



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Statistical Learning

2021-2-F8204B018-F8204B033M

Obiettivi formativi

Il recente sviluppo tecnologico consente agli scienziati e alle aziende di raccogliere sistematicamente dati di elevata complessità e dimensione, tuttavia la metodologia statistica tradizionale può risultare inadeguata ad affrontare le sfide dei *big data*.

- Introduzione al *deep learning* (3 cfu, Prof. Borrotti)
- Inferenza statistica moderna (3 cfu, Prof. Solari)

Lo studente, al termine del corso, avrà perfezionato le abilità di *problem solving* e di programmazione in R e sarà in grado di utilizzare metodi statistici avanzati per trarre conclusioni sui dati.

Contenuti sintetici

Introduzione al deep learning

Il modulo si pone l'obiettivo di introdurre i concetti fondamentali del *deep learning* permettendo la loro applicazione consapevole a problemi applicativi.

Inferenza statistica moderna

Il modulo si pone l'obiettivo di introdurre metodi statistici avanzati per affrontare il problema della molteplicità dei test, dell'inferenza post-selezione e dell'inferenza per dati ad elevata dimensionalità.

Programma esteso

Introduzione al deep learning

- Introduzione al Deep Learning
- Reti Neurali
- Reti Neurali Shallow e Deep
- Recurrent Neural Network
- Convolutional Neural Network

Inferenza moderna

- La scienza moderna è in crisi
- Il problema della molteplicità dei test
- Il problema dell'inferenza *post hoc*
- Il problema dell'inferenza per modelli ad elevata dimensionalità

Prerequisiti

Si consiglia la conoscenza degli argomenti trattati nei corsi *Probabilità e Statistica Computazionale M*, *Statistica Avanzata M* e nel modulo *Data Mining* del corso *Data Science M*.

Metodi didattici

Le lezioni si svolgono sia in aula che in laboratorio.

Nel periodo di emergenza Covid-19 le lezioni si svolgeranno in modalità da remoto asincrono, eventualmente con eventi in videoconferenza sincrona e/o in presenza fisica.

Modalità di verifica dell'apprendimento

La modalità di verifica si basa su due prove, una per ciascuna parte del corso

1. ____
2. Inferenza statistica moderna

ed una prova orale finale facoltativa. _____

Introduzione al deep learning

La modalità di verifica consiste nella presentazione di un progetto applicativo concordato con il docente. Il progetto si compone di

- Report
- File Rmarkdown contenente tutto il codice utilizzato per ottenere i risultati
- Slides della presentazione

La presentazione del progetto ha come obiettivo la verifica delle abilità di analisi e di comunicazione dei risultati. ____

Inferenza statistica moderna

La modalità di verifica è suddivisa in due parti:

1. Homework
2. Prova orale

Gli homework (parte 1.) devono essere consegnati prima dell'appello d'esame, rispettando le scadenze indicate. Agli studenti frequentanti viene data l'opportunità di svolgere gli homework in gruppi di massimo 3 persone.

Nel periodo di emergenza Covid-19 le prove d'esame saranno solo telematiche. Verranno svolte utilizzando le piattaforme WebEx e Moodle e nella pagina e-learning dell'insegnamento verrà riportato un link pubblico per l'accesso all'esame di possibili spettatori virtuali.

Testi di riferimento

Introduzione al deep learning

- _____
- _____
 - _____
 - _____
 - _____

Inferenza moderna

- Archivio del corso: <https://aldosolari.github.io/MI/>
- Efron, Hastie (2016) *Computer-Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press
- Wainwright (2019) *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press

Periodo di erogazione dell'insegnamento

Lingua di insegnamento

Le lezioni si svolgono in Italiano, tuttavia i libri di testo sono in lingua Inglese.
