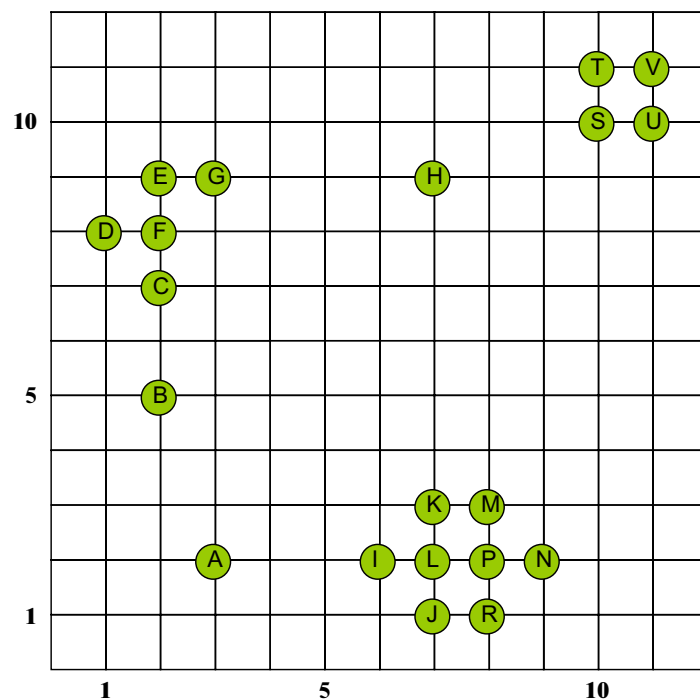


Knowledge Discovery in Databases
 SS 2011

Übungsblatt 9: Outlier Detection, Assoziationsregeln

Aufgabe 9-1 *Outlier Scores*

Gegeben sei der folgende 2-dimensionale Datensatz:



Verwenden Sie als Distanzfunktion zwischen den Punkten wieder die Manhattan-Distanz (L_1 -Norm)

Berechnen Sie (unter Ausschluss des Anfragepunktes bei der Berechnung der k NN):

- Den LOF-Wert für $k = 2$ für die Punkte H , L und B .
- Den LOF-Wert für $k = 4$ für die Punkte H , L und B .
- Die k NN-Distanz für $k = 2$ für alle Punkte.
- Die k NN-Distanz für $k = 4$ für alle Punkte.
- Die aggregierten k NN-Distanzen für $k = 2$ und $k = 4$ für alle Punkte.

Diskutieren Sie die Wahl von k für diesen Datensatz.

Aufgabe 9-2 *Apriori-Algorithmus*

Gegeben ist die Menge der Items $I = \{A, B, C, D, E, F, G, H, I, K, L, M\}$.

Weiterhin ist eine Menge von Transaktionen T laut folgender Tabelle gegeben:

Menge der Transaktionen T

Transaktions ID	gekaufte Items
1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

- (a) Bestimmen Sie zum minimalen Support von 30% die häufig auftretenden Itemsets. Verwenden Sie dazu den Apriori-Algorithmus. Geben Sie insbesondere die Kandidatenmengen nach den Join-Schritten und nach den Prune-Schritten an, sowie die häufig auftretenden Itemsets mit ihrem jeweiligen Support.
- (b) Bestimmen sie alle Assziationregeln, die man aus dem frequent Itemset $\{B E G H\}$ mit einer Konfidenz von mindestens 60% und dem Support 4 (Frequency 30%) bilden kann. Nutzen sie dabei die in der Vorlesung vorgestellte Monotonie aus.