

COURSE SYLLABUS

Statistical Learning M

2122-2-F8204B015

Obiettivi formativi

Statistical Learning è un'area della Statistica di recente sviluppo e caratterizzata da forti sovrapposizioni con il *Machine Learning*. Il corso si pone come obiettivo l'introduzione dei principali metodi di *Statistical Learning*, discutendone sia gli algoritmi che gli aspetti inferenziali.

Contenuti sintetici

Introduzione a metodi statistici avanzati, in particolare:

- modelli di regressione per dati ad elevata dimensionalità
- procedure di selezione delle variabili con garanzie inferenziali
- intervalli di previsione non parametrici

Programma esteso

- Prediction, Estimation, and Attribution.
- James-Stein estimation.

- Ridge regression.
- Splines.
- Additive models.

- Classical versus high-dimensional theory.

- Sparse Modeling and the Lasso.
- Best Subsets Selection.
- Data splitting for variable selection.
- Stability selection.
- Knockoff filter.
- Conformal prediction.

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 sia in laboratorio.

Modalità di verifica dell'apprendimento

La modalità di verifica consiste in una prova scritta in laboratorio. Lo studente (oppure il docente) può richiedere la prova orale.

La prova scritta ha come obiettivo la verifica delle abilità di analisi dei dati e di programmazione e la comprensione degli aspetti matematico-statistici degli argomenti trattati.

Testi di riferimento

- Efron, Hastie (2016) *Computer-Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press
- Hastie, Tibshirani, Friedman (2009). *The Elements of Statistical Learning*. Springer
- Hastie, Tibshirani, Wainwright (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press
- Lewis, Kane, Arnold (2019) *A Computational Approach to Statistical Learning*. Chapman And Hall/Crc.
- Shalizi (2021). *Advanced Data Analysis from an Elementary Point of View*.
- Wainwright (2019) *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press

Periodo di erogazione dell'insegnamento

Secondo semestre, primo periodo.

Lingua di insegnamento

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