

Unüberwachtes Lernen

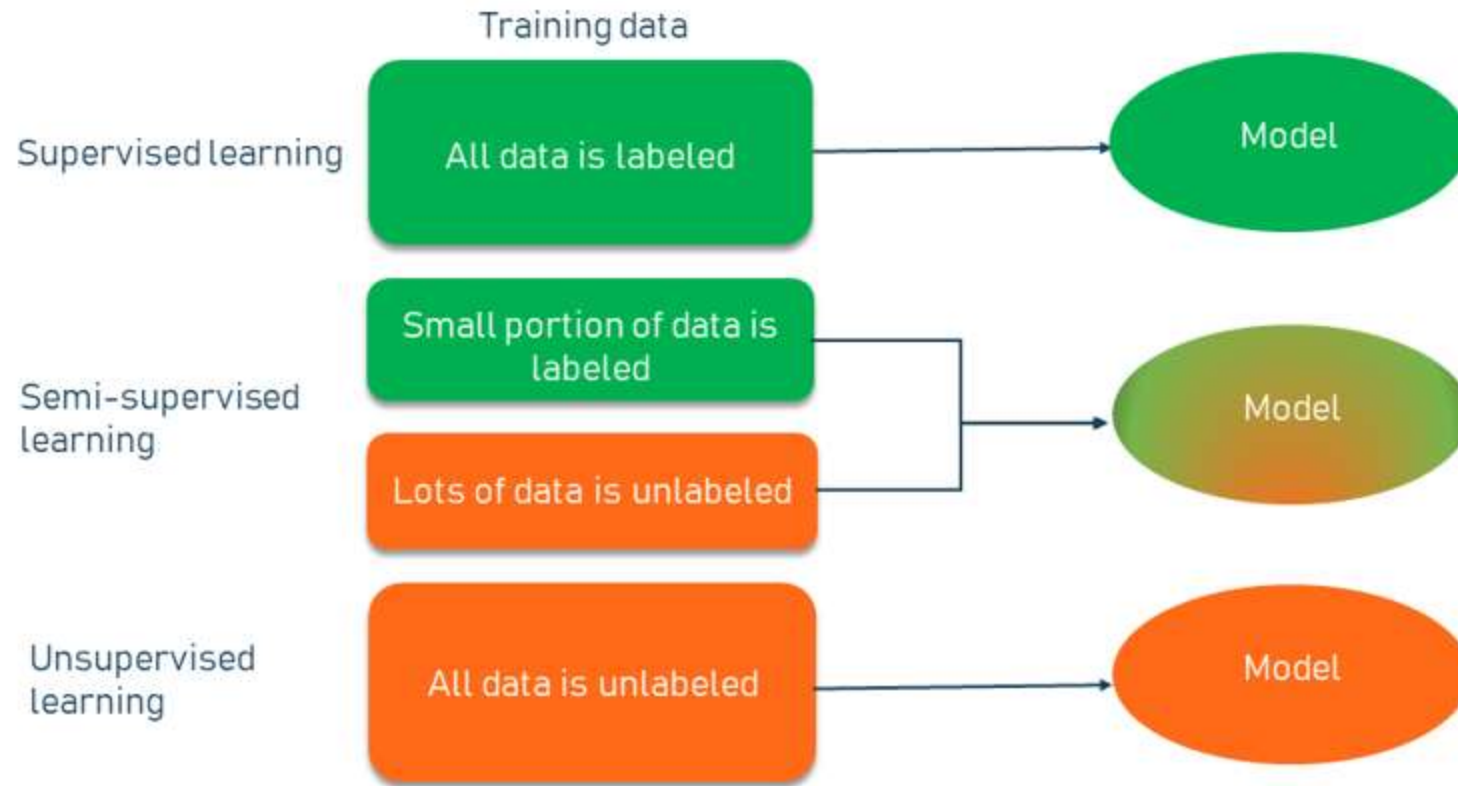
K-Means Algorithmus

Cihat Özkurt 11148632

WPF - DLML

Prof. Dr. Wolfgang Konen
Prof. Dr. Daniel Gaida

Die Unterschiede zwischen Überwachtes und Unüberwachtes Lernen



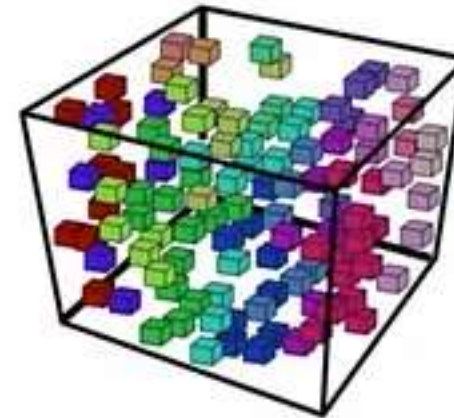
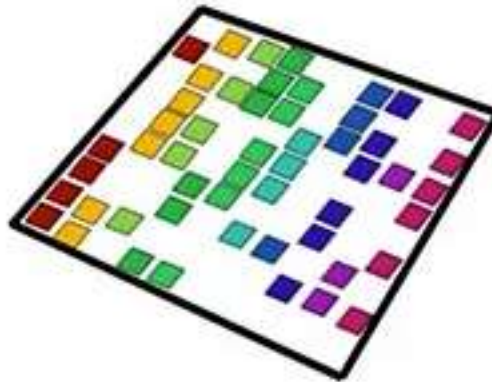
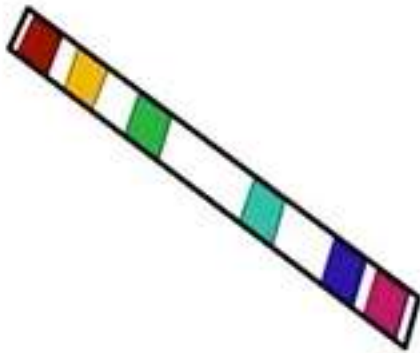
Die Unterscheiden

	Überwachtes Lernen	Unüberwachtes Lernen
Input Data	Gelabelt	nicht gelabelt
Lernziel	Klassifizierung und Regression	Zusammenhänge, Pattern
Rechenkomplexität	Niedrig	Hoch
Leistung	Genauer	Weniger genauer
Feedback	Hat explizites Feedback	kein explizites Feedback

