
Esercitazione Octave

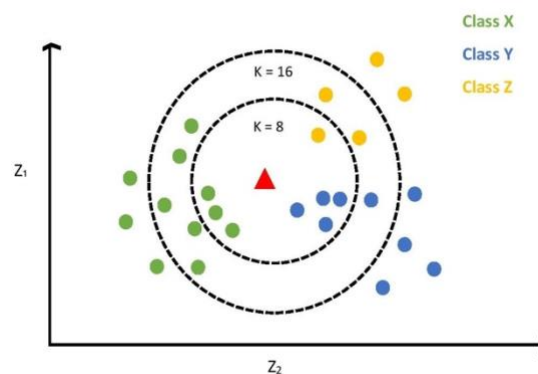
K-Nearest Neighbours

Intro

L'algoritmo K-Nearest Neighbours (K-NN) è un algoritmo di classificazione, cioè dato un elemento in input associa la classe di appartenenza. L'algoritmo si basa su una logica molto semplice. Dato un insieme di dati già classificato:

- viene calcolata la distanza fra il punto da classificare ed ogni punto del set di dati noto
- vengono presi i K elementi più vicini (dove K è un parametro dell'algoritmo configurabile)
- viene assegnata come classe di appartenenza la più ricorrente.

Nell'esempio in figura, dato il punto da classificare (indicato con il triangolo in rosso):



- per $K=8$ vengono presi gli 8 elementi più vicini
- viene assegnata la classe Y perché più ricorrente in questo sottoinsieme
- per $K=16$ vengono presi i 16 elementi più vicini
- viene assegnata la classe X perché più ricorrente in questo sottoinsieme

Questo algoritmo, concettualmente molto semplice, può essere utilizzato per classificare le malattie al cuore. ([See more](#))

Prerequisiti

In questa esercitazione avremo due set di dati, entrambi classificati, utilizzeremo uno di questi come set noto (training set) e l'altro come set da classificare (test set). Utilizzeremo dunque la classe del test set per verificare la corretta classificazione di ogni elemento.

Ogni elemento da classificare è caratterizzato da 2 coordinate (x, y) ed un terzo elemento che indica la classe di appartenenza (che può valere 1 o 2).

Implementazione

Per implementare l'algoritmo esegui i seguenti passi:

1. Calcolare la distanza euclidea fra il punto del test set ed ogni punto del training set.
2. Ordinare il vettore delle distanze in ordine crescente
3. Scegliere un valore per K
4. Prendere le classi dei K elementi più vicini
5. Trovare la classe più frequente. *Hint: utilizza la funzione [mode\(\)](#)*

Nota: i punti 1-4 sono uguali all'esercitazione precedente.

Una volta classificato l'elemento verificare che la classe sia corretta e calcolare l'accuratezza dell'algoritmo:

$$accuracy = \frac{hits}{\#elements}$$

Per disegnare il test set in un grafico potete usare i seguenti comandi

```
figure
scatter(train(:,1), train(:,2), 50, train(:,3), 'Filled')
colormap(jet(size(knnClassify2dTrain,1)))
grid on
```

Approfondimento

Calcola e riporta su un grafico l'andamento dell'accuratezza dell'algoritmo al variare di K.

Per disegnare il grafico richiesto utilizzare la funzione [plot\(\)](#).