

Data Mining e Scoperta di Conoscenza

Progetto 7

Implementazione dell'algoritmo di clustering gerarchico DBSCAN

1. Si estenda Weka con una nuova classe **weka.clusterers.DBSCAN.java**, che implementi l'algoritmo di clustering gerarchico **DBSCAN** i cui dettagli sono riportati nel seguente articolo "*A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*" fornito insieme al progetto.
2. Si applichi l'algoritmo implementato:
 - a. sul dataset UCI
 - b. sui dati transazionali forniti contestualmente al progetto valutando come varia l'accuratezza predittiva
 - c. rispetto agli altri algoritmi di clustering forniti in Weka.
 - d. al variare dei parametri dell'algoritmo.
 - e. Al variare delle metriche utilizzate.
3. Estendere Weka con la classe **weka.core.TransactionInstance** che consenta la rappresentazione ottimale dei dati transazionali.

NOTE PER L'ESECUZIONE DEL PROGETTO

1. Scrivi un rapporto di circa 10 pagine in cui
 - a. Descrivi analiticamente l'algoritmo che hai implementato.
 - b. Commenti le parti essenziali del codice Java che hai scritto, e metti in un'appendice l'intero codice
 - c. commenti e illustri graficamente e quantitativamente gli esperimenti effettuati.
2. Prepare delle slides Powerpoint (non più di 10 slides) in cui riassumi gli esiti del progetto