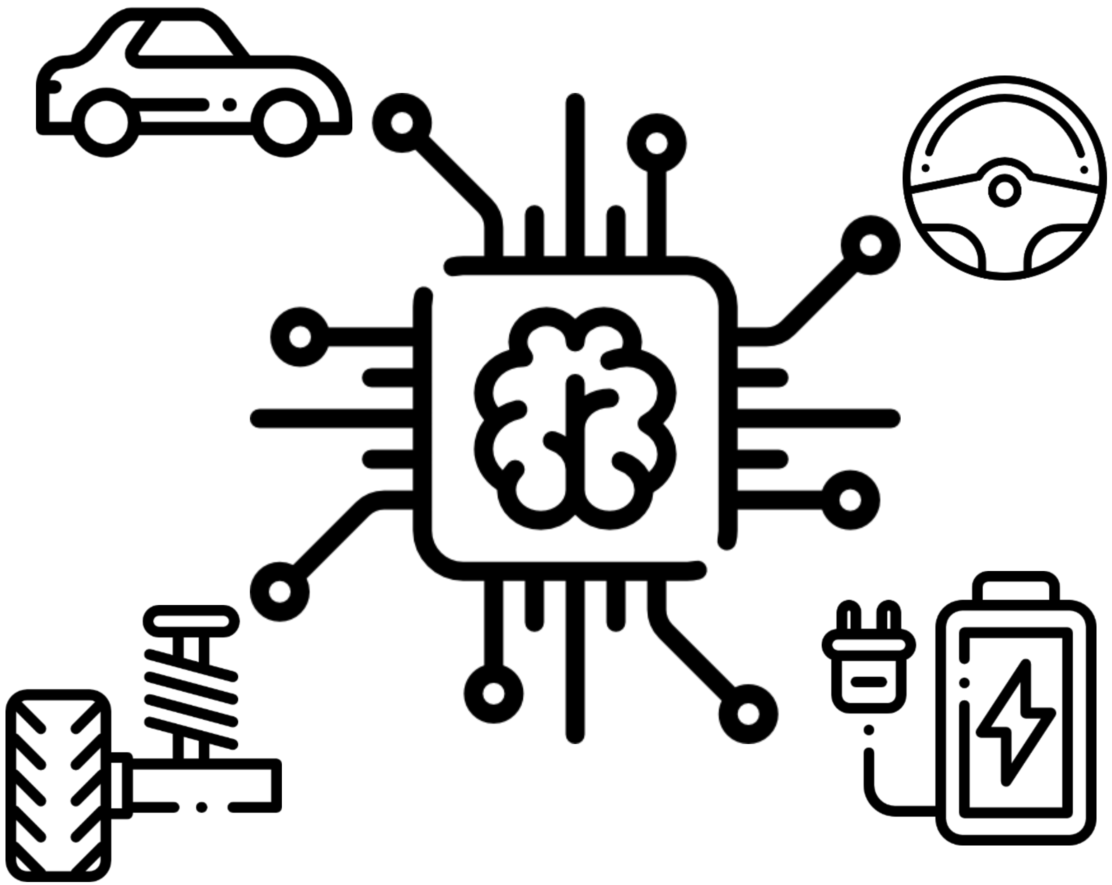
Lehrstuhl für Fahrzeugtechnik
Fakultät für Maschinenwesen
Technische Universität München

Artificial Intelligence in Automotive Technology

Johannes Betz / Prof. Dr.-Ing. Markus Lienkamp / Prof. Dr.-Ing. Boris Lohmann

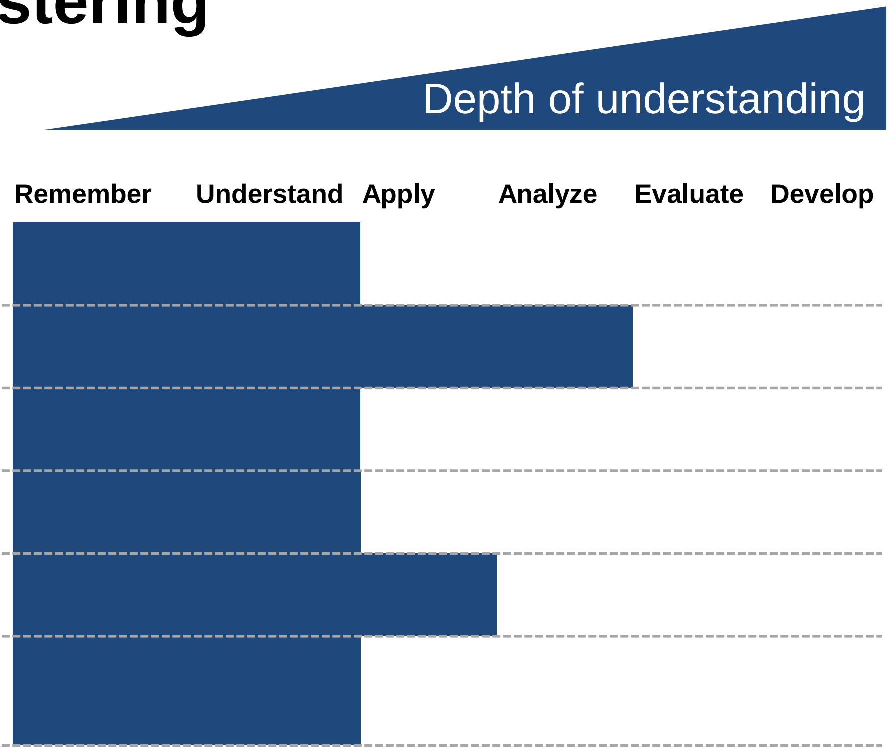


Lecture Overview

Lecture 16:15 – 17:45	Practice 17:45 – 18:30
1 Introduction: Artificial Intelligence 17.10.2019 – Johannes Betz	**Practice 1** 17.10.2019 – Johannes Betz
2 Perception 24.10.2019 – Johannes Betz	**Practice 2** 24.10.2019 – Johannes Betz
3 Supervised Learning: Regression 31.10.2019 – Alexander Wischnewski	**Practice 3** 31.10.2019 – Alexander Wischnewski
4 Supervised Learning: Classification 7.11.2019 – Jan Cedric Mertens	**Practice 4** 7.11.2019 – Jan Cedric Mertens
5 Unsupervised Learning: Clustering 14.11.2019 – Jan Cedric Mertens	**Practice 5** 14.11.2019 – Jan Cedric Mertens
6 Pathfinding: From British Museum to A* 21.11.2019 – Lennart Adenaw	**Practice 6** 21.11.2019 – Lennart Adenaw
7 Introduction: Artificial Neural Networks 28.11.2019 – Lennart Adenaw	**Practice 7** 28.11.2019 – Lennart Adenaw
8 Deep Neural Networks 5.12.2019 – Jean-Michael Georg	**Practice 8** 5.12.2019 – Jean-Michael Georg
9 Convolutional Neural Networks 12.12.2019 – Jean-Michael Georg	**Practice 9** 12.12.2019 – Jean-Michael Georg
10 Recurrent Neural Networks 19.12.2019 – Christian Dengler	**Practice 10** 19.12.2019 – Christian Dengler
11 Reinforcement Learning 09.01.2020 – Christian Dengler	**Practice 11** 09.01.2020 – Christian Dengler
12 AI-Development 16.01.2020 – Johannes Betz	**Practice 12** 16.01.2020 – Johannes Betz
13 Guest Lecture: VW Data:Lab 23.01.2020 –	

Objectives for Lecture 5: Clustering

After the lecture you are able to…



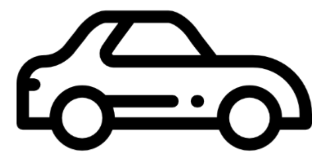
- … understand the concept of clustering and its association to pattern recognition.
- … analyze the quality of given clusters regarding to different criteria.
- … understand the workflow of unsupervised learning.
- … understand the concepts of different clustering methods together with their pros and cons.
- … implement, train and use a clustering method with python libraries.
- … identify if a problem belongs to regression, classification or clustering.

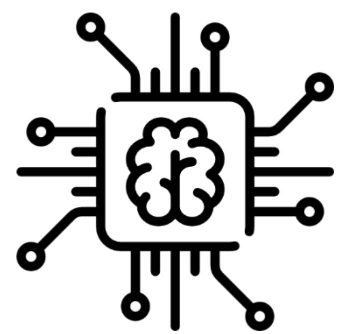
Supervised Learning: Classification
Johannes Betz / Prof. Dr. Markus Lienkamp / Prof. Dr. Boris Lohmann

(Jan Cedric Mertens, 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





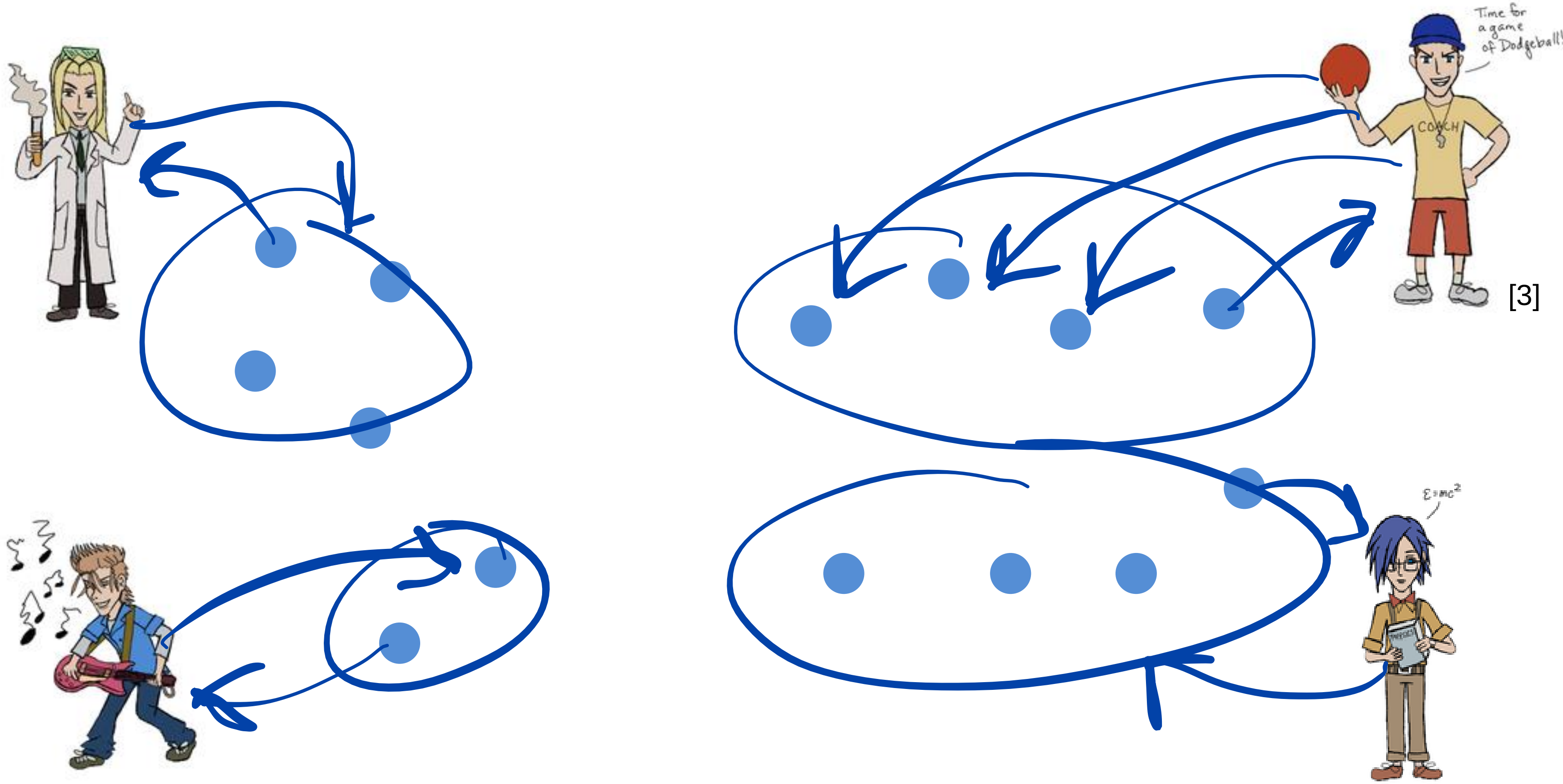
Clustering

"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]

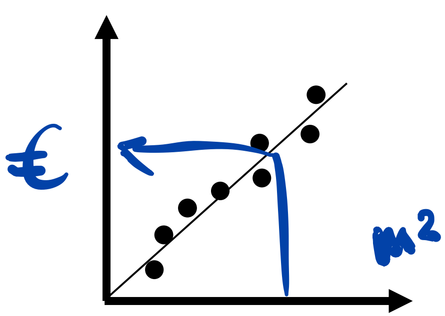
Kommentarfolie

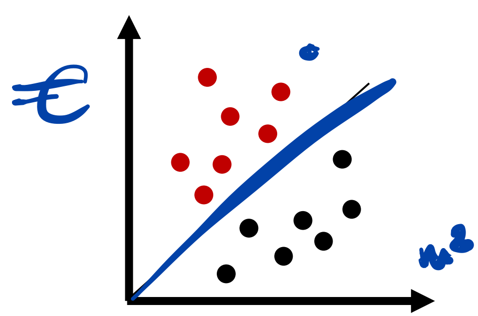
Clustering, independent of machine learning, deals with identifying similarities and differences between different data points and dividing the data into groups. In contrast to classification, where a model is learned and later applied to other data, clustering usually starts with all data present (without 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 preprocessing 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.

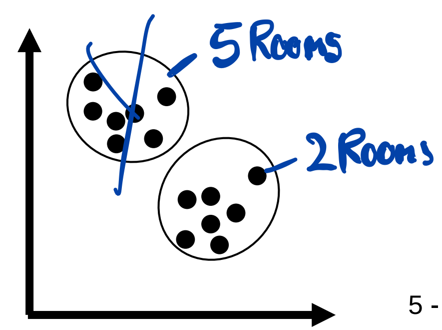
Method Overview

Pattern Recognition

Regression	Classification	**Clustering**
• Predict **continuous** valued output • Supervised	• Predict **discrete** valued output • Supervised	• Predict discrete valued output • **Unsupervised**

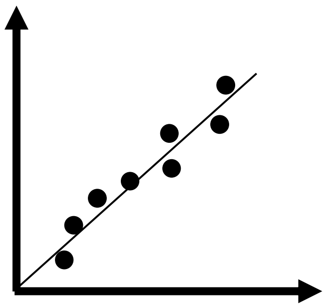






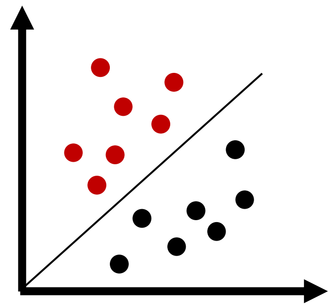
Method Overview

Regression



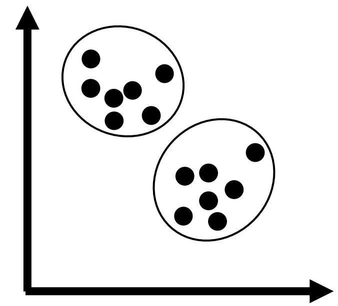
- House pricing
- Number of sales
- Persons weight

Classification



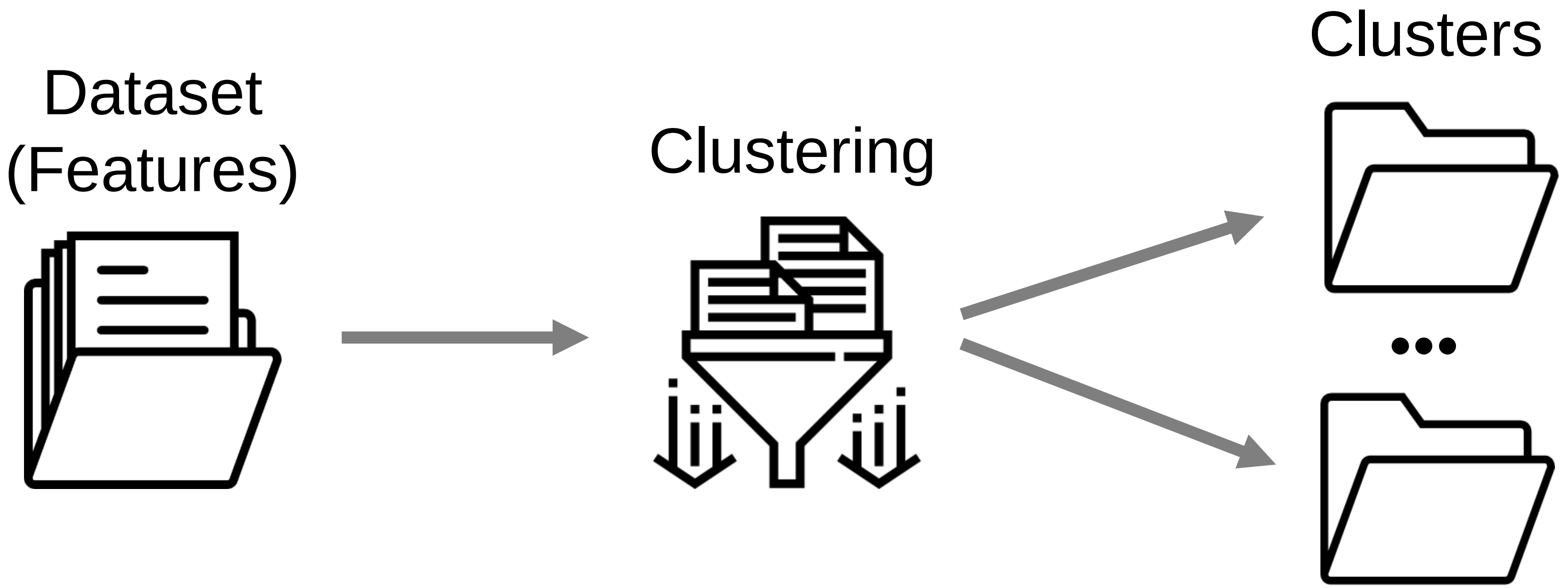
- Object detection
- Spam detection
- Cancer detection

