

ARCHITETTURE AVANZATE PER I SISTEMI DI ELABORAZIONE E PROGRAMMAZIONE

Costanza Ferrari, Michele De Fusco, Luca Cuconato

Relazione del progetto

Indice

| | | |
|----------|------------------------------------|----------|
| 1 | Introduzione | 2 |
| 2 | Le ottimizzazioni richieste | 3 |

Capitolo 1

Introduzione

La traccia del progetto richiedeva agli studenti, sostanzialmente, di implementare l'algoritmo KNN (K Nearest Neighbors) utilizzato come modello di machine learning e impiegato per la classificazione degli oggetti.

L'obiettivo è quello di trovare, per ogni elemento della collezione Query i k punti più simili presenti nel dataset dei dati.

La ricerca dei k elementi più simili viene fatta sulla base di una misurazione, la distanza euclidea, che va a calcolare la distanza fra vettori prendendo in considerazione ogni componente.

Il problema principale risiede nelle risorse che il calcolo della distanza euclidea, calcolata per ogni vettore presente in Q per tutti gli elementi presenti nel dataset, richiederebbe. Il tempo di esecuzione, sviluppando la distanza euclidea sarebbe $O(n * D)$ in cui n sono gli elementi del dataset e D sono

Capitolo 2

Le ottimizzazioni richieste