

NEAREST NEIGHBORS

mercoledì 8 novembre 2023

17:24

Consente di classificare gli elementi che stiamo analizzando

Esempio: Valutare se un' email è spam o no, la classificazione può essere effettuata tramite

Esistono due tipi di modelli:

Parametric

Non-parametric:

-Distanza

-

Rappresentazioni dati:

TR osservazioni che l'algoritmo può imparare

TE tutto quello che l'algoritmo ancora non ha visto

VA set di validazione

Ovviamente:

TR != TE

Decision Boundaries

K-nearest Neighbors:

L'oggetto prendere il colore dei vicini

Il miglior K si trova utilizzando il VA, in genere è dispari. Il K si trova dopo una serie di prove p migliore.

Metriche:

-Accurata:

Esempio:

A=100

B=90

$\text{Accuracy}(A, B) = |A * 2| / |B| = 90 / 100$

NB: Non è buona per datasets non bilanciati, in genere nei casi reali non è mai bilanciato

bilanciato

-Confusion matrix:

Falsi positivi

Falsi negativi

Veri positivi

Veri negativi

Righe per la veridicità, invece le colonne per predizione

Precision e recall

F1 score