Clustering



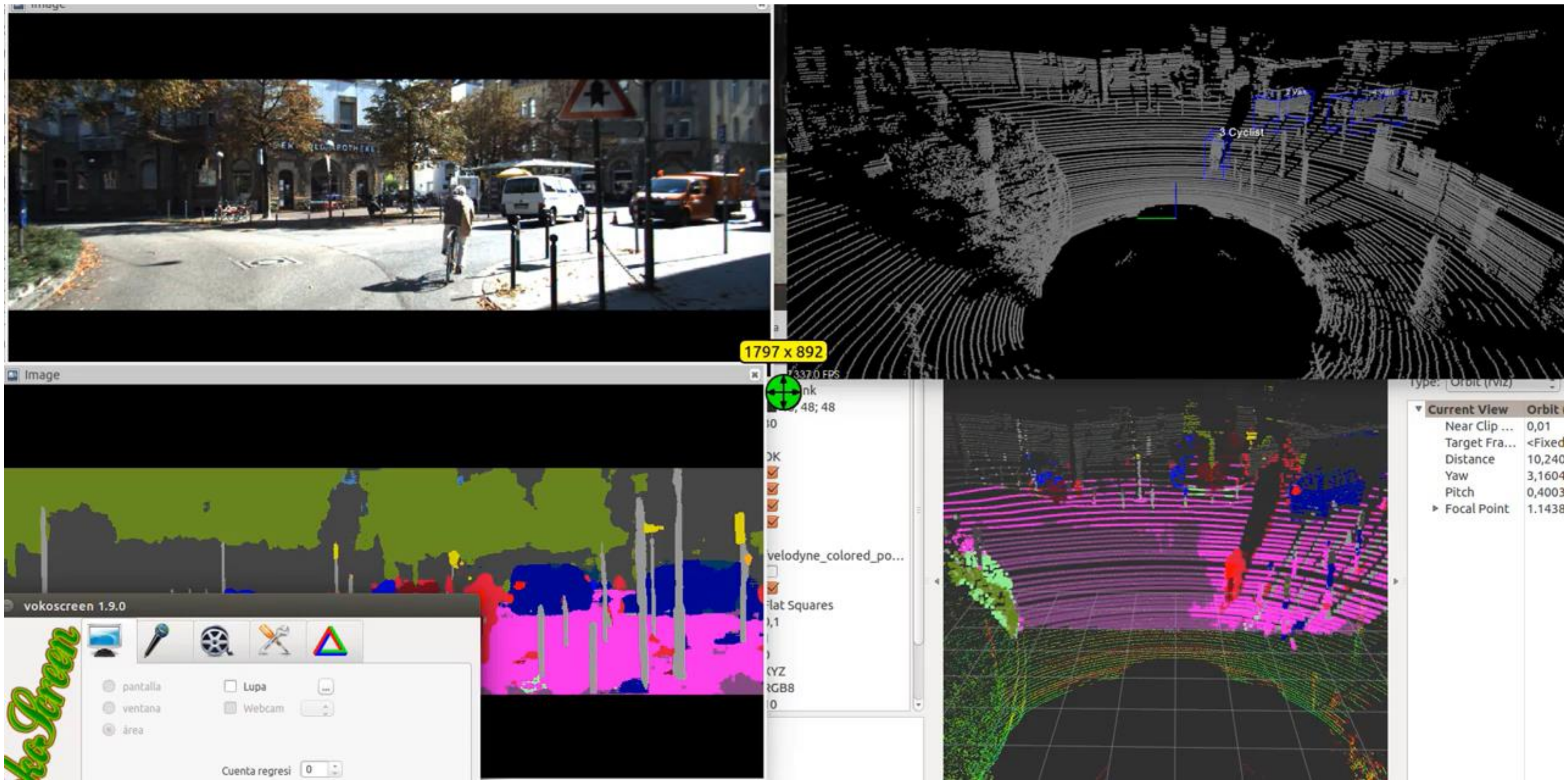
- Genome patterns
- Google news
- Point cloud (lidar) processing

General Approach



News (keywords, …)	Similarities?	Clusters
Genomes (size, …)	Similarities?	Clusters
Points (position, …)	Similarities?	Clusters

Clustering - Example



[4]

Cluster Points → Classifiy subset of cluster → Assign class to cluster

Clustering vs. Segmentation

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

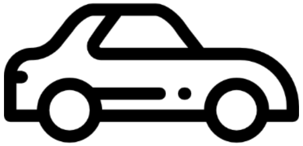
Supervised Learning: Classification

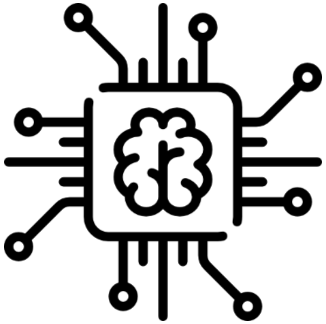
Johannes Betz / Prof. Dr. Markus Lienkamp / Prof. Dr. Boris Lohmann

(Jan Cedric Mertens, 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

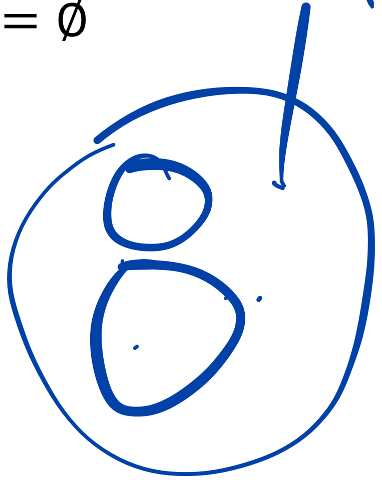




Formal Definition - Clustering

Rest is Noise

- Elements $e \in E$
- Cluster $c \in C,\ with\ c \subseteq E\ and\ \bigcup_{c \in C} = E\ and\ \bigcap_{c \in C} = \emptyset$
- Representative $r_c = mean(c)$



- $variability(c) = \sum_{e \in c} distance(r_c, e)^2 \geq 0$

- Clustering: $Minimize \sum_{c \in C} variability(c)$

With $k = |C|$

$1 < k < n = |E|$

Formal Definition - Distance

	1	2
	●	1
	1	2

Manhattan

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

squared Euclidian saves one squareroot

	1	$\sqrt{2}$
	●	1
	1	$\sqrt{2}$

Euclidian^2

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

	1	1
	●	1
	1	1

Chebyshev

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

Kommentarfolie

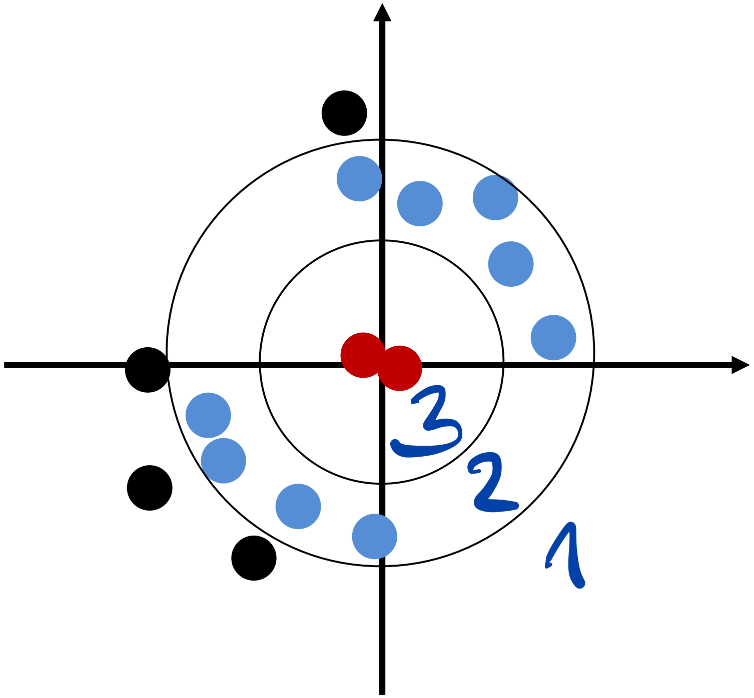
The distance is the most important measure in clustering, but it should be noted that not only spatial distance is meant, but distance in any numerically representable property (color, size, weight, ...).

In everyday life, the Euclidean distance is mostly meant when it comes to measuring how far two points are away from each 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

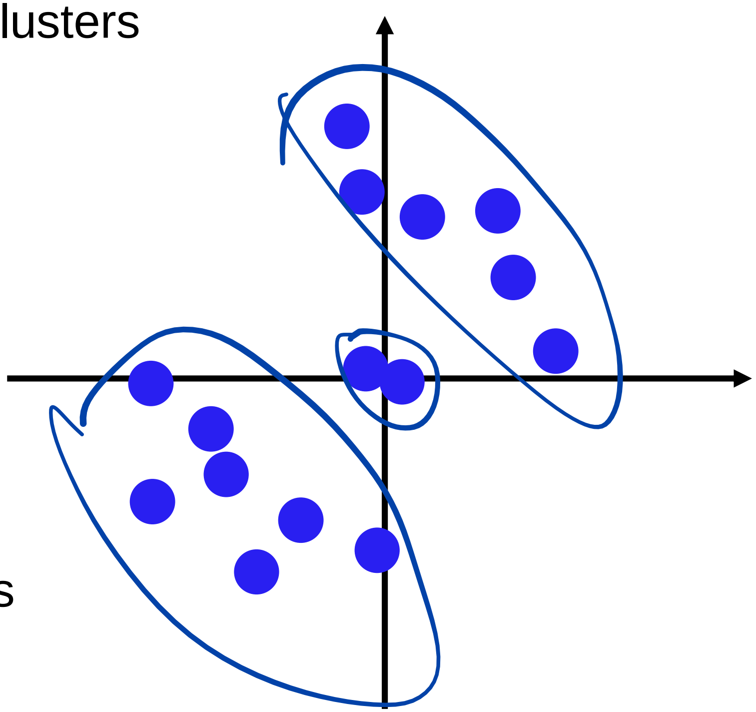


Clustering

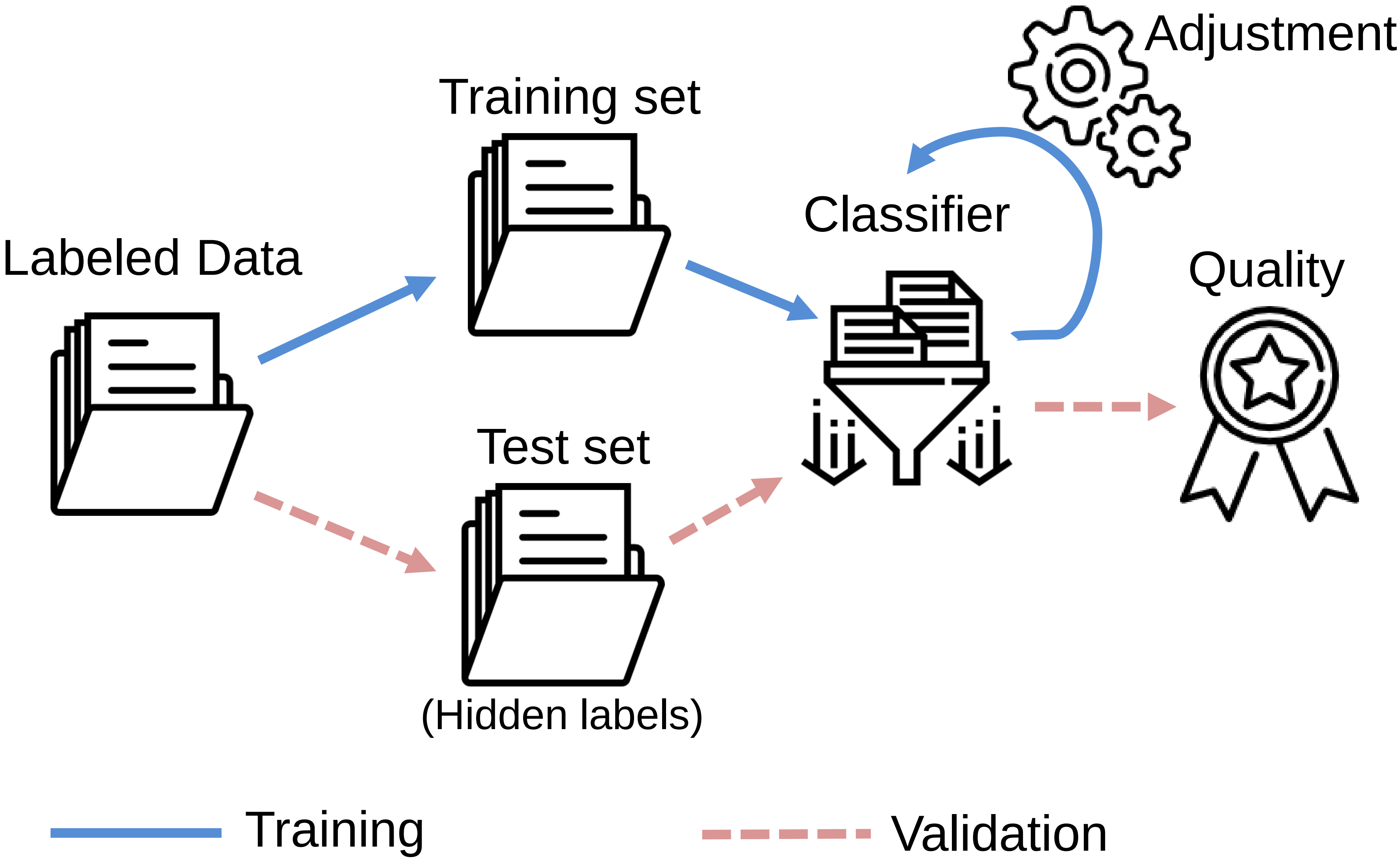
- Grouping a set of data objects into clusters
  - Cluster: a collection of elements
  - Similar to each other within the same cluster
  - Dissimilar to the objects in other clusters

- Difference to classification
  - No given clusters/classes
  - Unsupervised learning

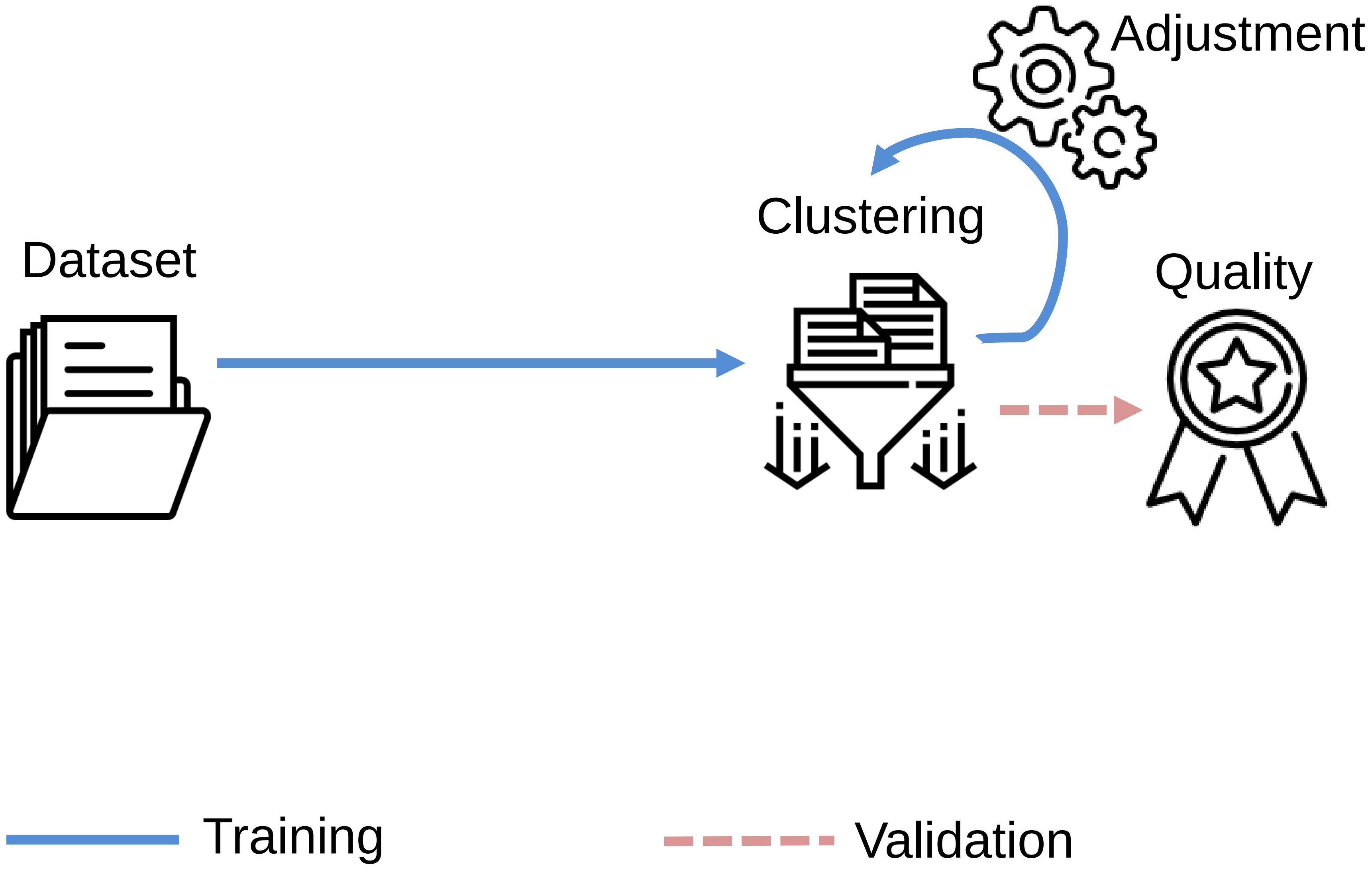
- Application
  - Get insights in large datasets
  - Preprocessing for other algorithms



