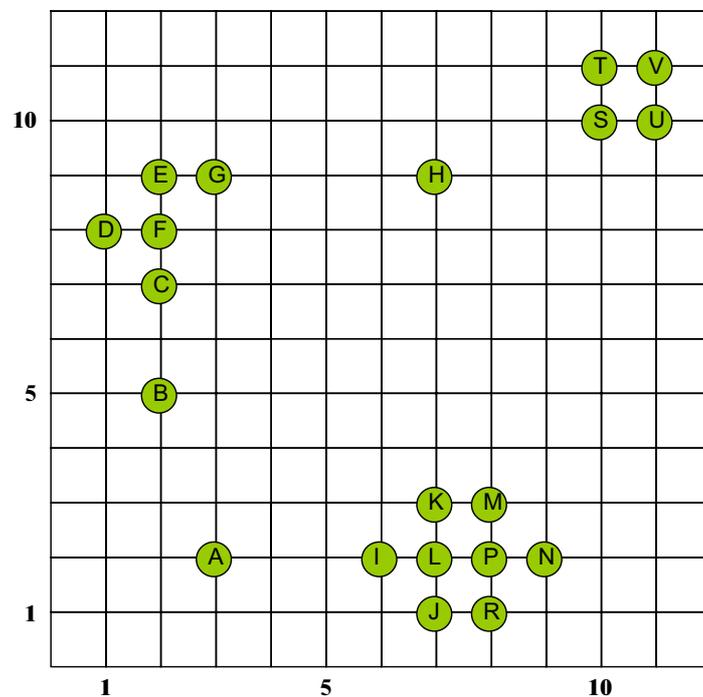


Knowledge Discovery in Databases  
 WS 2010/11

Übungsblatt 11: Outlier Detection

Aufgabe 11-1 *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.