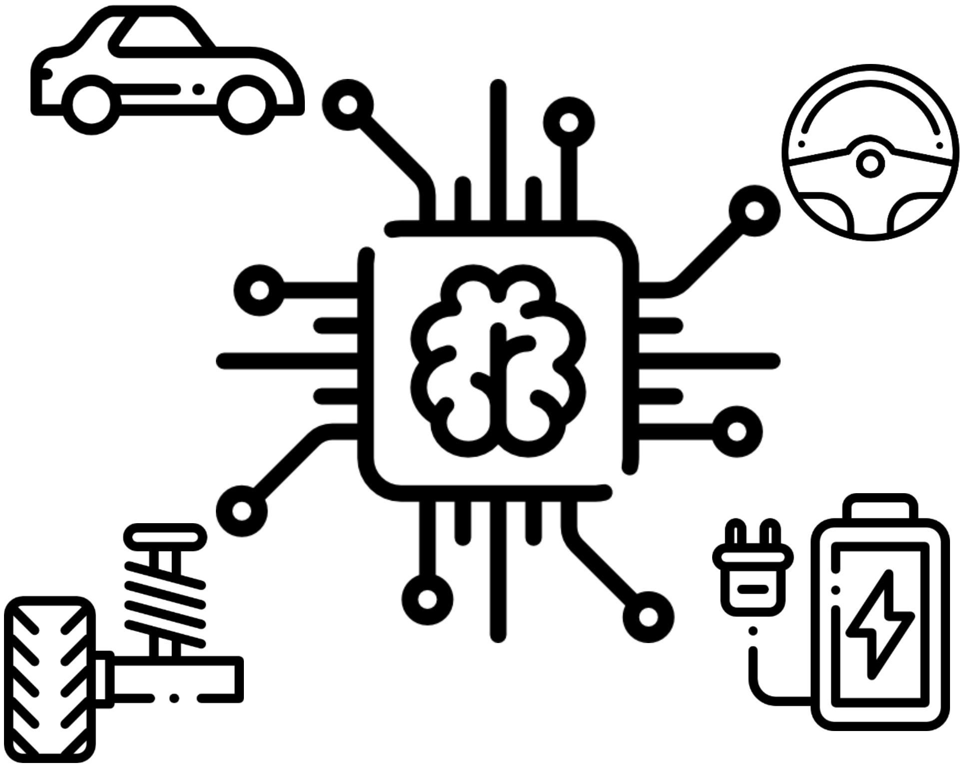
Institute of Automotive Technology
TUM School of Engineering and Design
Technical University of Munich

Artificial Intelligence in Automotive Technology

Maximilian Geißlinger / Fabian Netzler

Prof. Dr.-Ing. Markus Lienkamp



Lecture Overview

Lecture 16:15-17:45 | Practice 17:45-18:30	
1 Introduction: Artificial Intelligence	20.10.2022 – Maximilian Geißlinger
2 Perception	27.10.2022 – Sebastian Huber
3 Supervised Learning: Regression	03.11.2022 – Fabian Netzler
4 Supervised Learning: Classification	10.11.2022 – Andreas Schimpe
5 Unsupervised Learning: Clustering	17.11.2022 – Andreas Schimpe
6 Introduction: Artificial Neural Networks	24.11.2022 – Lennart Adenaw
7 Deep Neural Networks	08.12.2022 – Domagoj Majstorovic
8 Convolutional Neural Networks	15.12.2022 – Domagoj Majstorovic
9 Knowledge Graphs	12.01.2023 – Fabian Netzler
10 Recurrent Neural Networks	19.01.2023 – Matthias Rowold
11 Reinforcement Learning	26.01.2023 – Levent Ögretmen
12 AI-Development	02.02.2023 – Maximilian Geißlinger
13 Guest Lecture	09.02.2023 – to be announced

Objectives for Lecture 5: Clustering

After the lecture you are able to…

… understand the concept of clustering and its association with pattern recognition.

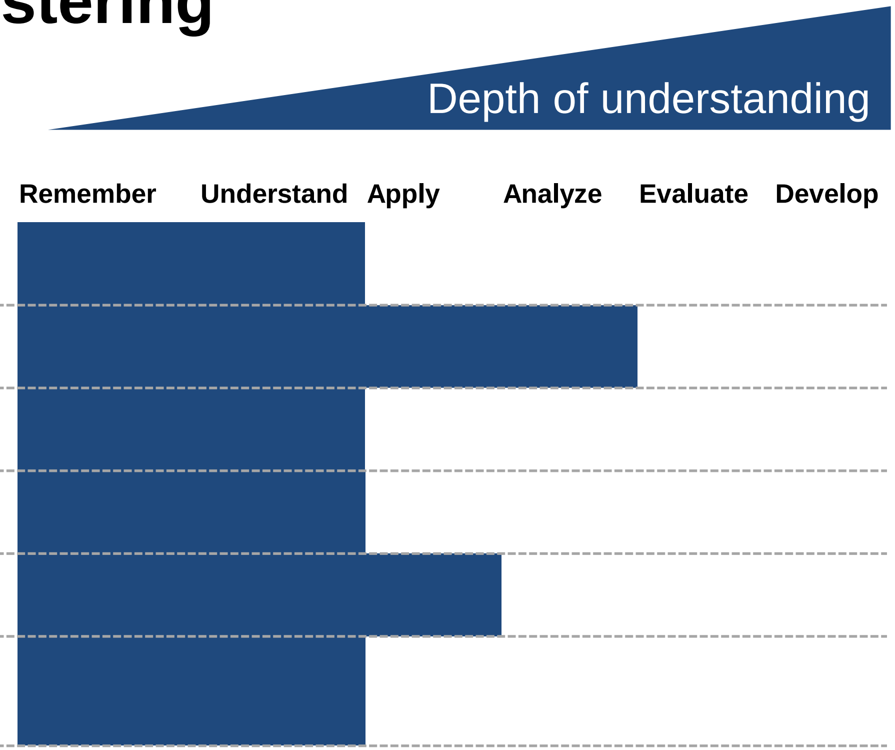
… analyze the quality of given clusters regarding to different criteria.

… understand the workflow of unsupervised learning.

… understand the concepts of different clustering methods together with their pro and cons.

… implement, train and use a clustering method with python libraries.

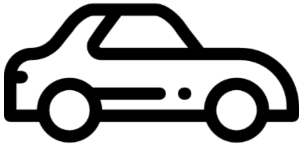
… identify if a problem belongs to regression, classification or clustering.

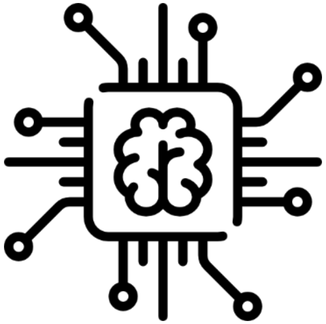


Unsupervised Learning: Clustering
Prof. Dr.-Ing. Markus Lienkamp

(Andreas Schimpe, M.Sc.)

Agenda





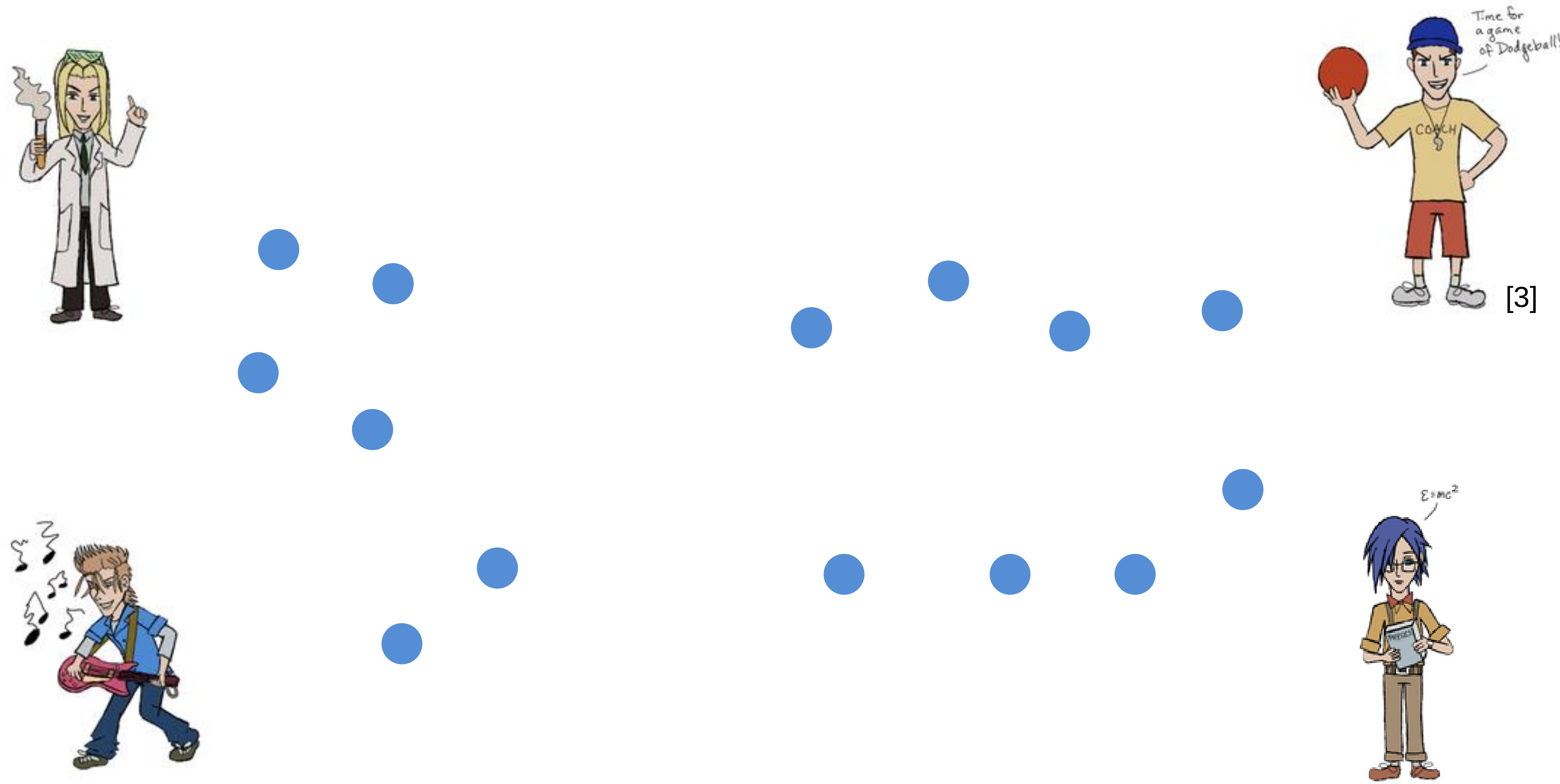
Clustering

“Grouping of similar things that are close together, sometimes surrounding something” [2]

[1]

Clustering

“Grouping of similar things that are close together, sometimes surrounding something” [2]



[3]

Additional Slide

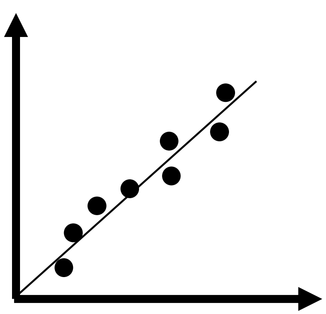
Clustering, independent of machine learning, deals with identifying similarities and differences between different data points and dividing the data into groups. In contrast to classification, where a model is learned and later applied to other data, clustering usually starts with all data present (without label) and then divides the data into different groups. It is not intended that later new data points will be added to the clusters (although this would not be a problem). The result of clustering is not a model, but a division of the data. In many cases, this is a pre-processing step, and helps with handling large data. If, for example, a lidar point cloud is created by an automated vehicle, the effort to examine each point is far too great. A first step is to cluster the points and identify which points belong to the same object. So one can continue working with a few objects instead of thousands of points.

Method Overview

Pattern Recognition

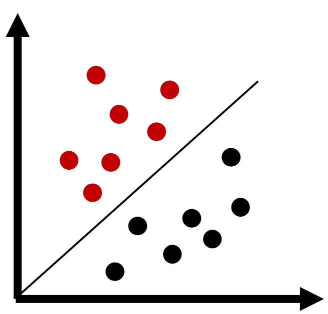
Regression

- Predict **continuous** valued output
- Supervised



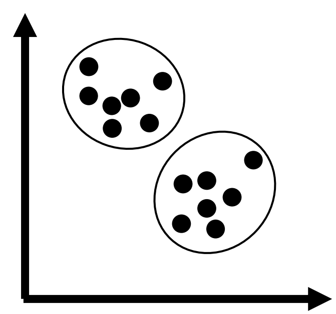
Classification

- Predict **discrete** valued output
- Supervised



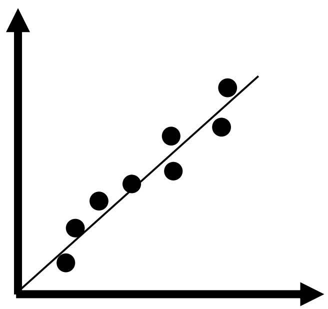
Clustering

- Predict discrete valued output
- **Unsupervised**



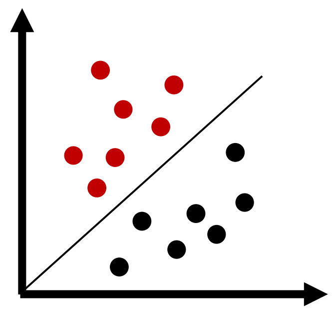
Method Overview

Regression



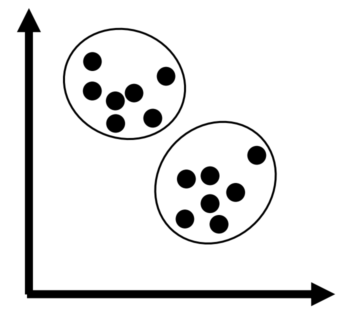
- House pricing
- Number of sales
- Persons weight

Classification



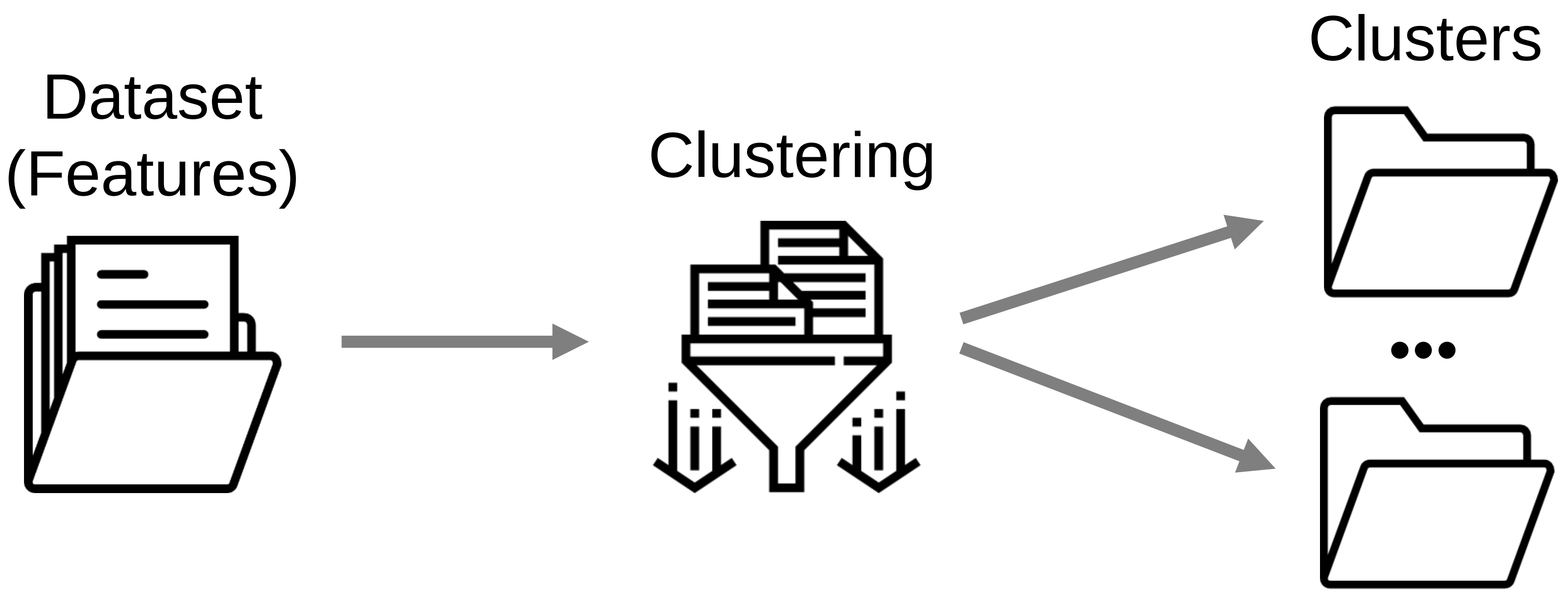
- Object detection
- Spam detection
- Cancer detection

Clustering



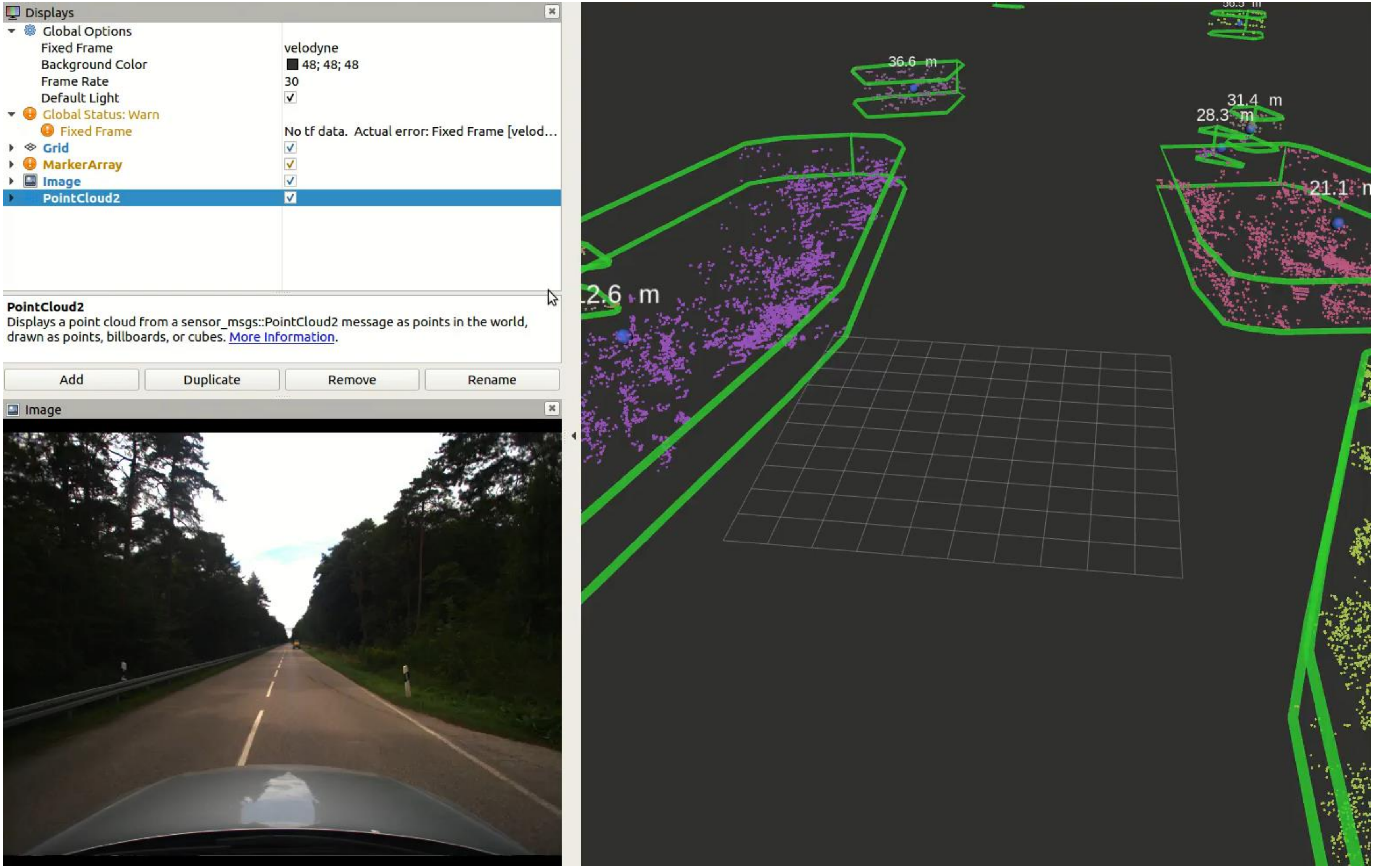
- Genome patterns
- Google news
- Point-cloud (Lidar) processing

General Approach



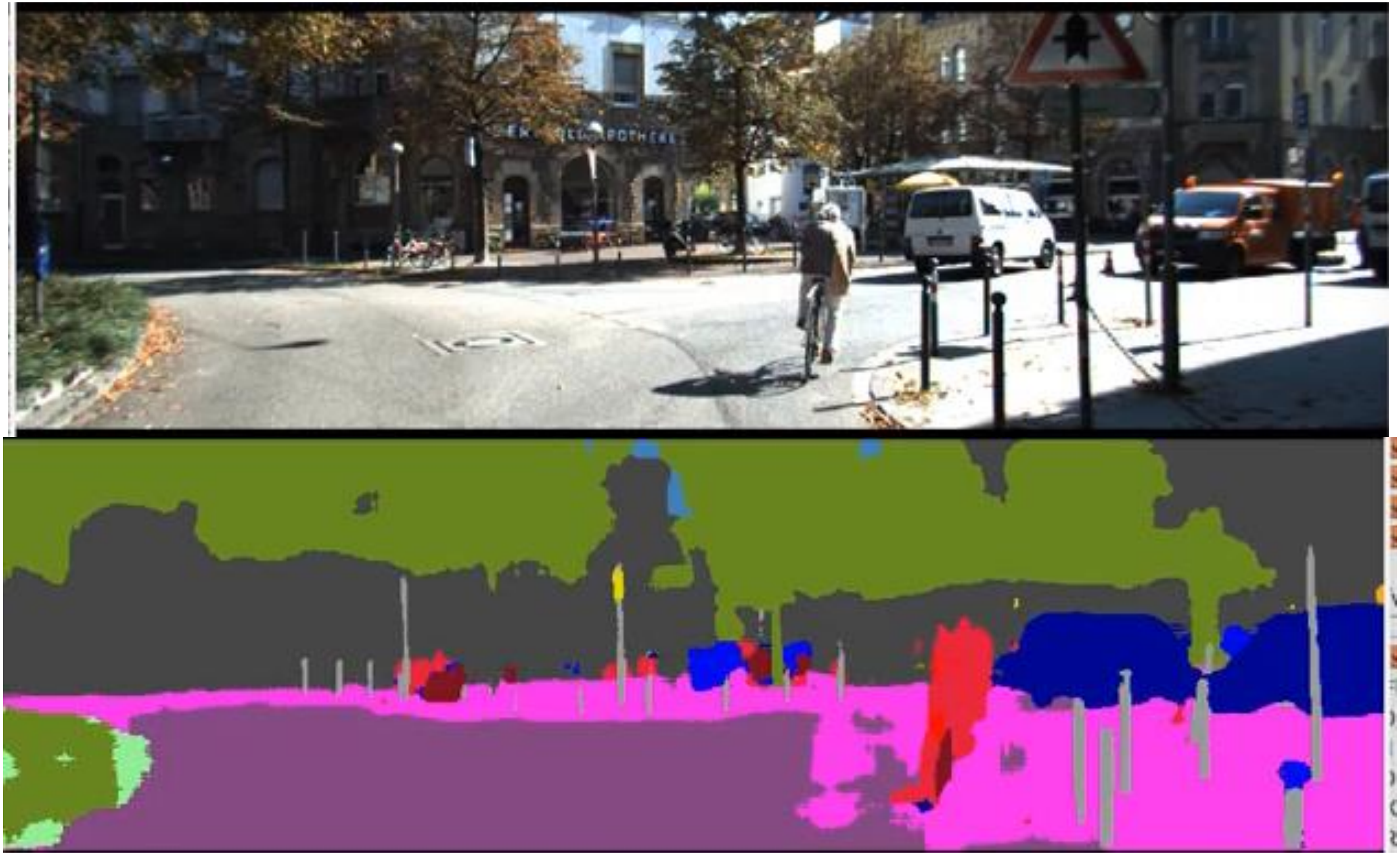
News (Keywords, …)	Similarities?	Clusters
Genomes (Pattern, …)	Similarities?	Clusters
Points (Position, …)	Similarities?	Clusters

Point-Cloud Clustering



(Autoware: Euclidean_Cluster_Detect)

Camera Image Segmentation



[4]

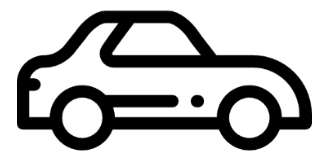
Clustering vs. Segmentation

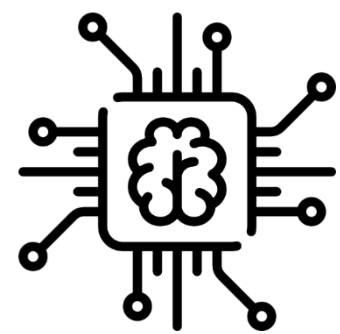
- Both terms are interchangable
- Statistical background: Clustering
- Business background: Segment
- Clustering produces segments and vice versa

Unsupervised Learning: Clustering
Prof. Dr.-Ing. Markus Lienkamp

(Andreas Schimpe, M.Sc.)