Supervised Learning - Classification



Unsupervised Learning - Clustering



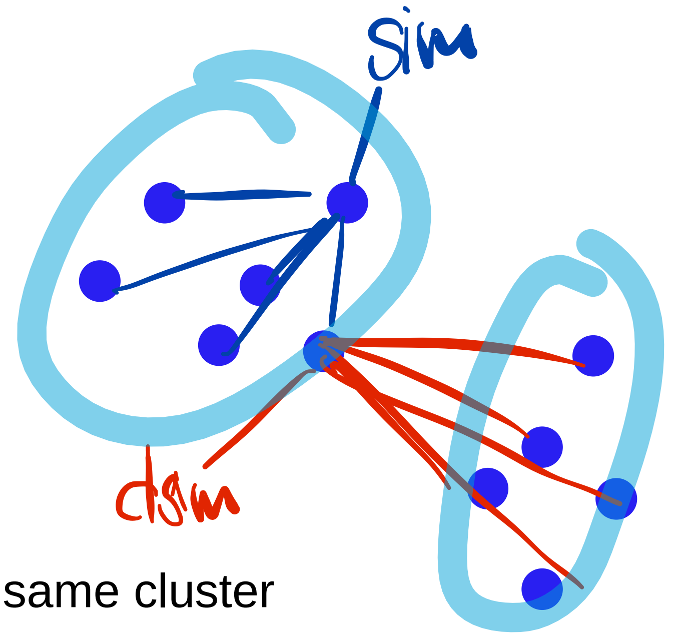
Kommentarfolie

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 depend on k
  - $k = 2$: very large distances
  - $k = n - 1$: very small distances

- Similarity $sim(o)$
  - Within a cluster: $o \in a \in C$
  - Average distance to all elements within the same cluster
  - $sim(o) = \frac{1}{|a|} \sum_{e \in a} distance(o, e)$
- Dissimilarity $dsim(o)$
  - To other clusters: $e \notin b \in C$
  - Average distance to all elements of the second closest cluster
  - $dsim(o) = \min_{c \neq a} \left(\frac{1}{|c|} \sum_{e \in c} distance(o, e) \right)$



Quality Measure of a Cluster

- Silhouette coefficient
  - $s(o) = \frac{dsim(o) - sim(o)}{\max\{sim(o), dsim(o)\}}$
  - $if\ sim(o) = dsim(o) = 0, then\ s(o) = 0$
  - $s(o)\ \in [-1, 1]$
  - $silh(c) = \frac{1}{|c|} \sum_{o \in c} s(o)$
  - $silh(E) = \frac{1}{|E|} \sum_{o \in E} s(o)$

dsim ≫ sim : silh → 1 ⊕

dsim ≈ sim : silh → 0

dsim ≪ sim : silh → -1 ⊖

silh ≥ 0.7 ⇒ ⊕

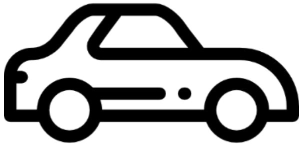
Supervised Learning: Classification

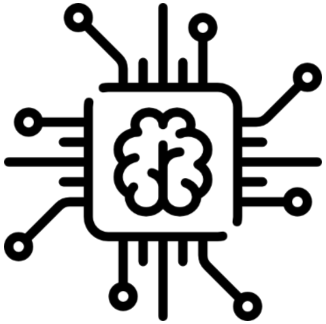
Johannes Betz / Prof. Dr. Markus Lienkamp / Prof. Dr. Boris Lohmann

(Jan Cedric Mertens, 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

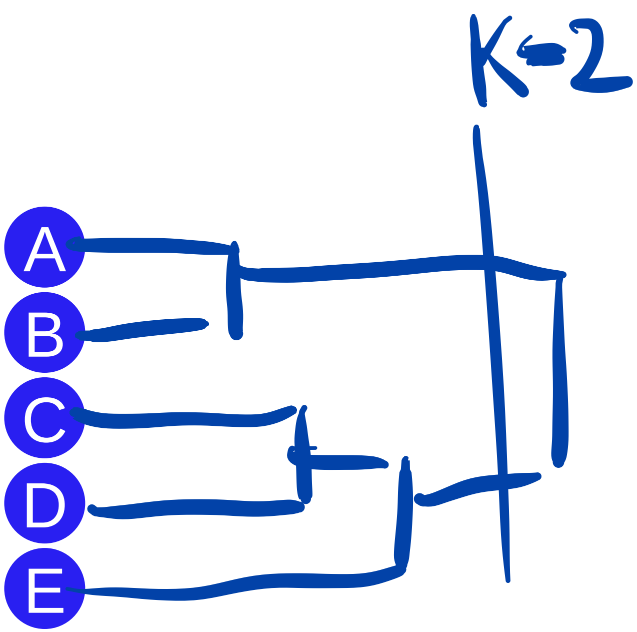


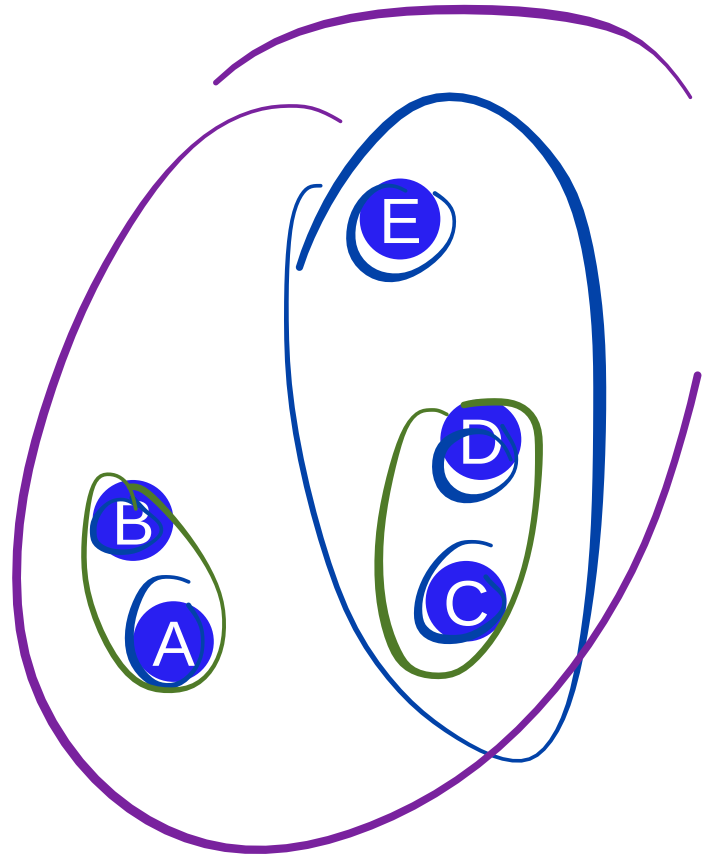


Hierarchical Clustering

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

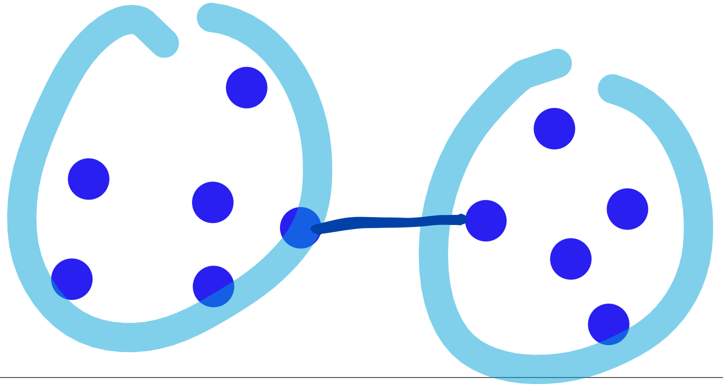
- Top down (divisive)/bottom up (agglomerative)



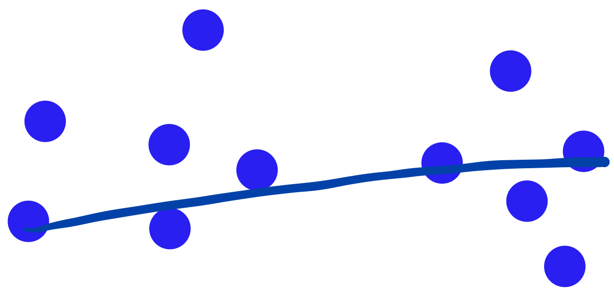


Distance Between Clusters

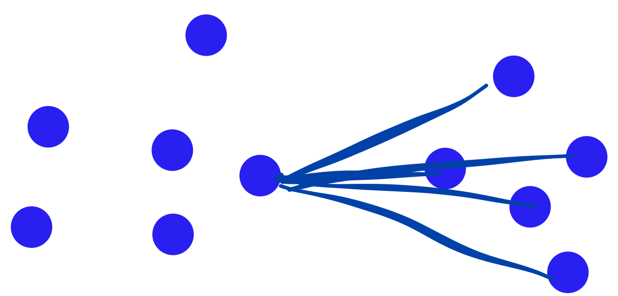
- Single Link
  - Smallest distance between two point of different clusters



- Complete Link
  - Largerst distance between two points of different clusters

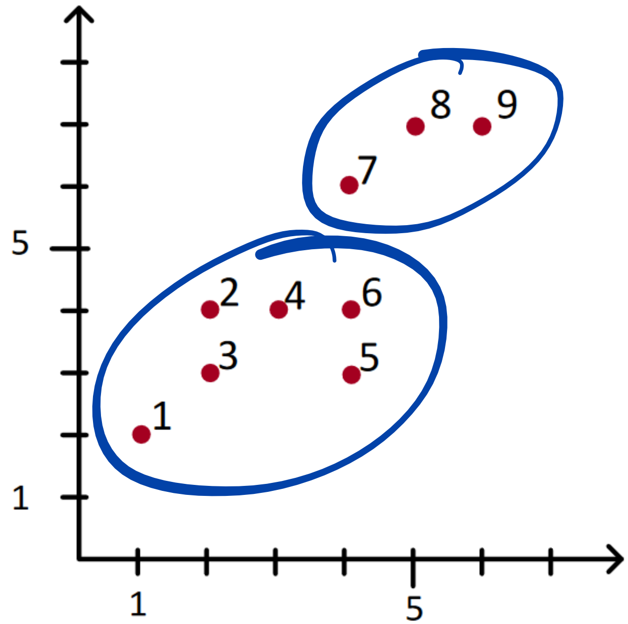


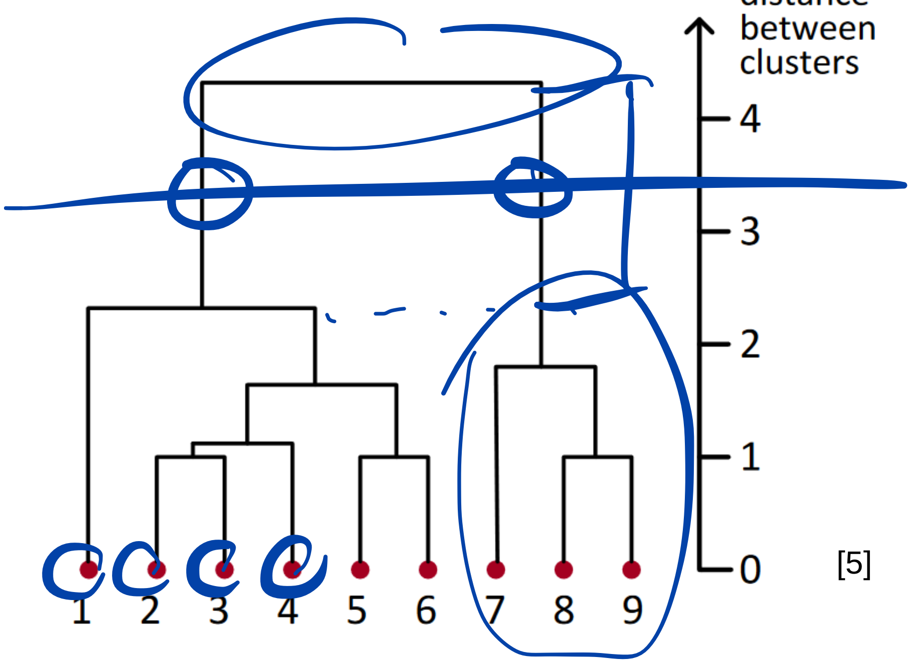
- Average Link
  - Average distance between all points of one cluster to all points of a different cluster



Dendrogram

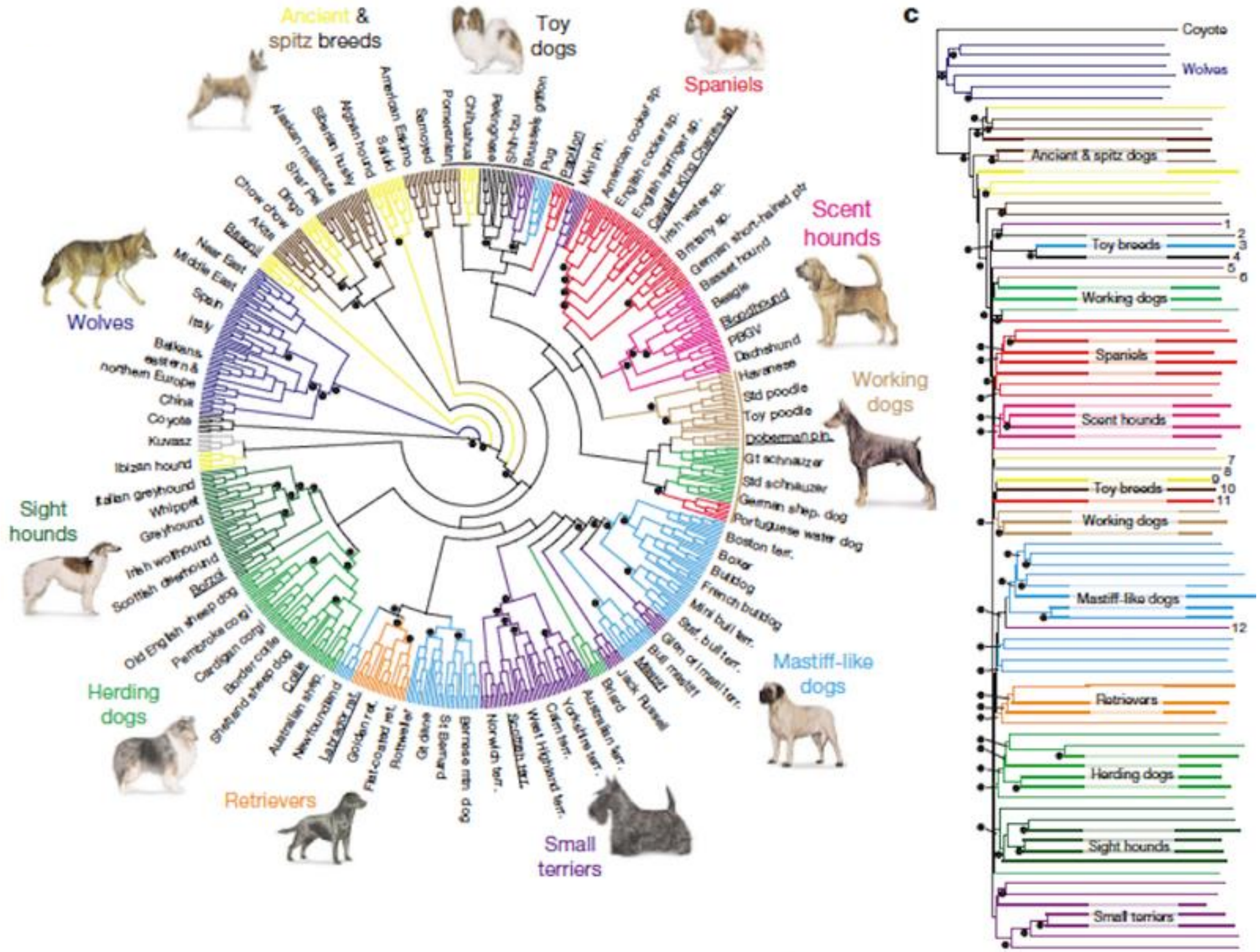
- Root: Cluster with all points
- Leaf: Cluster with one point
- Edges: Combine two clusters
- Depth: Distance between two combined clusters





[5]