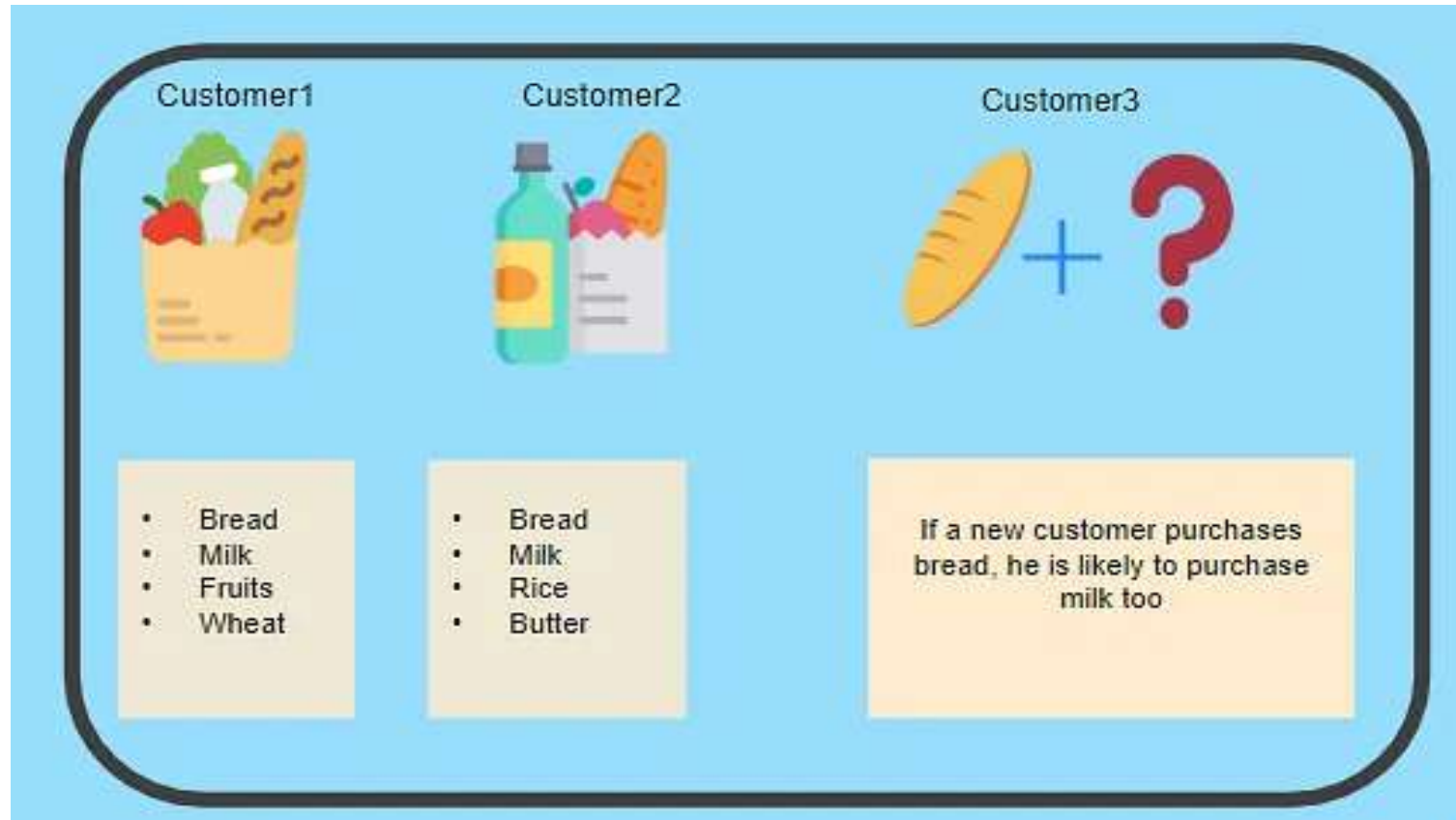
Aufgaben des unüberwachten Lernens

1. Dimensionreduktion

1 dimension: 10 positions 2 dimensions: 100 positions 3 dimensions: 1000 positions