Agenda





Formal Definition – Clustering

- Elements $e \in E$
- Cluster $\mathrm{c} \in C$ with $c \subseteq E$
  - Union $\bigcup_{c \in C} = E$
  - Intersection $\bigcap_{c \in C} = \emptyset$

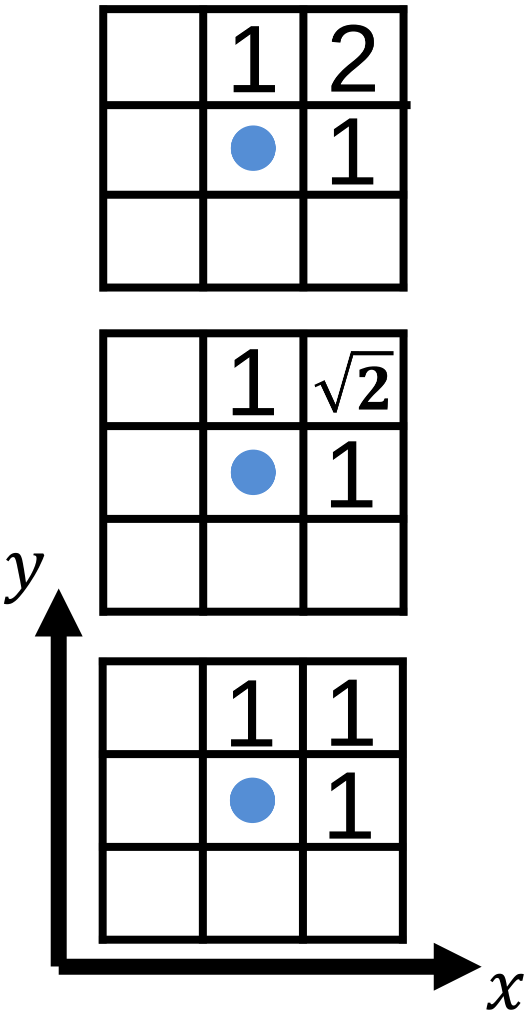
- Representative $r_c = \mathrm{mean}(c)$

- $variability(c) = \sum_{e \in c} \mathrm{distance}(r_c, e)^2$

- **Clustering Objective**

$$\text{minimize } \sum_{c \in C} variability(c)$$

Formal Definition - Distance



Manhattan

$$|x_1 - x_2| + |y_1 - y_2|$$

Euclidian

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Chebyshev

$$\max(|x_1 - x_2|, |y_1 - y_2|)$$

Additional Slide

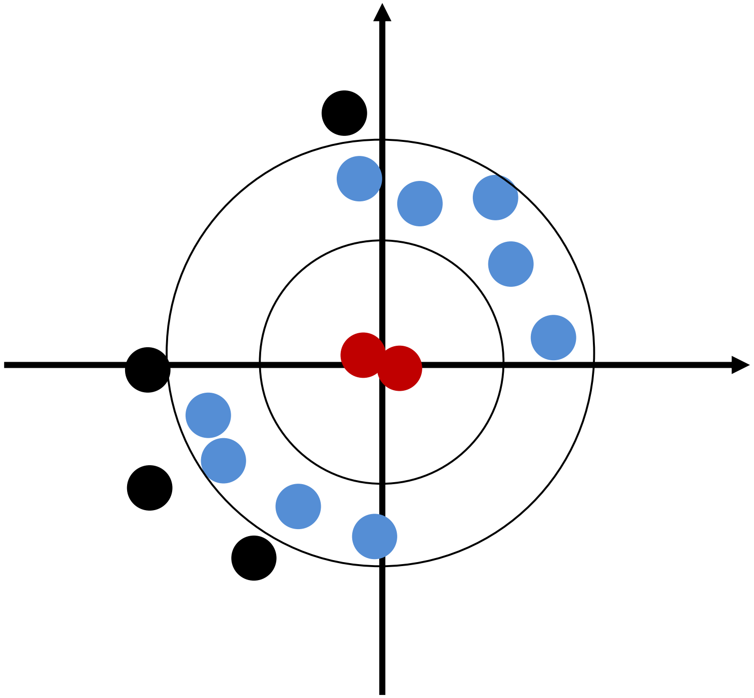
The distance is the most important measure in clustering, but it should be noted that not only the spatial distance is meant, but the distance in any numerically representable property (color, size, weight ...).

In everyday life mostly the Euclidean distance is meant, when it comes to measuring how far two points are away from one another. But there are several other possibilities as well.

Recap Classification

- Classification
  - Labeled training data (supervised)
  - Given classes

- Example: Dart
  - Shooting a target
  - 3 classes for points

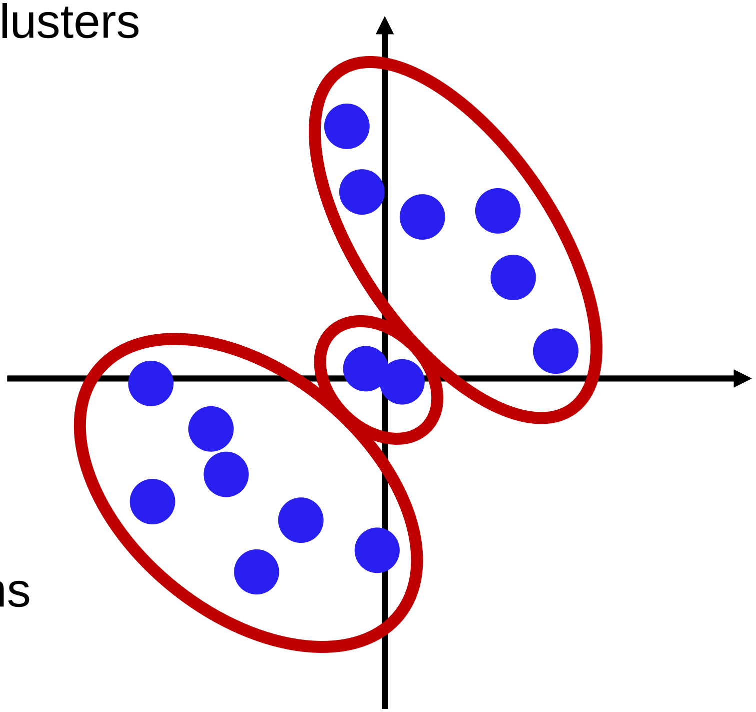


Clustering

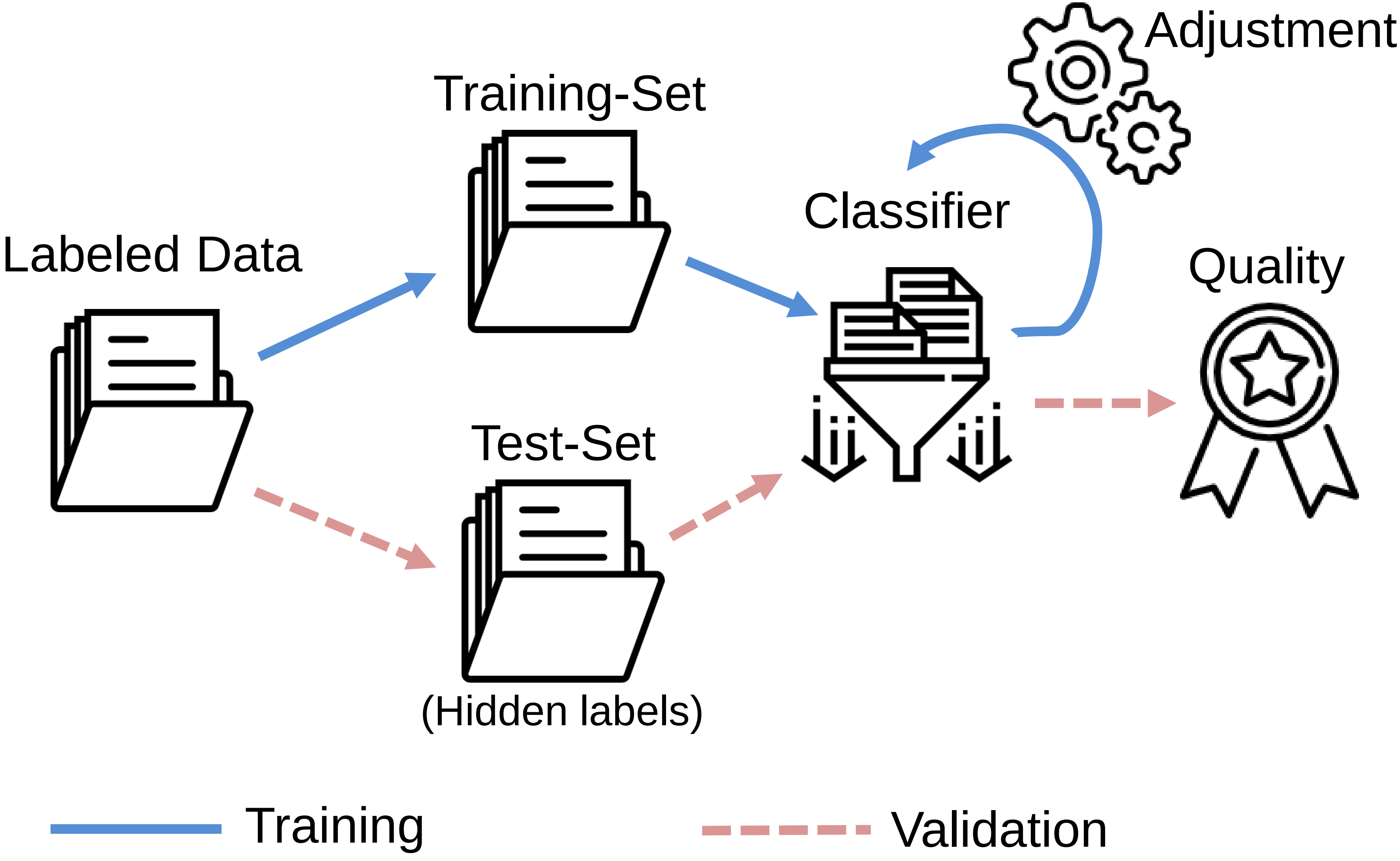
- Grouping a set of data objects into clusters
  - Cluster: a collection of elements
  - Similar to one another within the same cluster
  - Dissimilar to the objects in other clusters

- Difference to classification
  - No classes (labels) given
  - Unsupervised learning

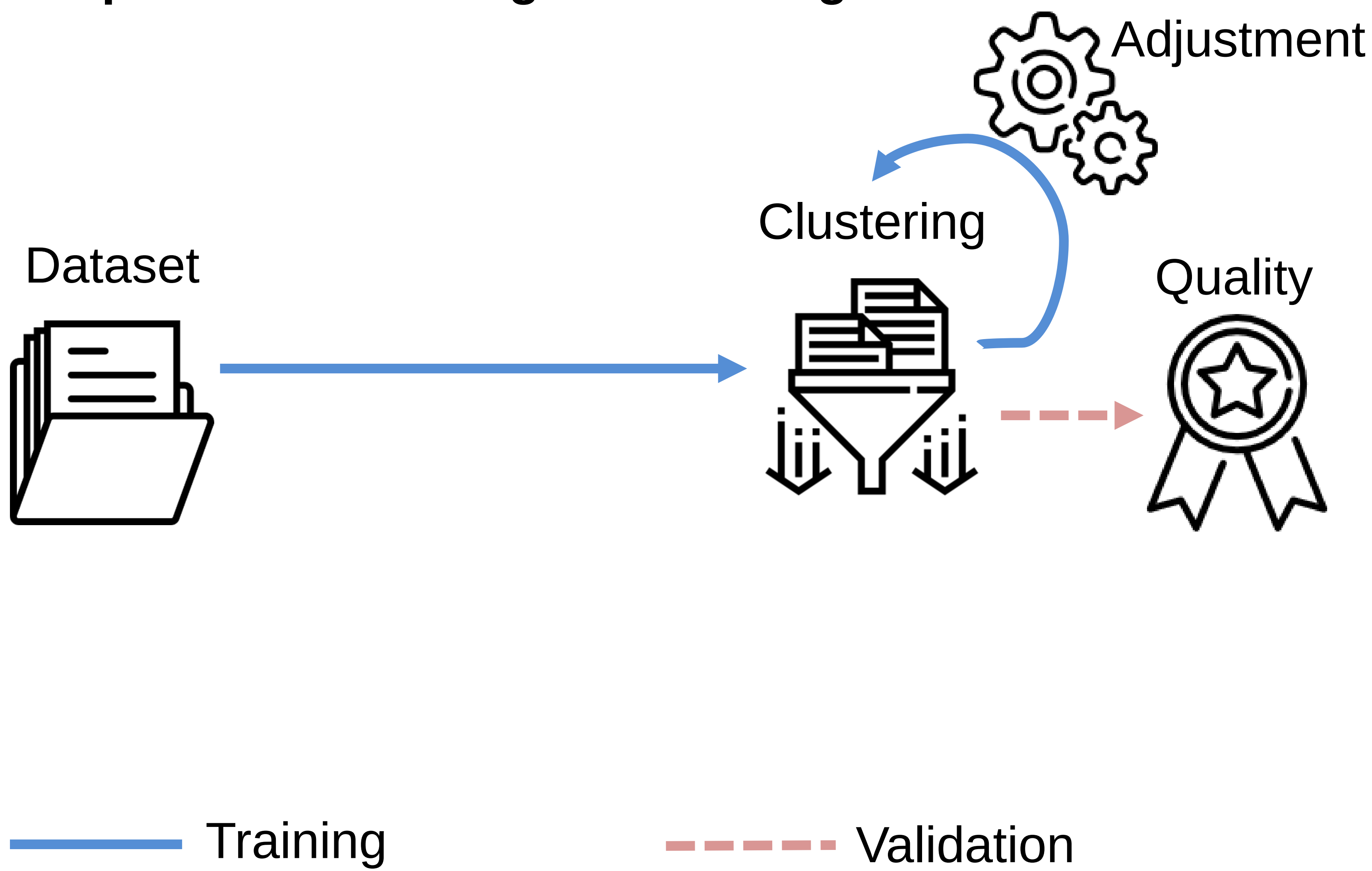
- Application
  - Get insights in large datasets
  - Pre-processing for other algorithms



Supervised Learning – Classification



Unsupervised Learning – Clustering

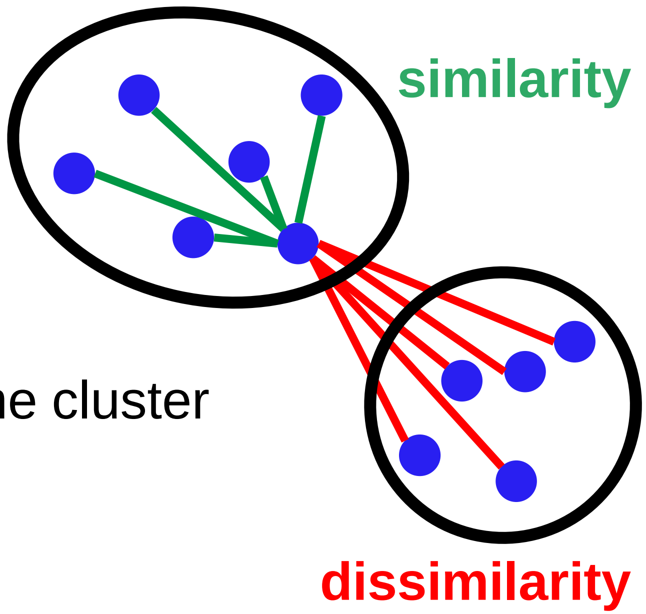


Additional Slide

Unsupervised learning has the lowest data requirements. No labels or subdivision into Training and Test Set are required. Therefore, no concrete knowledge about the data is necessary, but it is only necessary to know within which features similarities and differences should be searched. The clustering provides a division into different groups and afterwards the division can be evaluated.

Quality Measure of a Cluster

- **Distances to Representatives**
  Highly dependent on number of clusters k
  - $k = 2$: very large distances
  - $k = n - 1$: very small distances
- **Similarity**
  Average distance to all elements within same cluster
  - Element $o \in a \in C$
  - $sim(o) = \frac{1}{|a|} \sum_{e \in a} distance(o, e)$
- **Dissimilarity**
  Average distance to all elements of second closest cluster
  - Element $o \in a \in C$ and other cluster(s) $b \in C \backslash a$
  - $dsim(o) = \min_{b \in C \backslash a} \left(\frac{1}{|b|} \sum_{e \in b} distance(o, e) \right)$



Quality Measure of a Cluster

- **Silhouette Coefficient**

$$s(o) = \frac{dsim(o) - sim(o)}{\max\{sim(o), dsim(o)\}}$$

  - $s(o) \in [-1, 1]$
  - $sim(o) \gg dsim(o) \rightarrow \boldsymbol{s(o) = -1}$
  - $sim(o) = dsim(o) \rightarrow \boldsymbol{s(o) = 0}$
  - $sim(o) \ll dsim(o) \rightarrow \boldsymbol{s(o) = +1}$

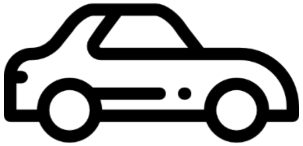
    ($silh(o) \geq 0.7$ desirable)

  - Mean of cluster c : $silh(c) = \frac{1}{|c|} \sum_{o \in c} s(o)$
  - Mean of whole data set E : $silh(E) = \frac{1}{|E|} \sum_{o \in E} s(o)$

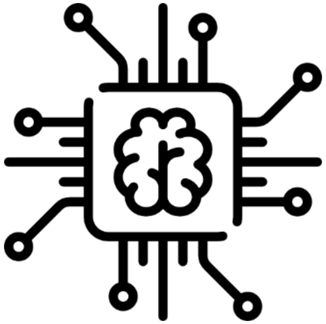
Unsupervised Learning: Clustering
Prof. Dr.-Ing. Markus Lienkamp

(Andreas Schimpe, M.Sc.)

Agenda

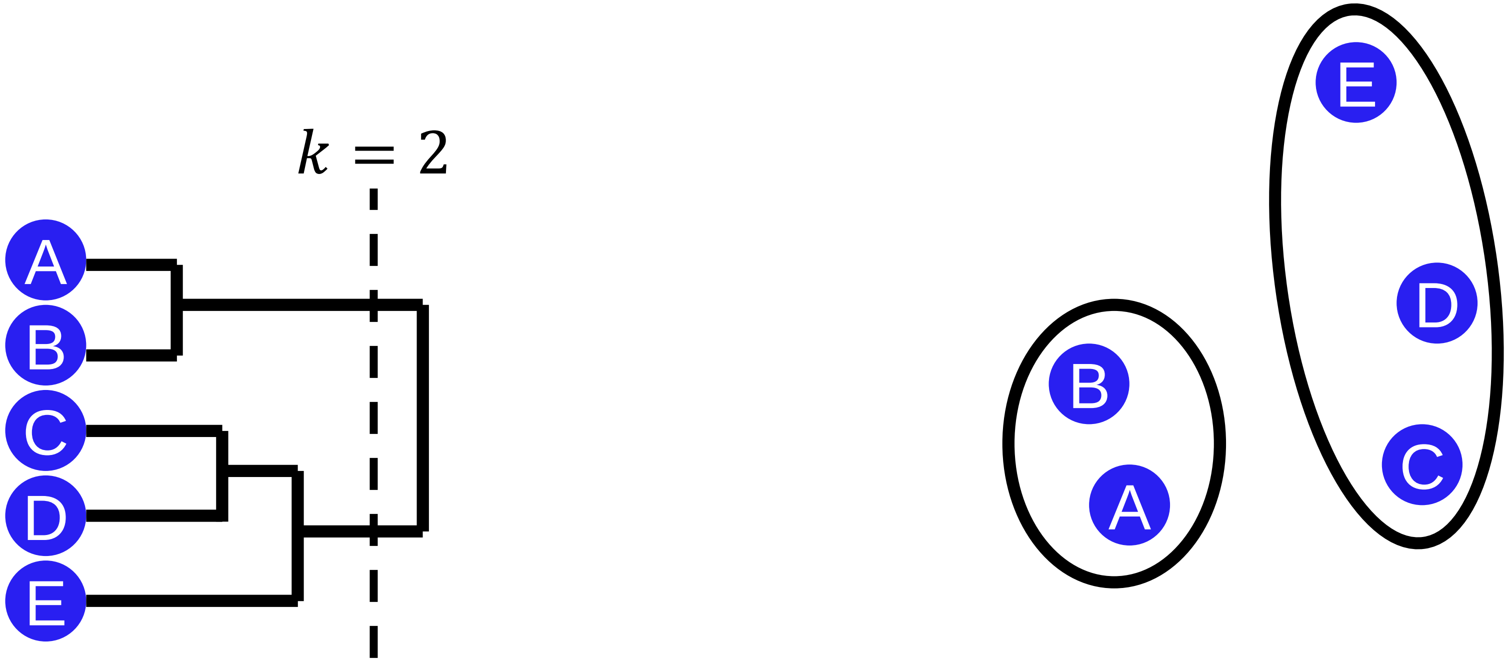


1. Chapter: Introduction
    - 1.1 Overview
    - 1.2 Training and Validation
2. Chapter: Methods
    - **2.1 Hierarchical Clustering**
    - 2.2 k-means
    - 2.3 DBSCAN
3. Chapter: Application
4. Chapter: Summary



Hierarchical Clustering

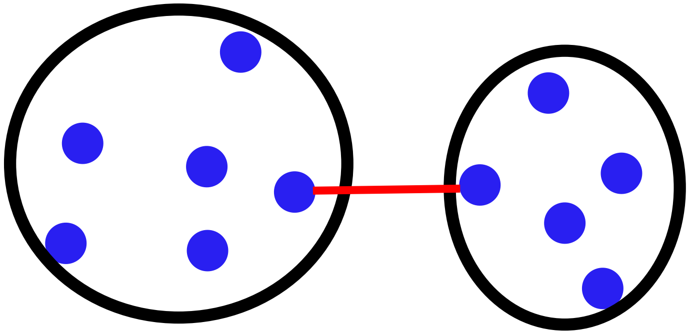
1. Start with one cluster per element
2. Combine two closest (most similar) clusters
3. Repeat until all elements are in one cluster