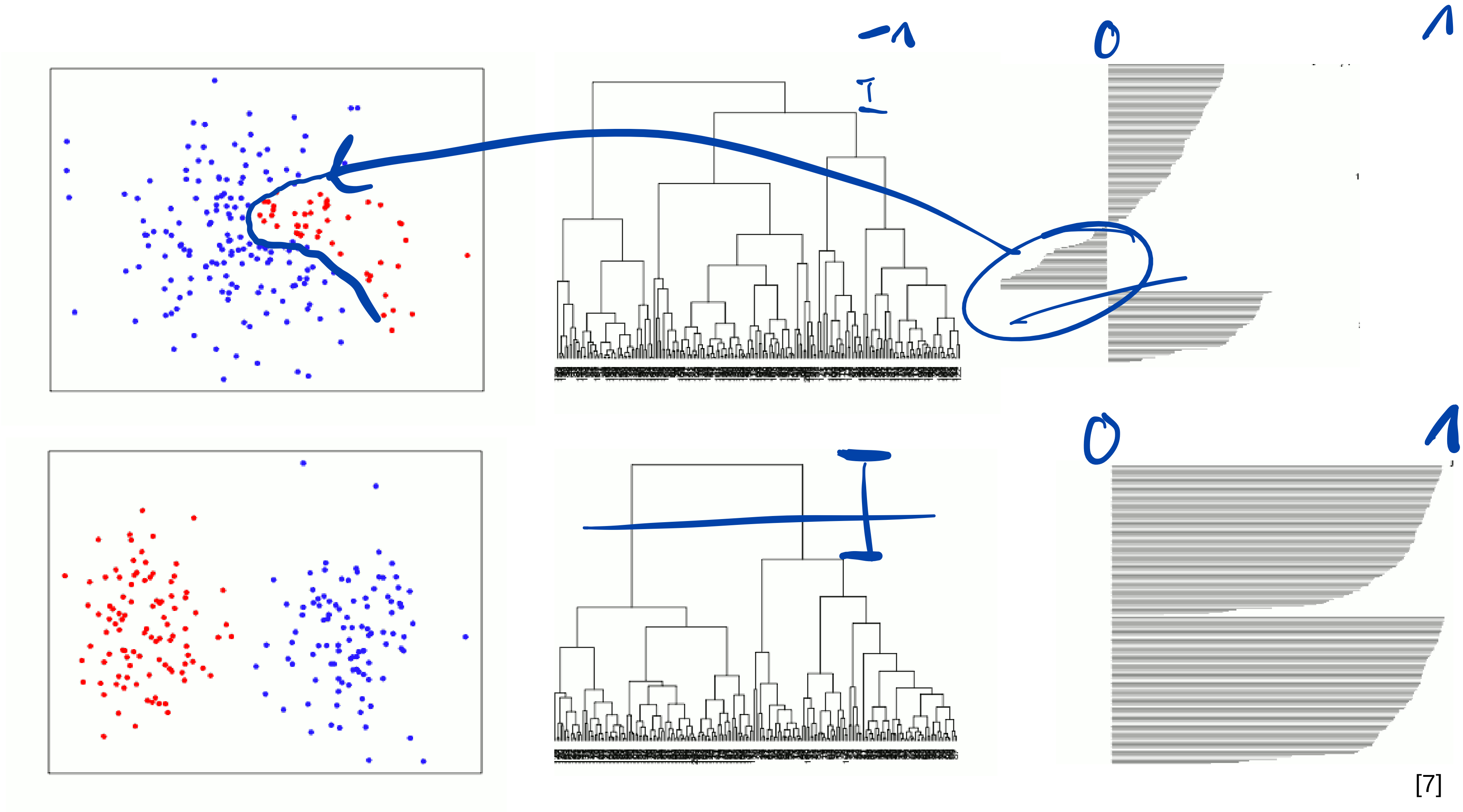
Dendrogram - Example



[6]

Hierarchical Clustering - Silhouette Coefficient

silh



[7]

Kommentarfolie

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.

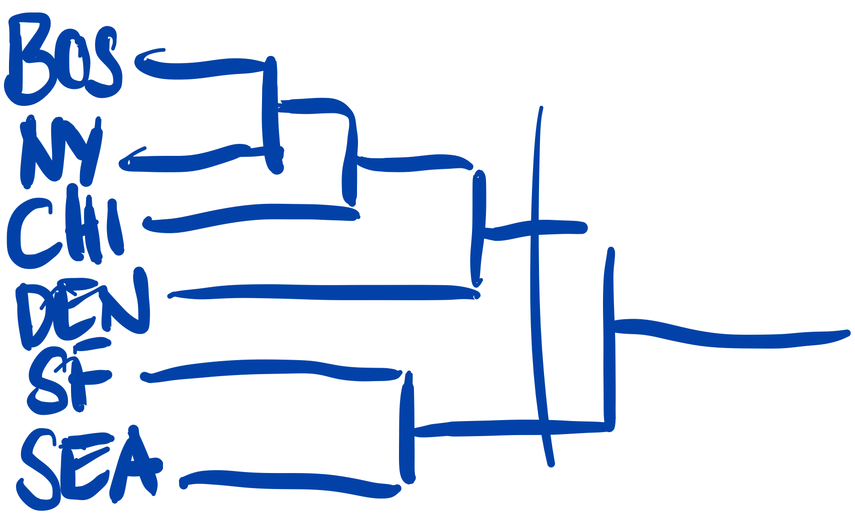
Below a suitable number of clusters has been chosen and the quality of the clusters is correspondingly much better.

Hierarchical Clustering - Example

CLUSTER1 CLUSTER2

	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

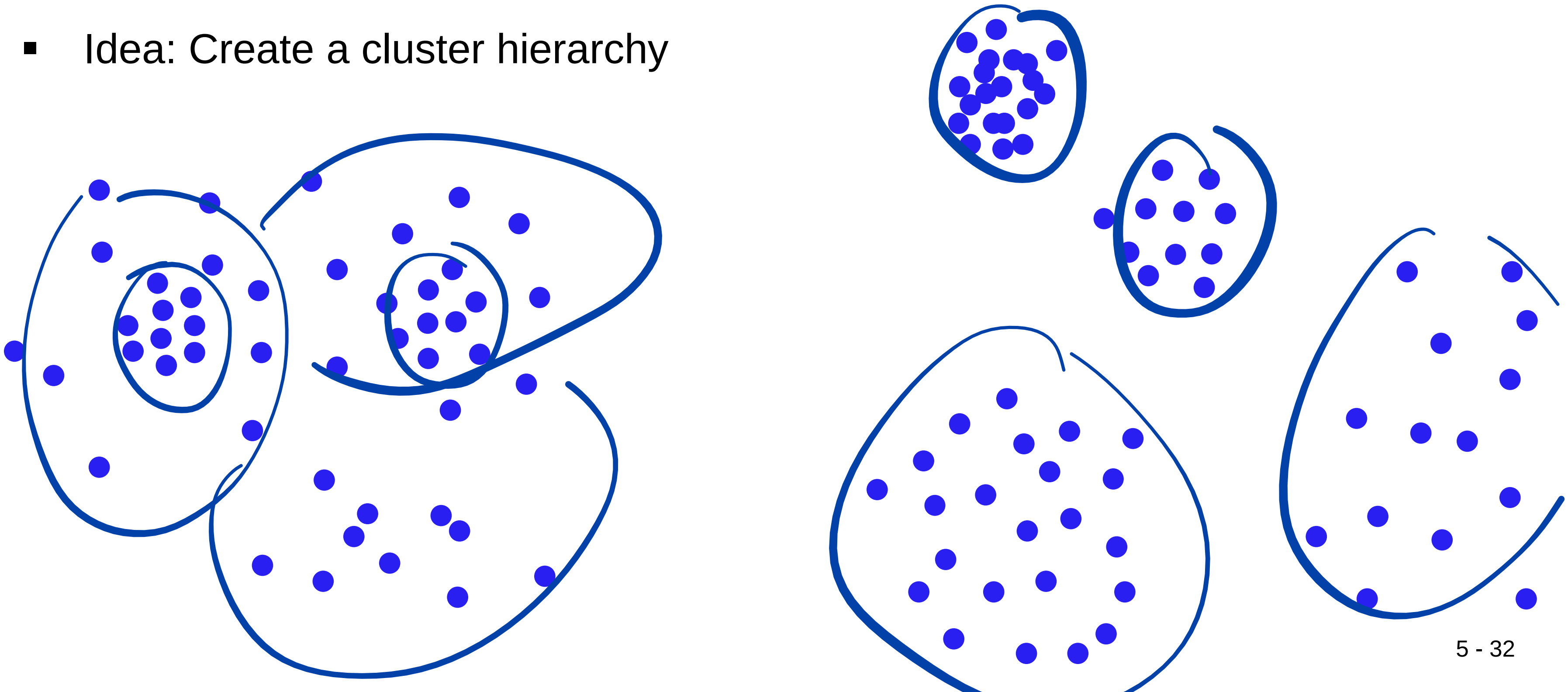
[8]



Hierarchical Clustering

- Problems
  - Clusters within clusters
  - Different cluster densities in one dataset

- Idea: Create a cluster hierarchy



Discussion Hierarchical Clustering

- Pro:
  - **Generic**: No cluster number or parameters must be defined
  - **Visualization**: E.g., dendrogram shows hierarchy
  - **Hierarchy**: Relationship between clusters
  - **Deterministic:** Generates always the same clusters

- Contra:
  - **Scalability:** Runtime $\mathcal{O}(n^3)$
  - **Choice:** The final cluster must be selected from the hierarchy

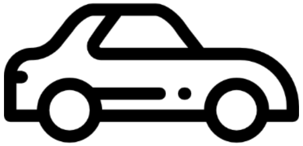
Supervised Learning: Classification

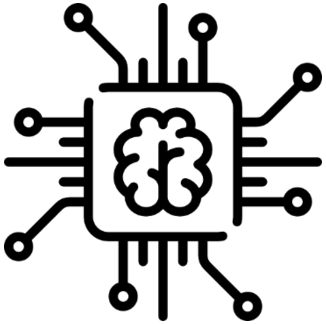
Johannes Betz / Prof. Dr. Markus Lienkamp / Prof. Dr. Boris Lohmann

(Jan Cedric Mertens, M.Sc.)

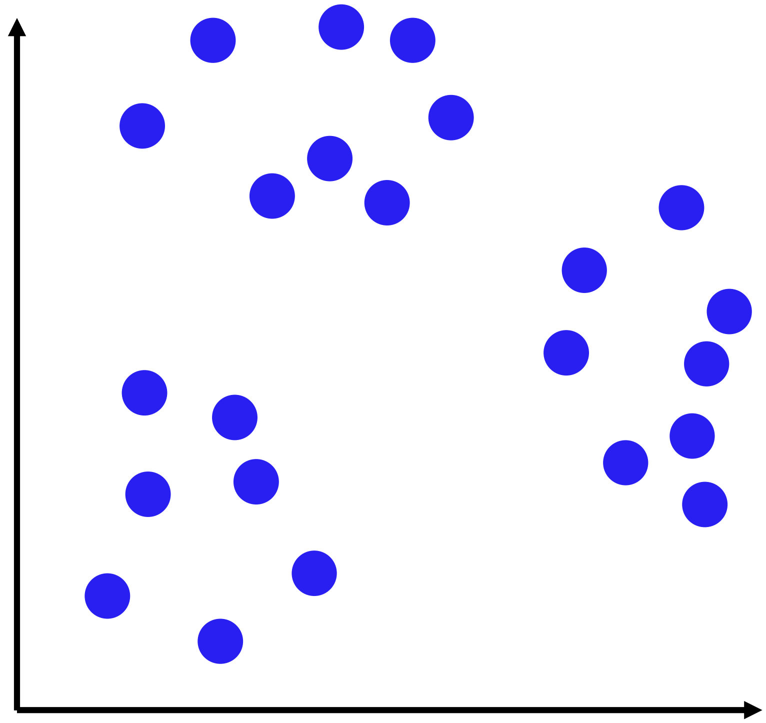
Agenda

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



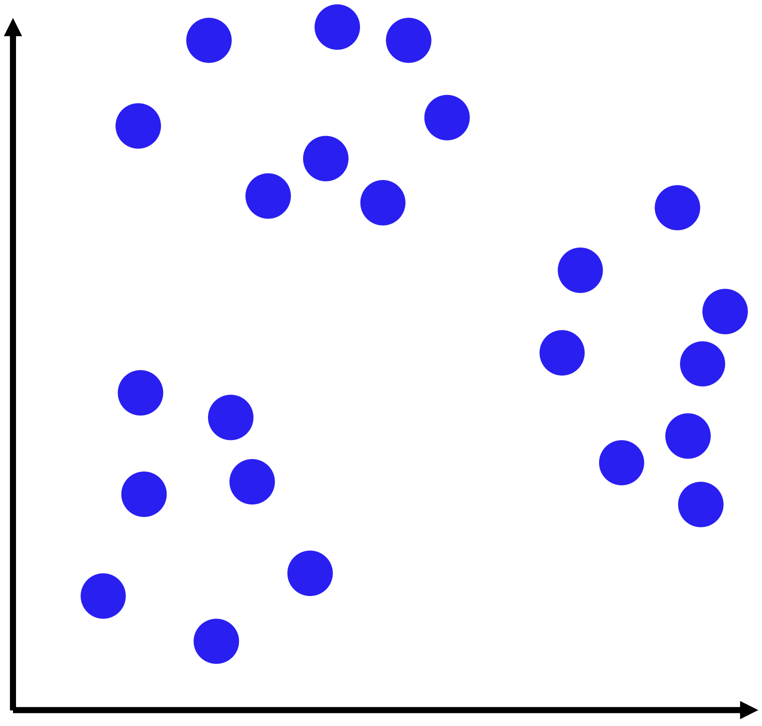
- Minimize squared distances to the cluster mean (variability)
- Minimize the summed variability of all clusters

  Large Sum → Poor clustering

  Minimal Sum → Optimal clustering
- Computationally challenging
  - NP-hard

$O(n^k)$ $P \stackrel{?}{=} NP$ $O(2^{nk})$

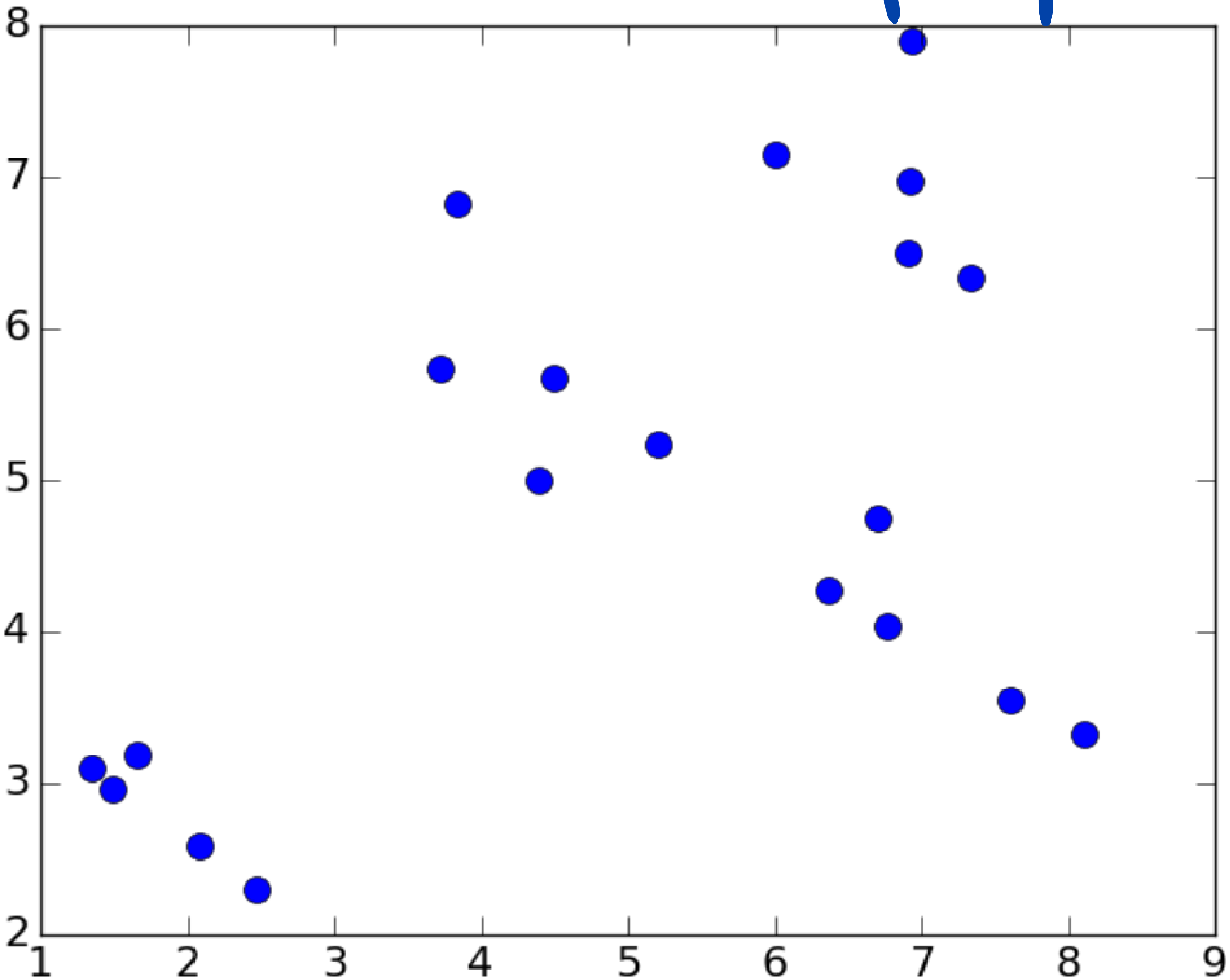
K-Means Algorithm (Lloyd)



- Given:
  - Number of desired clusters k
  - Dataset
- Initialization:
  - Choose k arbitrary representatives
- Repeat until stable:
  - Assign objects to nearest representative
  - Compute center of each cluster as new representative

