

Python, Pandas e Machine Learning

Programma completo (6+ ore)

Installiamo Python e Jupyter (10 minuti)

Iniziamo installando Python, Jupyter e le principali librerie che utilizzeremo.

- Installazione e primo accesso a Jupyter

Import dei dati con Python (45+ minuti)

Studiamo come importare all'interno dei DataFrame i dati provenienti da varie fonti come CSV, Excel, Json e file strutturati.

- Import di file CSV
- Caratteristiche dei DataFrame
- Import di file Excel
- Import di file JSON
- Importiamo un nuovo file CSV più complesso
- Approfondimento: import da file strutturato

Analisi dei dati dichiarative con Python (60+ minuti)

Impariamo a filtrare, aggregare, combinare, trasformare e aggiornare i dati all'interno di un DataFrame.

- Organizzare il codice in funzioni
- Filtrare i dati - parte 1
- Filtrare i dati - parte 2
- Aggregare i dati - metodo groupby
- Combinare i dati - metodo merge
- Combinare i dati - metodo concat
- Ordinare i dati e modificare la struttura di un DataFrame
- Modificare e trasformare i dati in un DataFrame

Esercitazione sul PreProcessing (40+ minuti)

Esercitiamoci su quanto appreso nel capitolo precedente implementando direttamente con Python le principali attività di pre-processing dei dati per il Machine Learning.

- Esercitazione: pre-processing dei dati
- Soluzione pt. 1: import del file Iris
- Soluzione pt. 2: filtrare e ordinare i dati
- Soluzione pt. 3: gestire i null - parte 1
- Soluzione pt. 4: gestire i null - parte 2
- Soluzione pt. 5: gestione variabile target e suddivisione in training e test
- Soluzione pt. 6: valorizzazione null
- Soluzione pt.7 : normalizzazione dati
- Soluzione pt.8: aggiunta colonna pesi e conversione dati in array di numpy

Machine Learning e programmazione a oggetti (60+ minuti)

Implementiamo da zero il Perceptron: il primo storico algoritmo di Machine Learning alla base delle reti neurali moderne. Introduciamo il paradigma della programmazione a oggetti.

- Ripasso sintassi di Python
- Implementazione del Perceptron - parte 1
- Implementazione del Perceptron - parte 2
- Perceptron e programmazione a oggetti - parte 1
- Perceptron e programmazione a oggetti - parte 2
- Perceptron e programmazione a oggetti - parte 3

Librerie per il Machine Learning e classificazione supervisionata (100+ minuti)

Entriamo nel mondo della libreria Scikit-Learn per addestrare algoritmi completi di Machine Learning comprensivi di analisi esplorativa dei dati, pre-processing, creazione di pipeline, tecniche di convalida e ottimizzazione dei parametri. In questo modulo svolgeremo insieme un **progetto completo di Machine Learning** per creare un documento che potrai allegare al tuo CV come prova tangibile delle skill acquisite.

- Machine Learning con Scikit-Learn - parte 1
- Machine Learning con Scikit-Learn - parte 2
- Installazione Seaborn
- Analisi esplorativa dei dati - calcolo dei principali indici statistici

- Analisi esplorativa - rappresentazioni grafiche con Seaborn
- Machine Learning sul dataset Titanic - studio delle variabili
- Machine Learning sul dataset Titanic - encoding e normalizzazione
- Creazione di Pipeline e convalida k-fold
- Ottimizzazione dei parametri
- Matrice di confusione
- Creiamo un portfolio da Data Analyst

Algoritmi di Regressione e Clustering (40+ minuti)

Applichiamo il Machine Learning e Scikit-Learn a problemi di regressione e clustering dei dati.

- Introduzione alla regressione e prime analisi del dataset Diamonds
- Pre-processing del dataset Diamonds
- Applicazione di algoritmi di regressione lineare
- KnRegressor e analisi dei residui
- Clustering dei dati

Torna alla pagina del corso <https://www.yimp.it/python-pandas-e-machine-learning/>