Distance between Clusters

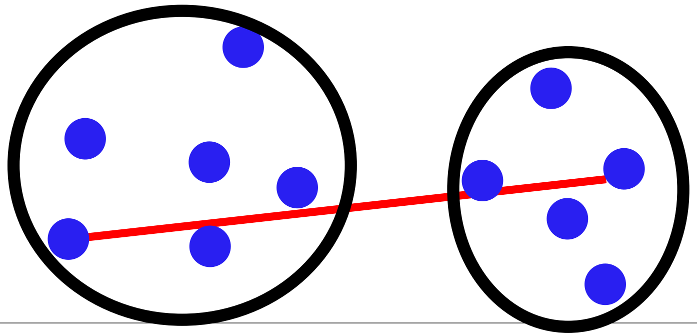
Single Link

- Smallest distance between two point of different clusters



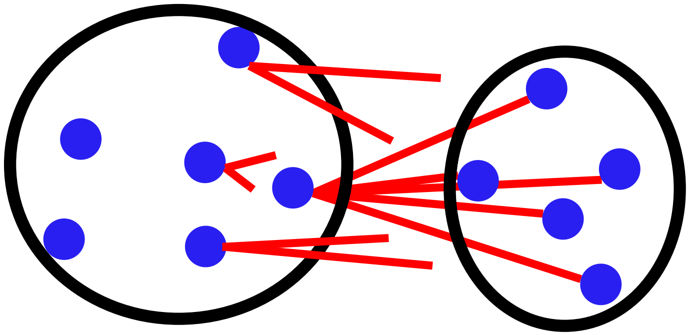
Complete Link

- Largest distance between two points of different clusters



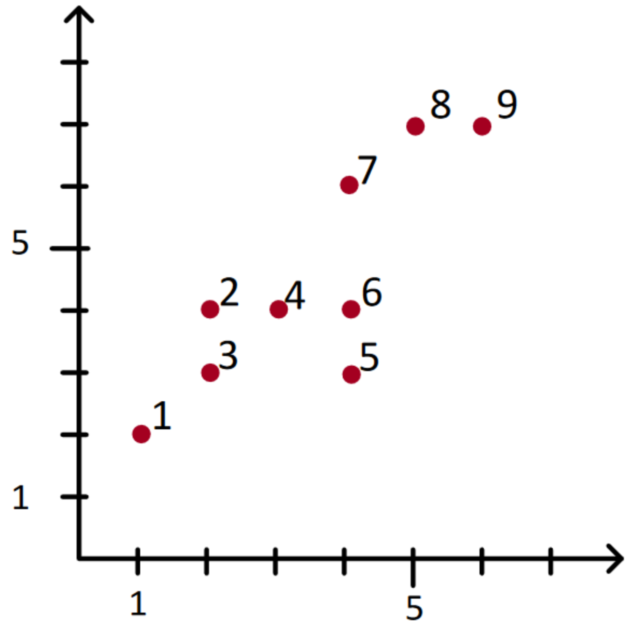
Average Link

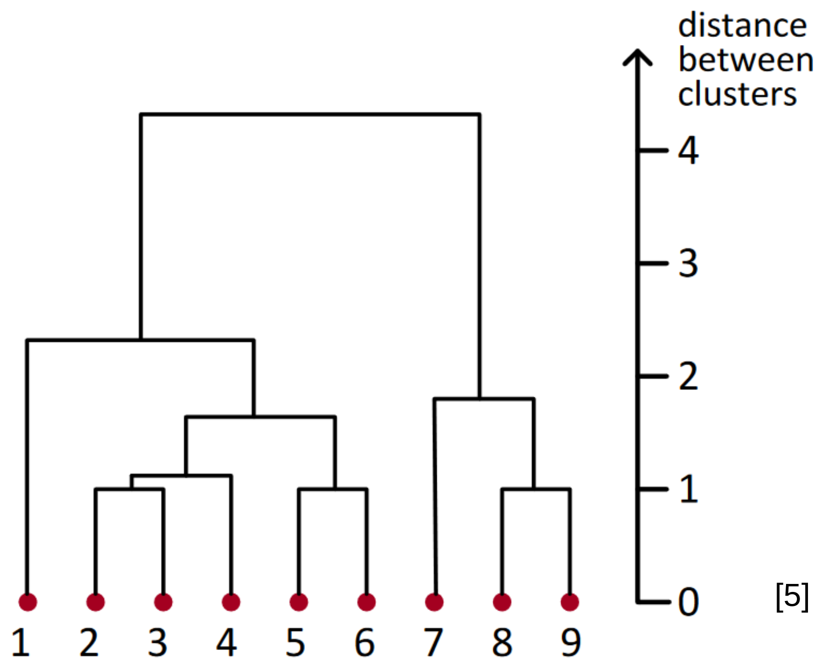
- Average distance between all points of one cluster to all points of a different cluster



Dendrogram

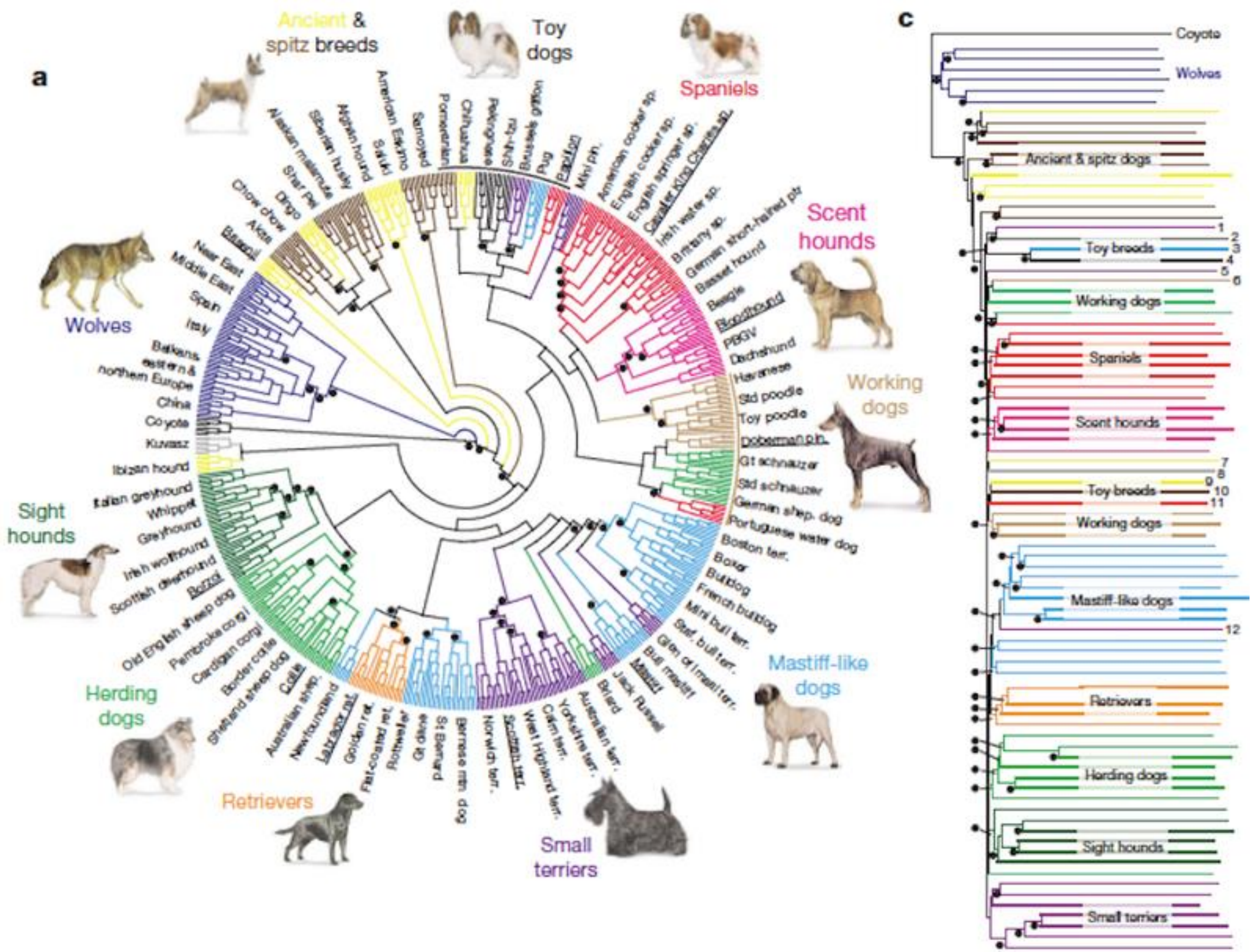
- Root: Cluster with all points
- Leaf: Cluster with one point
- Edges: Combine two clusters
- Depth: Distance between two combined clusters





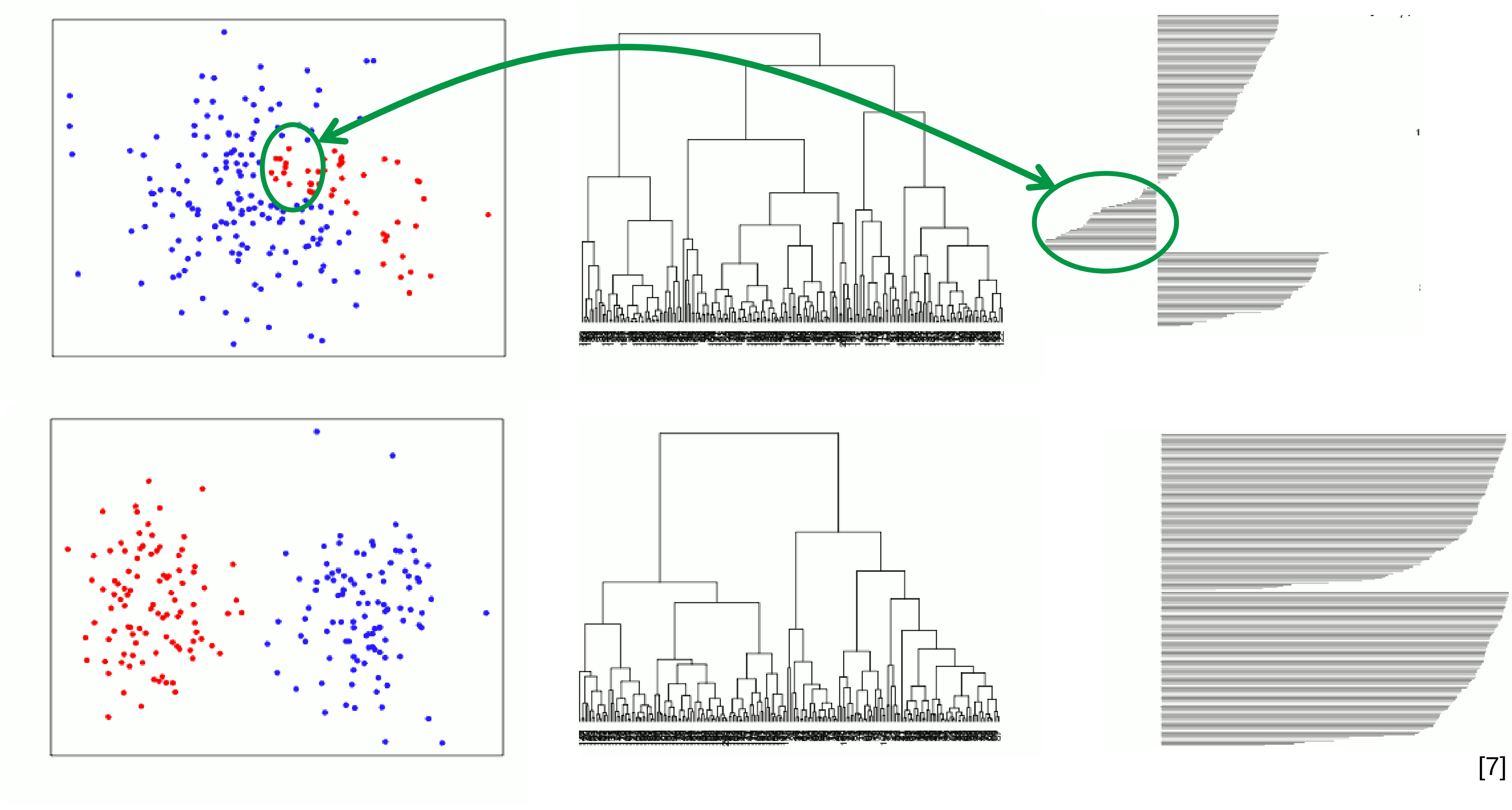
[5]

Dendrogram – Example



[6]

Hierarchical Clustering – Silhouette coefficient



[7]

Additional Slide

Above we try to divide a cloud of points into two clusters. Since only one cluster would be necessary here, we already see in the dendrogram that the distances between the elements of different clusters are small. It can be seen even more clearly in the silouette coefficient, which becomes negative.

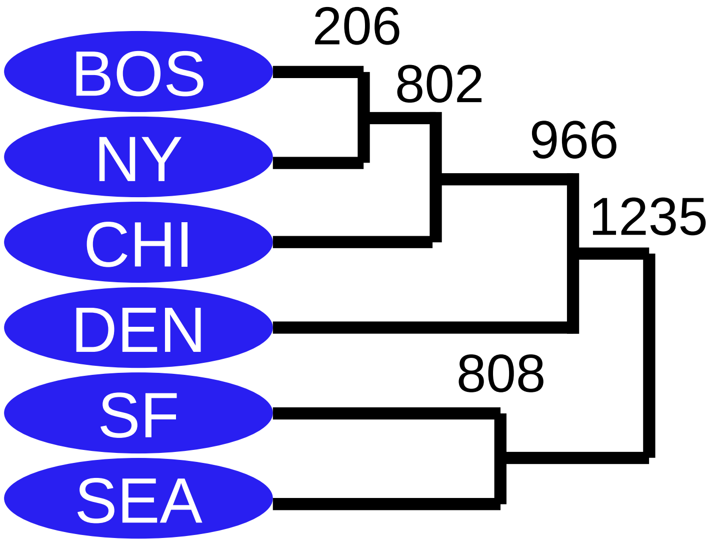
Below a suitable number of clusters has been chosen and the quality of the clusters is correspondingly much better.

Hierarchical Clustering – Example

	BOS	NY	CHI	DEN	SF	SEA
BOS	0	206	963	1949	3095	2979
NY		0	802	1771	2934	2815
CHI			0	966	2142	2013
DEN				0	1235	1307
SF					0	808
SEA						0

[8]

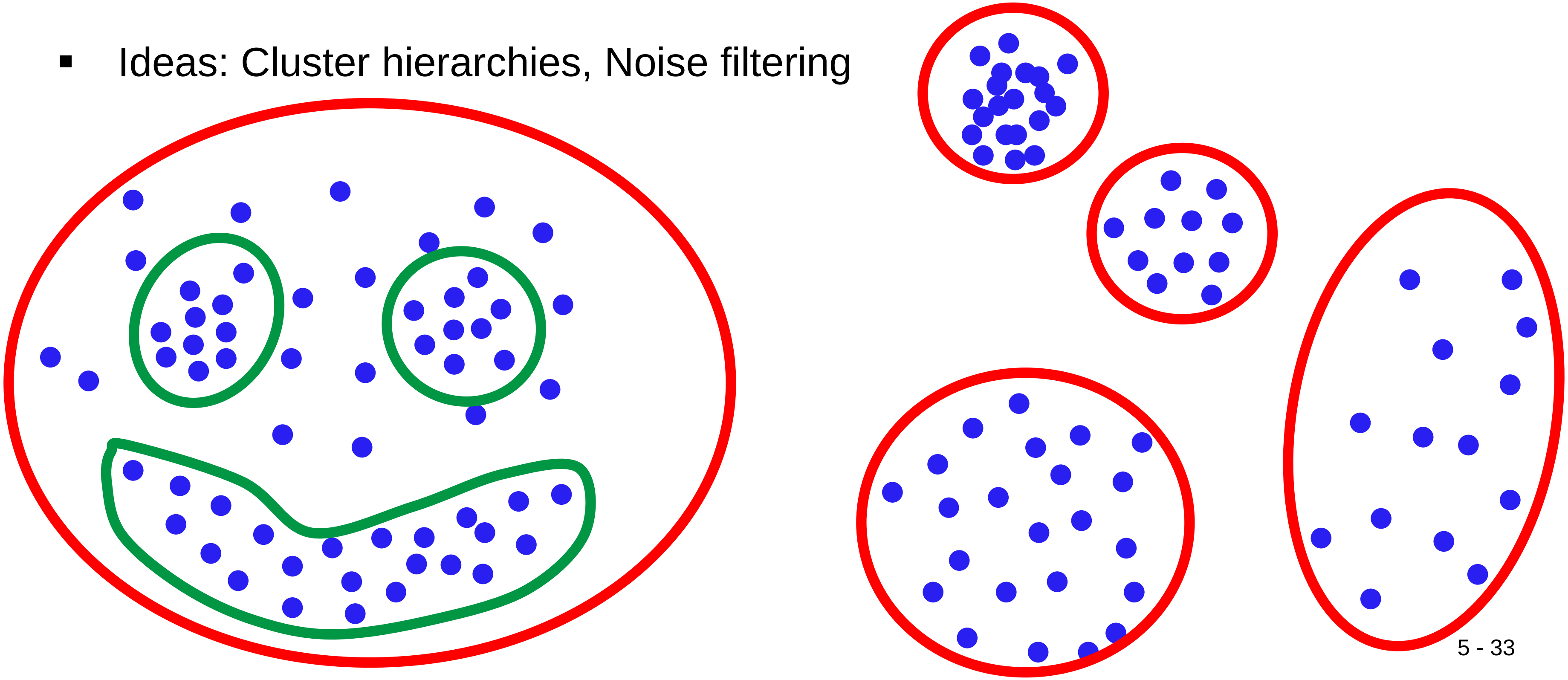
Single Link



Hierarchical Clustering

- Problems
  - Clusters within clusters
  - Variable densities within data set

- Ideas: Cluster hierarchies, Noise filtering



Discussion Hierarchical Clustering

Pro

- **Generic**: No cluster number or other parameters must be defined
- **Visualization**: Dendrogram shows hierarchy
- **Hierarchy**: Relationship between clusters
- **Deterministic:** Generates always same clusters

Contra

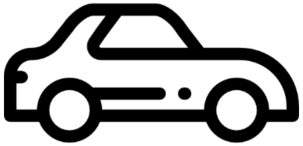
- **Scalability:** Runtime $\mathcal{O}(n^3)$
- **Choice:** Final clustering result (number of clusters) must be chosen from hierarchy

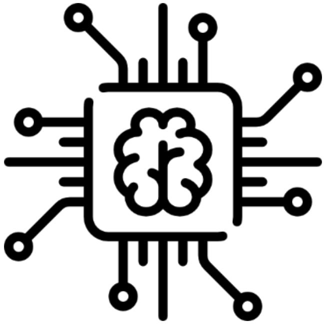
Unsupervised Learning: Clustering
Prof. Dr.-Ing. Markus Lienkamp

(Andreas Schimpe, M.Sc.)

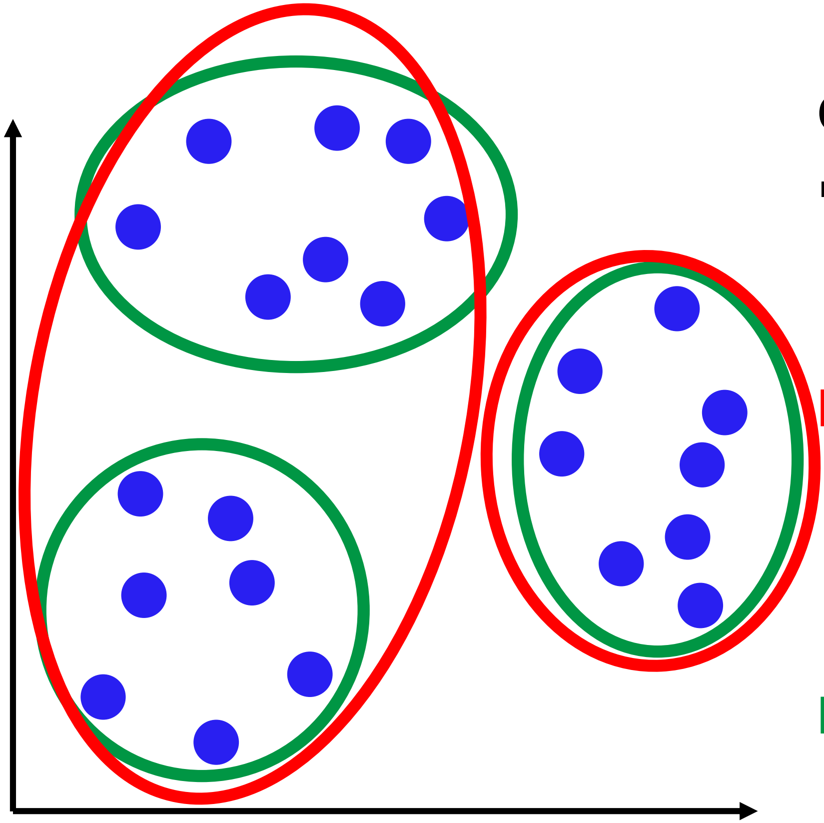
Agenda

1. Chapter: Introduction
    1.1 Overview
    1.2 Training and Validation
2. Chapter: Methods
    2.1 Hierarchical Clustering
    2.2 k-means
    2.3 DBSCAN
3. Chapter: Application
4. Chapter: Summary





K-Means – Basic Idea



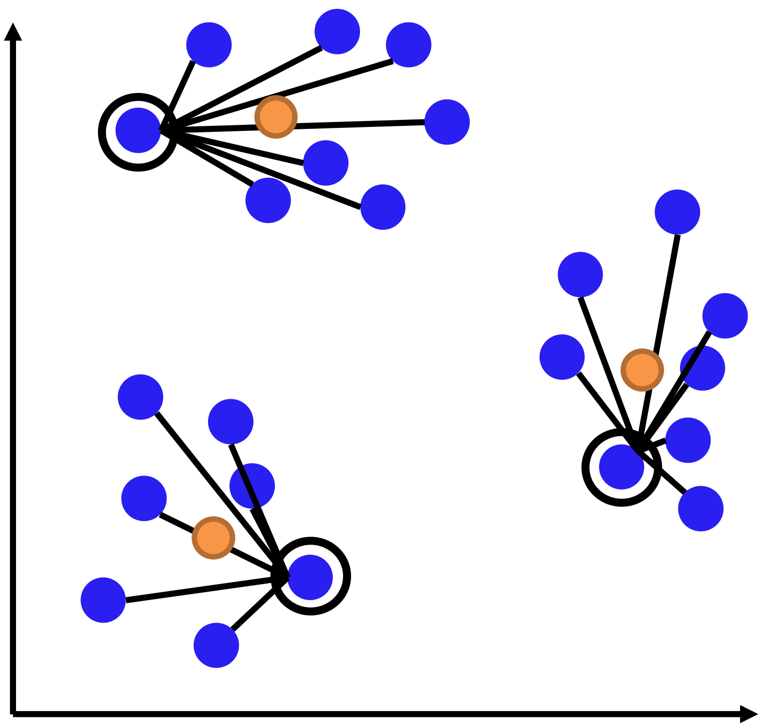
Objective

- Given desired number of clusters, minimize cluster variability

Large sum of cluster variabilities
→ Poor clustering
→ or bad choice for *k*

Minimal sum of cluster variabilities
→ Optimal clustering

K-Means Algorithm (Lloyd)