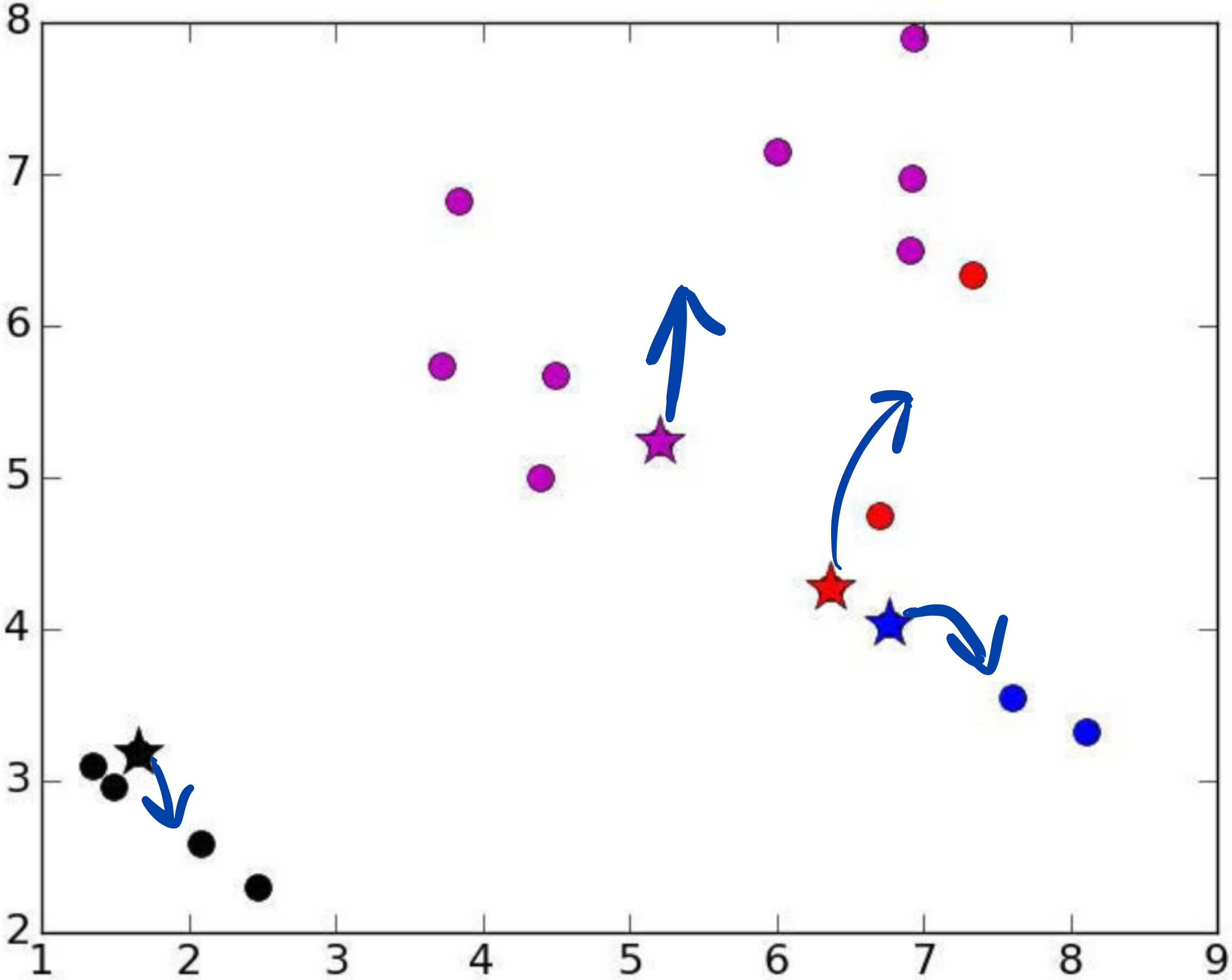
K-Means Algorithm (Lloyd)

K=4



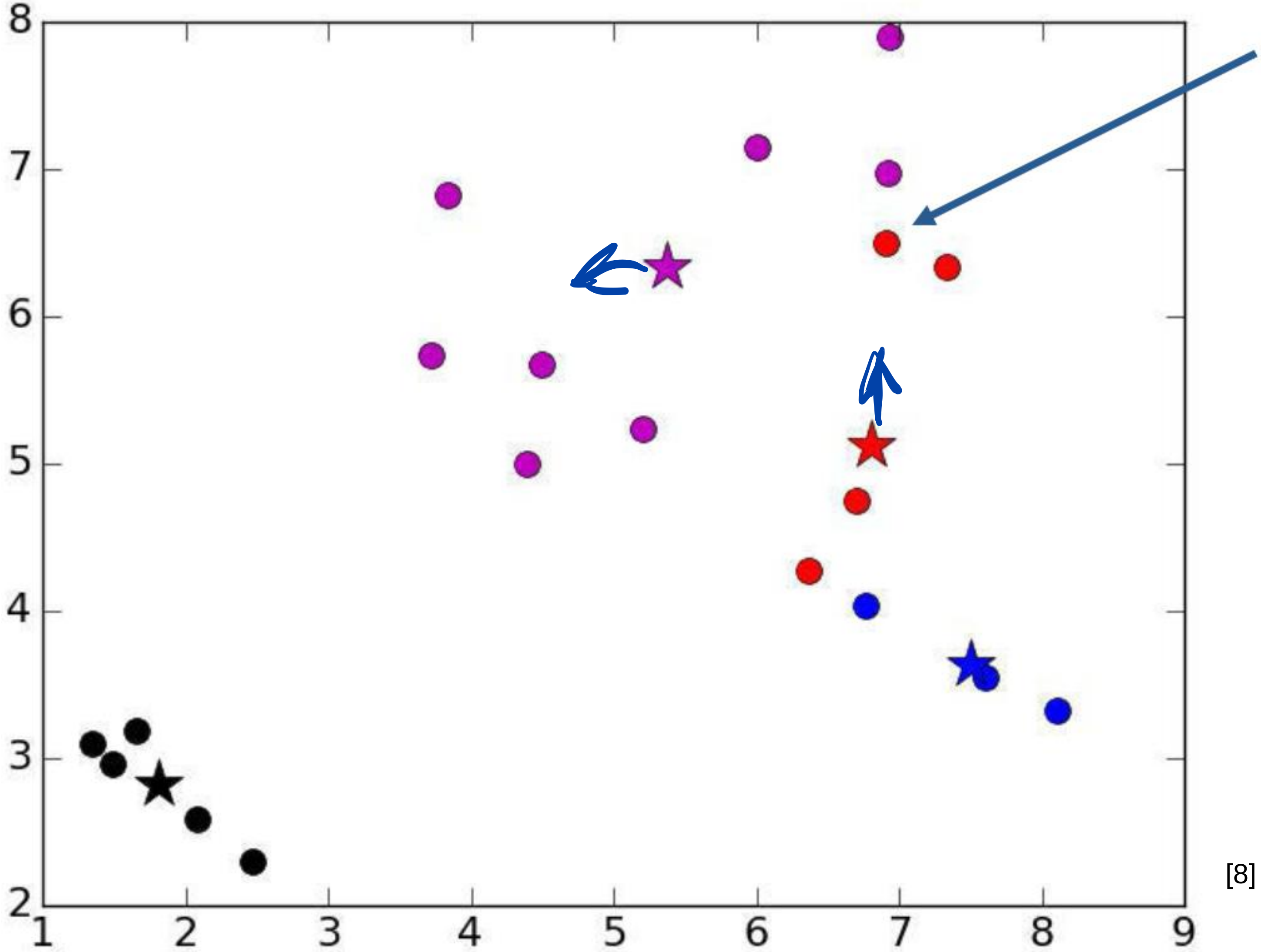
[8]

K-Means Algorithm (Lloyd) – Initialization



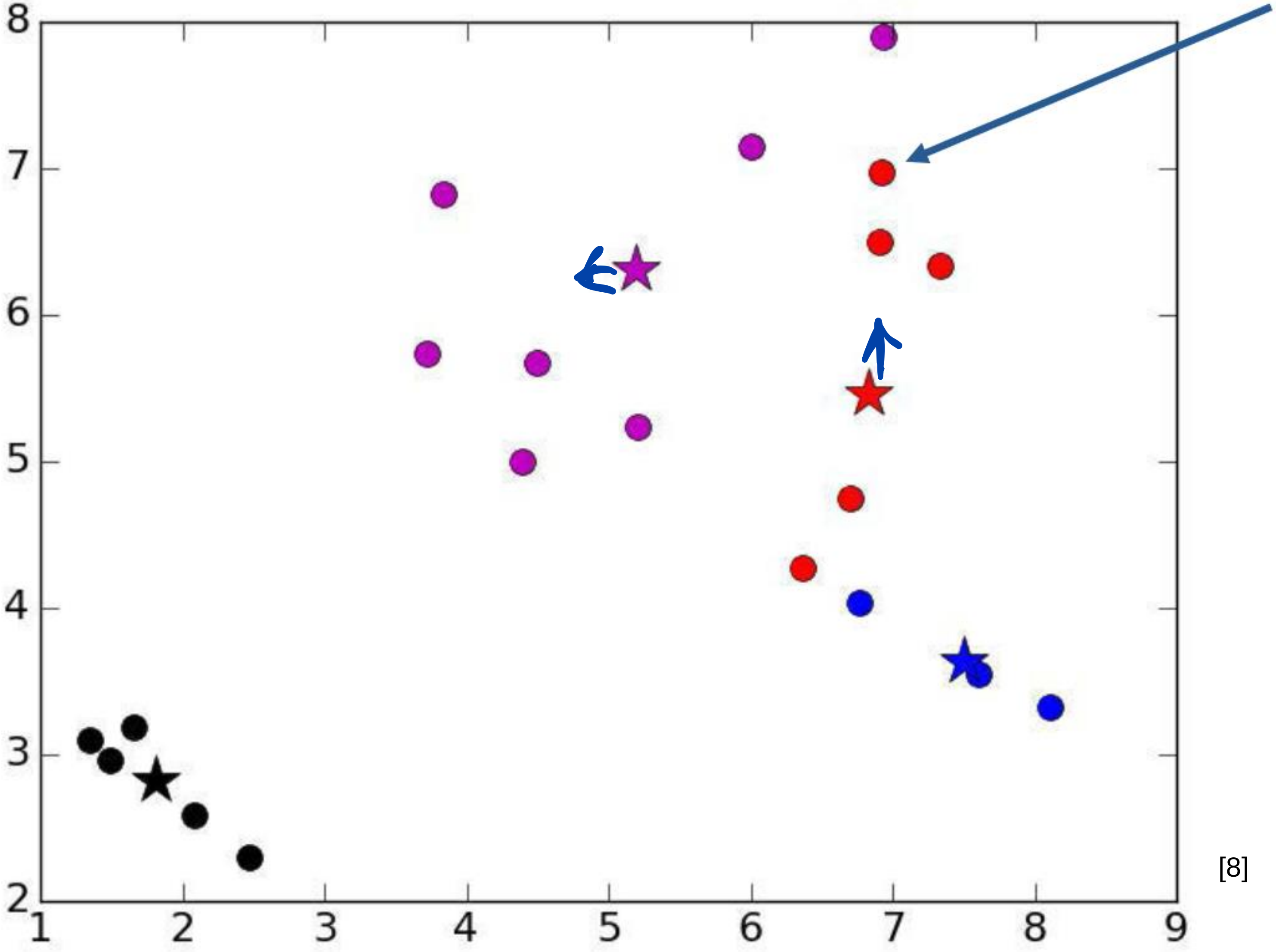
[8]

K-Means Algorithm (Lloyd) – Iteration 1



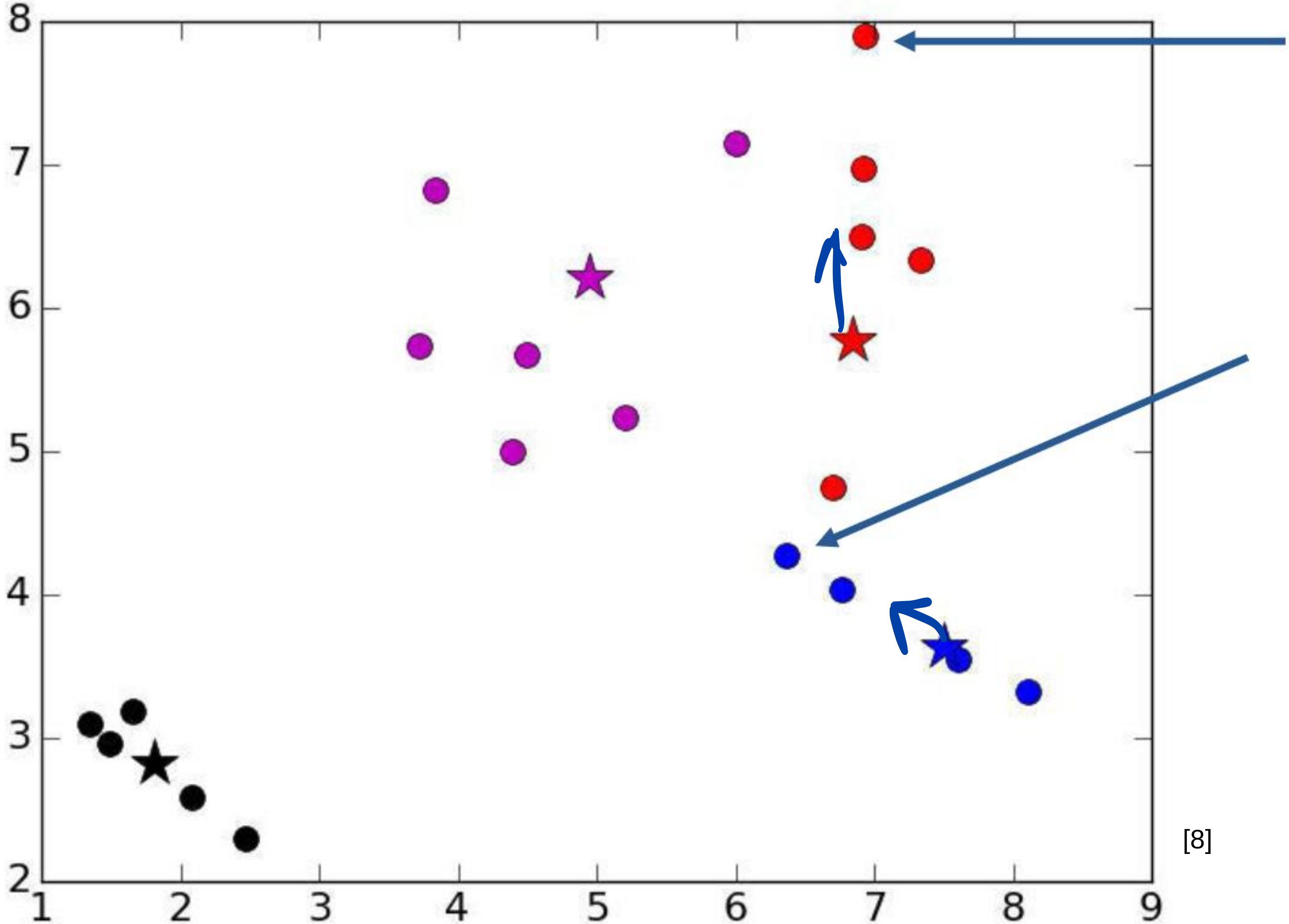
[8]

K-Means Algorithm (Lloyd) – Iteration 2



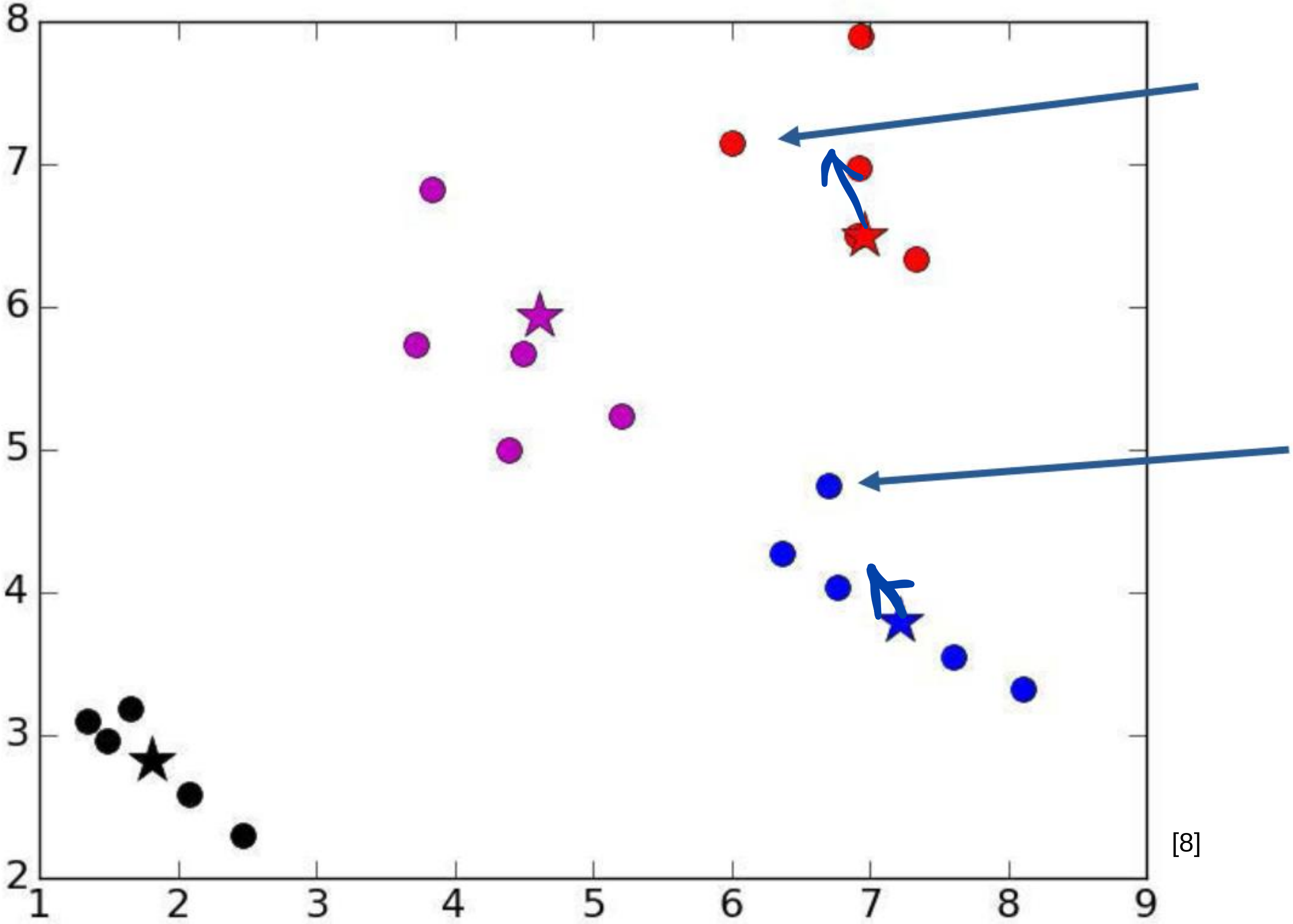
[8]

