

Data and Web Mining

Appunti

Lorem ipsum

Raccolta domande teoriche

Quali sono le differenze tra gli algoritmi di supervised learning e quelli di unsupervised learning?

Lorem ipsum

Descrivere l'algoritmo K-nn e spiegare cosa succede cambiando il parametro k

Lorem ipsum

Descrivere l'algoritmo per la costruzione di un Decision Tree e spiegare le metriche e i vari parametri

Lorem ipsum

Cosa cambia in un decision Tree utilizzato in un task di classificazione a un decision Tree utilizzato per un task di regressione?

Lorem ipsum

Descrivere un algoritmo efficiente per costruire un decision Tree

Lorem ipsum

Descrivere la regressione lineare e le metriche

Lorem ipsum

Cos'è l'overfitting?

Lorem ipsum

Differenza tra bias e variance?

Lorem ipsum

Descrivere la logistic regression

Lorem ipsum

Descrivere SVM, il suo funzionamento e le metriche

Lorem ipsum

Cosa sono train validation e test

Lorem ipsum

Spiegare k-fold cross-validation

Lorem ipsum

Cos'è lo stratified sampling e quando si utilizza

Lorem ipsum

Cosa significa fare tuning dei parametri?

Lorem ipsum

Cos' è un classificatore baseline?

Lorem ipsum

Come funziona il modello naive bayes

Lorem ipsum

Spiegare i metodi principali per processare testo

Lorem ipsum

Spiegare bias-variance decomposition

Lorem ipsum

Spiegare quando è utile usare bagging e quando invece è utile usare boosting

Lorem ipsum

Spiegare adaboost

Lorem ipsum

Spiegare random forest

Lorem ipsum

Come è possibile usare la random forest per stimare la similarità

Lorem ipsum

Come è possibile usare la random forest per identificare gli outliers?

Lorem ipsum

Come è possibile usare la random forest per rimpiazzare i valori mancanti?

Lorem ipsum

Come è possibile fare feature selection con la random forest?

Lorem ipsum

Come si esegue il feature engineering

Lorem ipsum

Cos'è la confusion matrix

Lorem ipsum

Quali sono le principali metriche di prestazione di un modello

Lorem ipsum

Spiegare cos'è la ROC curve (e AUC) / Come si valuta un classificatore binario

Lorem ipsum

Overfitting vs underfitting

Lorem ipsum

A cosa servono le association rules

Lorem ipsum

Spiegare apriori

Lorem ipsum

Spiegare fp growth

Lorem ipsum

Spiegare cos'è k-shingles

Lorem ipsum

Spiegare il Min-Hashing

Lorem ipsum

Spiegare Min-Hash Signatures e LSH

Lorem ipsum

Spiegare sim-hashing

Lorem ipsum

Spiegare le misure di qualità dei recommender systems

Lorem ipsum

Spiegare le varie tecniche usate in recommender system

Lorem ipsum

Quali sono le principali misure di similarità?

Lorem ipsum

Spiegare il problema della dimensionalità

Lorem ipsum

Come è possibile risolvere il problema della dimensionalità tecniche

Lorem ipsum

Cosa significa fare clustering

Lorem ipsum

Descrivere i vari tipi di clustering

Lorem ipsum

Approcci al clustering

Lorem ipsum

Come funziona k-means

Lorem ipsum

Come funziona k-means++

Lorem ipsum

Come funziona k-medoid (PAM)

Lorem ipsum

Spiegare come funziona HAC e le varie misure (vedi anche complessità)

Lorem ipsum

A cosa serve il dendrogramma

Lorem ipsum

Spiegare clustering divisivo

Lorem ipsum

Spiegare DBScan

Lorem ipsum

Spiegare la valutazione del clustering intrinseca ed estrinseca / Cos'è il silhouette coefficient

Lorem ipsum

Cos'è l'hard clustering e il soft clustering?

Lorem ipsum

Spiegare fuzzy C-means.

Lorem ipsum

Spiegare SOM(self organizing map)

Lorem ipsum

Spiegare cos'è un ANN e come funziona

Lorem ipsum

Spiegare le varie funzioni di attivazione

Lorem ipsum

Cos'è la loss function per un ANN

Lorem ipsum

Cosa apprende un ANN

Lorem ipsum

Cos'è una rete convoluzionale e come funziona

Lorem ipsum

A cosa servono i filtri

Lorem ipsum

Per quali task può essere usata un ANN e invece una convolutional network?

Lorem ipsum

Come risolvere l'overfitting in un ANN

Lorem ipsum

Vedi web search and ranking

Lorem ipsum

Raccolta esercizi pratici

Dato un dataset, trovare la radice del DT usando GINI Index

Lorem ipsum

Dato un dataset, trovare la radice del DT usando Information Gain

Lorem ipsum

Dato un dataset, predire la classe di decisione per la nuova istanza specificata

Lorem ipsum

Data una matrice Transaction ID - Items, trovare l'item set di candidati X con A-Priori Algorithm, considerando 2 come supporto minimo. Calcolare anche la confidenza degli elementi dei candidati risultanti, con confidenza minima pari al 60%

Lorem ipsum

Data una matrice Transaction ID - Items, trovare il Frequent Pattern con l'algoritmo FP_Growth

Lorem ipsum

Filtri convoluzionali

Lorem ipsum