Input

- Number of desired clusters k
- Dataset

Initialization

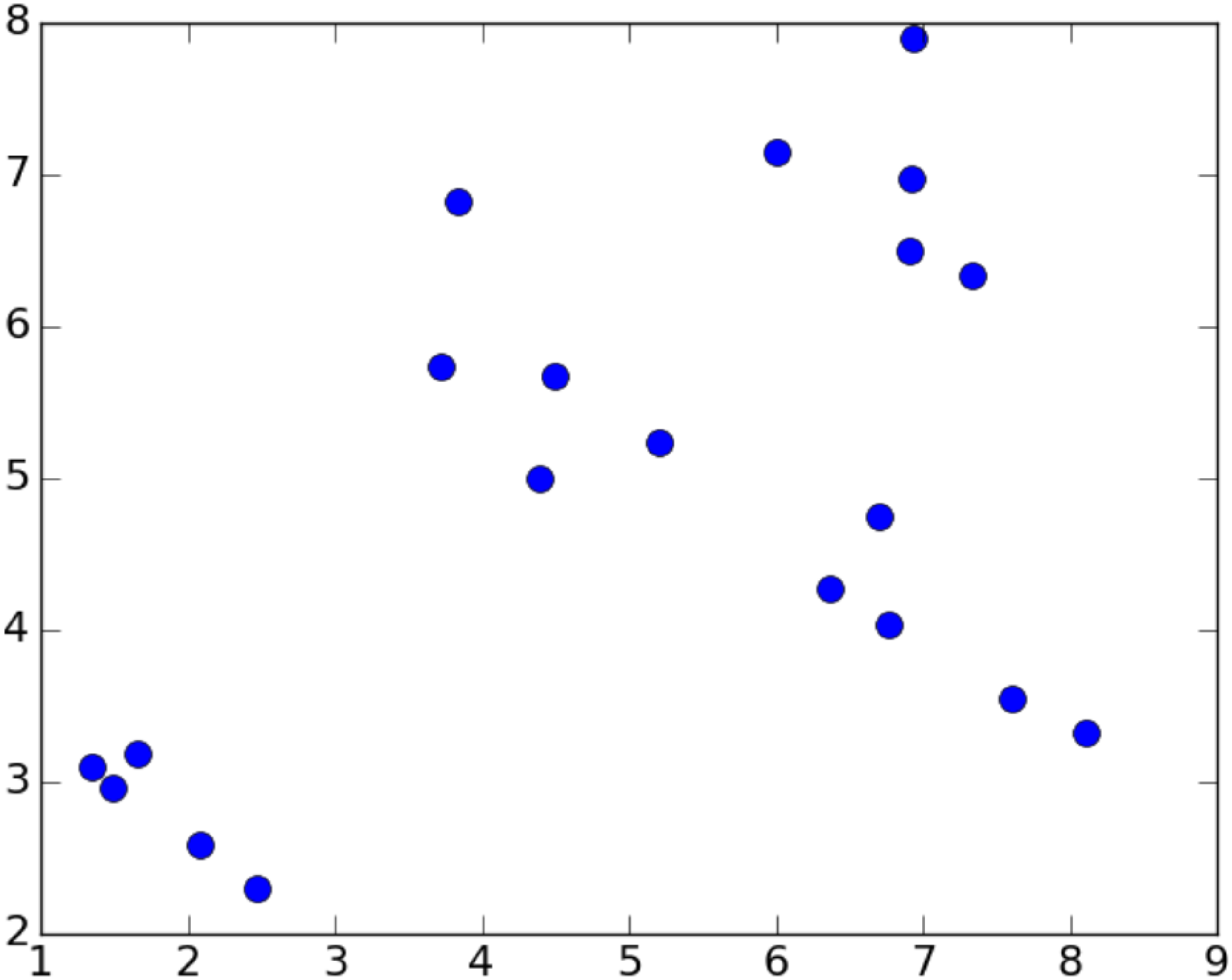
- Choose k arbitrary representatives

$k = 3$

Repeat until stable

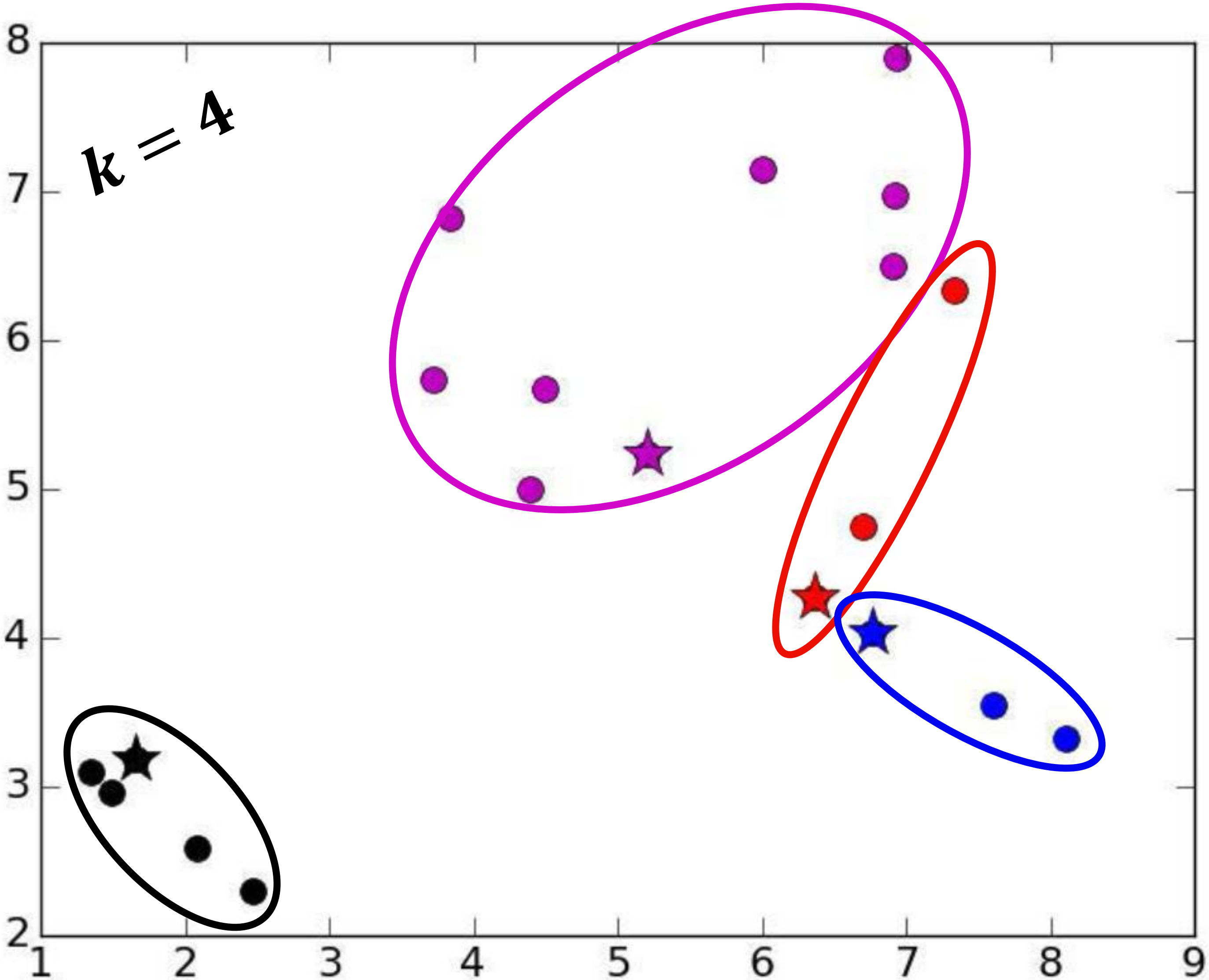
- Assign objects to nearest representative
- Compute center of each cluster as new representative

K-Means Algorithm (Lloyd)



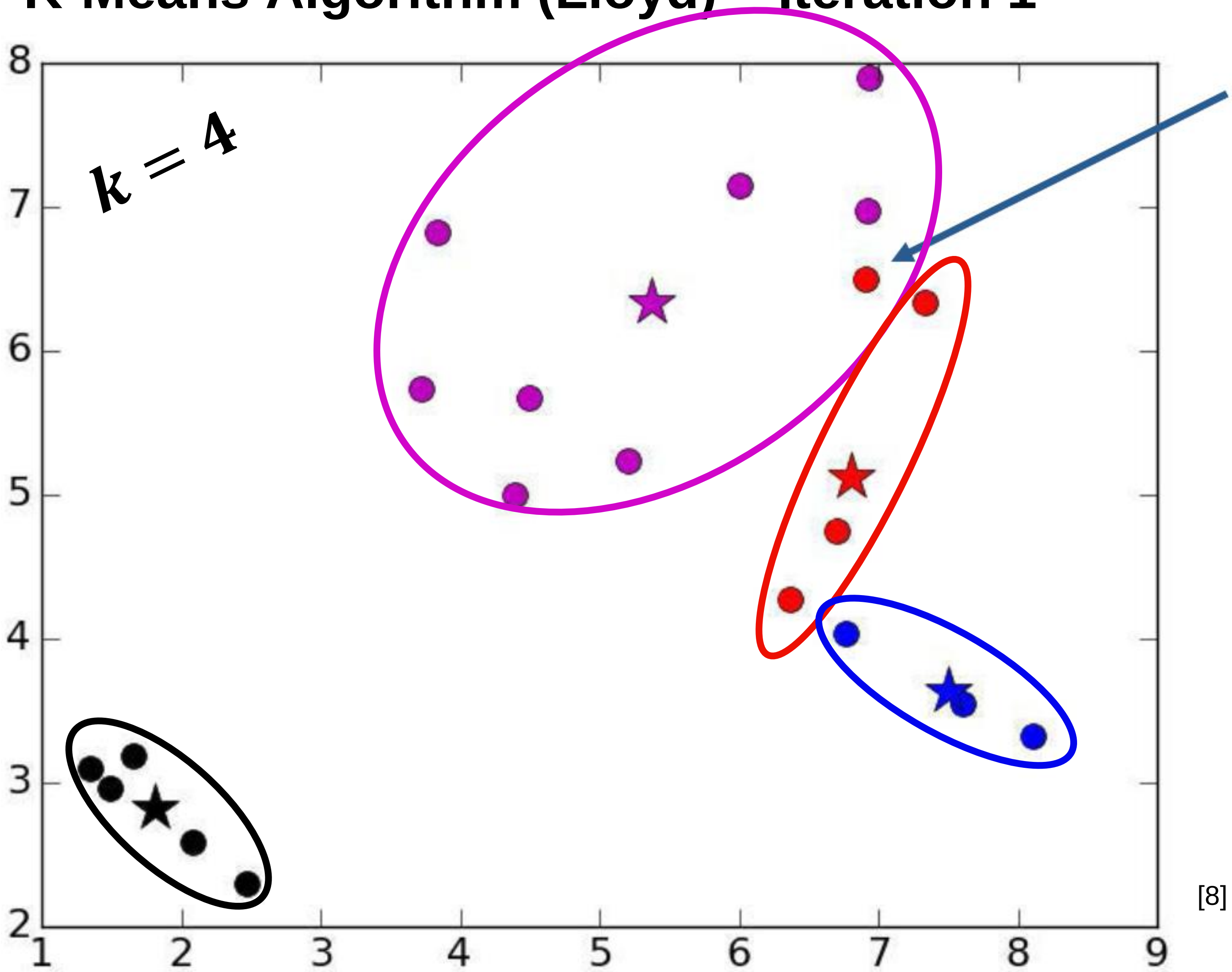
[8]

K-Means Algorithm (Lloyd) – Initialization



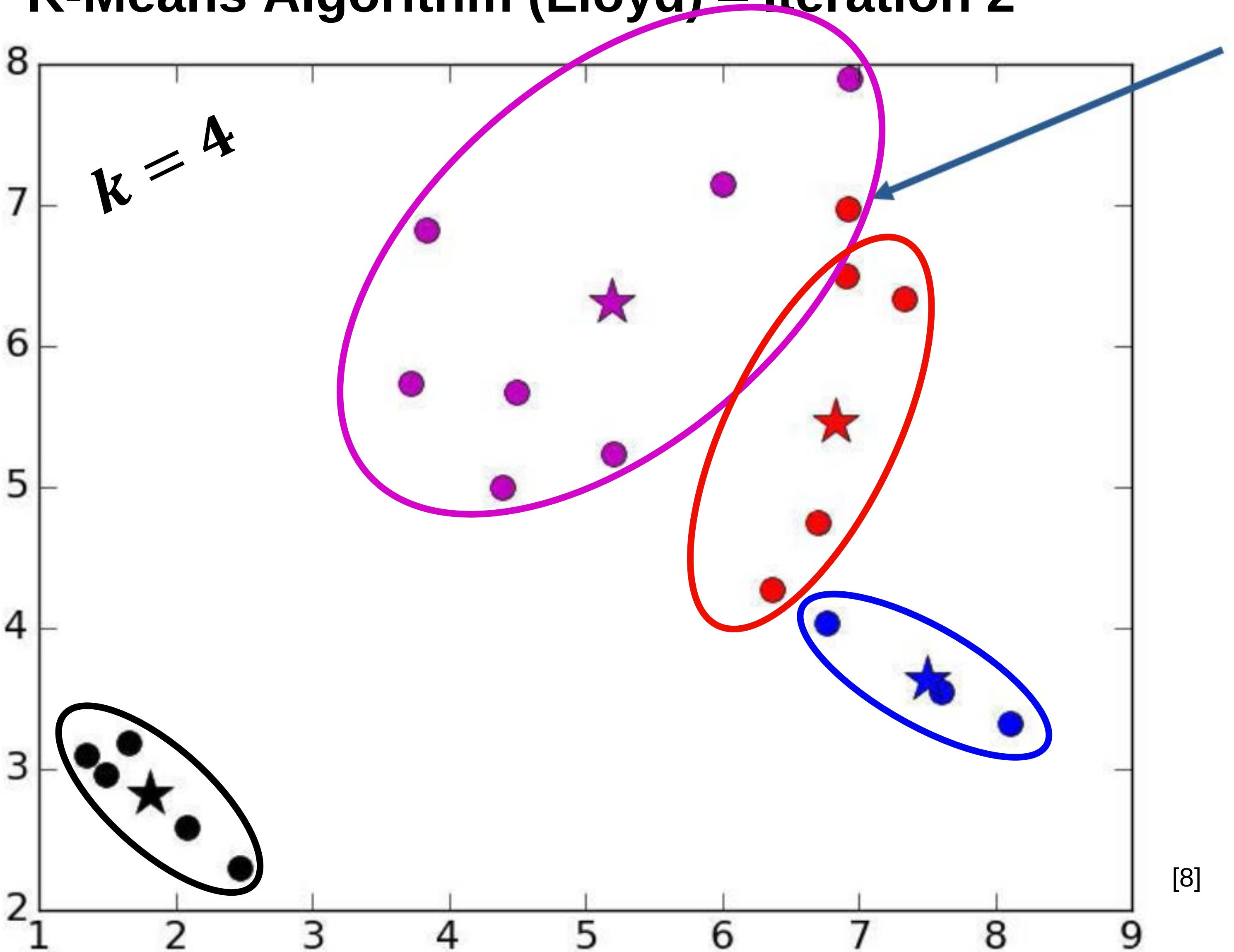
[8]

K-Means Algorithm (Lloyd) – Iteration 1



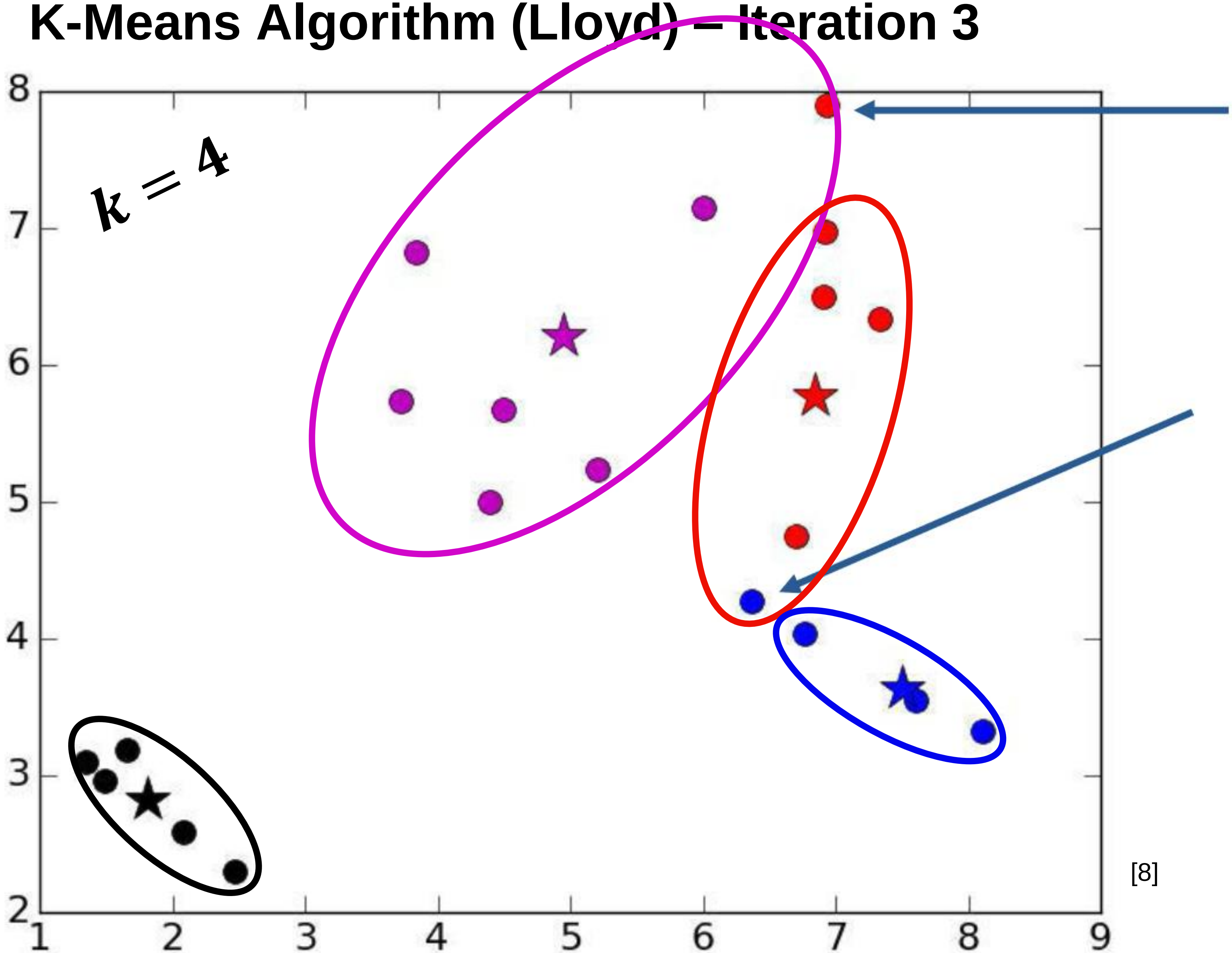
[8]

K-Means Algorithm (Lloyd) – Iteration 2



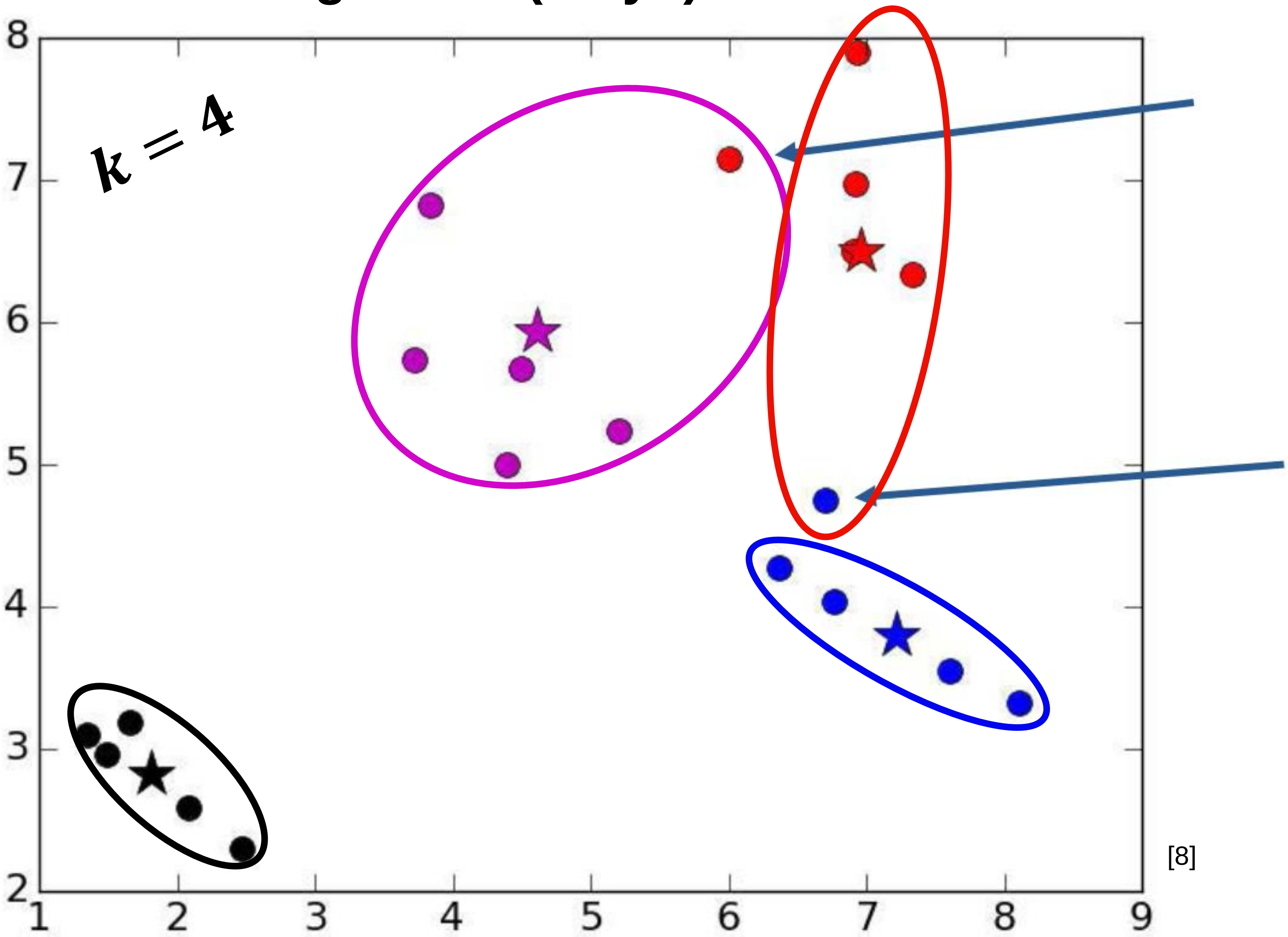
[8]

K-Means Algorithm (Lloyd) – Iteration 3



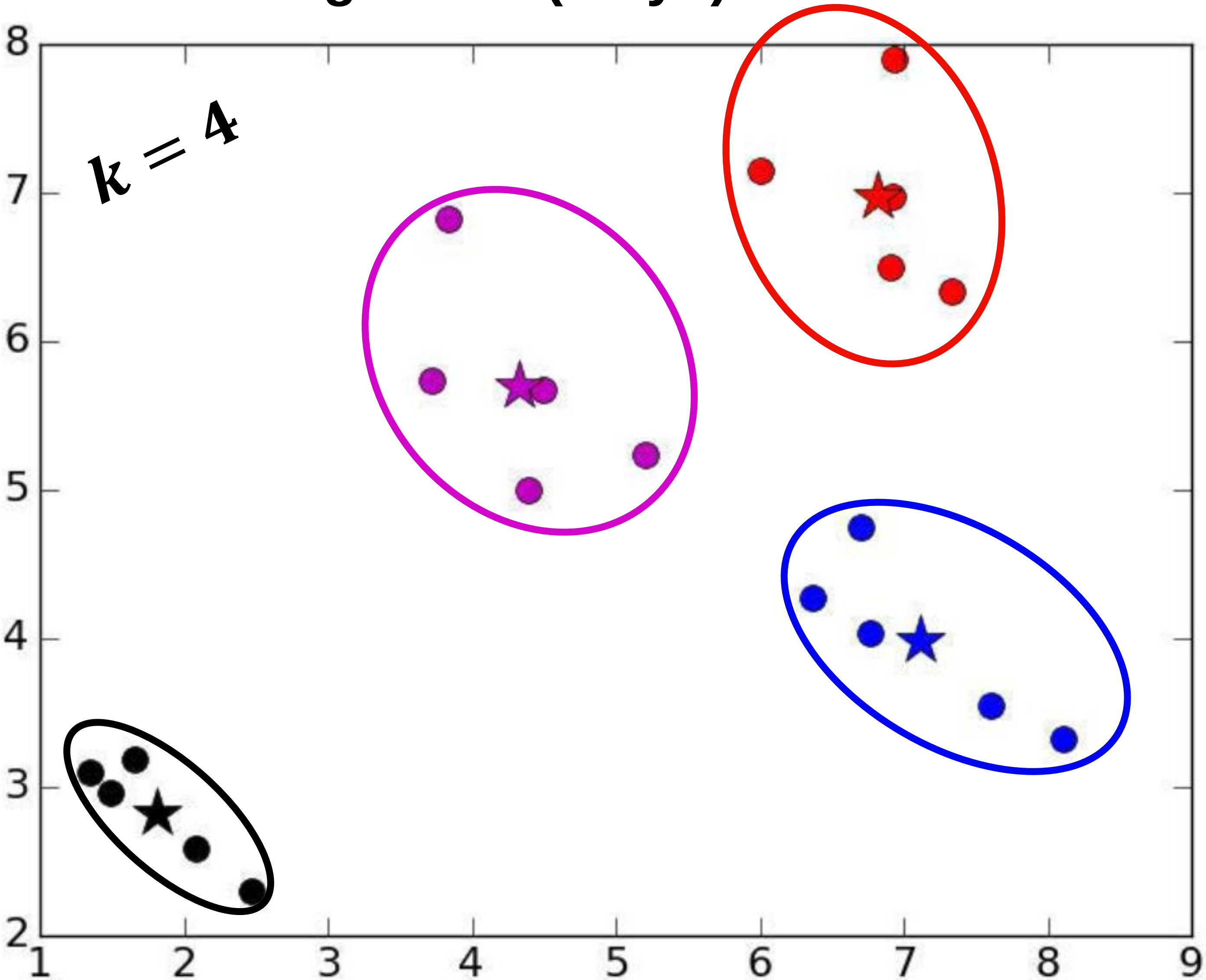
[8]