K-Means Algorithm (Lloyd) – Iteration 3



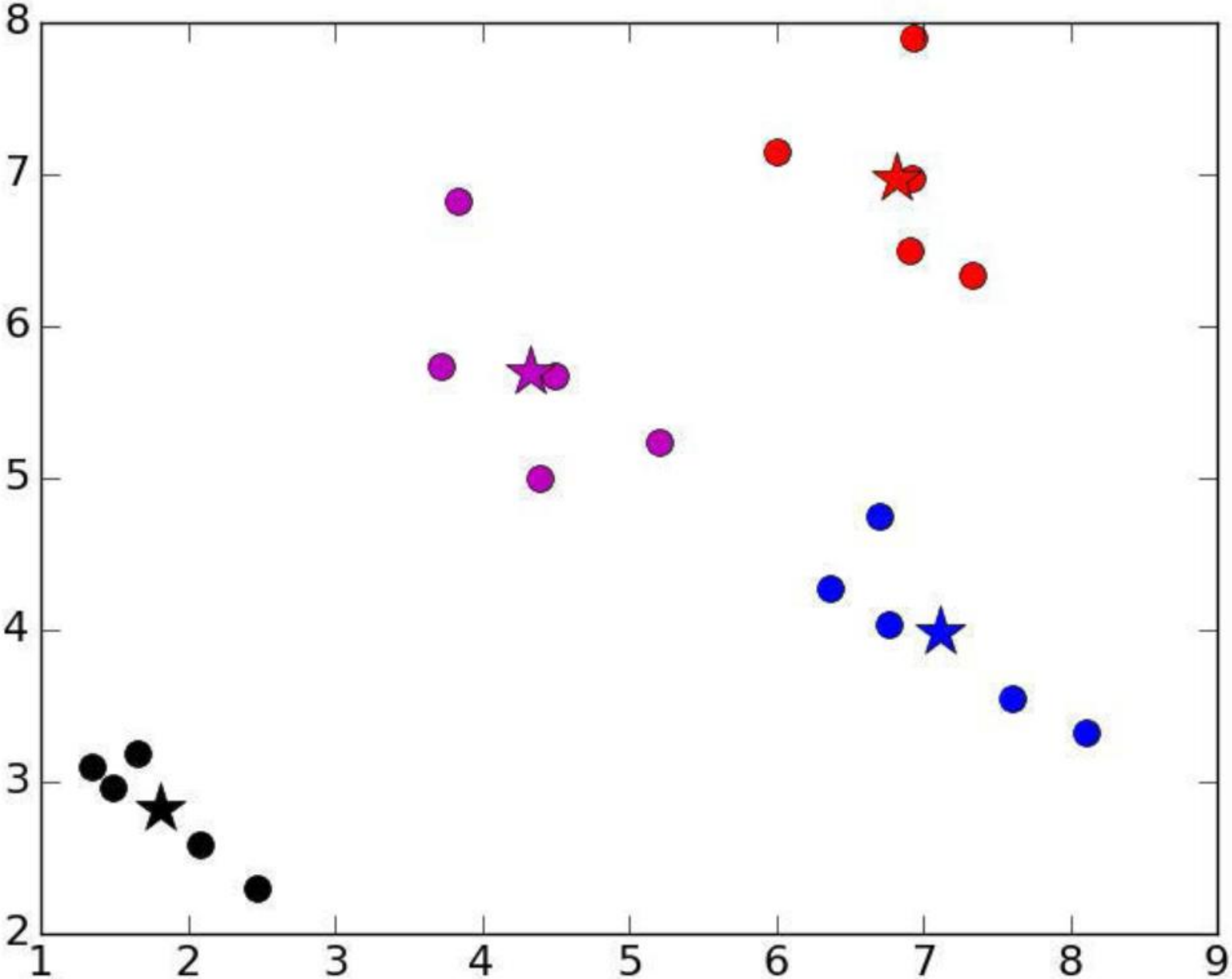
[8]

K-Means Algorithm (Lloyd) – Iteration 4



[8]

K-Means Algorithm (Lloyd) – Iteration 5



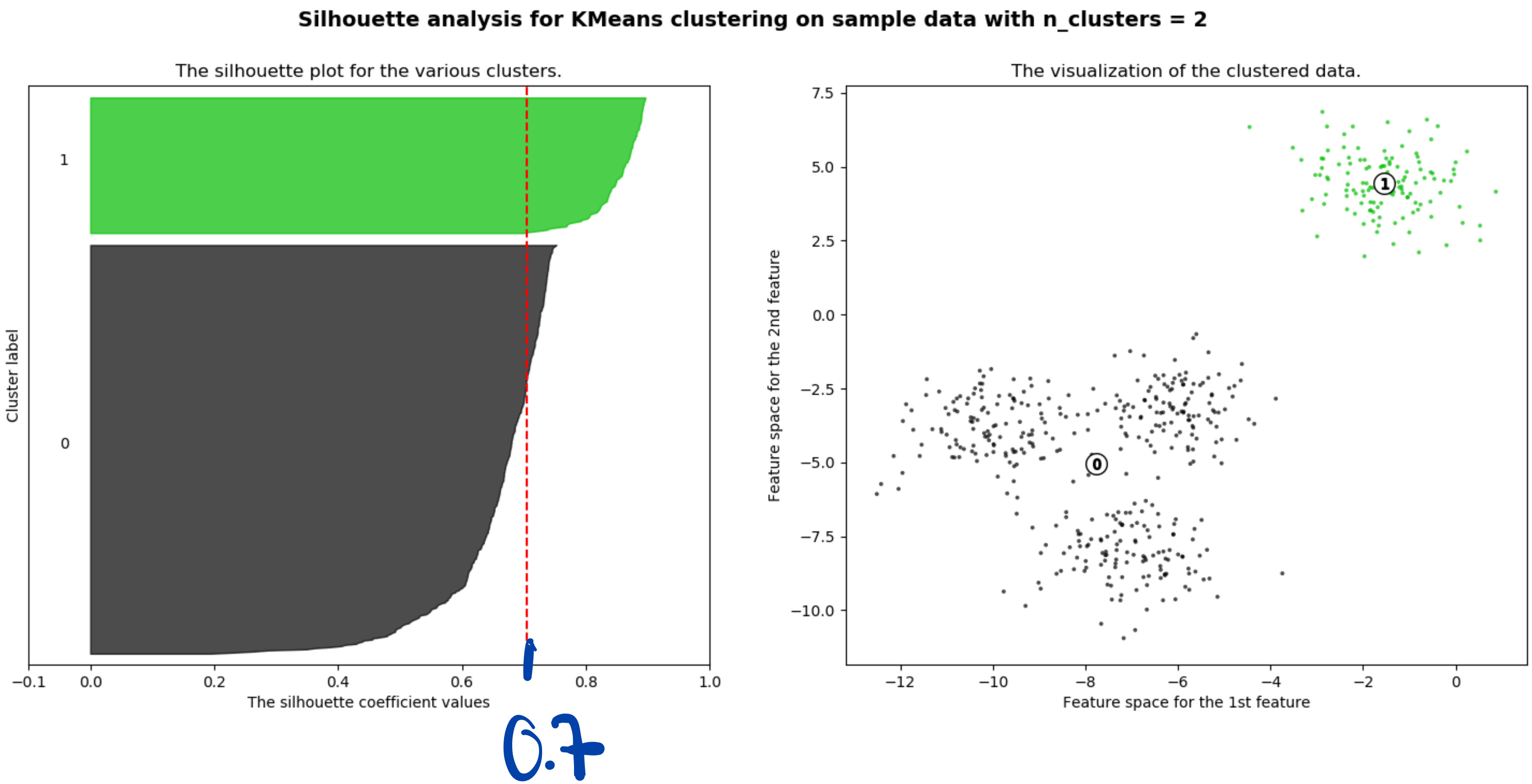
[8]

K-Means Algorithm – How to Choose k?

- A priori knowledge of an expert
  - "There are five different types of bacteria": k = 5

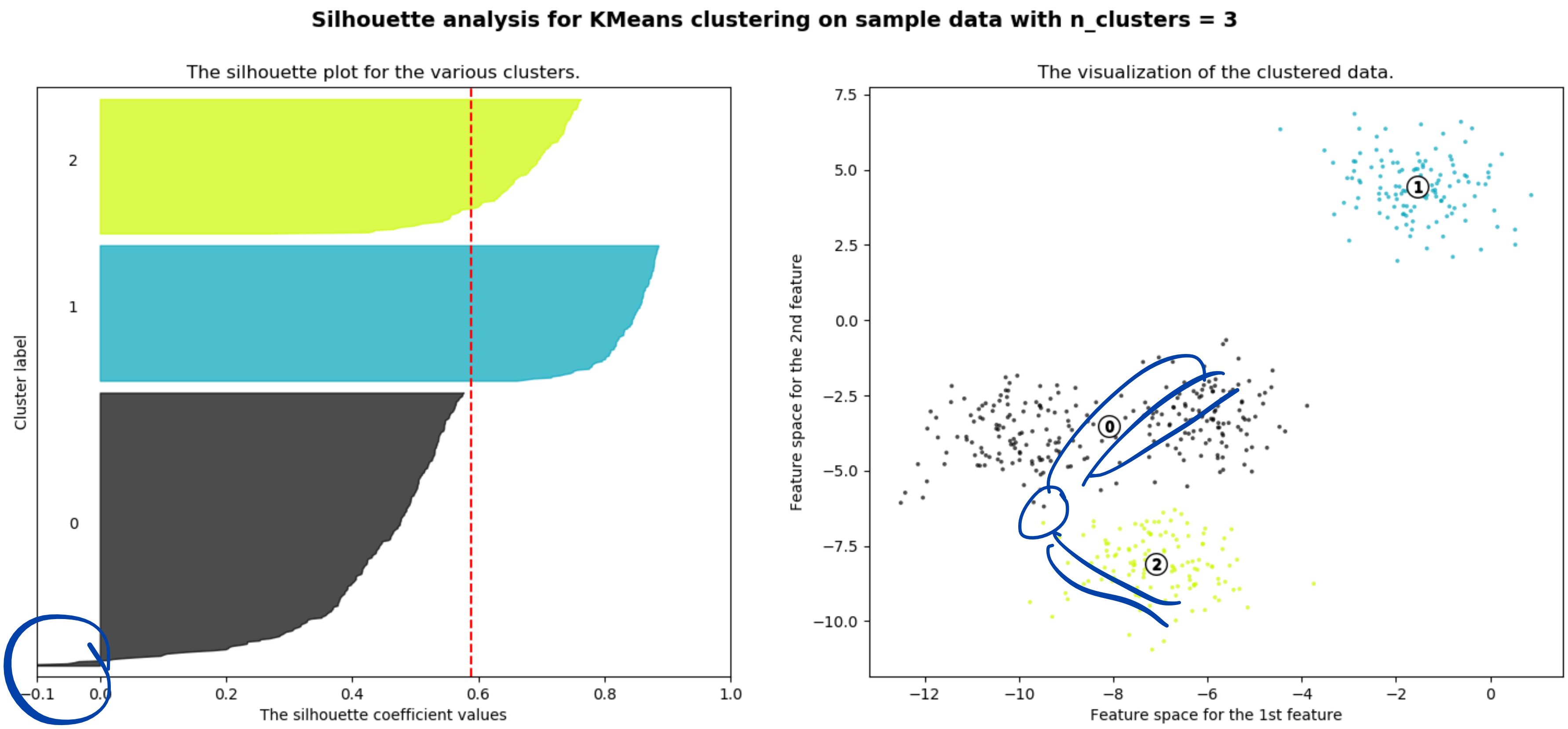
- Search for a good k
  - Naïve approach: Brute Force with k = 2 … n-1
  - Run hierarchical clustering on subset of data

K-Means Algorithm – How to Choose k?



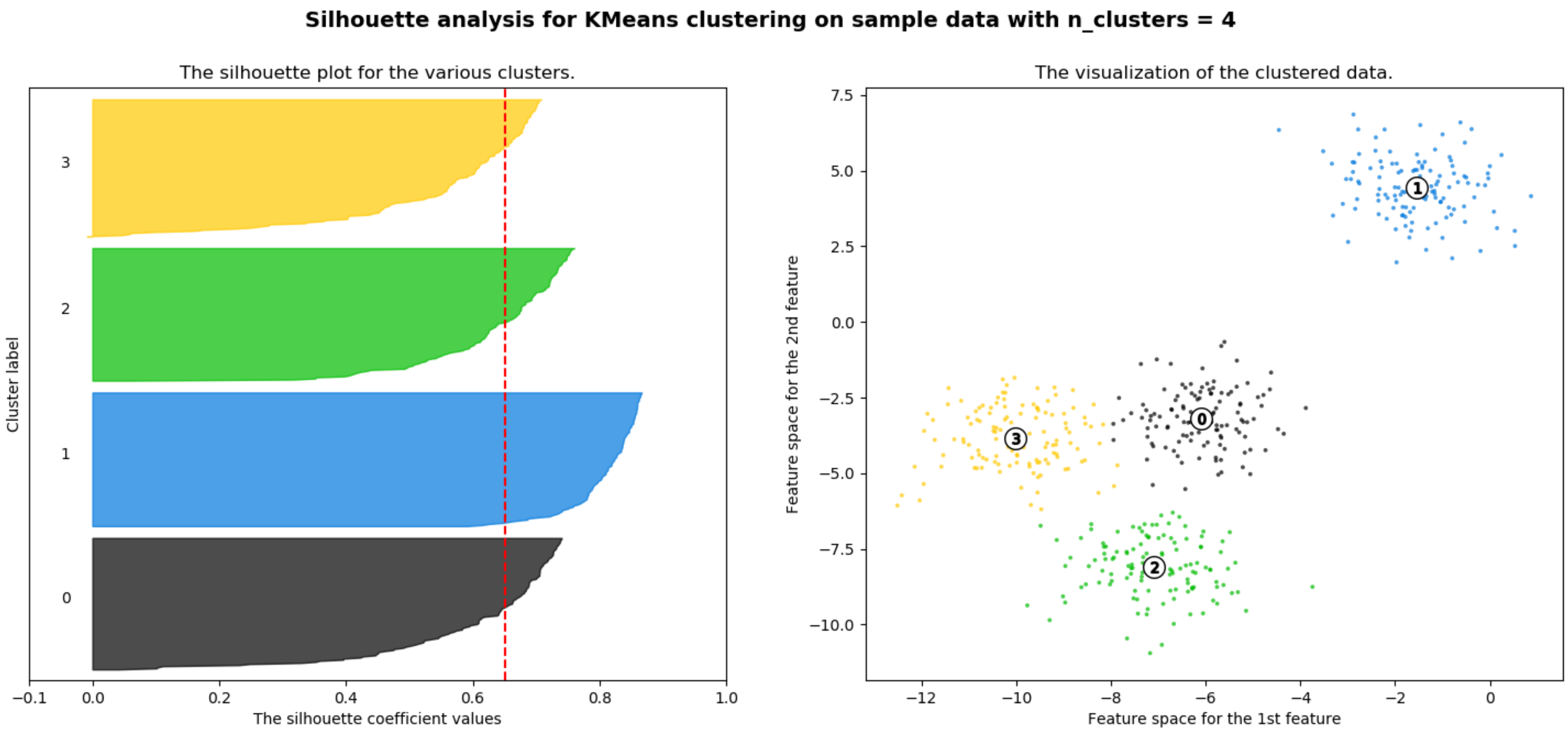
[9]

K-Means Algorithm – How to Choose k?