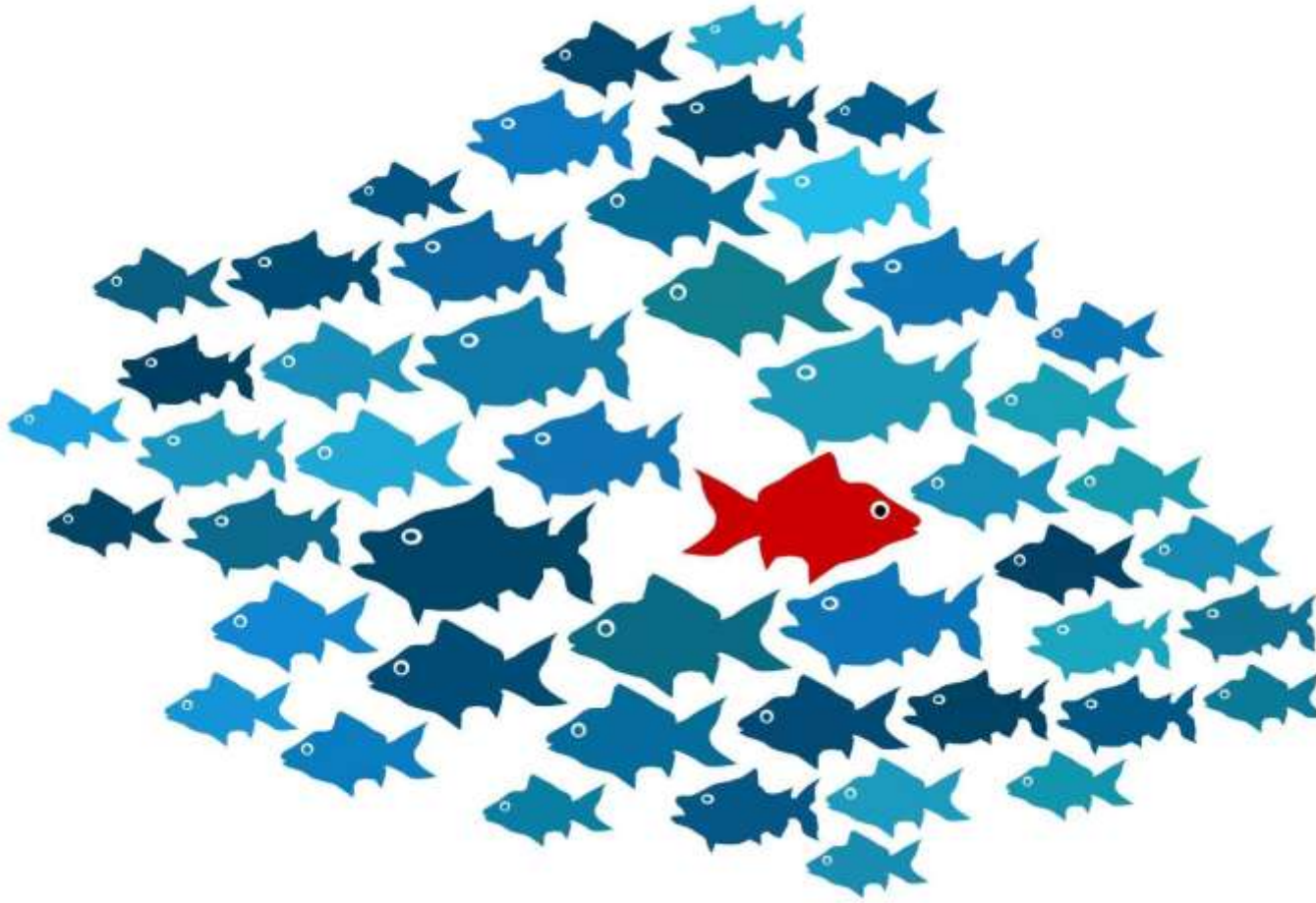


2. Assoziationen



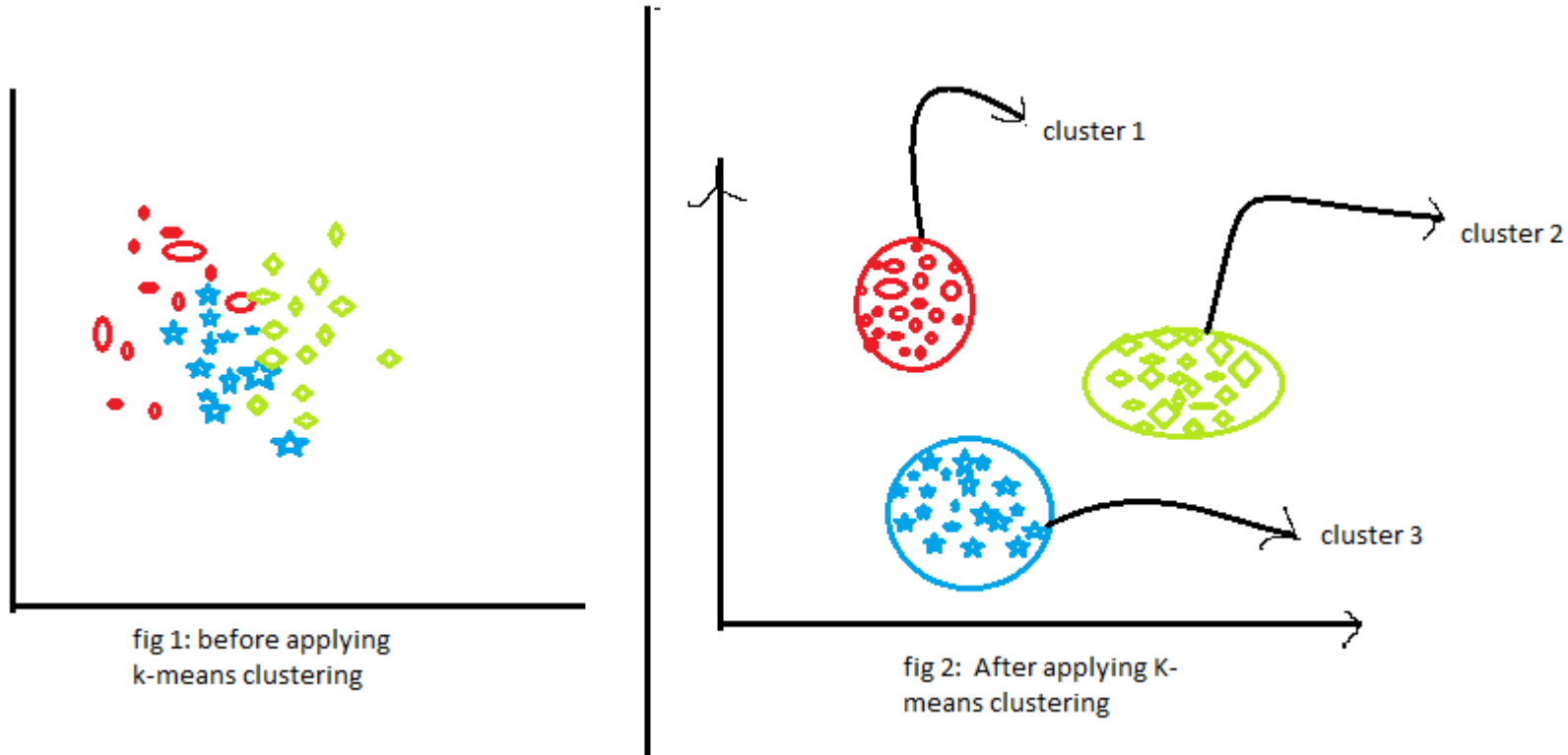
- Youtube, Amazon Recommendation System

3. Anomalieerkennung



- Erkennung von Betrug, Fehlerdiagnose