K-Means Algorithm (Lloyd) – Iteration 4



[8]

K-Means Algorithm (Lloyd) – Iteration 5



[8]

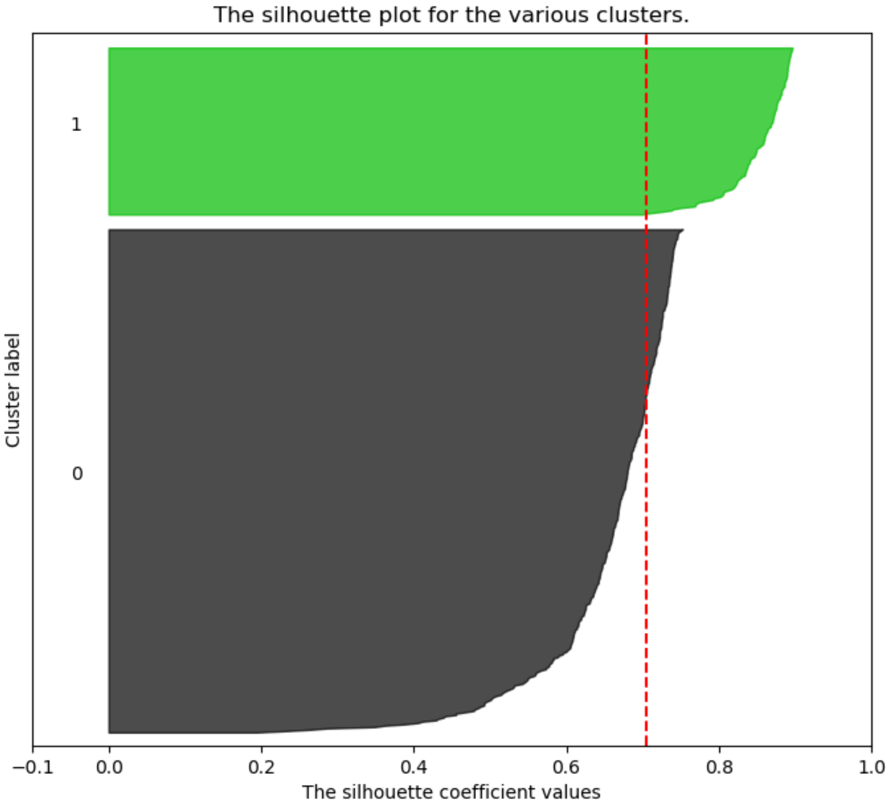
K-Means Algorithm – How to choose k?

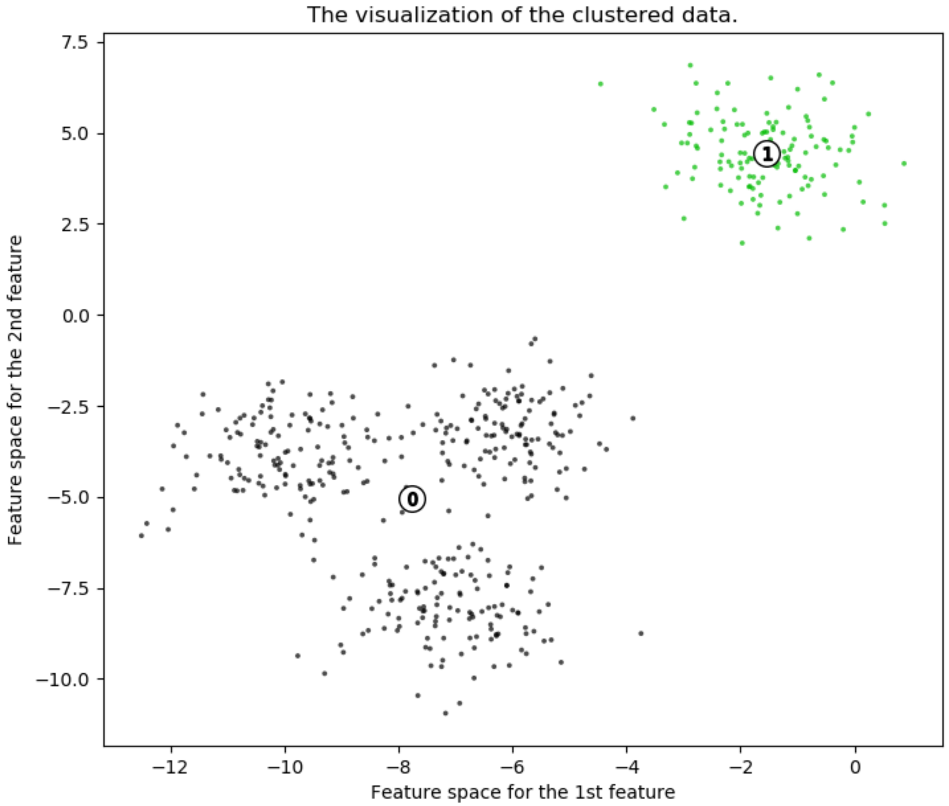
- A priori knowledge of an expert
  - „There are five different types of bacteria“ → $k = 5$
- Search for a good k
  - Naïve approach: Brute Force, iterating $k = 2, 3, 4, \ldots$
  - Run hierarchical clustering on subset of data beforehand

K-Means Algorithm – How to choose k?

Avg. Silh. Coeff.

$$k = 2$$



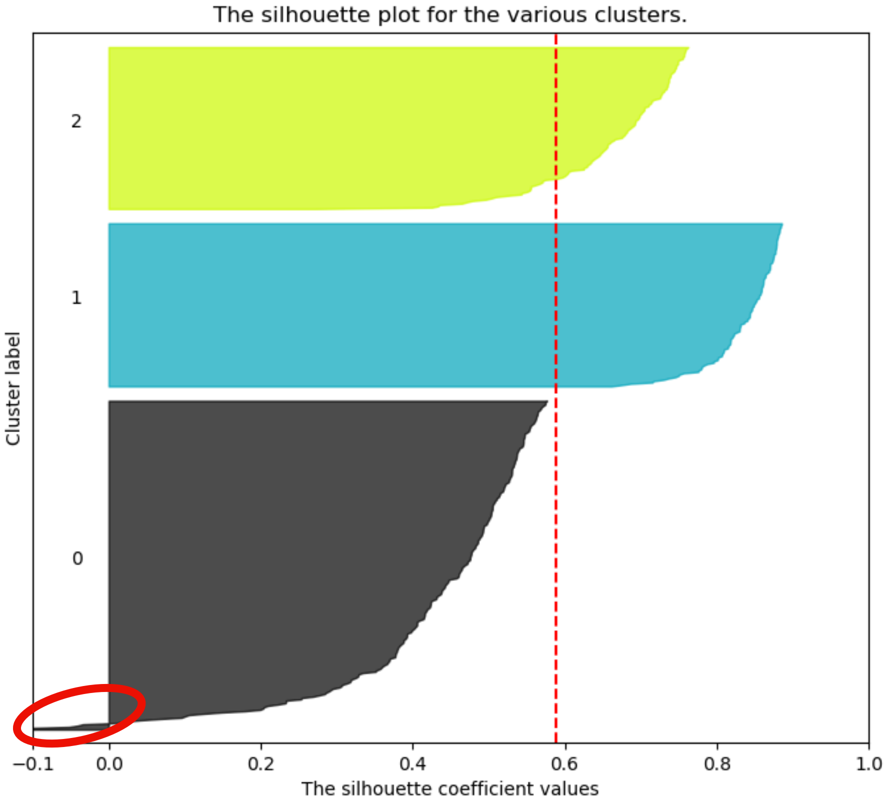


[9]

K-Means Algorithm – How to choose k?

Avg. Silh. Coeff.

$$k = 3$$

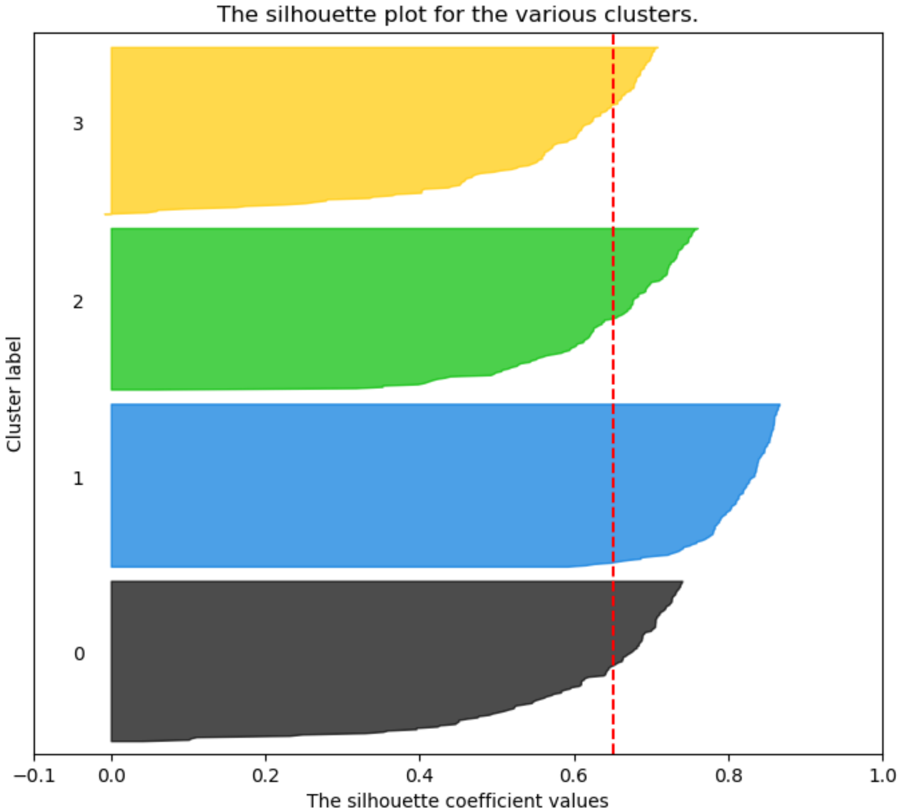


[9]

K-Means Algorithm – How to choose k?

Avg. Silh. Coeff.

$$k = 4$$

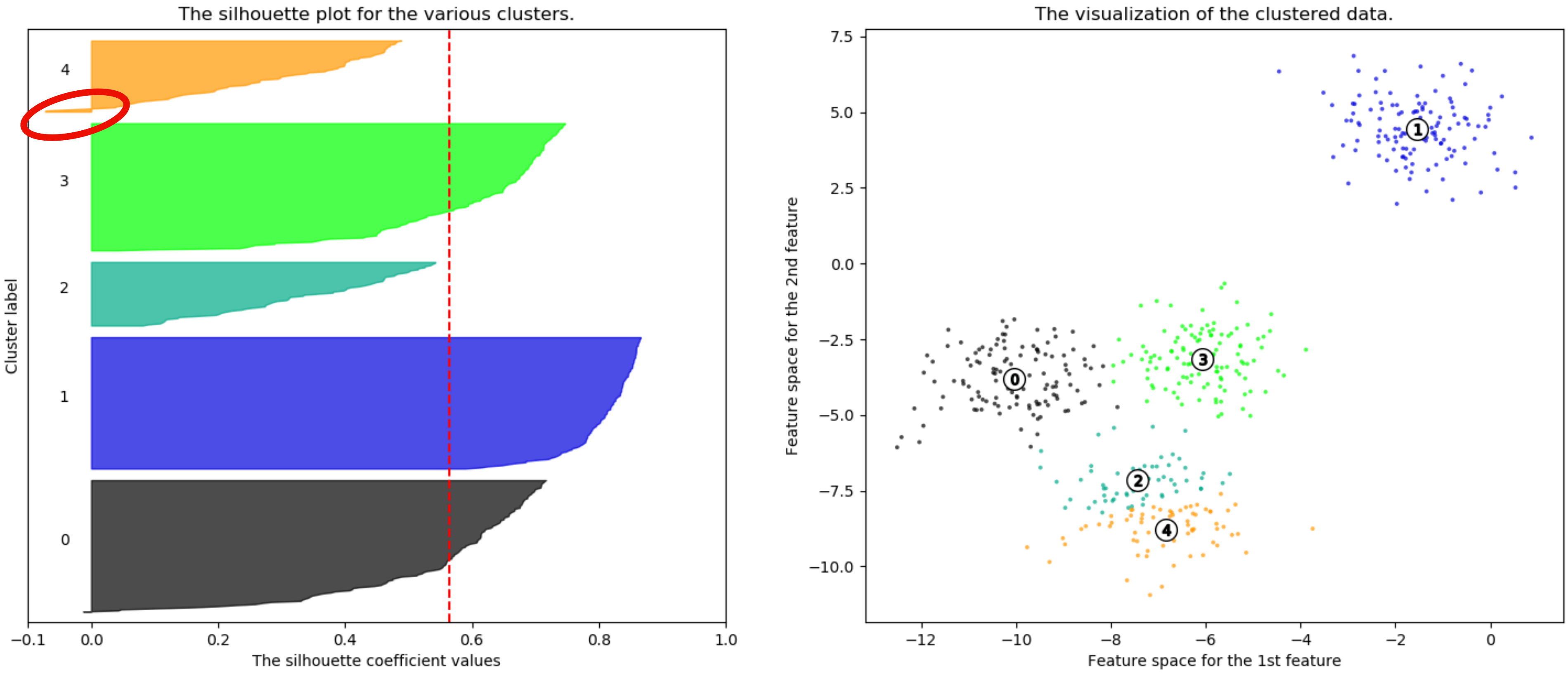


[9]

K-Means Algorithm – How to choose k?

Avg. Silh. Coeff.

$$k = 5$$

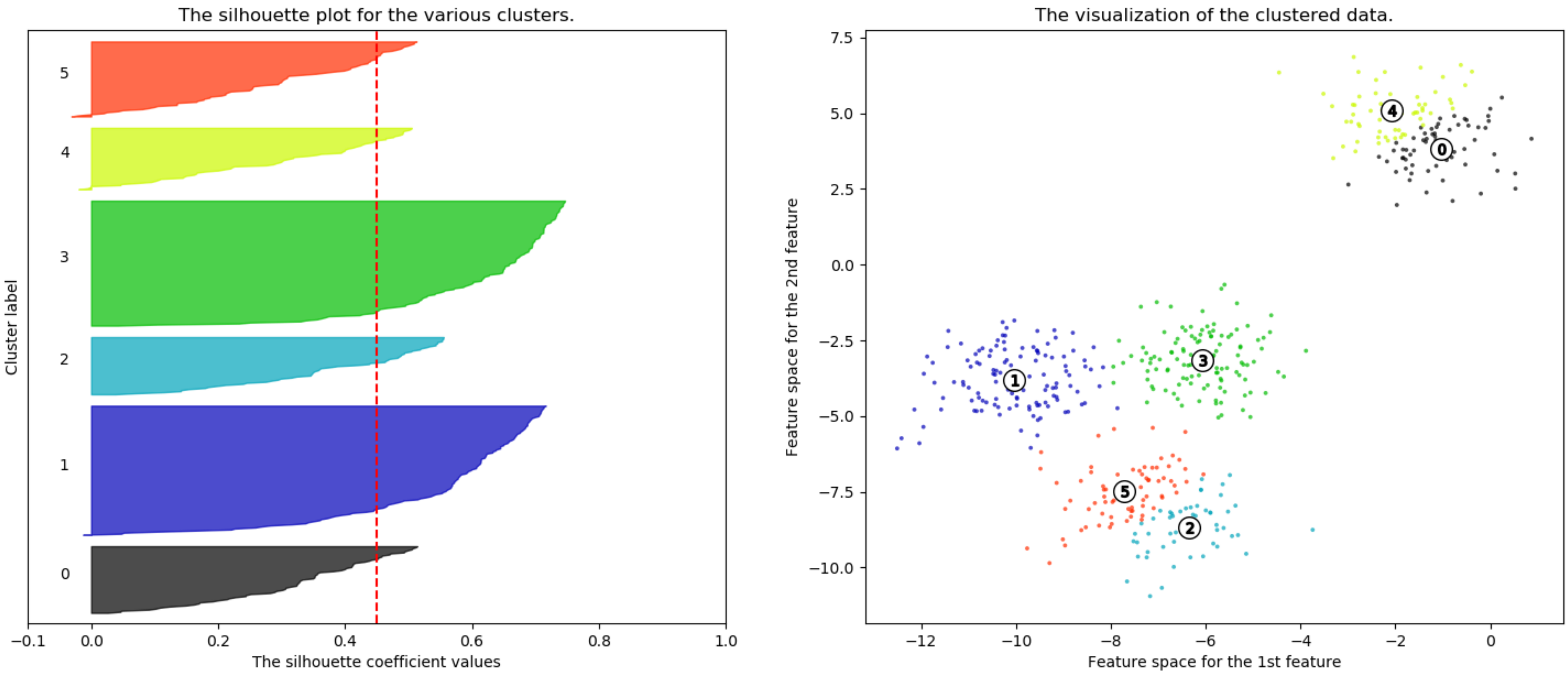


[9]

K-Means Algorithm – How to choose k?

Avg. Silh. Coeff.

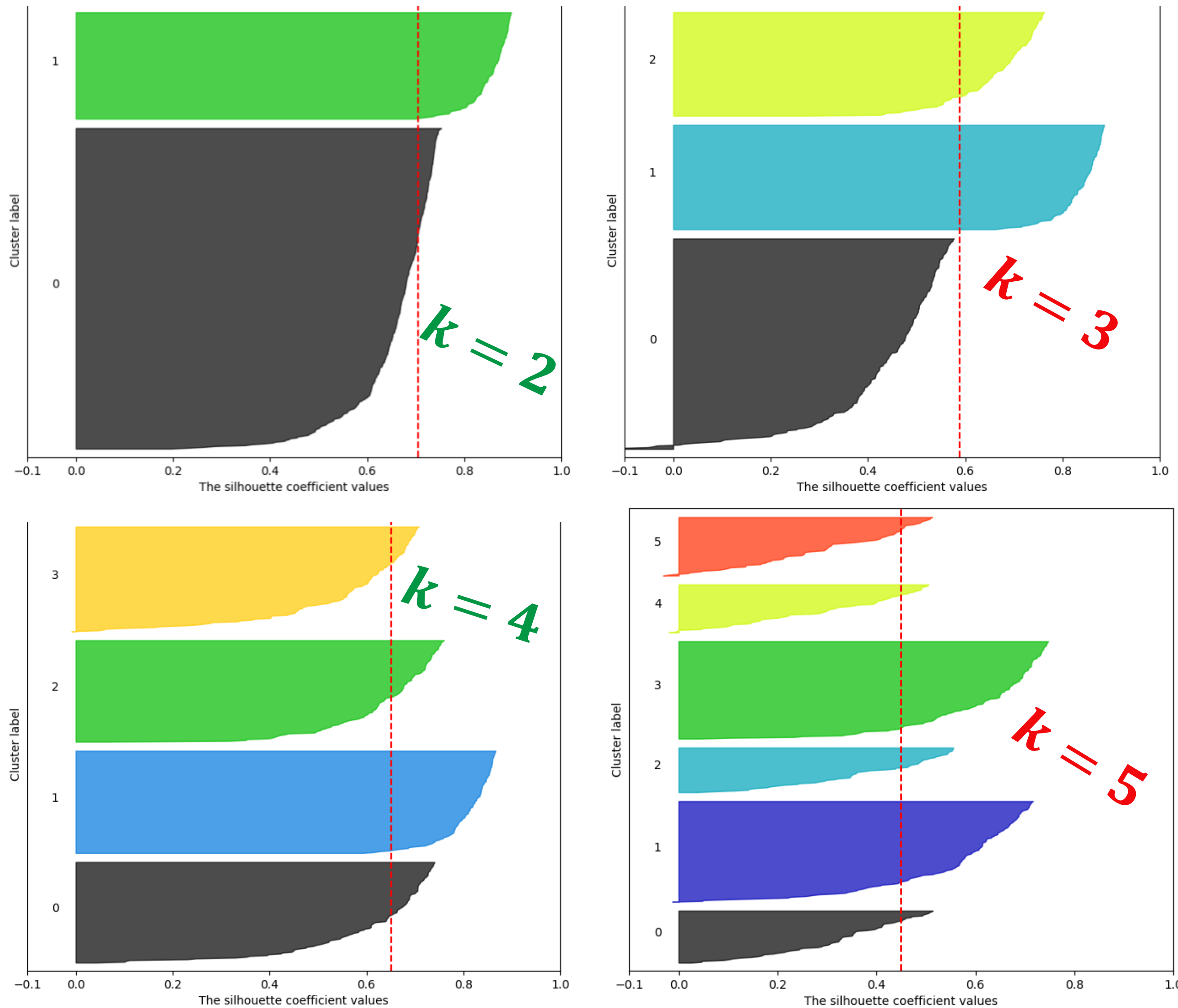
$$k = 6$$



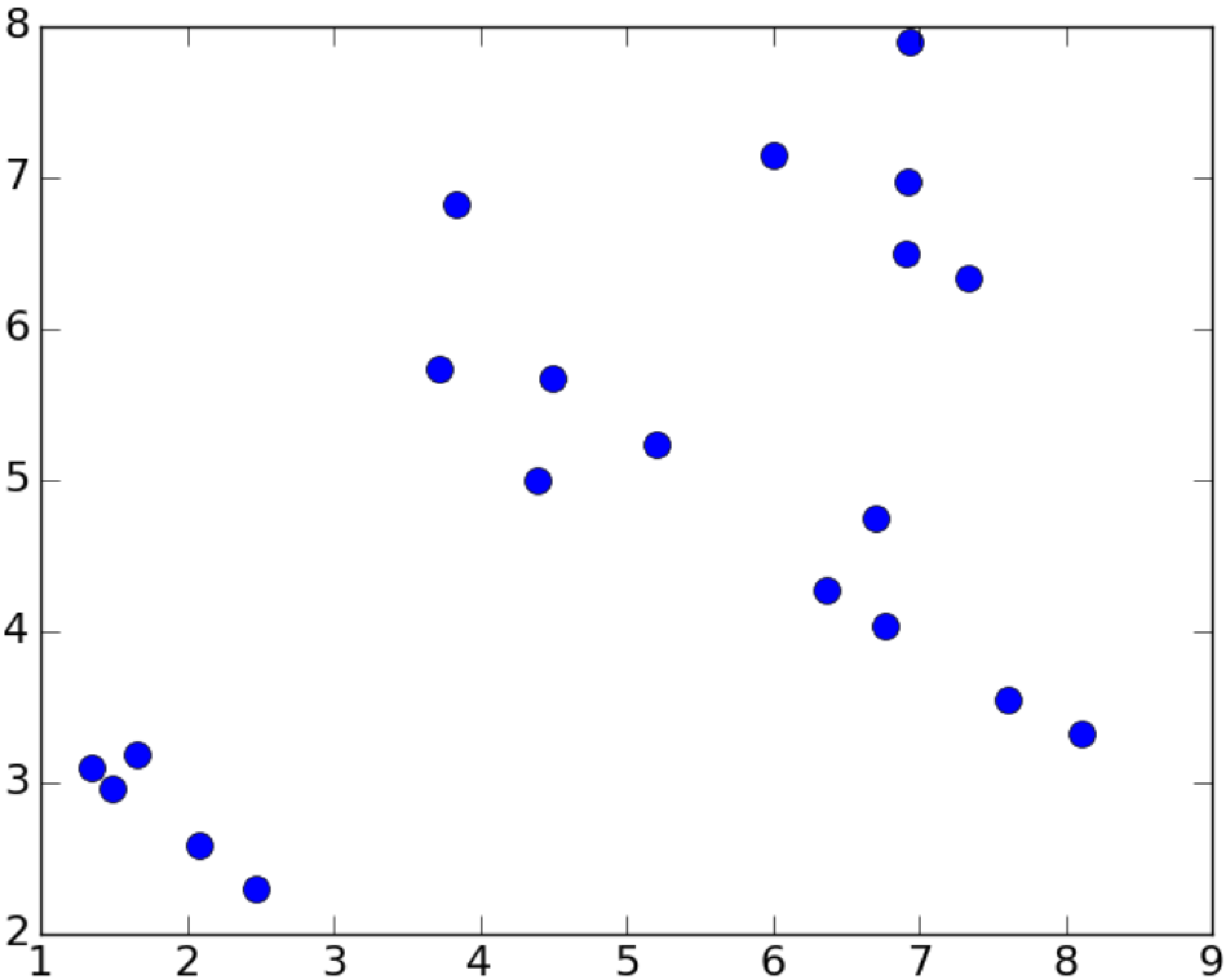
[9]

K-Means Algorithm – How to choose k?