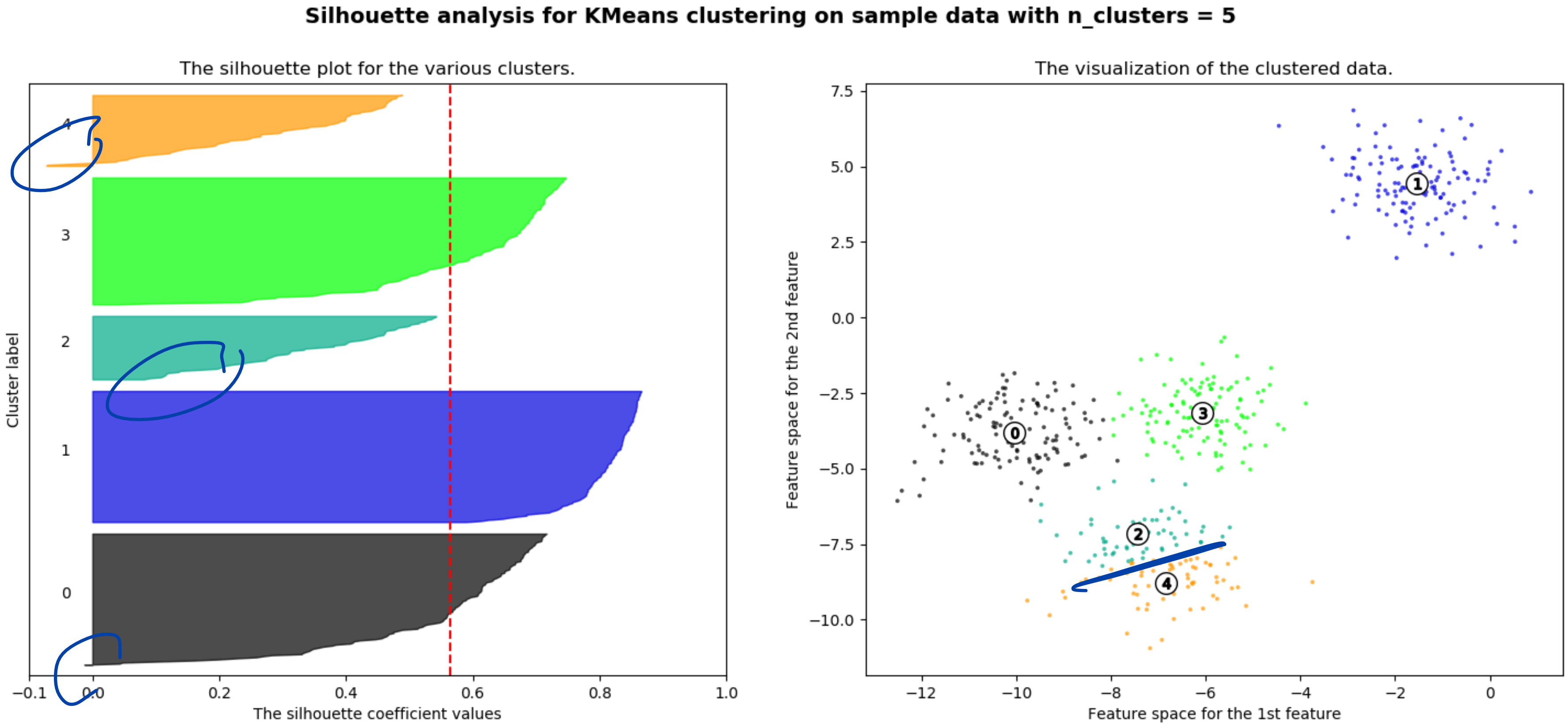
[9]

K-Means Algorithm – How to Choose k?



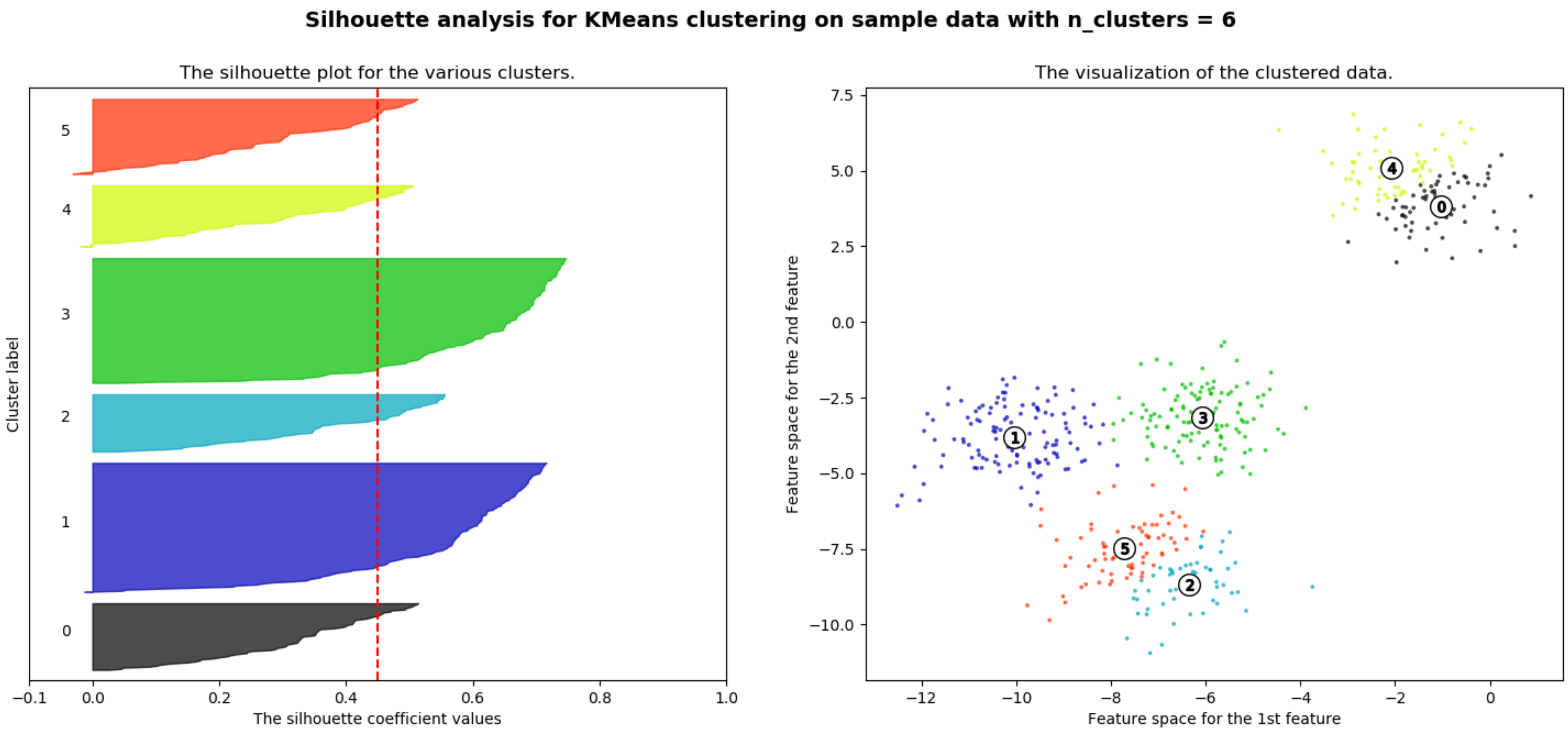
[9]

K-Means Algorithm – How to Choose k?



[9]

K-Means Algorithm – How to Choose k?

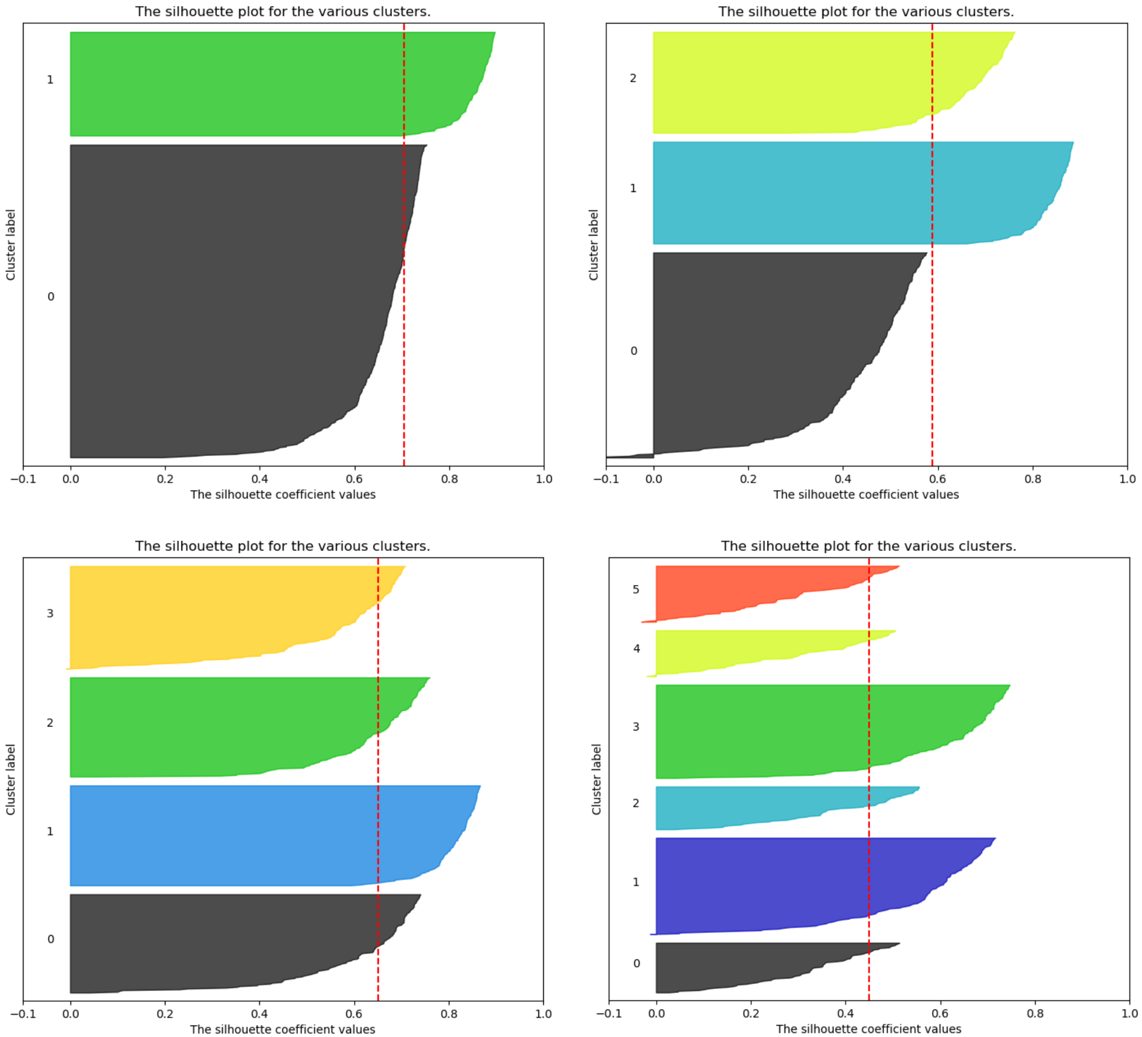


[9]

K-Means Algorithm – How to Choose k?

k = 2

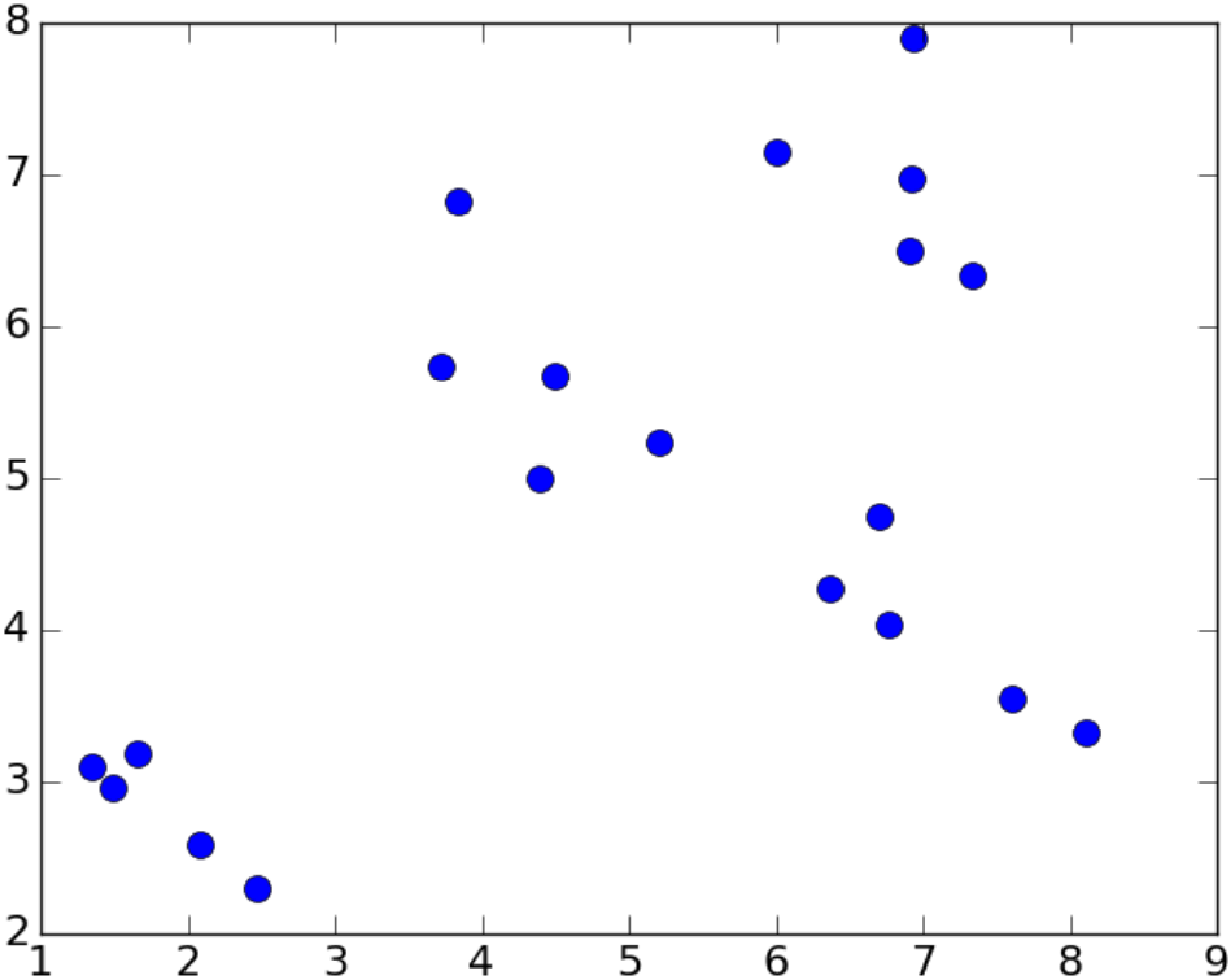
~~k = 3~~



k = 4

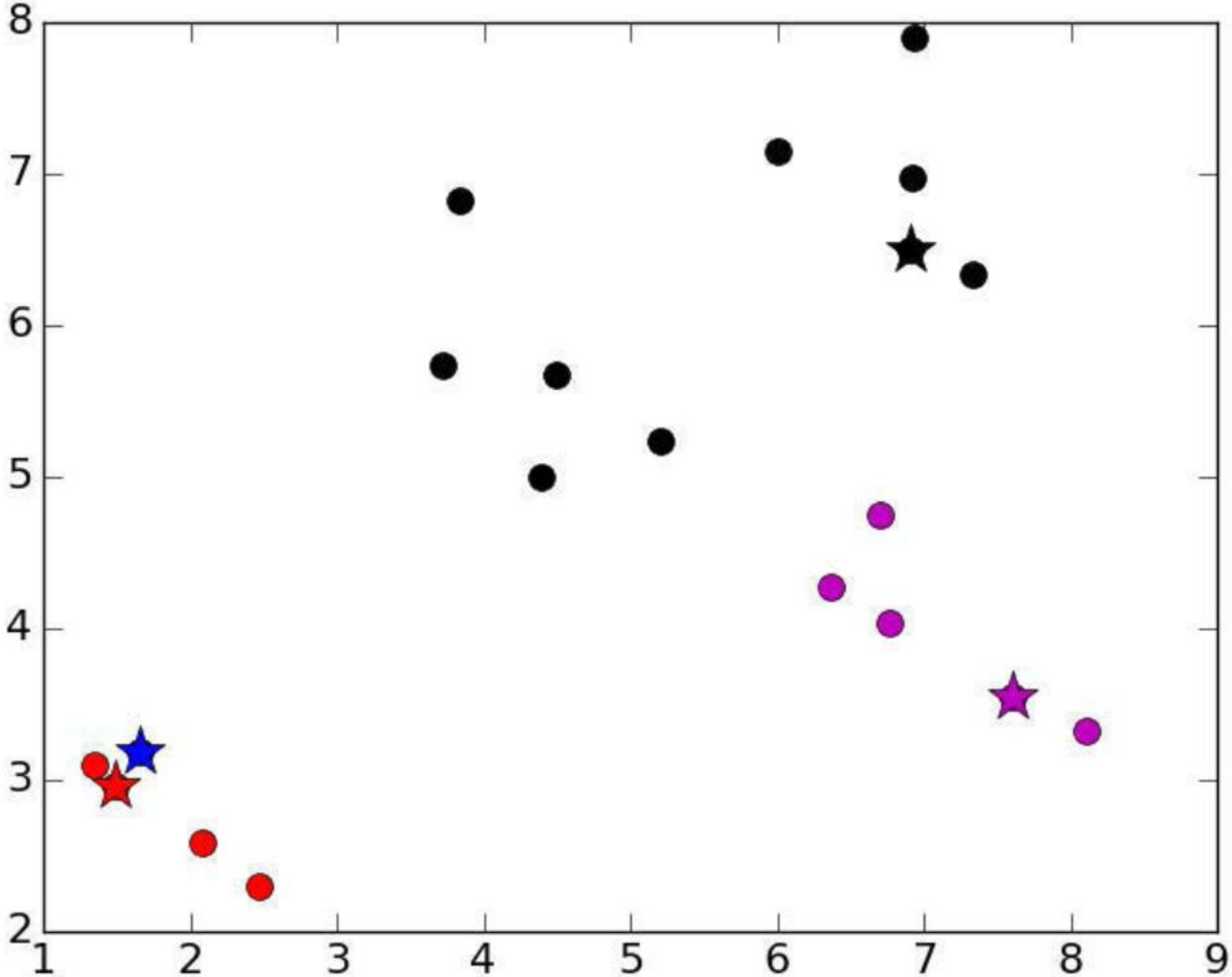
~~k = 5~~

K-Means Algorithm – How to Handle Randomness?



