiglia la conoscenza degli argomenti trattati negli insegnamenti di “Probabilità e Statistica Computazionale M”, “Statistica Avanzata M” e del modulo “Machine Learning” dell’insegnamento “Data Science M”.

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 47 ore di didattica saranno così suddivise:

- 30 ore di lezione svolte in modalità erogativa in presenza;
- 12 ore di attività di laboratorio svolte in modalità interattiva da remoto o in presenza;
- 5 ore di case study.

Modalità di verifica dell'apprendimento

L'esame si compone di un esame scritto/pratico da svolgere in laboratorio.

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

I. Goodfellow, Y. Bengio, A. Courville (2016) Deep Learning. MIT Press.
F. Chollet, J.J. Allaire (2018) Deep Learning with R. Manning.

Altro materiale verrà suggerito durante il corso.

Periodo di erogazione dell'insegnamento

Primo semestre, secondo ciclo.

Lingua di insegnamento

Le lezioni saranno svolte in italiano. Il materiale e i libri di testo sono in Inglese.
Le lezioni potranno essere svolte in inglese se si rendesse necessario.

Sustainable Development Goals

ISTRUZIONE DI QUALITÀ