Avg. Silh. Coeff.

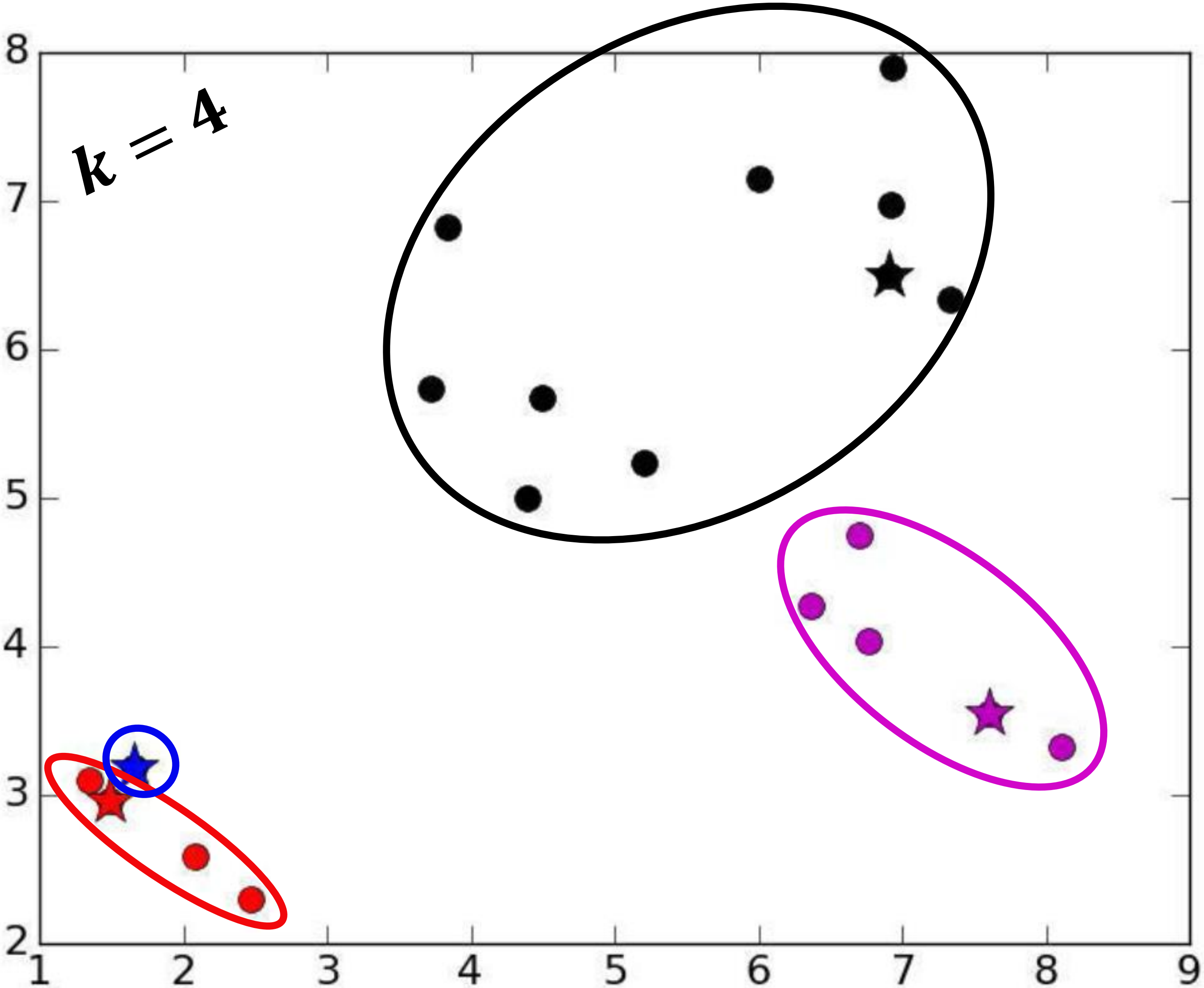


K-Means Algorithm – How to handle randomness?



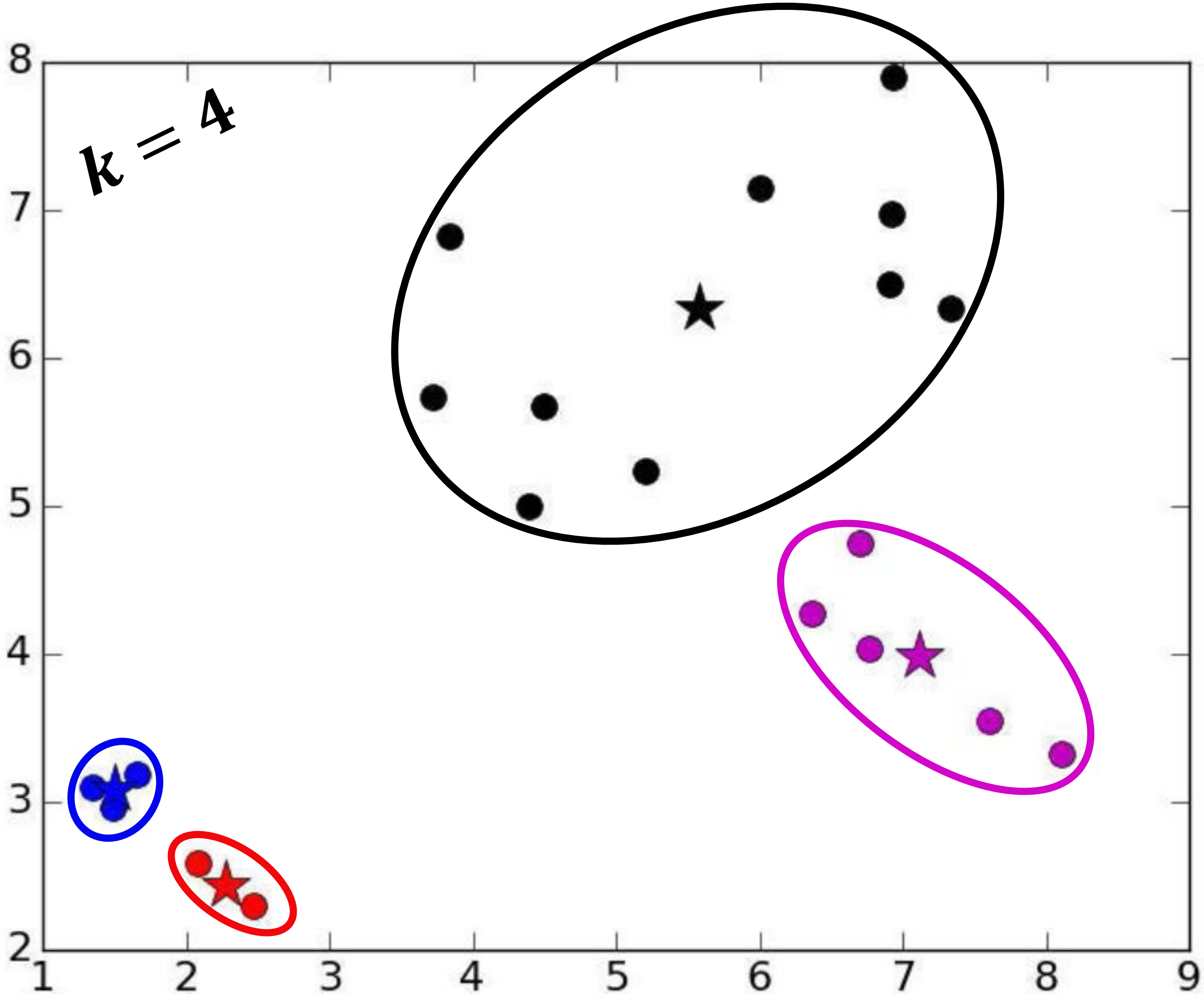
[8]

K-Means Algorithm – How to handle randomness?



[8]

K-Means Algorithm – How to handle randomness?



[8]

K-Means Algorithm – How to handle randomness?

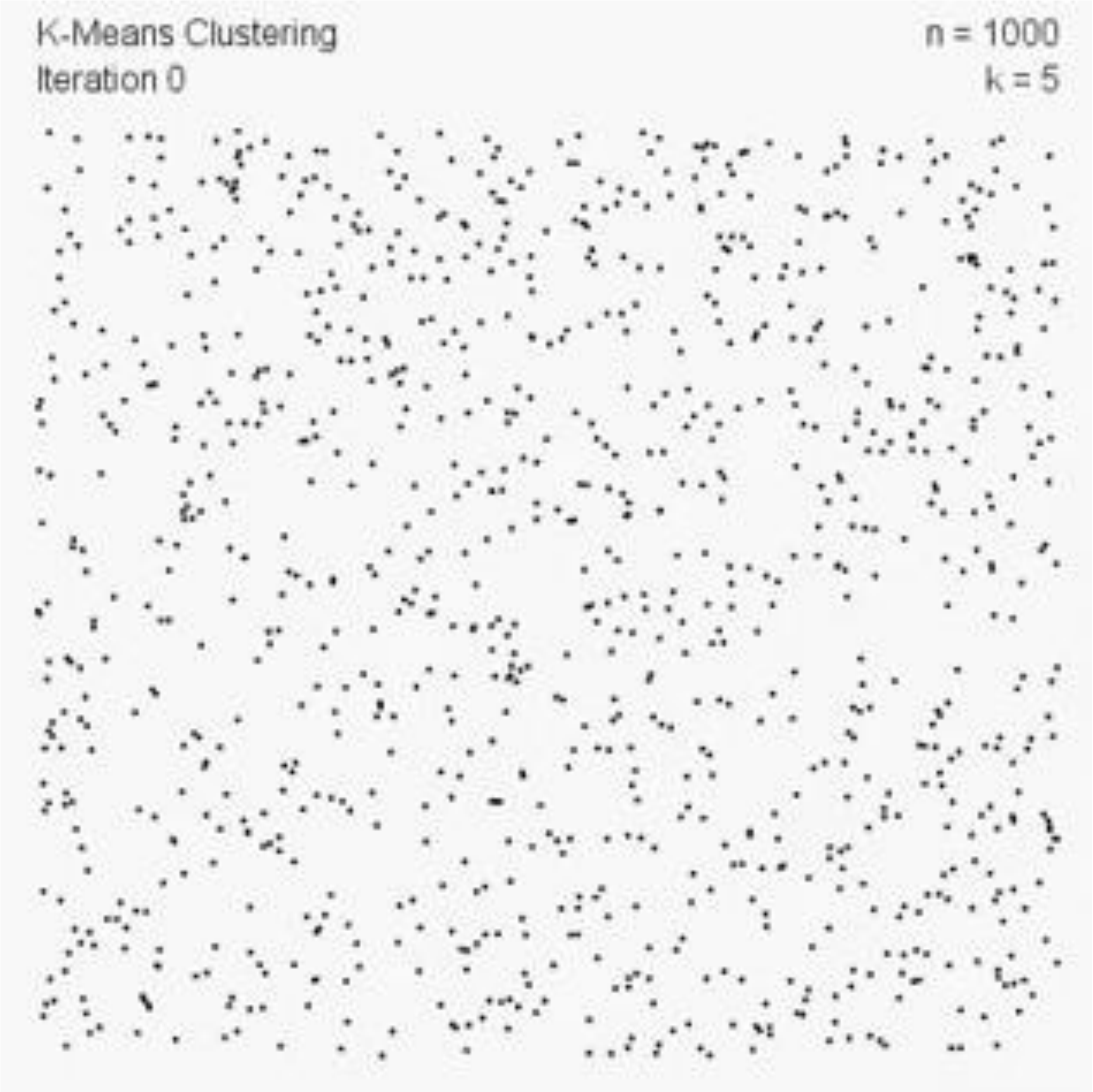
Naïve approach

- Given large data set E, get a small random subset D from E
- Cluster D and use found representatives for initialization

Improved approach

- Given large data set E, get m small random subsets A ... $M \subset E$
- Cluster A to M and save representatives R_A ... R_M
- Cluster merged set $AM = A \cup \cdots \cup M, m$ times with R_A ... R_M as initial representatives
- Select representation from (R_A ... R_M), which yielded best clustering result on AM, use as initial representation for E

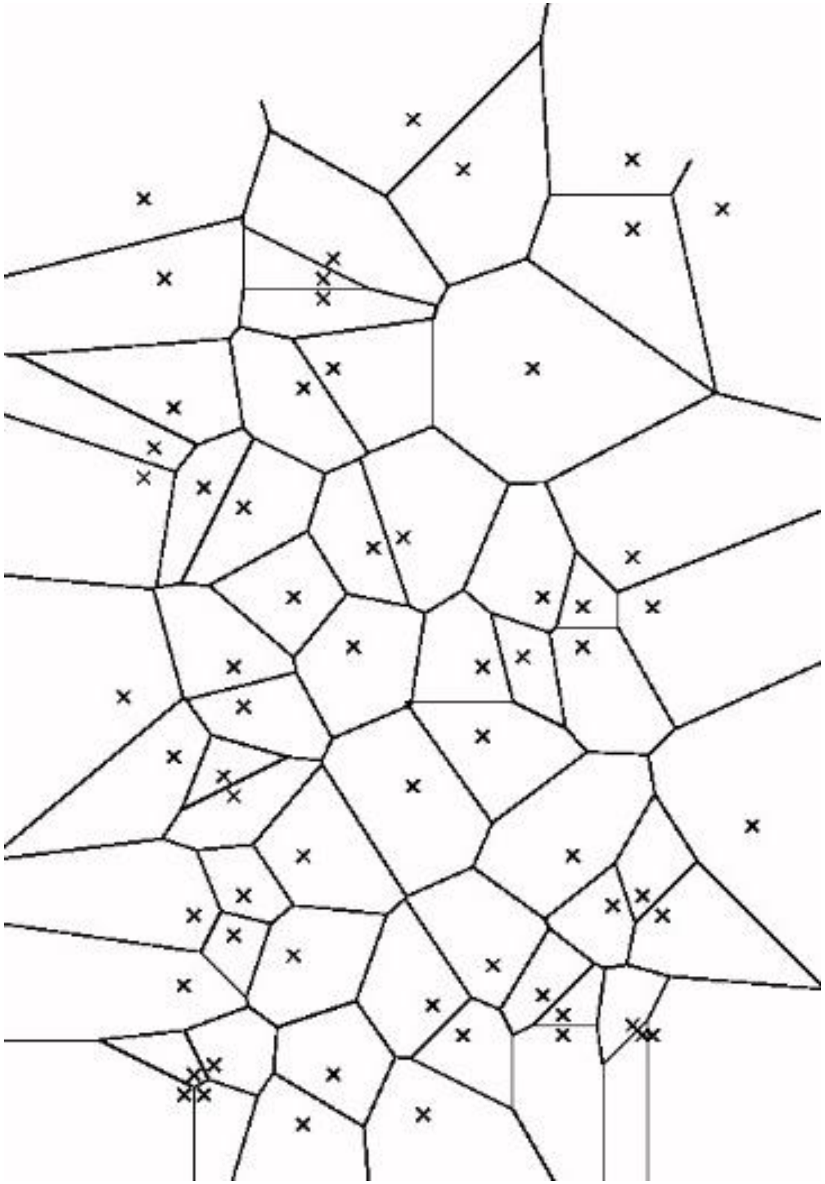
K-Means Example



[10]

Voronoi Model

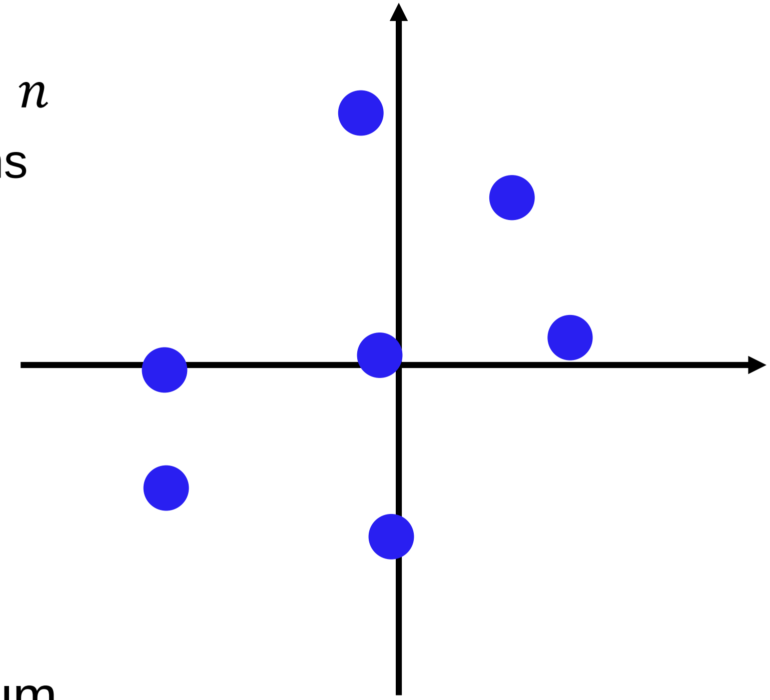
- The Voroni diagramm partitions space in convex Voroni cells for each point
- Voroni cell for a point covers the area which nearest to this data point



Discussion K-Means

Pro

- **Efficiency**: $\mathcal{O}(tkn)$ with typically $k, t \ll n$
  - n = #objects, k = #cluster, t = #iterations
- **Implementation:** Easy to use



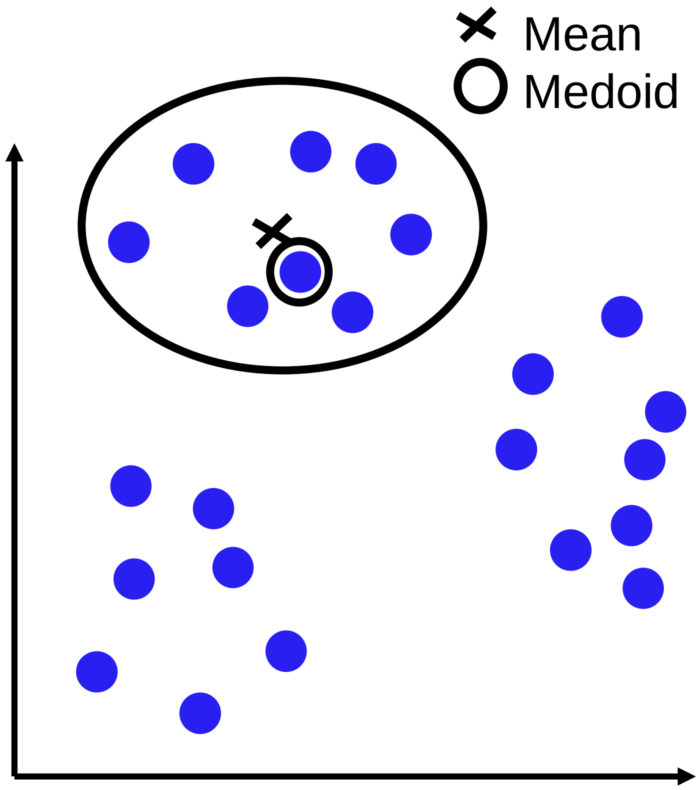
Contra

- **Applicability**: mean must exist
- **Noise:** Sensitive to outliers
- **Specification:** k must be defined
- **Initialization:** Might run in local optimum
- **Cluster Form:** Convex space partitions

Additional Slide

K-means is a widely used method for clustering where the number of clusters must be specified. The basic idea is to define k representants in the data and then assign the elements of the data to the representatives which are closest to them. After each step, the mean value of all elements in a cluster is determined and selected as the new representative of the cluster. Thus it can be that the affiliation of one element to the respective cluster changes. Once the change is stagnated, k-means is complete. It is important to choose suitable representatives when initializing, otherwise the algitmus can get stuck in a local minima.

Variants – K-Medoids, K-Median Clustering



- Representative: Mean → Object from cluster
  - Means do not always exist
- Variants for representative:
  - Medoid: Existing object in “middle”
  - Median: Median of (sorted) cluster
- Influence of outliers increases if squared distance is used → use normal distance

Discussion k-Means, k-Medoid & k-Median

	K-means	K-medoid	K-median
Data	Numerical data to compute mean	metric	ordered data
Efficiency	High $O(tkn)$	Low $O(tk(n-k)^2)$	High $O(tkn)$
Sensitivity to outliers	High	Low	Low

Pro

- **Implementation:** Easy to use

Contra

- **Specification:** k must be defined
- **Cluster Form:** Convex space partitions
- **Initialization:** Might run in local optimum

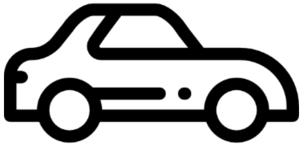
n : #objects, k : #cluster, t : #iterations

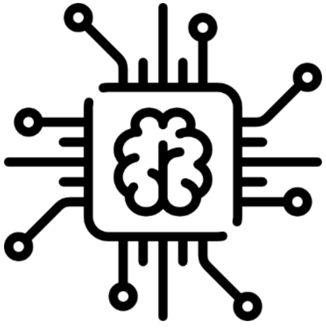
Unsupervised Learning: Clustering
Prof. Dr.-Ing. Markus Lienkamp

(Andreas Schimpe, M.Sc.)

