

Descrizione del progetto

Il progetto svolto per il corso di Metodi avanzati di programmazione si occupa di eseguire data mining ossia l'estrazione automatica di conoscenza in voluminose basi di dati, nel nostro caso estrarremo i dati da un database SQL. Effettueremo il data mining tramite il clustering e nello specifico tramite l'algoritmo *k-means*.

I dati vengono considerati come una collezione di transazioni dove ogni transazione è un vettore di coppie attributo-valore. Il nostro scopo è dunque calcolare e memorizzare i centroidi (ovvero dei valori attorno al quale si concentra la collezione di dati) dei cluster e successivamente minimizzare la distanza tra una transazione nuova e la rappresentazione dei cluster (centroidi) per scoprire il cluster di appartenenza. Utilizzeremo un approccio client-server, nel primo abbiamo la parte che interagisce con l'utente tramite inserimenti e stampe, mentre nel server avvengono tutte le effettive operazioni logiche

A livello logico vengono inizializzati i k segmenti (come insiemi vuoti) vengono inizializzati dei centroidi, si scelgono k transazioni (centroidi) in maniera **CASUALE** e le si inseriscono nei segmenti: un centroide per segmento. Viene assegnata ciascuna transazione al suo cluster (Si sceglie di spostare la transazione nel cluster che minimizza tale distanza.). Vengono ricalcolati i centroidi dei cluster, ci sono transazioni che hanno cambiato cluster di appartenenza? Se sì allora continuo a riassegnare le transazioni e via scorrendo fino a quando non avremo cluster stabili

Nel progetto con l'estensione avremo anche una rappresentazione visiva dei dati tramite la libreria javafx