4. Clustering (Gruppierung)

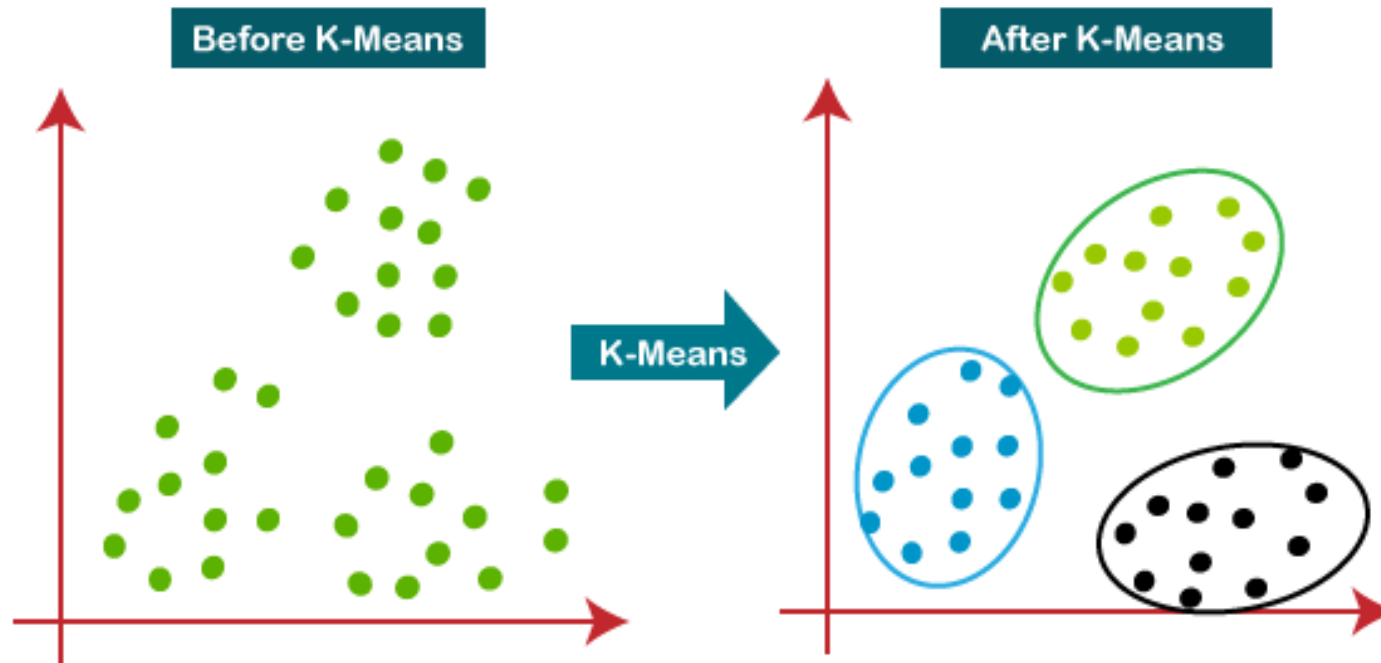


Customer Segmentation, Bild- und Objekterkennung

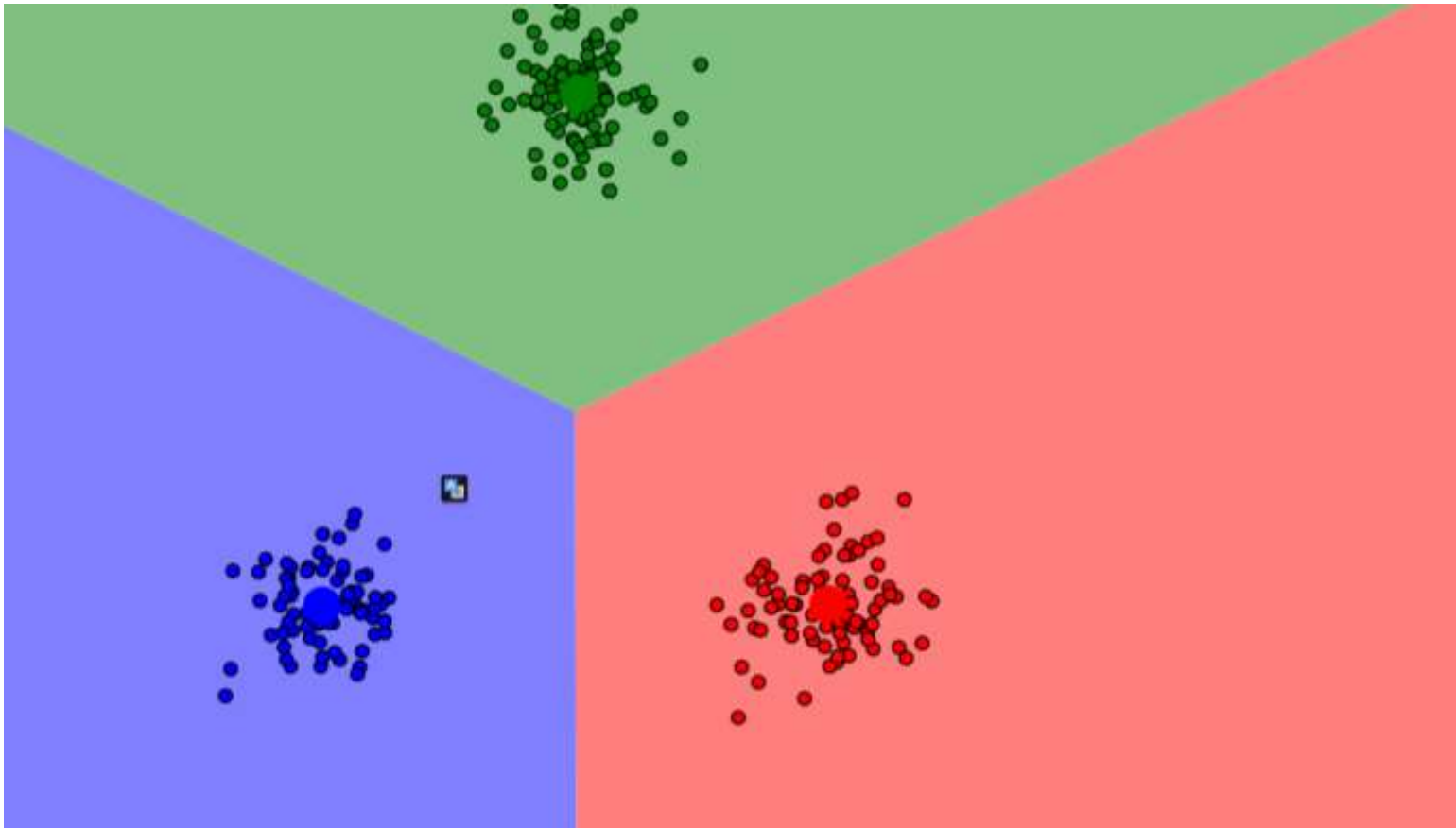
Clustering Algorithmen