Agenda

1. Chapter: Introduction
   - 1.1 Overview
   - 1.2 Training and Validation
2. Chapter: Methods
   - 2.1 Hierarchical Clustering
   - 2.2 k-means
   - **2.3 DBSCAN**
3. Chapter: Application
4. Chapter: Summary





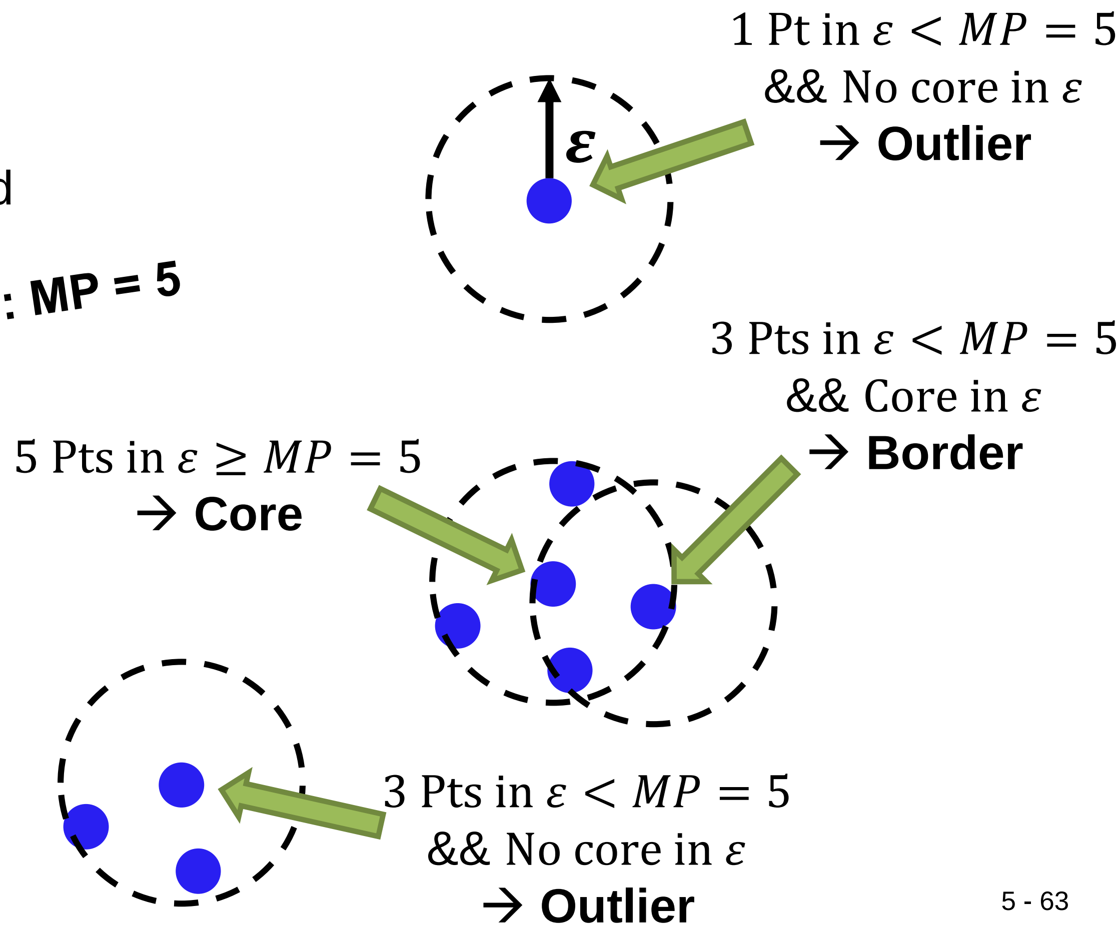
Density-Based Clustering: DBSCAN

- Density-Based Spatial Clustering Application with Noise

- Two parameters
  - $\mathcal{E}$-radius neighborhood
  - Minimum Points (MP)

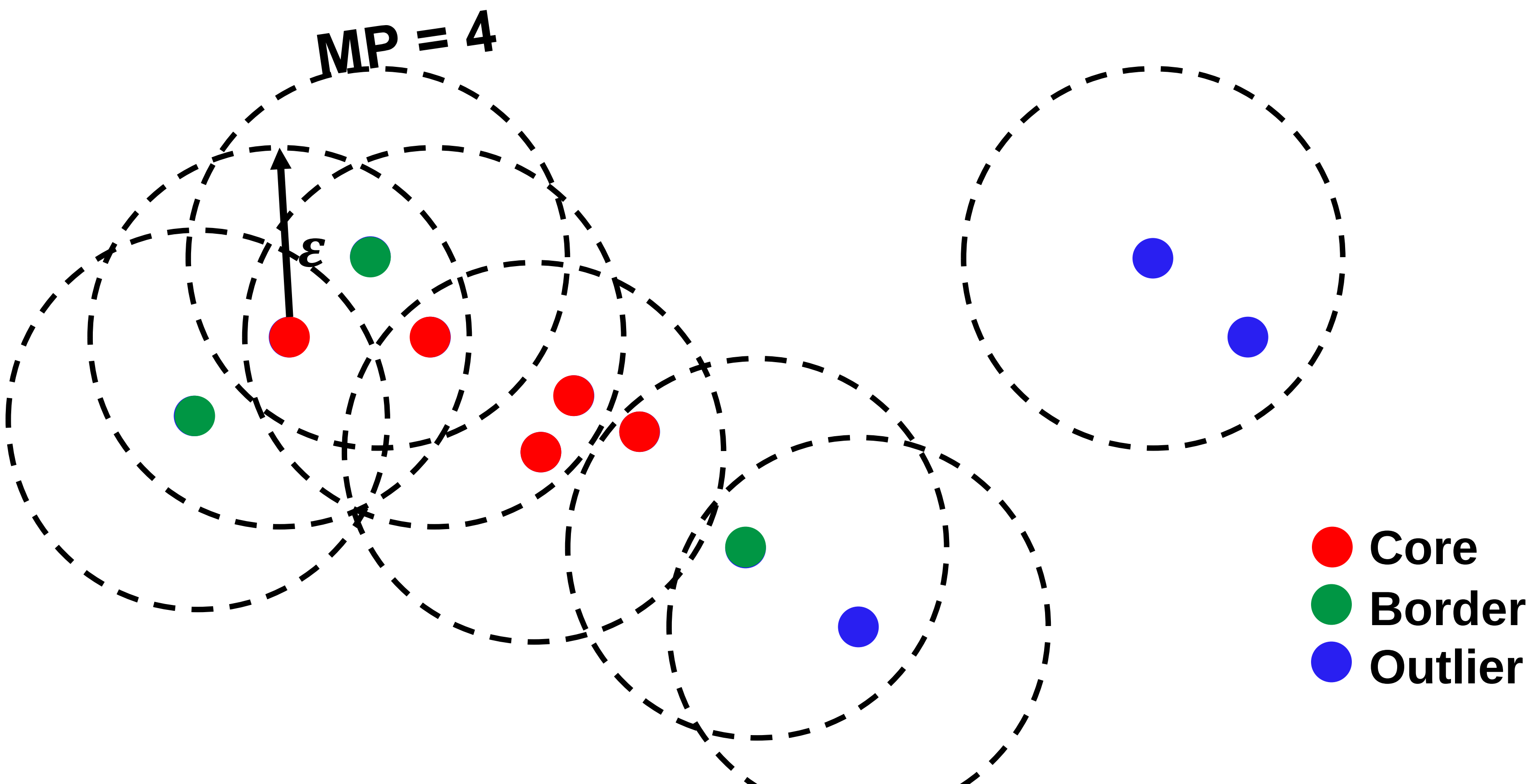
Example: MP = 5

- Three Point-classes
  - Core
  - Border
  - Outlier



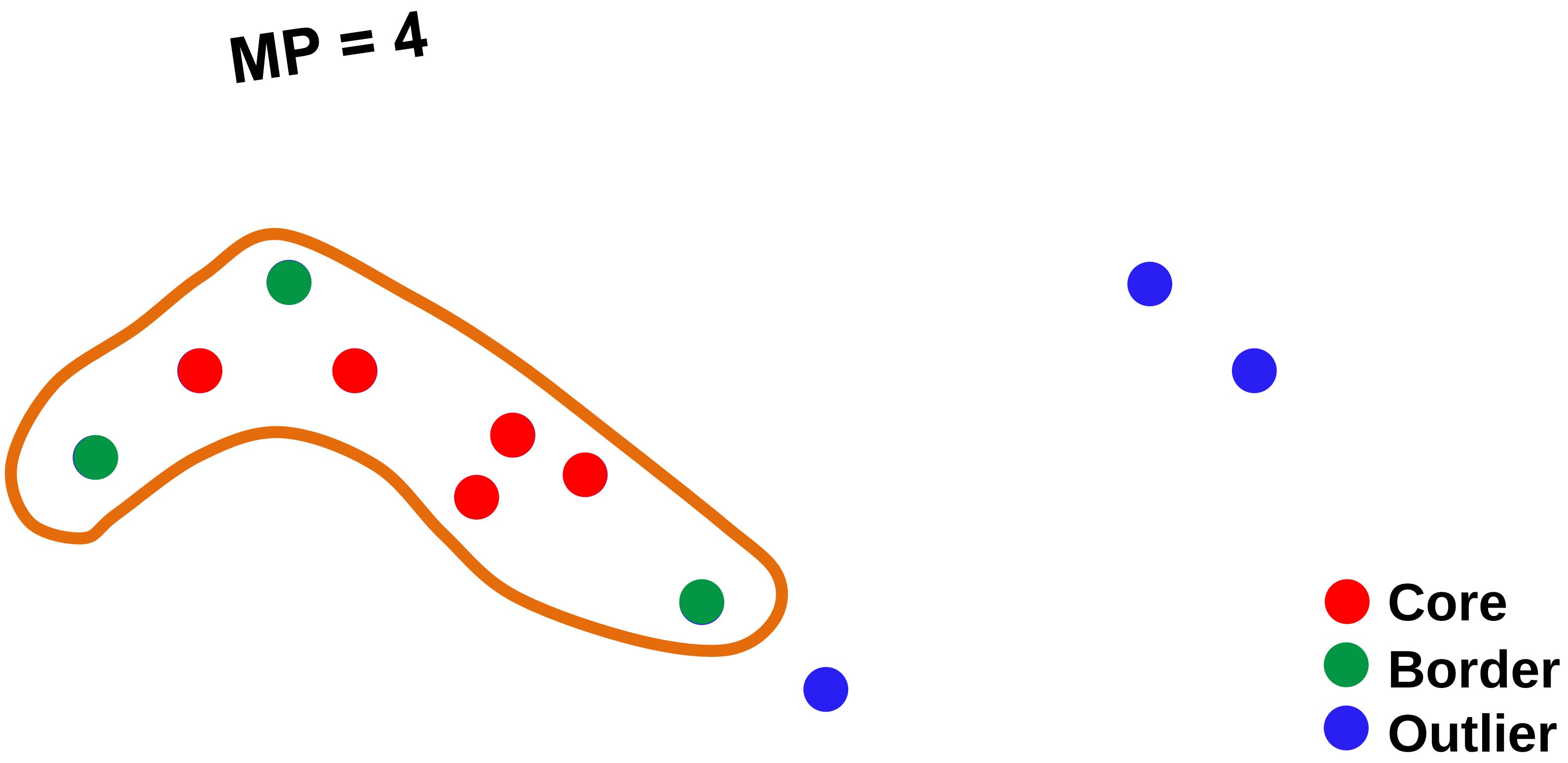
DBSCAN – Density Reachability

- p_n is „reachable“ from p_1, if there is a path p_1 ... p_n where each p_i on the path must be a core point, except for p_n

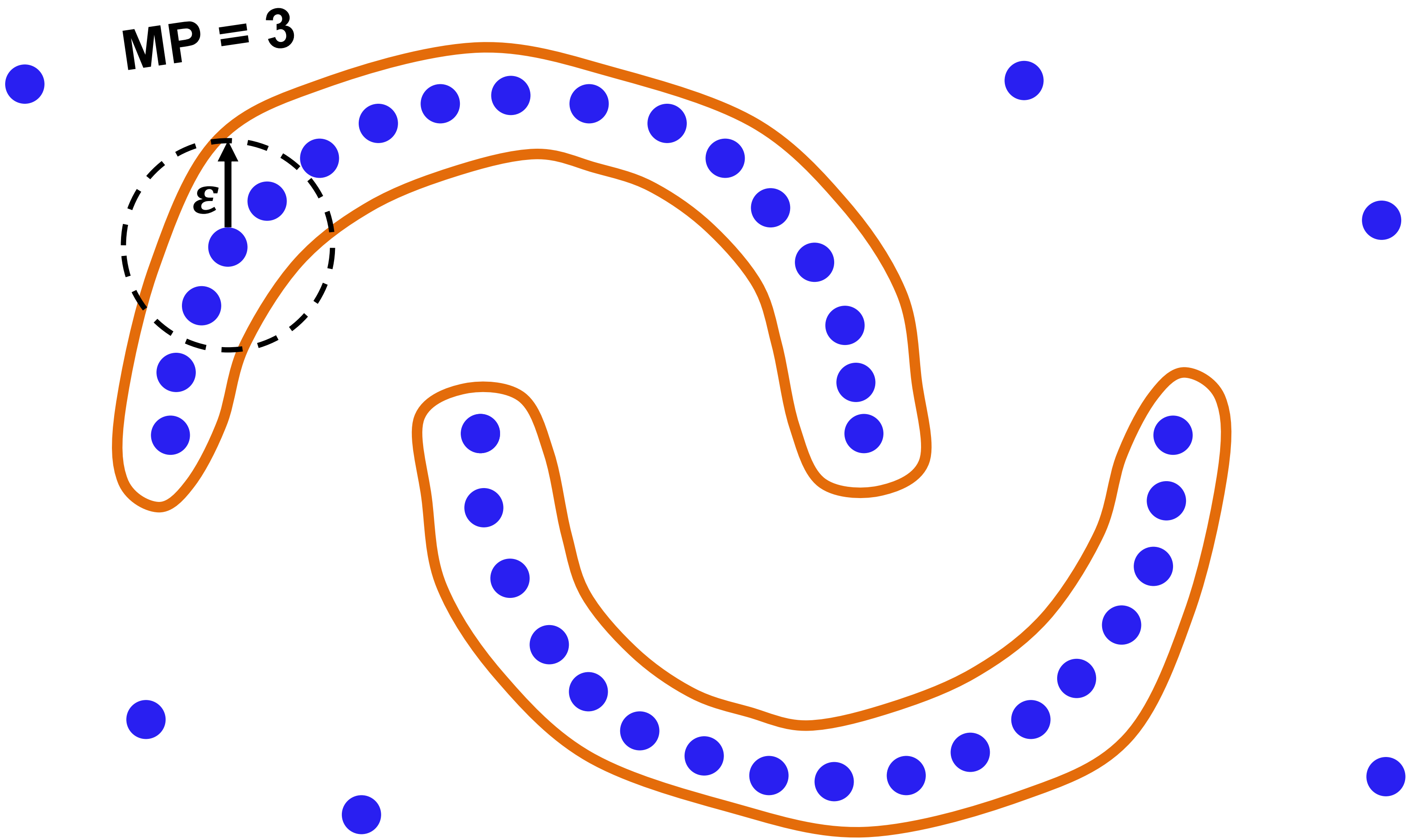


DBSCAN – Density Reachability

- p_n is „reachable“ from p_1, if there is a path p_1 ... p_n where each p_i on the path must be a core point, except for p_n



DBSCAN – Example



Discussion DBSCAN

Pro

- **Cluster Form:** Arbitrary (convex and non-convex) space partitions
- **Specification:** k is determined automatically
- **Noise:** Separates clusters from noise
- **Efficiency**: $\mathcal{O}(n^2)$

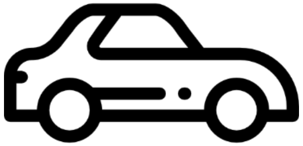
Contra

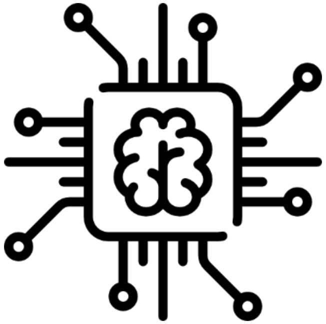
- **Specification:** Parameters difficult to determine
- **Sensitivity:** Very sensitive to parameter changes

Unsupervised Learning: Clustering
Prof. Dr.-Ing. Markus Lienkamp

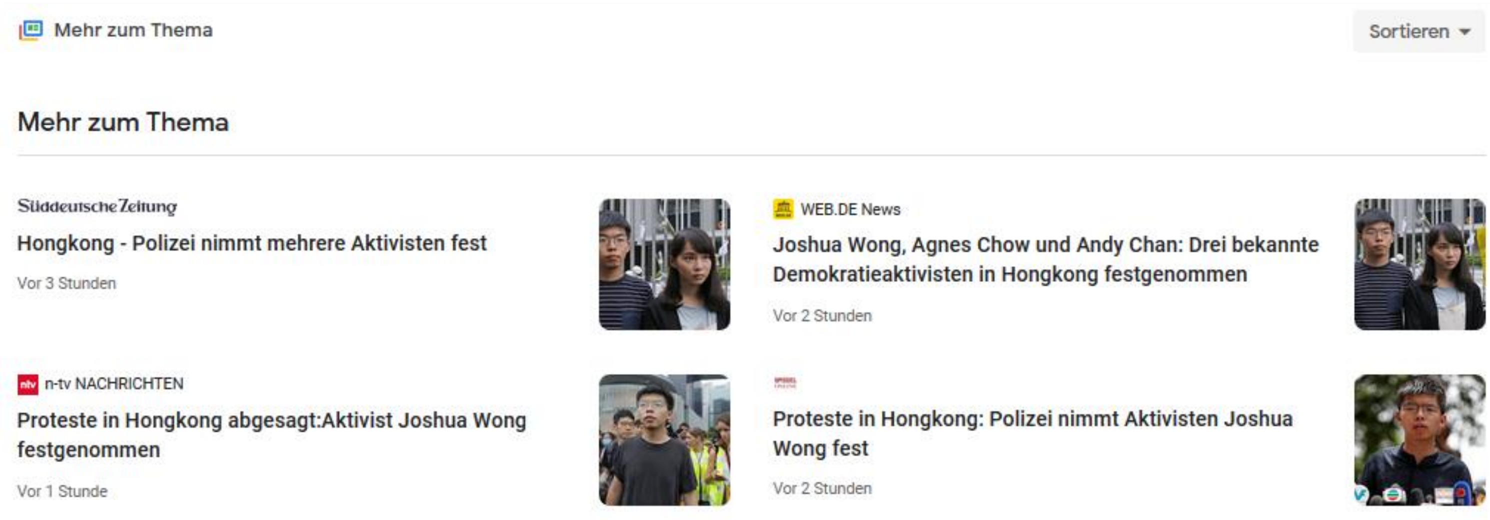
(Andreas Schimpe, M.Sc.)

Agenda



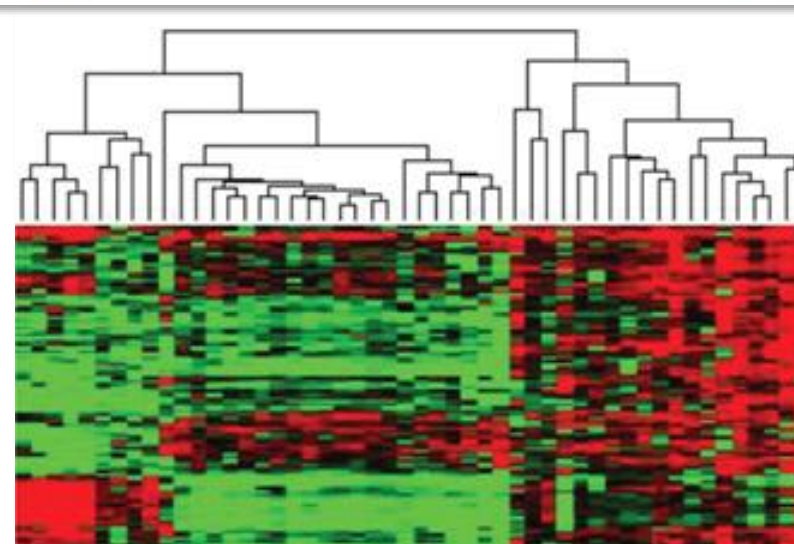


Applications



Google News

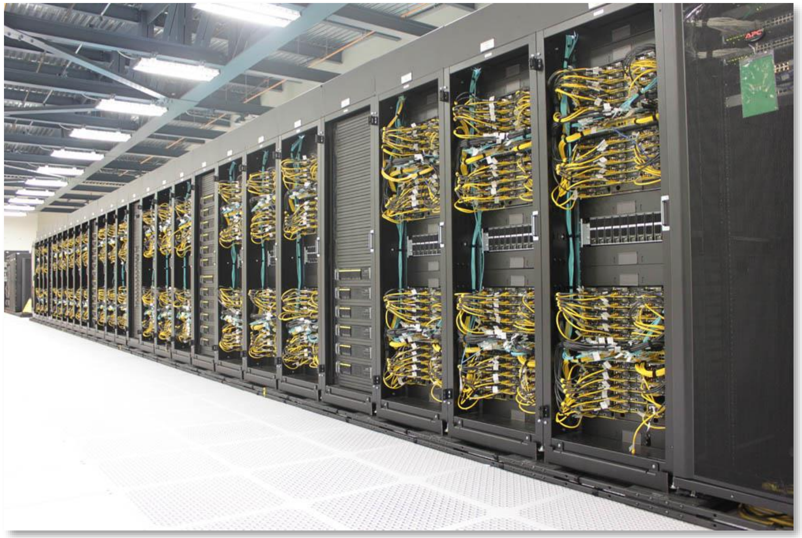
Genome Pattern



[14]

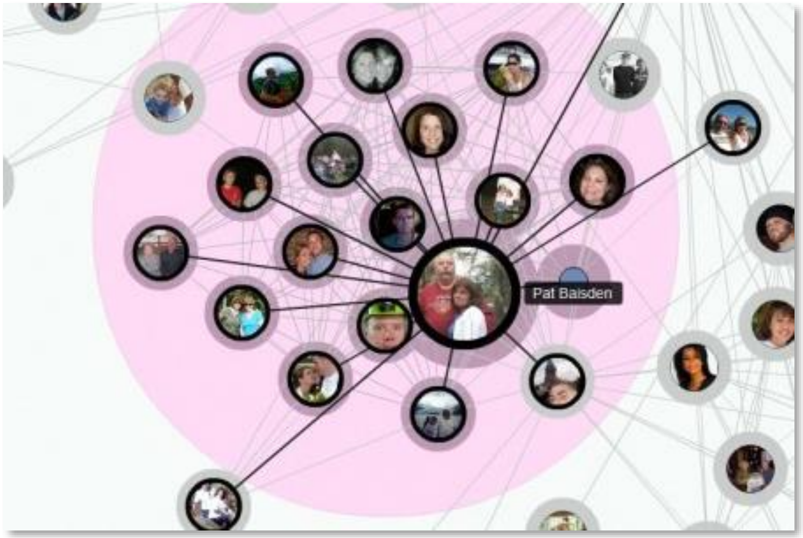
People

Applications



[11]

Computing Cluster



[12]

Sozial Network

Market Segmentation

[13]

Applications

- Customer Clustering
  - Amazon: Product suggestion (personalized advertisment)
  - Netflix: Movie suggestion

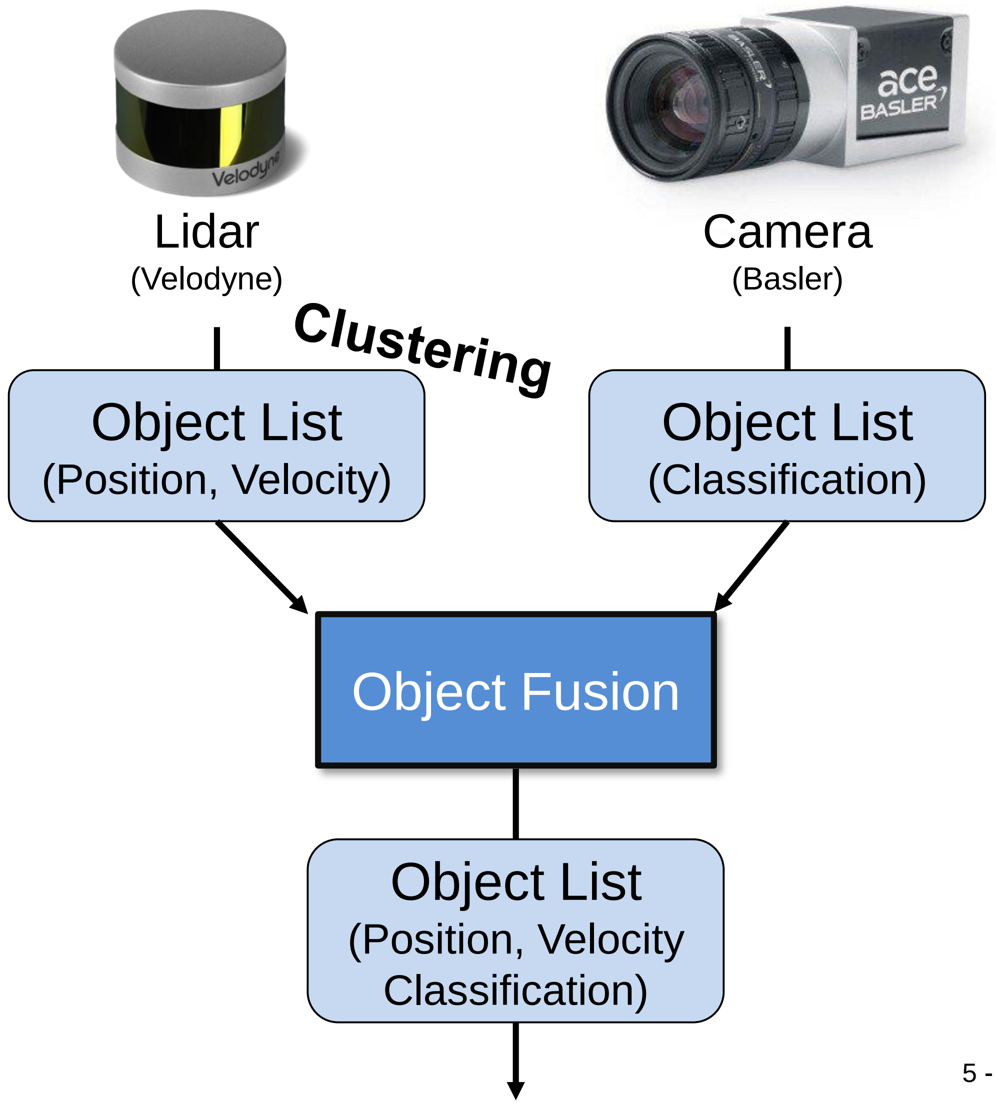
  - Netflix 1,000,000 $ challenge from 2006

