

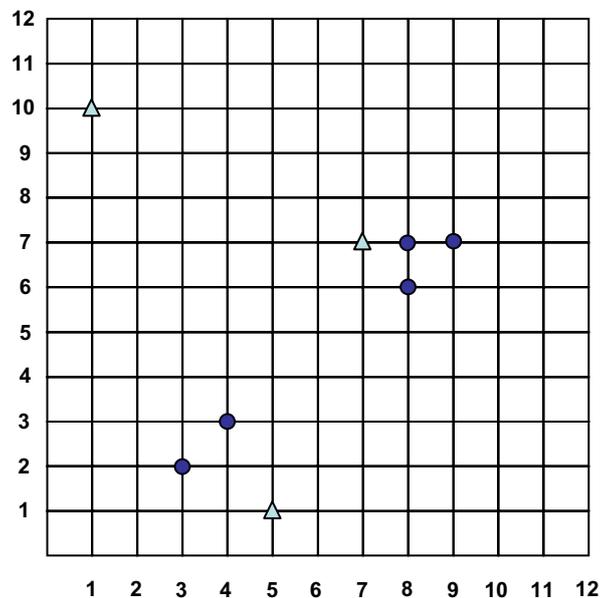
Managing Massive Multiplayer Online Games
SS 2015

Übungsblatt 8: KDD Methoden

Besprechung: 18.06.2015 und 22.06.2015

Aufgabe 8-1 Clusterung mit Varianzminimierung

Gegeben sei erneut folgender Datensatz mit 8 Punkten (2-dimensionalen Featurevektoren).



Im folgenden sollen vollständige Partitionierungen des Datensatzes in $k = 2$ Cluster berechnet werden. Als Distanzfunktion zwischen den Punkten soll dabei die Manhattan-Distanz (L_1 -Norm) verwendet werden.

- (a) Erzeugen Sie eine Partitionierung in $k = 2$ Cluster mit dem einfachen Verfahren "Clustering durch Varianz Minimierung". Die initiale Partitionierung der Daten ist durch die Dreiecke und Punkte gegeben (die Dreiecke bilden einen initialen Cluster, genauso die Punkte). Beschreiben Sie jede Aktion des Algorithmus. Zeichnen Sie nach jedem Schritt die Zentroiden ein und markieren Sie die Punkte anhand ihrer Clusterzugehörigkeit. Denken Sie daran, bei der Zuordnung zu den Zentroiden die L_1 -Norm zu verwenden.

Tipp: Hierzu können Sie die Vorlage auf der letzten Seite benutzen, die Sie am besten mehrmals kopieren.

- (b) Zeigen Sie, dass dieses Verfahren von der Wahl des initialen Clusterings abhängt.

Aufgabe 8-2 *Frequent Itemset Mining*

Betrachten Sie ein Multiplayer Battle Arena Game(MOBA) bei dem die Spieler jeweils eine von 9 Heldenklassen wählen können. Es gibt folgende Klassen: Druiden(D), Hexer(H), Jäger(J), Krieger (K), Mönch (M), Priester(P), Ritter (R), Schurke(S), Zauberer(Z). Ein Team besteht dabei aus 5 Spielern. Gegeben sei nun folgende Datenbank aus Teamzusammensetzungen:

Gruppe	Klassen
1	D H K P Z
2	D J K P S
3	D J M P R
4	H K M P Z
5	D J M R Z
6	D M P S Z
7	D H J M P
8	D J M P Z
9	K M P S Z
10	D J K P Z
11	J K M P Z
12	D K M P R

Bestimmen Sie alle Teilteams die mindestens 5 mal vorkommen. Verwenden Sie hierzu den Apriori Algorithmus zur Bestimmung häufiger Itemsets.

Aufgabe 8-3 *Levenshtein Distanz*

Bestimmen Sie die Levenshtein Distanz der Sequenzen *BANANE* und *ANANAS*.