UNIVERSITÀ DI PISA

DATA ANALYTICS

FABRIZIO LOMBARDINI

Anno accademico	2023/24
CdS	INGEGNERIA DELLE TELECOMUNICAZIONI
Codice	908II
CFU	6

Moduli	Settore/i	Tipo	Ore	Docente/i
DATA ANALYTICS	ING-INF/03	LEZIONI	60	FABRIZIO LOMBARDINI FRANCO MARIA NARDINI

Obiettivi di apprendimento

Conoscenze
Lo scopo generale del corso è quello di far acquisire allo studente familiarità con la caratterizzazione, la simulazione e l'analisi statistica dei dati e dei segnali aleatori, in particolare di interesse per l'ingegneria delle telecomunicazioni, e con i principi di apprendimento automatico da grandi quantità di dati.
Una prima parte dell'insegnamento tratta i temi basilari dell'analisi dei dati con metodi statistici classici, quali la misura sperimentale ovvero stima di densità di probabilità e di indici statistici e la regressione lineare, e la teoria dei segnali aleatori tempo discreti, con relativi metodi base di stima per la loro caratterizzazione in potenza, sia nel dominio del tempo che della frequenza.
Una seconda parte del corso verterà sull'apprendimento automatico di modelli ovvero machine learning. Si introdurranno i principi dell'apprendimento automatico da grandi quantità di dati sia in ottica supervisionata che non supervisionata. Si analizzerà poi la teoria di apprendimento con particolare enfasi sul processo di learning e sulla metodologia sperimentale che guida l'apprendimento negli algoritmi analizzati.
Acquisizione di ulteriori conoscenze pratiche nell'ambito della simulazione e analisi statistica dei dati e dei segnali è ottenuta con riferimento al linguaggio Matlab, e nell'ambito della regressione e classificazione con machine learning tramite Python.

Modalità di verifica delle conoscenze
I docenti svolgeranno un certo numero di esercitazioni con il linguaggio Matlab e Python, fornendo poi i codici e risultati tipici. Le esercitazioni riguarderanno la generazione di dati e segnali secondo predefiniti modelli statistici, e l'implementazione software degli algoritmi di stima e di machine learning più importanti studiati nel corso. Lo studente può eseguire autonomamente gli algoritmi per una comprensione più profonda della teoria.

Capacità
Lo studente sarà in grado di strutturare algoritmi Matlab per risolvere un determinato problema di simulazione o analisi sperimentale di dati o segnali aleatori e sarà in grado di valutare le prestazioni dell'algoritmo, e di utilizzare librerie Python per risolvere problemi di regressione e classificazione tramite machine learning.

Modalità di verifica delle capacità
I docenti proporranno codici Matlab e Python per implementare gli algoritmi più importanti studiati nel corso in un certo numero di esercizi tipo. Lo studente può eseguire autonomamente gli algoritmi per capire meglio il loro comportamento e le loro prestazioni.

Comportamenti
Lo studente acquisirà l'abilità di affrontare un problema descrivendolo con un modello matematico e risolvendolo con precisione, e di utilizzare propriamente librerie di algoritmi di machine learning.

Modalità di verifica dei comportamenti
La verifica di quello che gli studenti hanno imparato si ha durante le lezioni, poi durante i ricevimenti individuali e infine attraverso l'esame finale.

Prerequisiti (conoscenze iniziali)
Conoscenze di base di teoria della probabilità, variabili aleatorie e vettori aleatori, dell'analisi dei segnali deterministici e dei sistemi tempo continuo, e dei segnali aleatori tempo continuo (propedeuticità consigliata: Segnali e Sistemi).

Programma (contenuti dell'insegnamento)

Teoria e simulazione di fenomeni aleatori: metodi di simulazione di variabili aleatorie, e di vettori aleatori (Gaussiani); segnali aleatori tempo discreto, con simulazione di semplici esempi; campionamento di segnali aleatori, e filtraggio di segnali aleatori tempo discreto.
Basi di Data Analytics: misura sperimentale (stima) di probabilità e densità di probabilità; stime campionarie di momenti di variabili aleatorie, intervallo di confidenza; stimatori base per l'analisi in potenza di segnali aleatori tempo discreto.
Apprendimento automatico (machine learning): approcci di classificazione (a singola e multipla etichetta), regressione (lineare e non lineare) e relative tecniche allo stato dell'arte quali: support vector machines, foreste di alberi di decisione, reti neuronali artificiali. Overfitting, underfitting, insiemi di train/validation e test su cui ottimizzare la prestazione dei modelli appresi, algoritmi di ottimizzazione e loro differenze, funzioni di errore e relativi parametri che ne orientano il funzionamento.
Esercitazioni sul Matlab: codifica e pratica di simulatori statistici e stimatori (di variabili e vettori aleatori, e di segnali aleatori tempo discreto).
Esercitazioni Python: stack di librerie per data science (NumPy, Pandas, Seaborn), e codifica, allenamento, ottimizzazione e verifica prestazione di algoritmi di machine learning per classificazione e regressione (in Scikit-learn, Pytorch).

Bibliografia e materiale didattico

Materiale fornito dal docente (testo di riferimento Teoria dei Fenomeni Aleatori, Galati-Pavan, ed. Texmat).

Modalità d'esame

Prova orale finale. Durante la prova, lo studente sarà valutato sulla base della sua capacità di discutere i contenuti del corso con competenza, consapevolezza critica, e correttezza di espressione. Nella prova allo studente viene anche chiesto di risolvere due esercizi basati sul linguaggio Matlab e Python al fine di dimostrare la capacità di mettere in pratica i principi di base della teoria analizzati durante il corso.

Ultimo aggiornamento 30/10/2023 20:33