- K-Means
- Db-Scan
- Hierarchische Clusteranalyse

K-Means Algorithmus



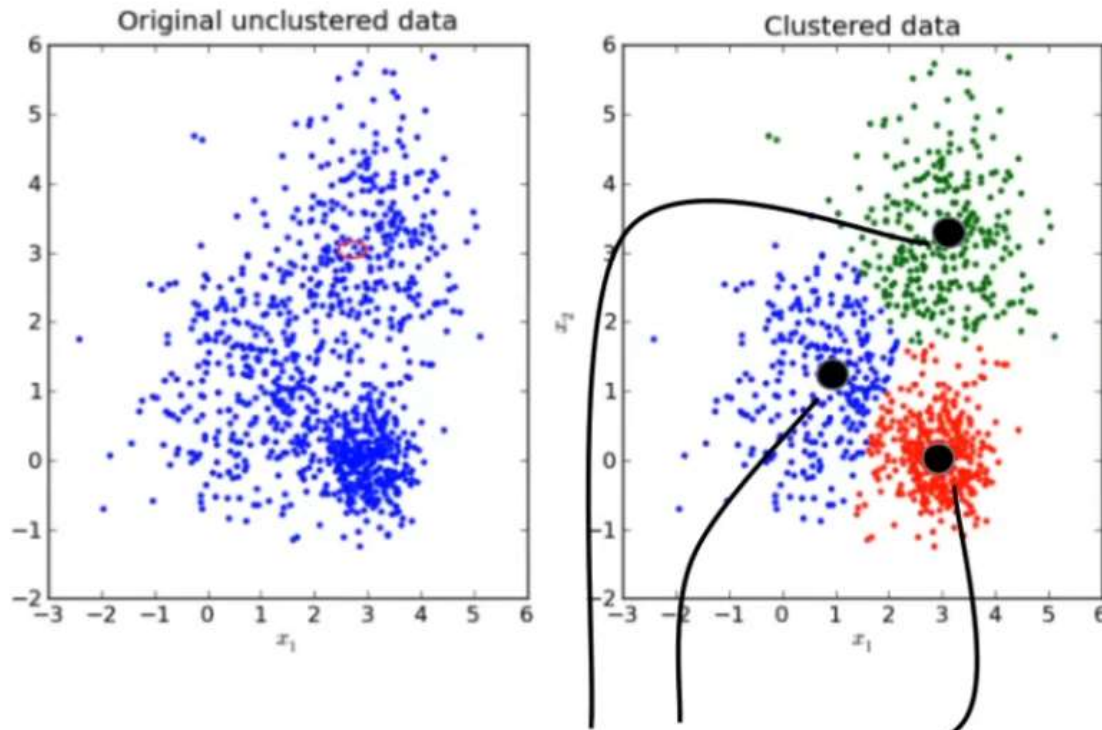
K-Means Algorithmus



- Jeder Cluster besteht in sich aus ähnlichen Daten, aber