UNIVERSITÀ DI PISA

ANALISI DEI DATI

MARCO ROMITO

Anno accademico	2022/23
CdS	MATEMATICA
Codice	699AA
CFU	6

Moduli	Settore/i	Tipo	Ore	Docente/i
ANALISI DEI DATI	MAT/06	LEZIONI	42	MARCO ROMITO

Obiettivi di apprendimento

Conoscenze
Al termine del corso lo studente avrà acquisito conoscenze di metodi di ``statistical learning'', nelle parti di previsione, inferenza, e implementazione

Modalità di verifica delle conoscenze
Lo studente sarà valutato riguardo la sua abilità di valutare il modello statistico più opportuno per affrontare un problema di ``statistical learning'', e di tradurlo in una soluzione algoritmica.

Capacità
Al termine del corso lo studente

- sarà in grado di formulare il modello statistico più opportuno per l'analisi quantitativa di dati,
- saprà implementare l'analisi formulata per mezzo di un software statistico,
- sarà in grado di trarre conclusioni e formulare previsioni sul problema esaminato.

Modalità di verifica delle capacità
Il soggetto delle prove d'esame sarà l'analisi e implementazione di modelli statistici.

Comportamenti
Il corso permetterà di gestire l'analisi quantitativa di tabelle di dati mediante metodi statistici.

Modalità di verifica dei comportamenti
Durante le sessioni di esame lo studente sarà esaminato sulle fasi di analisi statistica di una tabella di dati, dal riconoscimento del modello più efficace alla sua implementazione e previsione.

Prerequisiti (conoscenze iniziali)
Ci si aspetta che lo studente conosca i concetti e le idee di base della probabilità e della statistica, e di alcuni argomenti di base riguardanti l'analisi dei dati (regressione multivariata, analisi delle componenti principali, metodi autoregressivi per serie storiche). Ci si aspetta inoltre una conoscenza di base di R o python.

Indicazioni metodologiche
Il corso prevede lezioni frontali per la parte teorica. La parte implementativa è svolta parzialmente durante il corso, parzialmente come lavoro autonomo dello studente, attraverso lo svolgimento di progetti focalizzati via via su differenti argomenti scelti tra i contenuti del corso da svolgere in piccoli gruppi.

Programma (contenuti dell'insegnamento)
Introduzione al ``statistical learning'', apprendimento supervisionato e non supervisionato. Modelli di apprendimento, classi di predittori, apprendimento probabilmente approssimativamente corretto (PAC). Dimensione di Vapnik-Chervonenkis. Esame di alcuni semplici esempi (regressione lineare, varianti non lineari, k-nearest-heighbour). Metodi di valutazione dei modelli (cross-validation, bootstrap, criteri di informazione). Problemi di classificazione mediante regressione logistica, analisi discriminante, support vector machines. Metodi basati su alberi

e foreste. Discesa gradiente stocastica e reti neurali. Alcuni problemi di apprendimento non supervisionato.

Bibliografia e materiale didattico

S. Ben-David, S. Shalev-Shwartz: Understanding Machine Learning: From Theory To Algorithms
T. Hastie, R. Tibshirani, J. Friedman: The elements of Statistical Learning
J. Gareth, D. Witten, T. Hastie, R. Tibshirani: An introduction to statistical learning

Indicazioni per non frequentanti

La frequenza del corso è fortemente consigliata.

Modalità d'esame

La prova d'esame consiste in una prova orale.
Gli studenti frequentanti in alternativa potranno cimentarsi nella realizzazione di un progetto personale di analisi dei dati su un problema fornito dal docente.

Pagina web del corso

http://people.dm.unipi.it/romito/Teaching/2023/data

Altri riferimenti web

Il materiale del corso sarà reso disponibile sulla pagina e-learning del corso

Ultimo aggiornamento 30/08/2022 18:44