
Data Mining WS 2012/13

Institut für Neuroinformatik

Dr. F. Schwenker

3. Aufgabenblatt (Abgabe am 11.12.2012 zur Übung)

7. Aufgabe (5): (Austauschverfahren)

Bestimmen Sie eine Clusterung $\mathcal{C} = \{C_1, C_2\}$ (mit 2 Clustern) für die Punkte

$$X := \{x^1, \dots, x^6\} = \{(1, 0), (2, 0), (0, 1), (2, 2), (3, 3), (2, 3)\} \subset \mathbb{R}^2.$$

nach dem Austauschverfahren. Die Anfangscluster seien $C_1 = \{(2, 0), (3, 3), (2, 2)\}$ und $C_2 = X \setminus C_1$.

Bestimmen Sie die Werte $D_{\text{var}}(C_1, C_2)$ nach jedem Punktaustausch, achten sie darauf, dass die Cluster stets mindestens einen Datenpunkt enthalten, also $n_{\text{min}} = 1$

8. Aufgabe (5): (Austauschverfahren in R / matlab)

Implementieren sie den Austausch-Algorithmus in `matlab` oder `R` Berechnen sie eine Clustereung mit 3 Clustern C_1, C_2, C_3 für den Datensatz `80X` (diesen finden Sie auf der Webseite zur Veranstaltung).

Bestimmen Sie $D_{\text{var}}(C_1, C_2, C_3)$ nach jedem Datenpunkt (also auch wenn die Clusterzugehörigkeiten und damit natürlich auch $D_{\text{var}}(C_1, C_2, C_3)$ nicht verändert wurden) und plotten Sie den Verlauf von D_{var} bis zur Konvergenz (also t_{max} hinreichend groß setzen).

Initialisieren Sie die drei Cluster (mit möglichst gleicher Mächtigkeit) zufällig mit Datenpunkten (z.B. mit der `matlab`-Funktion `randperm`) und setzen Sie $n_{\text{min}} = 1$.

Drucken sie ihre Resultate in einem geeigneten Format aus und fügen sie ihre implementierte Lösung hinzu.