

Maschinelles Lernen

Aufgabenblatt 08

Prof. Dr. Christoph Böhm
Hochschule München

6. Juni 2024

Aufgabe 8.1 (2-Means Clustering). Gegeben seien die Datenpunkte in Tabelle 1.

Datenpunkt	$\mathbf{x}_1^{(i)}$	$\mathbf{x}_2^{(i)}$
$\mathbf{x}^{(1)}$	1	4
$\mathbf{x}^{(2)}$	1	3
$\mathbf{x}^{(3)}$	0	4
$\mathbf{x}^{(4)}$	5	1
$\mathbf{x}^{(5)}$	6	2
$\mathbf{x}^{(6)}$	4	0

Tabelle 1: Daten für das 2-Means Clustering.

1. Zeichnen Sie die Datenpunkte in ein Koordinatensystem ein.
2. Ordnen Sie den Punkten die initialen Cluster 0, 1, 1, 1, 0 und 0 zu.
3. Berechnen Sie die Cluster-Zentren.
4. Berechnen Sie die Abstände der Datenpunkte von den beiden Zentren und ordnen Sie die Punkte den entsprechenden Clustern zu. Verwenden Sie als Abstandsmaß die euklidische Distanz (Sie dürfen auch deren Quadrat verwenden).
5. Wiederholen Sie die Schritte bis die Zuordnung konvergiert ist.
6. Zeichnen Sie alle Datenpunkte mit ihrer Cluster-Zugehörigkeit und die beiden Cluster-Zentren wiederum in ein Koordinatensystem ein.

Aufgabe 8.2 (Hierarchisches Clustering). Gegeben sei die Unähnlichkeitsmatrix

$$d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \begin{bmatrix} & 0.3 & 0.4 & 0.7 \\ 0.3 & & 0.5 & 0.8 \\ 0.4 & 0.5 & & 0.45 \\ 0.7 & 0.8 & 0.45 & \end{bmatrix}.$$

1. Führen Sie eine hierarchische Clusteranalyse anhand von $d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ mit *complete linkage* durch und zeichnen Sie das entstehende Dendrogramm.
2. Führen Sie eine hierarchische Clusteranalyse anhand von $d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ mit *single linkage* durch und zeichnen Sie das entstehende Dendrogramm.

Aufgabe 8.3 (PCA und K-Means Praxisbeispiel). In dieser Aufgabe führen Sie zunächst eine Hauptkomponentenanalyse auf den Iris-Daten durch, um schließlich ein K-Means Clustering durchzuführen. Die Hoffnung ist, auch ohne die Labels, die einzelnen Iris Spezies zu erkennen.

1. Laden Sie den Iris Datensatz und führen mit Hilfe von `sklearn.decomposition.PCA` und `PCA.fit` eine PCA mit zwei Hauptkomponenten auf die Iris-Features durch.
2. Geben Sie die Hauptkomponenten mit Hilfe von `PCA.components_` und den Anteil der erklärten Varianz mit Hilfe von `PCA.explained_variance_ratio_` aus. Interpretieren Sie die Ergebnisse.
3. Transformieren Sie die Iris Features mit Hilfe von `PCA.transform` und erstellen Sie einen Scatterplot des Ergebnisses. Wie viele Cluster machen Sie per Auge aus?
4. Clustern Sie die transformierten Features mit Hilfe von `sklearn.cluster.KMeans` mit drei Clustern. Erstellen Sie zwei Scatterplots der transformierten Features einmal in den Farben der gefundenen Cluster `KMeans.predict` und einmal der tatsächlichen Spezies aus dem Iris Datensatz. Interpretieren Sie das Ergebnis.