

Clusteranalyse

Bei der Clusteranalyse (Clustering) wird eine Gruppe von Objekten in Untergruppen aufgeteilt, die als Cluster bezeichnet werden [1]. Objekte, die ähnliche Eigenschaften besitzen, werden demselben Cluster zugeordnet. Die Objekte innerhalb eines Clusters sollen sich möglichst stark von den Objekten anderer Cluster unterscheiden. [2].

k-Means- und k-Medoid-Algorithmus

Beim k-Means-Algorithmus handelt es sich um einen beliebten und einfachen Clustering-Algorithmus [3], der praktisch häufig zur Anwendung kommt [4]. Er gehört in die Kategorie der partitionierenden Verfahren. Diese sind dadurch gekennzeichnet, dass bei der Zerlegung einer Gruppe von Objekten in k Cluster keines der Cluster leer sein darf. Zusätzlich wird jedes Objekt genau einem Cluster zugeteilt. [2].

Jedes Cluster wird durch einen Prototyp repräsentiert. Der Prototyp ist anschaulich betrachtet der Mittelpunkt eines Clusters. Ein Objekt gehört demjenigen Cluster an, zu dessen Mittelpunkt es die kleinste Distanz aufweist. [5]. Man bezeichnet den Mittelpunkt eines Clusters als Centroid [2].

Abb.1 zeigt eine grafische Darstellung des k-Means-Algorithmus. Als Eingabe wird ein ganzzahliger Wert k benötigt, der die Anzahl der zu bildenden Cluster anzeigt [4, 5]. Der Algorithmus kann wie folgt beschrieben werden:

- 1) Wähle zufällig k Objekte als potenzielle Centroide aus.
- 2) Ordne jedes Objekt dem Centroiden zu, zu dem es den kleinsten Abstand hat, sodass Cluster entstehen.
- 3) Berechne für jedes Cluster den Centroid neu.
- 4) Wiederhole die Schritte 2 und 3 bis sich die Centroide nicht mehr ändern. [5].

Auf dem Arbeitsblatt beschäftigen sich die Lernenden anhand einer Variante des k-Means-Algorithmus mit dem unüberwachten Lernen. Er zeichnet sich dadurch aus, dass die Clusterprototypen keine Centroide, sondern Medoide sind. Im Gegensatz zu Centroiden sind Medoide immer Objekte aus der einzuteilenden Gruppe. Medoide liegen in einem Cluster am zentralsten. Jedes Objekt wird dem Cluster zugeordnet, zu dessen Medoid es den kleinsten Abstand hat. [2]. Die Abfolge der Schritte ist bei k-Means und k-Medoid identisch [1]. Eine Darstellung der Schritte beim k-Medoid-Algorithmus liefert Abb.2. Welche Cluster gebildet werden, hängt von den anfangs ausgewählten Centroiden bzw. Medoiden ab. Es gibt keine Garantie dafür, dass der Algorithmus die optimalste Aufteilung bestimmt. [2].

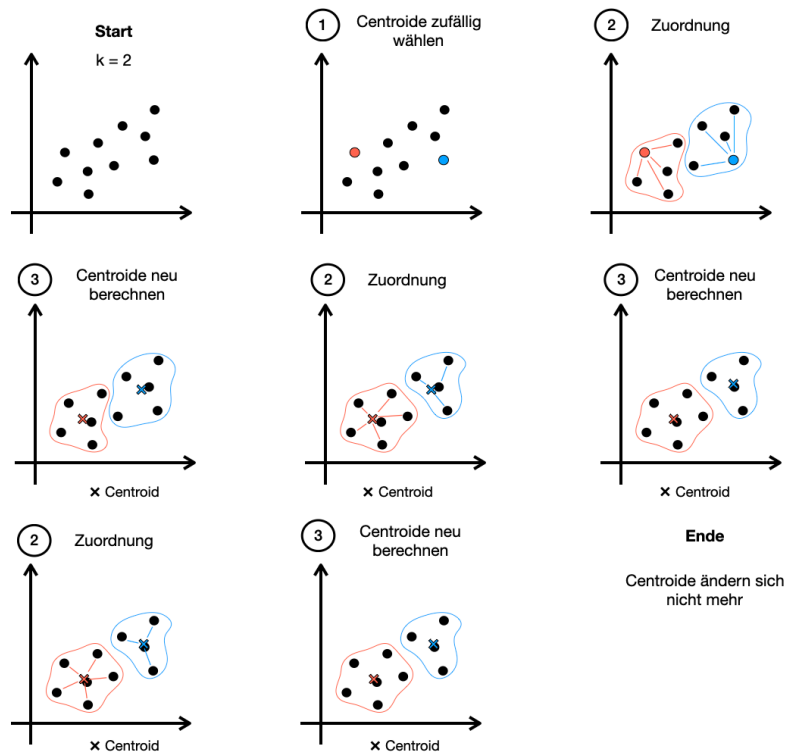


Abbildung 1: Grafische Darstellung des k-means-Algorithmus. Abbildung in Anlehnung an [2], S.53.

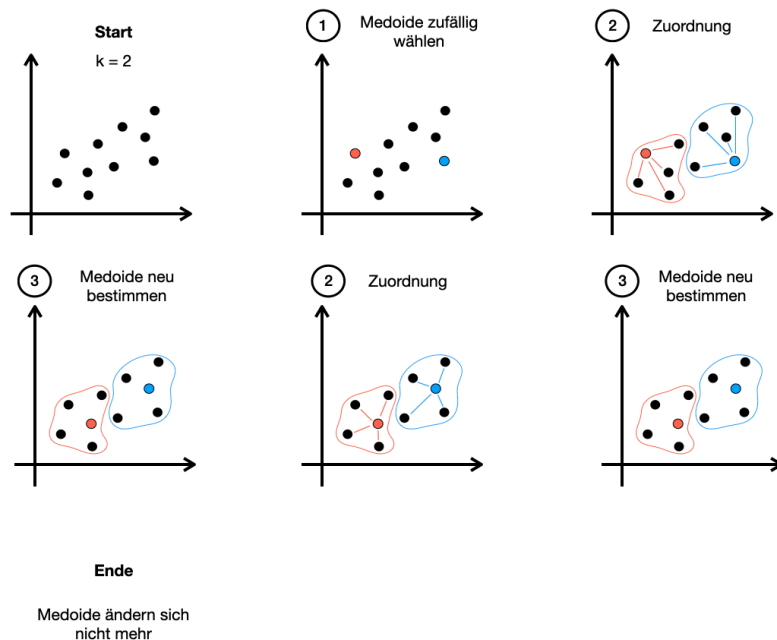


Abbildung 2: Grafische Darstellung des k-medoid-Algorithmus. Abbildung in Anlehnung an [2], S.53.

Literaturquellen

- [1] T. Jo, Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning, Cham: Springer, 2021.
- [2] M. Ester und J. Sander, Knowledge Discovery in Databases: Techniken und Anwendungen, Berlin und Heidelberg: Springer, 2000.
- [3] H. Aust, Das Zeitalter der Daten: Was Sie über Grundlagen, Algorithmen und Anwendungen wissen sollten, Berlin und Heidelberg: Springer, 2021.
- [4] L. Wuttke, Praxisleitfaden für Künstliche Intelligenz in Marketing und Vertrieb: Beispiele, Konzepte und Anwendungsfälle, Wiesbaden: Springer Gabler, 2021.
- [5] D. Müller, „k-Means-Algorithmus,“ in *Wie Maschinen lernen: Künstliche Intelligenz verständlich erklärt*, Wiesbaden, Springer, 2019, pp. 81-88.