

CORSO STATISTICA PER IL MACHINE LEARNING: FORMAZIONE PRATICA CON PYTHON

Durata: 3 giorni

PROGRAMMA

PARTE 1: STATISTICA DESCRITTIVA

MODULO 1: LA NATURA DELLA PROBABILITÀ E DELLA STATISTICA

- La natura della statistica, statistica descrittiva, statistica inferenziale
- Dati della popolazione e dati del campione
- Tecniche di campionamento: casuale semplice, sistematico, stratificato

MODULO 2: DISTRIBUZIONI DEI DATI E GRAFICI

- Variabili qualitative e quantitative
- Distribuzione di frequenza, cumulativa, categoriale, raggruppata
- Distribuzione empirica dei dati e sua rappresentazione grafica (grafico a barre, istogramma, poligono di frequenza, curva KDE)

MODULO 3: MISURE DI TENDENZA CENTRALE, FORMA E VARIABILITÀ

- Media, media pesata, media troncata, mediana, moda
- Forma della distribuzione: modalità, simmetria, asimmetria
- Range, varianza, deviazione standard, coefficiente di variazione
- Teorema di Chebyshev e regola empirica

MODULO 4: MISURE DI POSIZIONE E TRASFORMAZIONE DELLE VARIABILI

- Percentili, quartili, intervallo interquartile (IQR), Five-Number Summary
- Identificazione degli outlier e boxplot
- Standardizzazione (Z-score), normalizzazione min-max, trasformazione logaritmica, radice quadrata, reciproca

CORSO STATISTICA PER IL MACHINE LEARNING: FORMAZIONE PRATICA CON PYTHON

Durata: 3 giorni

PARTE 2: PROBABILITÀ, VARIABILI CASUALI E DISTRIBUZIONI CAMPIONARIE

MODULO 5: PROBABILITÀ

- Esperimento, evento, esito, spazio campionario
- Relazioni tra eventi: complemento, intersezione, unione, mutua esclusività
- Interpretazioni della probabilità: classica, empirica, soggettiva (bayesiana)
- Regole di addizione, moltiplicazione, probabilità condizionata, teorema di Bayes, legge dei grandi numeri
- Odds "a favore" e "contro"

MODULO 6: REGOLE DI CONTEGGIO

- Regola fondamentale del conteggio
- Permutazioni (semplici, parziali, con ripetizioni) e combinazioni

MODULO 7: DISTRIBUZIONI DI PROBABILITÀ DISCRETE

- Variabili casuali discrete, funzione di massa di probabilità (PMF)
- Distribuzione uniforme discreta, bernoulli, binomiale, geometrica, poisson

MODULO 8: DISTRIBUZIONI DI PROBABILITÀ CONTINUE

- Funzione di Densità di Probabilità (PDF) e Funzione di Distribuzione Cumulativa (CDF)
- Distribuzione uniforme continua, normale, esponenziale
- Distribuzione campionaria della media, t di Student, chi-quadrato, F di Fisher

PARTE 3: STATISTICA INFERENZIALE

MODULO 9: STIMA STATISTICA E INTERVALLI DI CONFIDENZA

- Parametri e statistiche, bias, varianza dello stimatore, errore standard
- Intervalli di confidenza per la media (sigma nota e sconosciuta) e per una proporzione

MODULO 10: TEST DI IPOTESI

- Ipotesi nulla e alternativa, errori di tipo I e tipo II
- Z-Test e T-Test per una media, test di ipotesi per una proporzione
- Confronto tra due medie (Z-test, t-Test Pooled/Nonpooled, U-Test di Mann-Whitney, campioni accoppiati)
- Test di ipotesi per due proporzioni

CORSO STATISTICA PER IL MACHINE LEARNING: FORMAZIONE PRATICA CON PYTHON

Durata: 3 giorni

MODULO 11: POTENZA STATISTICA ED EFFECT SIZE

- Calcolo delle probabilità per l'errore di tipo II
- Potenza statistica e fattori che la influenzano
- Cohen's d e interpretazione della dimensione dell'effetto

PARTE 4: REGRESSIONE, CORRELAZIONE E ANOVA

MODULO 12: COVARIANZA E CORRELAZIONE LINEARE

- Covarianza della popolazione e campionaria, matrice di covarianza
- Coefficiente di correlazione lineare di Pearson, limiti e corretto utilizzo

MODULO 13: REGRESSIONE LINEARE

- Metodo dei minimi quadrati, interpretazione geometrica
- Outlier, osservazioni influenti, assunzioni del modello, diagnostica dei residui
- Coefficiente di determinazione (r^2)

MODULO 14: ANALISI DELLA VARIANZA (ANOVA)

- Assunzioni dell'ANOVA, confronto con il t-test
- Variabilità entro e tra i gruppi, statistica F, interpretazione dei risultati

CORSO STATISTICA PER IL MACHINE LEARNING: FORMAZIONE PRATICA CON PYTHON

Durata: 3 giorni

OVERVIEW DEL CORSO

Il corso Statistica per il Machine Learning ti guida passo passo nella costruzione delle basi statistiche e probabilistiche necessarie per lavorare con consapevolezza nel Machine Learning e nell'analisi dei dati. Attraverso un approccio pratico e orientato al codice, imparerai a descrivere e interpretare i dati, ragionare in condizioni di incertezza e applicare gli strumenti dell'inferenza statistica: dalla statistica descrittiva ai test di ipotesi, dalla correlazione alla regressione lineare.

Grazie a 156 esempi pratici in Python, il corso ti permette di passare dalla teoria alla sperimentazione diretta sui dati, costruendo competenze che tornano utili ogni volta che si valuta un modello di Machine Learning o si interpretano i suoi risultati.

A CHI È RIVOLTO IL CORSO?

Il corso è rivolto a:

- Data Analyst e Data Scientist
- Sviluppatori software che vogliono avvicinarsi al Machine Learning
- Professionisti che si occupano di Business Intelligence
- Chi lavora con l'analisi dei dati e vuole basi statistiche più solide

COSA SAPRAI FARE ALLA FINE DEL CORSO?

Al termine del corso i partecipanti saranno in grado di:

- Descrivere e sintetizzare un dataset con distribuzioni di frequenza, misure di tendenza centrale e di variabilità
- Riconoscere outlier e applicare trasformazioni come standardizzazione (Z-score) e normalizzazione Min-Max
- Applicare le regole fondamentali della probabilità, inclusi il Teorema di Bayes e la Legge dei Grandi Numeri
- Lavorare con le principali distribuzioni di probabilità discrete e continue (Binomiale, Poisson, Normale, Esponenziale e altre)
- Costruire intervalli di confidenza e condurre test di ipotesi su medie e proporzioni
- Interpretare correttamente p-value, errori di Tipo I e II, potenza statistica ed Effect Size
- Studiare le relazioni tra variabili con covarianza, correlazione, regressione lineare e ANOVA
- Applicare ogni concetto direttamente in Python, con librerie come NumPy, SciPy, Pandas e Matplotlib