die Cluster sind einander nicht ähnlich

<https://www.naftaliharris.com/blog/visualizing-k-means-clustering/>

K-Means Algorithmus

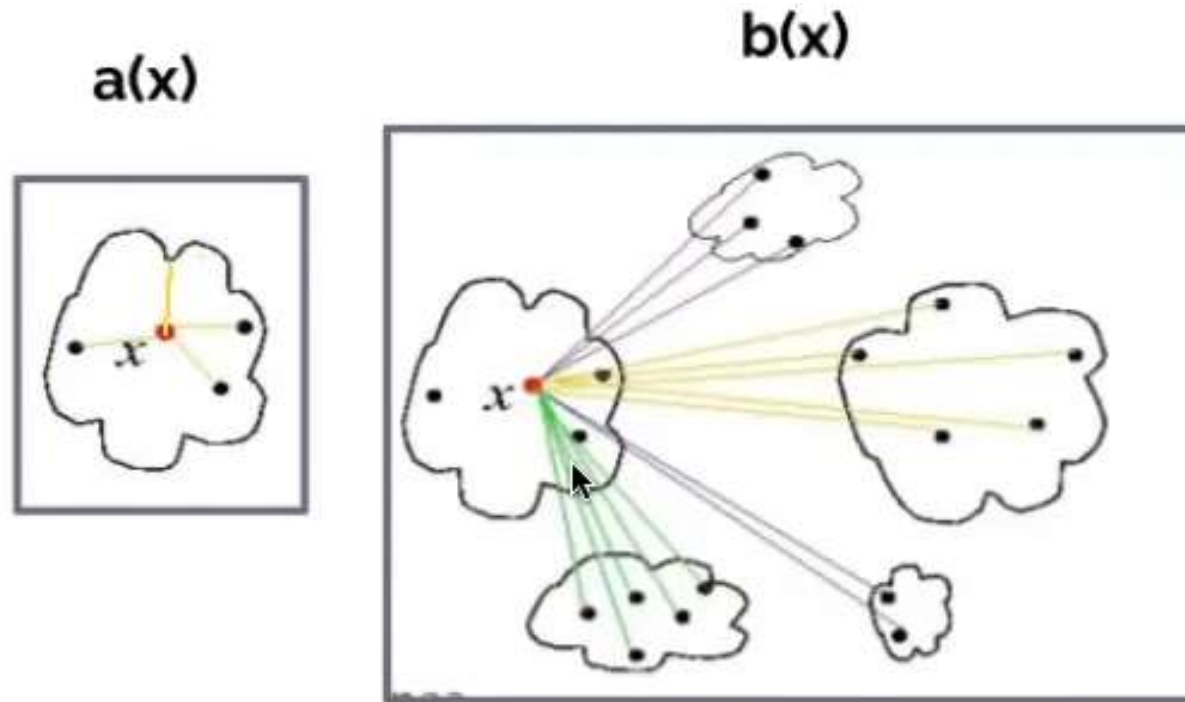


Centroids

- K-means ist ein iterativer Algorithmus, der sich letztendlich einer Lösung nähert.
- Das Ziel ist es, Datenpunkte innerhalb jedes Clusters ähnlich und untereinander verschiedene Cluster zu machen.