Clustering for Automotive Technology

- Traffic analysis
  - Collect mobility data of cars
    or density of certain regions
  - Use cluster algorithm to identify different groups
    - e.g., commuter, points of interest
  - Extract generalization of trajectories and traffic flow
  - Use knowledge for city planning and to identify bottlenecks

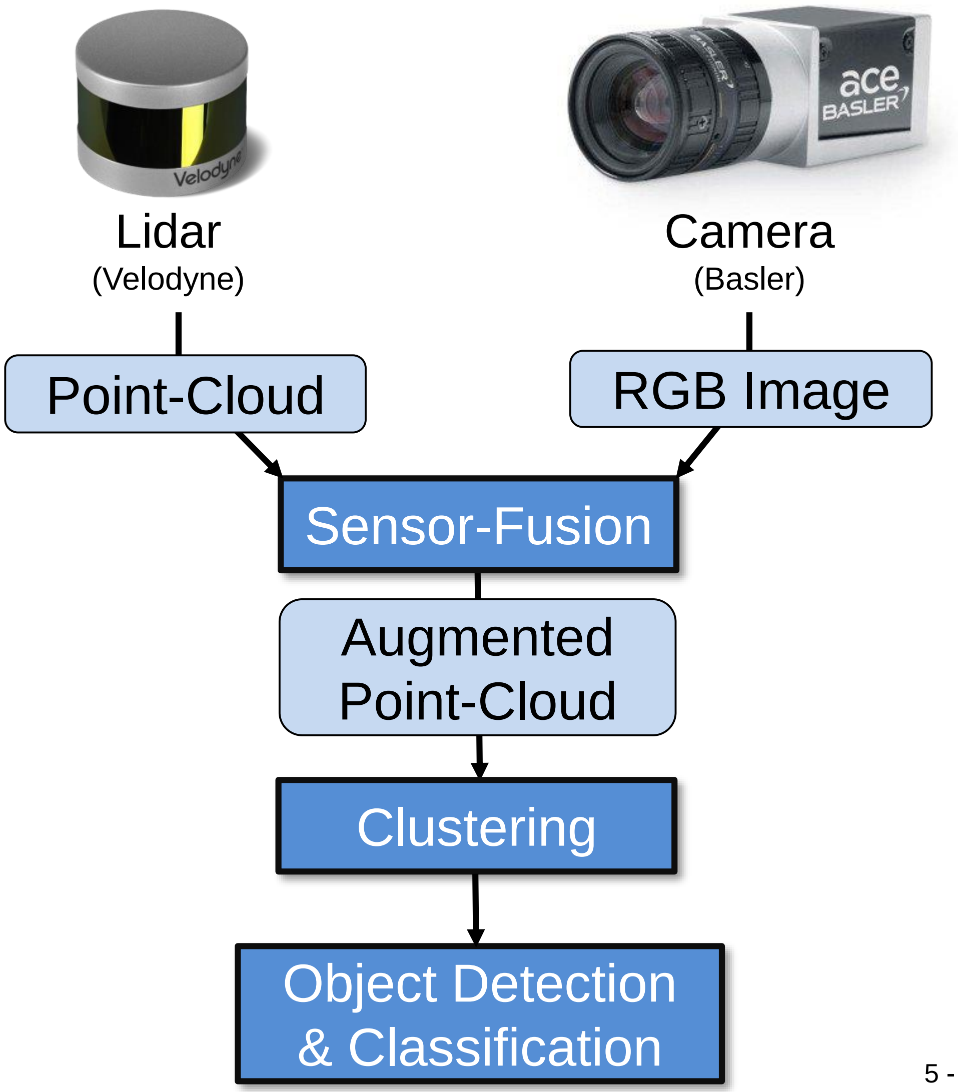
Clustering for Automotive Technology

- High Level Fusion
  - Object detection and classification based on limited data (only from one sensor at a time)
  - Fusion based on processed object lists (potential information loss)



Clustering for Automotive Technology

- Low Level Fusion
  - Overlay Lidar point-cloud data with camera image
  - Find clusters in augmented point-cloud
  - Object detection and classification based on fused raw data

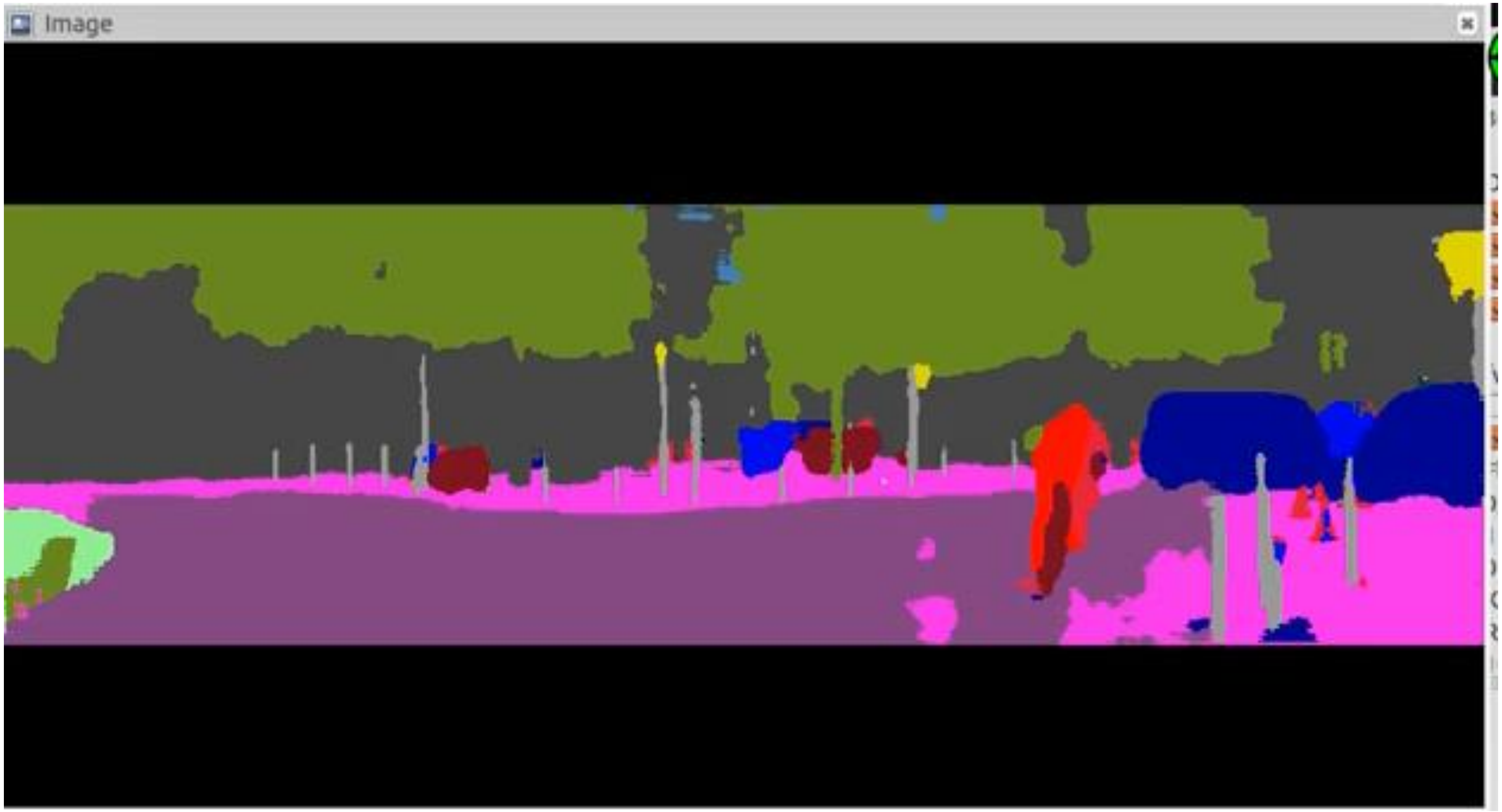


Object Detection and Classification Pipeline

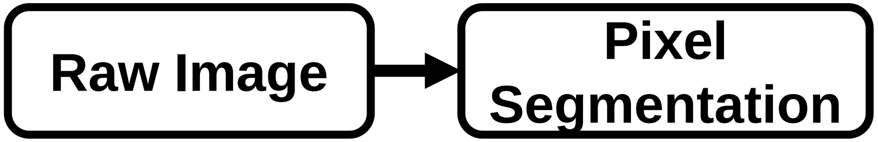
[4]

Raw Image

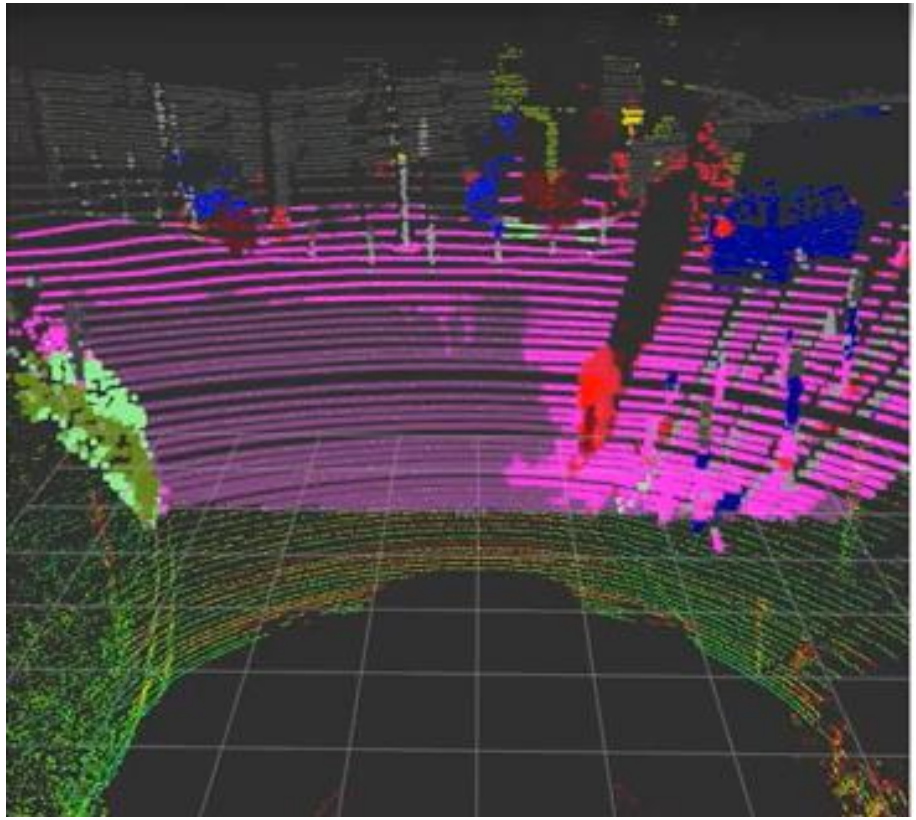
Object Detection and Classification Pipeline



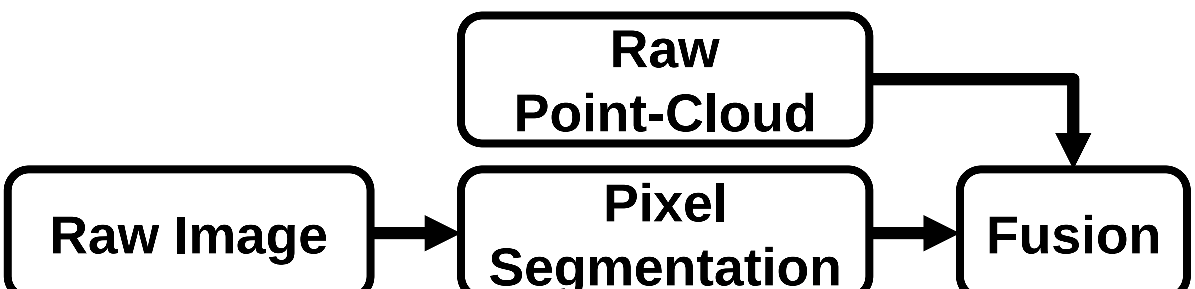
[4]



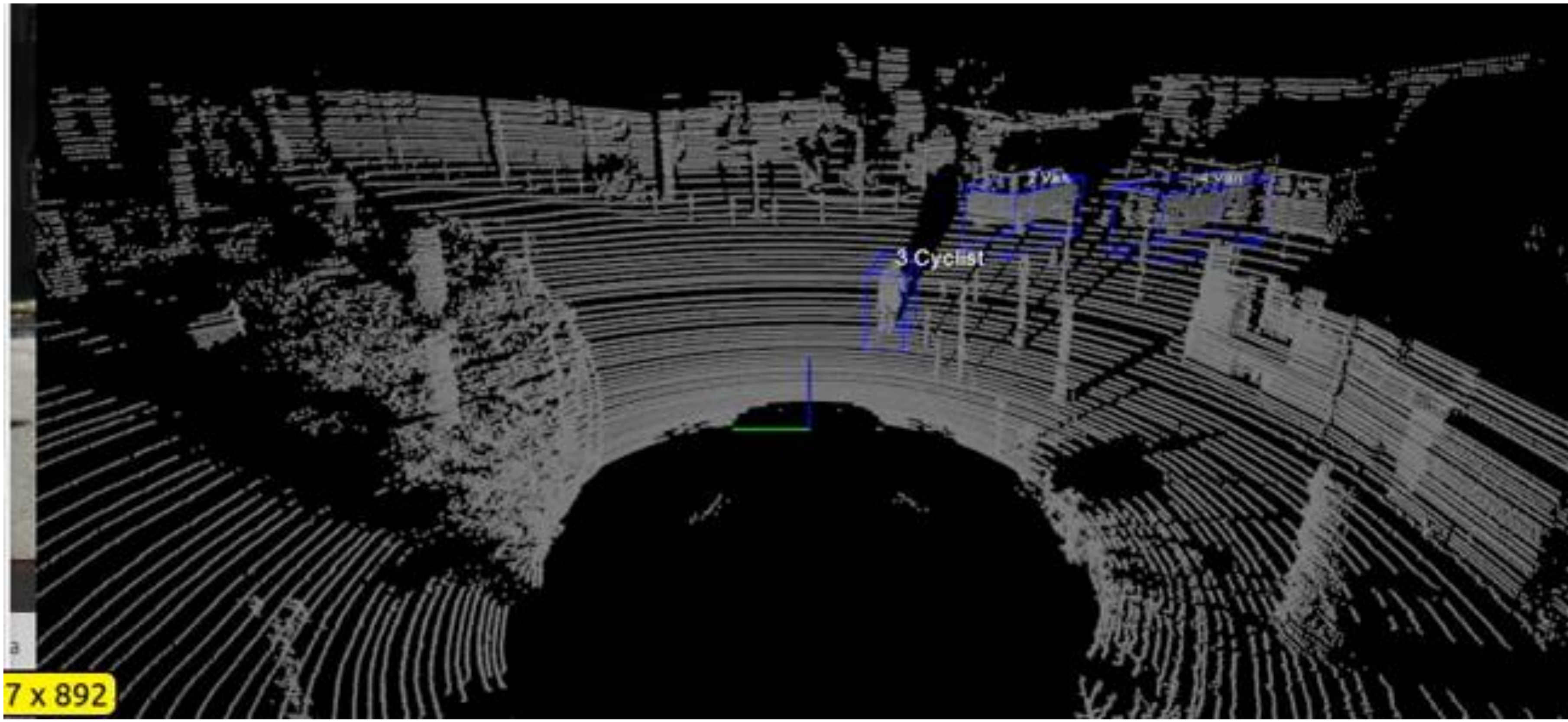
Object Detection and Classification Pipeline



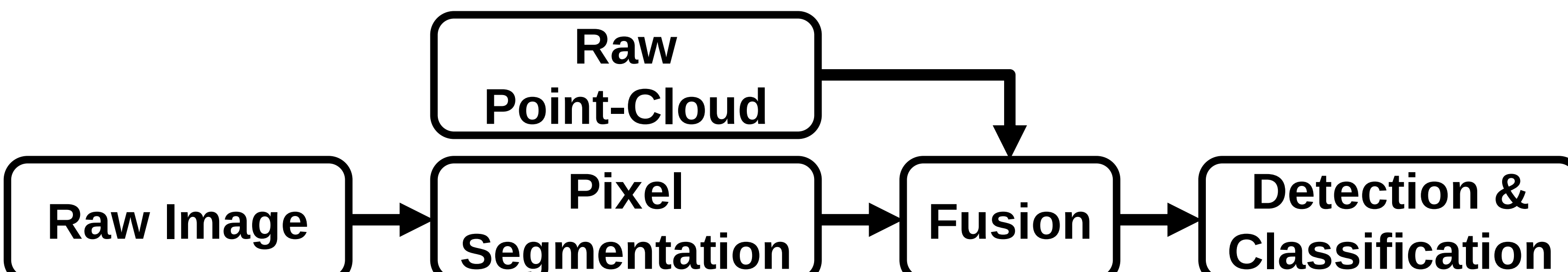
[4]



Object Detection and Classification Pipeline



[4]

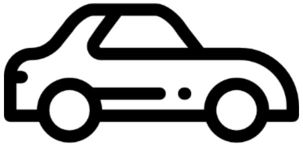


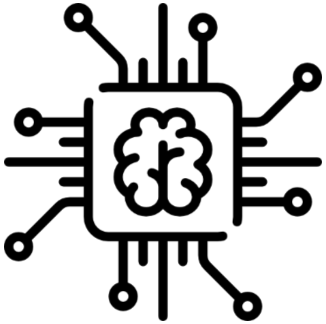
Unsupervised Learning: Clustering
Prof. Dr.-Ing. Markus Lienkamp

(Andreas Schimpe, M.Sc.)

Agenda

1. Chapter: Introduction
   - 1.1 Overview
   - 1.2 Training and Validation
2. Chapter: Methods
   - 2.1 Hierarchical Clustering
   - 2.2 k-means
   - 2.3 DBSCAN
3. Chapter: Application
4. **Chapter: Summary**



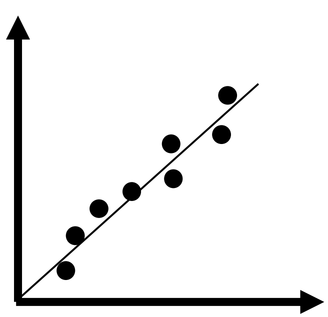


Summary

Pattern Recognition

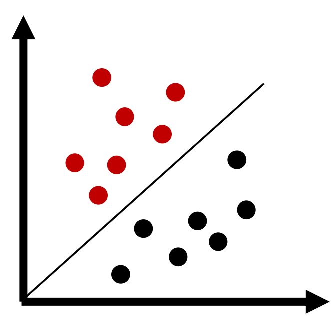
Regression

- Predict **continuous** valued output
- Supervised



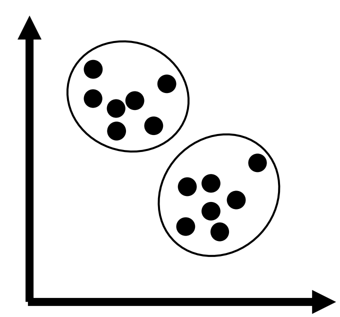
Classification

- Predict **discrete** valued output
- Supervised



Clustering

- Predict discrete valued output
- **Unsupervised**



Summary

What did we learn today:

- **Clustering** is about finding structures/groups in a dataset.
- Clustering is an **optimization problem** (Minimize variability).
- Elements **within** a cluster are **similar**.
- Elements from **different** clusters are **dissimilar**.
- The **distance** can be used to express similarity.
- Clustering is an **unsupervised** method. No labels are required.
- The **silhouette** can be used to express the **quality** of clustering.
- Segmentation and clustering are **interchangeable** terms.
- The concepts of **hierarchical clustering**, **k-means** and **DBSCAN**.
- Hierarchical clustering builds a **dendrogram**.
- The number of desired **clusters** can be selected **afterwards.**

Summary

What did we learn today:

- **K-means** is a fast, but greedy and non-deterministic algorithm.
- The **number of clusters** must be selected **beforehand**.
- Only **convex space** partitions can be generated.
- DBSCAN is a **density-based** method and can deal with **noise**.
- Elements are classified as **core**, **border** or **outlier** elements.
- **Arbitrary** (convex and non-convex) **shapes** can be clustered.
- Clustering is applied as **pre-processing** or to find **coherences**.
- Wide range of **clustering applications**, but rarely as stand alone.
- Experts or classification methods **give clusters meaning afterwards**.

Sources

- [1] https://dailyillini.com/news/2017/09/28/students-reflect-race-affects-classroom-participation/
- [2] https://dictionary.cambridge.org/dictionary/english/cluster
- [3] https://www.deviantart.com/gttorres/art/Not-Another-High-School-Story-256629151
- [4] https://www.youtube.com/watch?v=xXWLXfMugkM
- [5] http://www.dbs.ifi.lmu.de/Lehre/KDD/WS1718/04_Clustering-3.pdf
- [6] http://www.instituteofcaninebiology.org/how-to-read-a-dendrogram.html
- [7] https://de.wikipedia.org/wiki/Silhouettenkoeffizient
- [8] https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-0002-introduction-to-computational-thinking-and-data-science-fall-2016/lecture-slides-and-files/MIT6_0002F16_lec12.pdf
- [9] http://docs.w3cub.com/scikit_learn/auto_examples/cluster/plot_kmeans_silhouette_analysis/
- [10] https://www.youtube.com/watch?v=BVFG7fd1H30
- [11] https://computing.llnl.gov/tutorials/linux_clusters/
- [12] http://www.messersmith.name/wordpress/2009/10/12/visualizing-your-facebook-network-of-friends/
- [13] http://www.businessstudynotes.com/marketing/principle-of-marketing/discuss-market-segmentation-and-market-targetting/
- [14] http://www.discoveryandinnovation.com/BIOL202/notes/images/cluster_analysis.jpg
- [15] https://www.nytimes.com/2017/05/25/automobiles/wheels/lidar-self-driving-cars.html

Acknowledgment

- **Machine Learning (Stanford/Coursera)**
  - Andrew Ng

  https://www.coursera.org/learn/machine-learning

- **Knowledge Discovery in Databases I (LMU)**
  - Prof. Dr. Peer Kröger

  http://www.dbs.ifi.lmu.de/cms/studium_lehre/lehre_master/kdd1718/index.html

- **Introduction to Computational Thinking and Data Science (MIT)**
  - Prof. Eric Grimson

  https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-0002-introduction-to-computational-thinking-and-data-science-fall-2016