

Knowledge Discovery in Databases II
SoSe 2009

Übungsblatt 8: Multirepräsentiertes Clustering und Minimal Matching Distanz

Besprechung am 2.7.2009

Aufgabe 8-1 *Multirepräsentiertes Clustering*

Gegeben sei ein Datensatz X , so dass jeder Punkt durch 2 zweidimensionale Vektoren repräsentiert wird.

$$A = (0, 1); (3, 0)$$

$$B = (-1, -1); (2, 0)$$

$$C = (0, 0); (3, 1)$$

$$D = (0, -3); (-2, 2)$$

$$E = (2, 1); (-2, -3)$$

Wir wollen auf diesem Datensatz multirepräsentiertes Clustering mittels DBSCAN durchführen.

- (a) Wie unterscheidet sich multirepräsentiertes Clustering von gewöhnlichem Clustering? Welche besonderen Schwierigkeiten sind damit verbunden?
- (b) Es sei $MinPoints = 3$. Für welche Werte von ϵ_1, ϵ_2 sind die Objekte C und D Kernobjekte nach
 - der Vereinigungsmethode?
 - der Schnittmethode?

Aufgabe 8-2 *Kuhn-Munkres Algorithmus*

Gegeben sei folgende Kostenmatrix K . Führen Sie ein Hungarian Matching auf diesen Matrizen durch:

$$K_1 = \begin{bmatrix} 2 & 1 & 3 \\ 6 & 5 & 4 \\ 7 & 8 & 9 \end{bmatrix};$$

$$K_2 = \begin{bmatrix} 90 & 75 & 75 & 80 \\ 35 & 85 & 55 & 65 \\ 125 & 95 & 90 & 105 \\ 45 & 110 & 95 & 115 \end{bmatrix};$$

Aufgabe 8-3 *Hausdorff-Distanz*

Zeigen Sie, dass die Hausdorff-Distanz alle drei Eigenschaften einer Metrik erfüllt.