Silhouette Score als Fehlermetrik

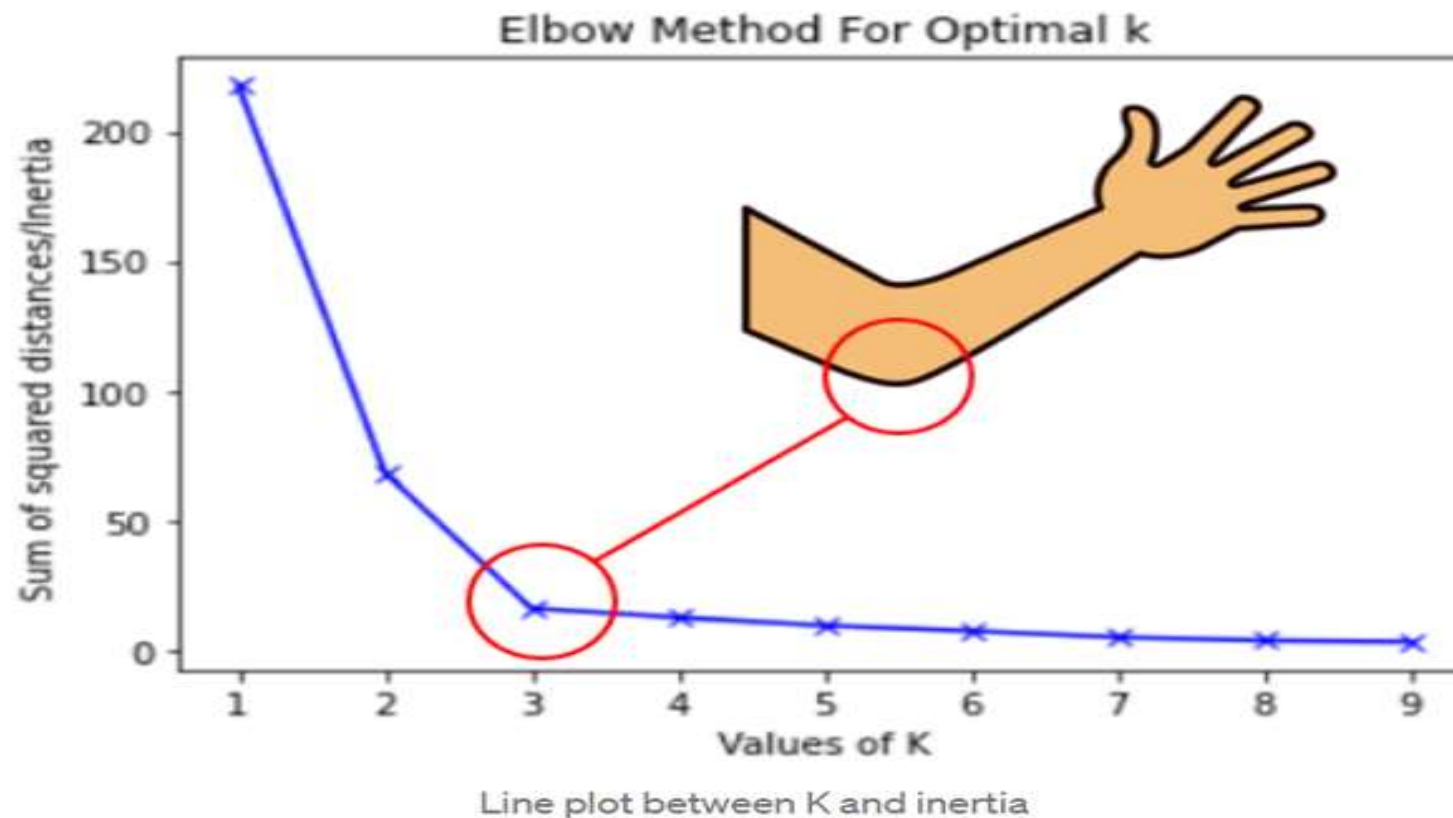
- Der Silhouettenkoeffizient liegt zwischen -1 und 1



$$s = \frac{b - a}{\max(a, b)}$$

Optimal Nummer der Clusters Finden

Elbow Method



Inertia Formula

$$\sum_{i=1}^N (x_i - C_k)^2$$

C = centroids

X = einen Datenpunkt im Datensatz

Quellenverweise:

1. <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>
2. <https://av-eks-blogoptimized.s3.amazonaws.com/62725cluster0>
3. <https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning>
4. <https://www.simplilearn.com/tutorials/machine-learning-tutorial/supervised-and-unsupervised-learning>
5. <https://www.altexsoft.com/blog/semi-supervised-learning/>
6. <https://www.analyticsvidhya.com/blog/2020/10/a-simple-explanation-of-k-means-clustering/>
7. <https://www.kaggle.com/>
8. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>