

Knowledge Discovery in Databases  
SS 2013

Übungsblatt 3: Clusteranalyse – PAM, EM und DBSCAN

**Aufgabe 3-1 PAM**

Zeigen Sie, dass der Algorithmus PAM konvergiert.

**Aufgabe 3-2 Zuweisung im EM-Algorithmus**

Gegeben sei eine Datenmenge mit 100 Punkten, die drei Gausscluster  $A$ ,  $B$  und  $C$  und den Punkt  $p$  enthält.

Der Cluster  $A$  enthält 30% aller Punkte und ist repräsentiert durch den Mittelwert aller seiner Punkte

$$\mu_A = (2, 2) \text{ und die Kovarianzmatrix } \Sigma_A = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}.$$

Der Cluster  $B$  enthält 20% aller Punkte und ist repräsentiert durch den Mittelwert aller seiner Punkte

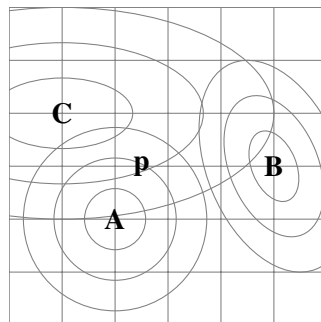
$$\mu_B = (5, 3) \text{ und die Kovarianzmatrix } \Sigma_B = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}.$$

Der Cluster  $C$  enthält 50% aller Punkte und ist repräsentiert durch den Mittelwert aller seiner Punkte

$$\mu_C = (1, 4) \text{ und die Kovarianzmatrix } \Sigma_C = \begin{pmatrix} 16 & 0 \\ 0 & 4 \end{pmatrix}.$$

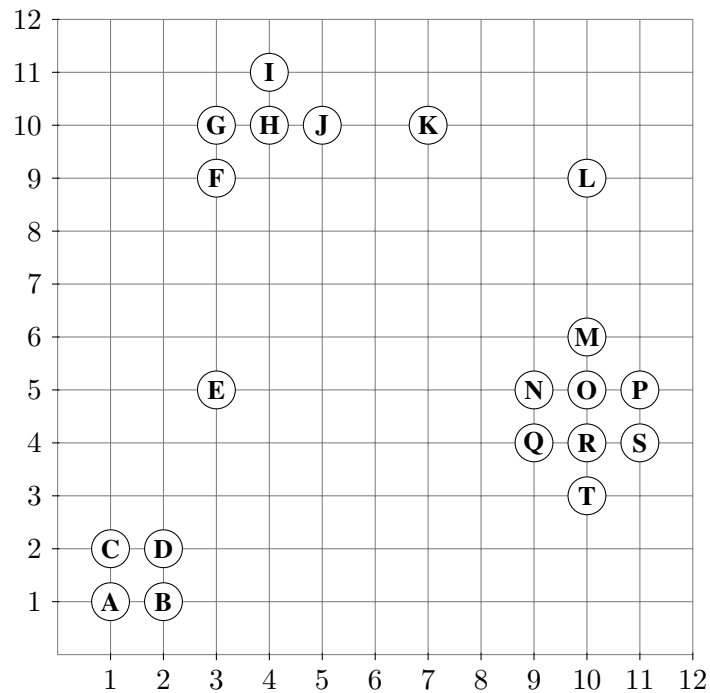
Der Punkt  $p$  ist durch die Koordinaten  $(2.5, 3.0)$  gegeben. Geben Sie die beiden Wahrscheinlichkeiten an, mit der  $p$  zu den Clustern  $A$ ,  $B$  bzw.  $C$  gehört.

Achtung: Folgende Skizze ist nur zu Veranschaulichungszwecken gedacht und nicht maßstabsgetreu!



### Aufgabe 3-3 DBSCAN

Gegeben sei folgender Datensatz:



Als Distanzfunktion verwenden Sie die (einfachere) Manhattan-Distanz:

$$L_1(x, y) = |x_1 - y_1| + |x_2 - y_2|$$

Führen Sie den Algorithmus DBSCAN auf dem Datensatz durch, und notieren Sie, welche Punkte Kernpunkte, Randpunkte und Noise sind.

Verwenden Sie folgende Parameterisierungen:

- Radius  $\varepsilon = 1.1$  and  $minPts = 2$
- Radius  $\varepsilon = 1.1$  and  $minPts = 3$
- Radius  $\varepsilon = 1.1$  and  $minPts = 4$
- Radius  $\varepsilon = 2.1$  and  $minPts = 4$
- Radius  $\varepsilon = 4.1$  and  $minPts = 5$
- Radius  $\varepsilon = 4.1$  and  $minPts = 4$

Bei  $minPts = 2$ , was passiert mit Randpunkten?

Sie können ihre Berechnungen / Implementierung mit ELKI verifizieren.