

Übungsblatt 17

Übung 17.1 Welche Cluster ergibt der K -Means Algorithmus mit $K = 2$ für die folgenden Testpunkte:

$$\begin{array}{lll} x_1 = (1, 2) & x_2 = (2, 2) & x_3 = (4, 2) \\ x_4 = (1, 1) & x_5 = (2, 1) & x_6 = (4, 1) \end{array}$$

- für die Startwerte $\mu_1 = x_2$ und $\mu_2 = x_5$?
- für die Startwerte $\mu_1 = x_2$ und $\mu_2 = x_3$?

Übung 17.2 Zerlegen Sie die Punkte der vorherigen Aufgabe in $K = 2$ Cluster

- mit dem Single-Link hierarchischen Algorithmus.
- mit dem Complete-Link hierarchischen Algorithmus.

Welches praktische Problem bei der Anwendung der Algorithmen tritt in diesem Fall auf?

Übung 17.3 Das Ground-Truth Clustering G und zwei Test-Clusterings C_1, C_2 von sechs Messpunkten sei:

$$\begin{aligned} G &= \{1, 2\}, \{3, 4, 5\}, \{6\} \\ C_1 &= \{1, 2\}, \{3, 4\}, \{5, 6\} \\ C_2 &= \{1, 2\}, \{3, 4\}, \{5\}, \{6\} \end{aligned}$$

- Berechnen Sie den *Rand-Index* und den *Jaccard-Koeffizienten*
- Berechnen Sie das *F-Measure*

als Kennzahlen für den Abstand der Test-Clusterings vom Ground-Truth Clustering.

Übung 17.4 Zeigen Sie, dass der Jaccard-Koeffizient J immer kleiner oder gleich dem Rand-Index R ist, d.h.

$$J = \frac{a}{a + b + c} \leq \frac{a + d}{a + b + c + d} = R$$