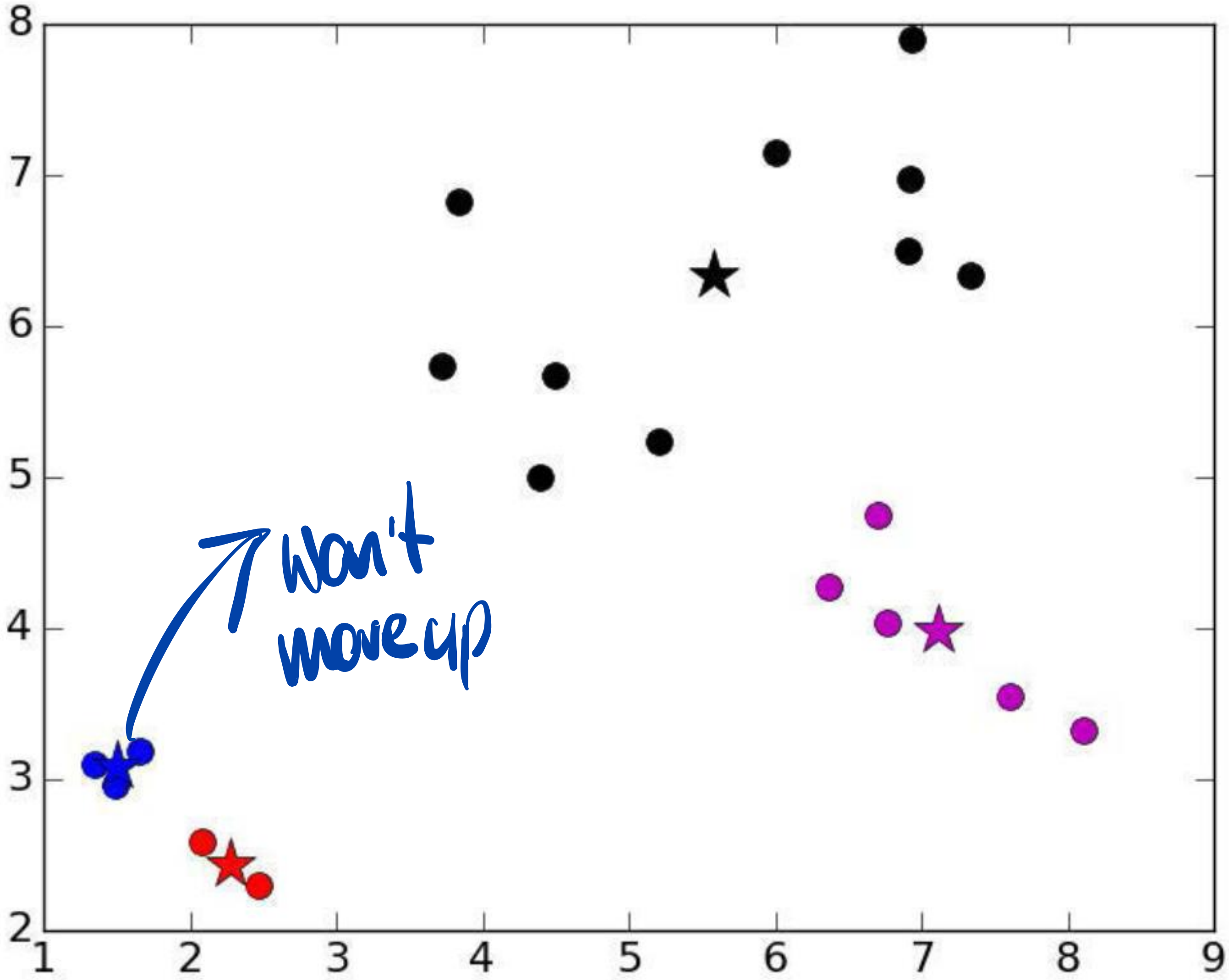
[8]

K-Means Algorithm – How to Handle Randomness?



[8]

K-Means Algorithm – How to Handle Randomness?



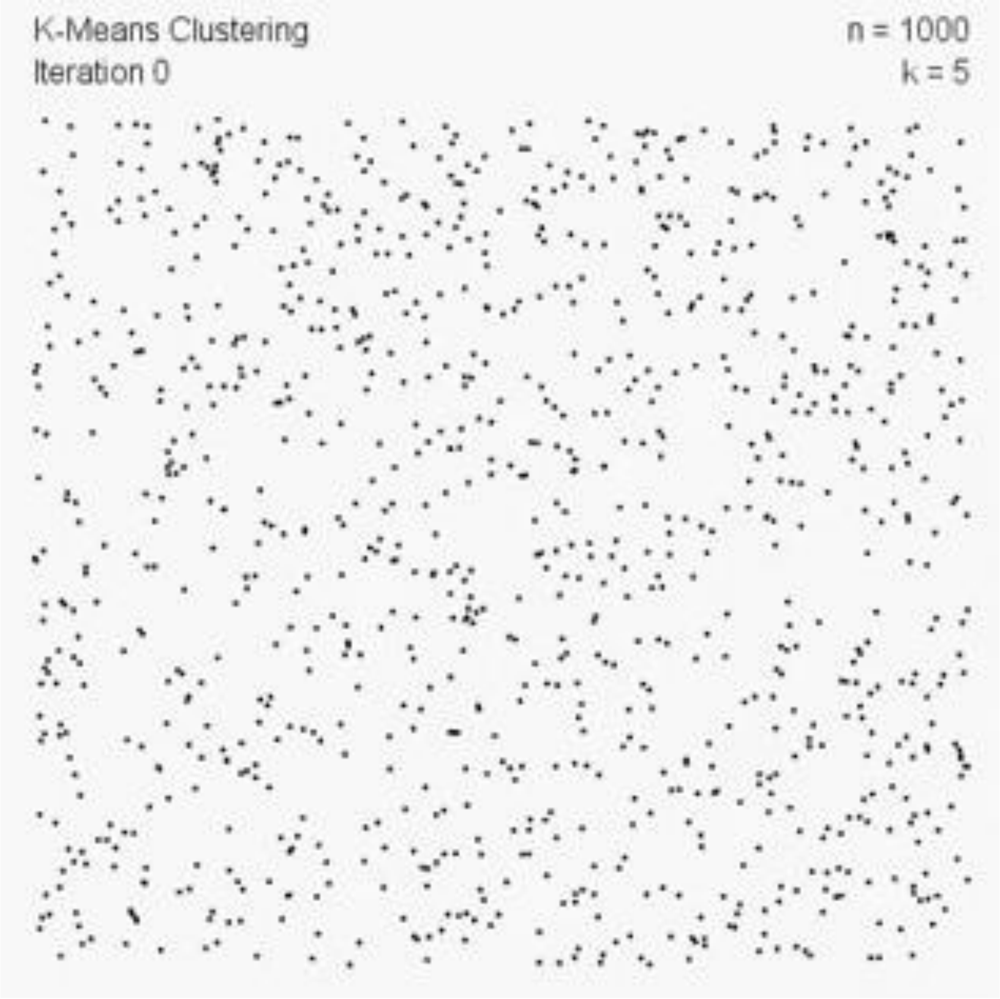
[8]

K-Means Algorithm – How to Handle Randomness?

- Naïve approach
  - Get a small random subset D from E
  - Cluster D and use found representatives for initialization

- Improved approach
  - Get m small random subsets A ... $M \subset E$
  - Cluster A to M and save representatives R_A ... R_M
  - Cluster the merged set $AM = A \cup \cdots \cup M, m$ times with R_A ... R_M as inititial representatives
  - Use the represenation (R_A ... R_M) of the best clustering of AM as initial representation for E

K-Means Example

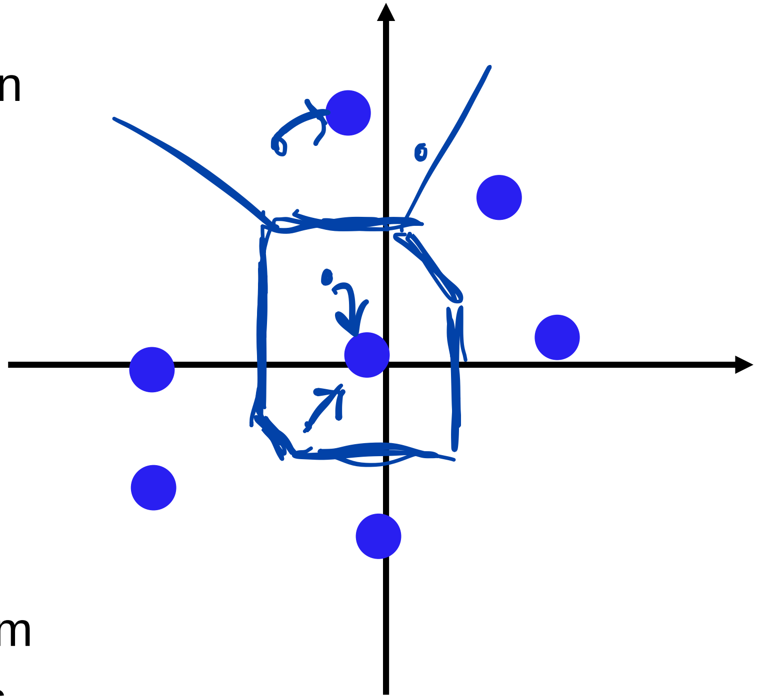


[10]

Discussion K-Means

- Pro:
  - **Efficiency**: $\mathcal{O}(tkn)$ with typically k,t << 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



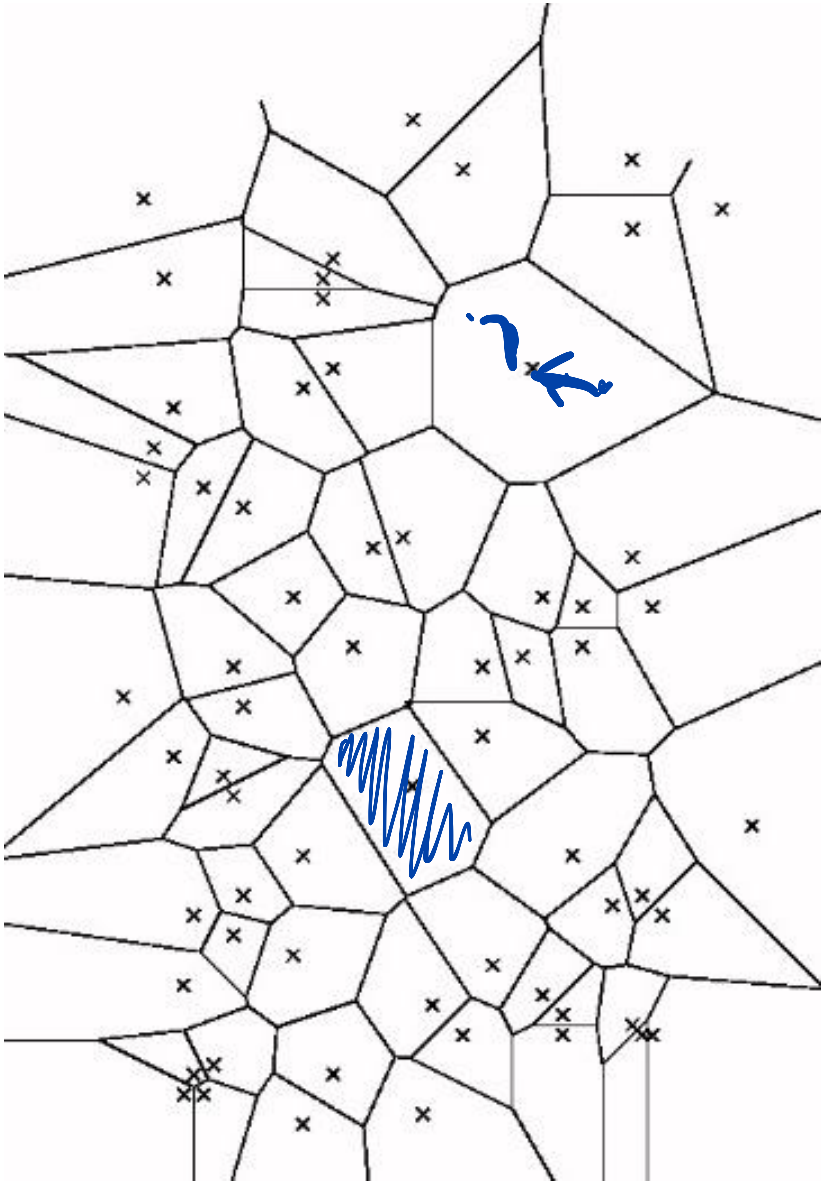
Kommentarfolie

K-means is a widely used method for clustering where the number of clusters must be specified beforehand. 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 algorithm can get stuck in a local minima.

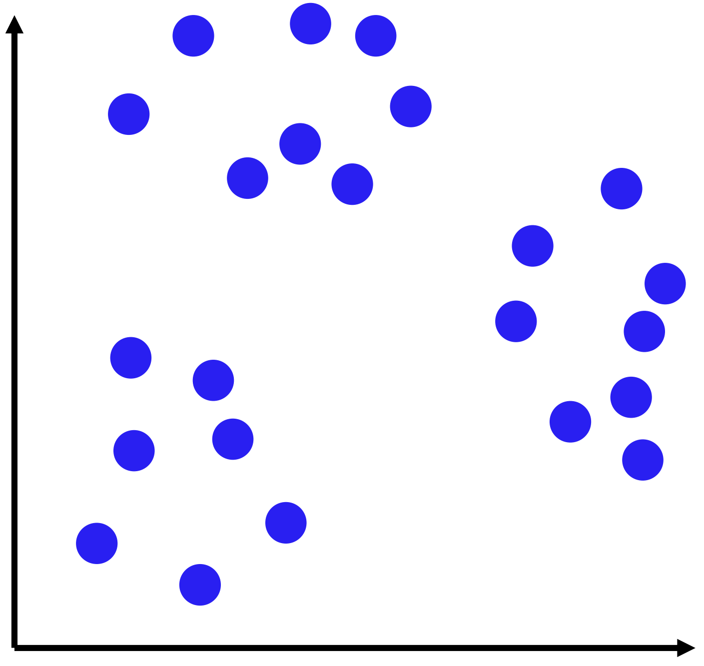
Voronoi Model

- The Voroni diagramm partiones the space in Voroni cells for each point p

- The Voroni cell for point p covers the area which nearest data point is p



Variants - K-Medoids, K-Median Clustering



- Representative: Mean → object from cluster
  - Means do not always exist
- Distance: Squared distance → normal distance
  - Influence of outliers is reduced
- Two variants for representative:
  - Medoid: Object in the "middle"
  - Median: Median of the cluster
- Basic idea:
  - Minimal distance between the objects of a cluster to its representative

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

	K-means	K-medoid	K-median
data	Numerical data (mean)	metric	ordered attributed data
efficiency	High O(tkn)	Low O(tk(n-k)^2)	High O(tkn)
Senitivity to outlivers	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

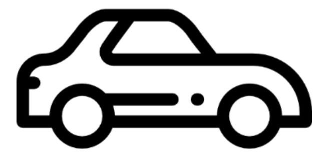
Supervised Learning: Classification

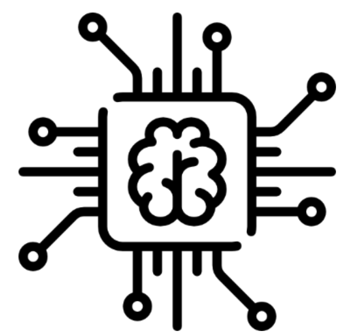
Johannes Betz / Prof. Dr. Markus Lienkamp / Prof. Dr. Boris Lohmann

(Jan Cedric Mertens, 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



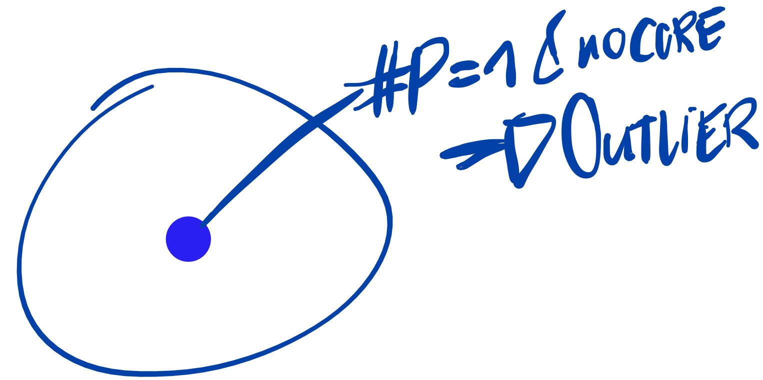


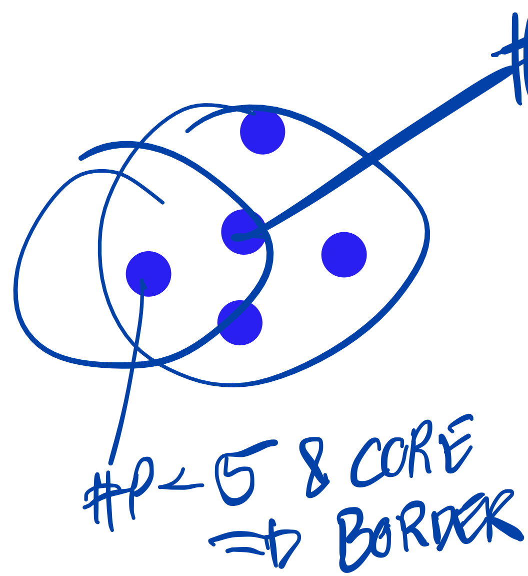
Density Based Clustering - DBSCAN

- Density-Based Spatial Clustering Application with Noise

- Two parameters
  - ε-radius neighborhood = 1
  - Minimum Points = 5

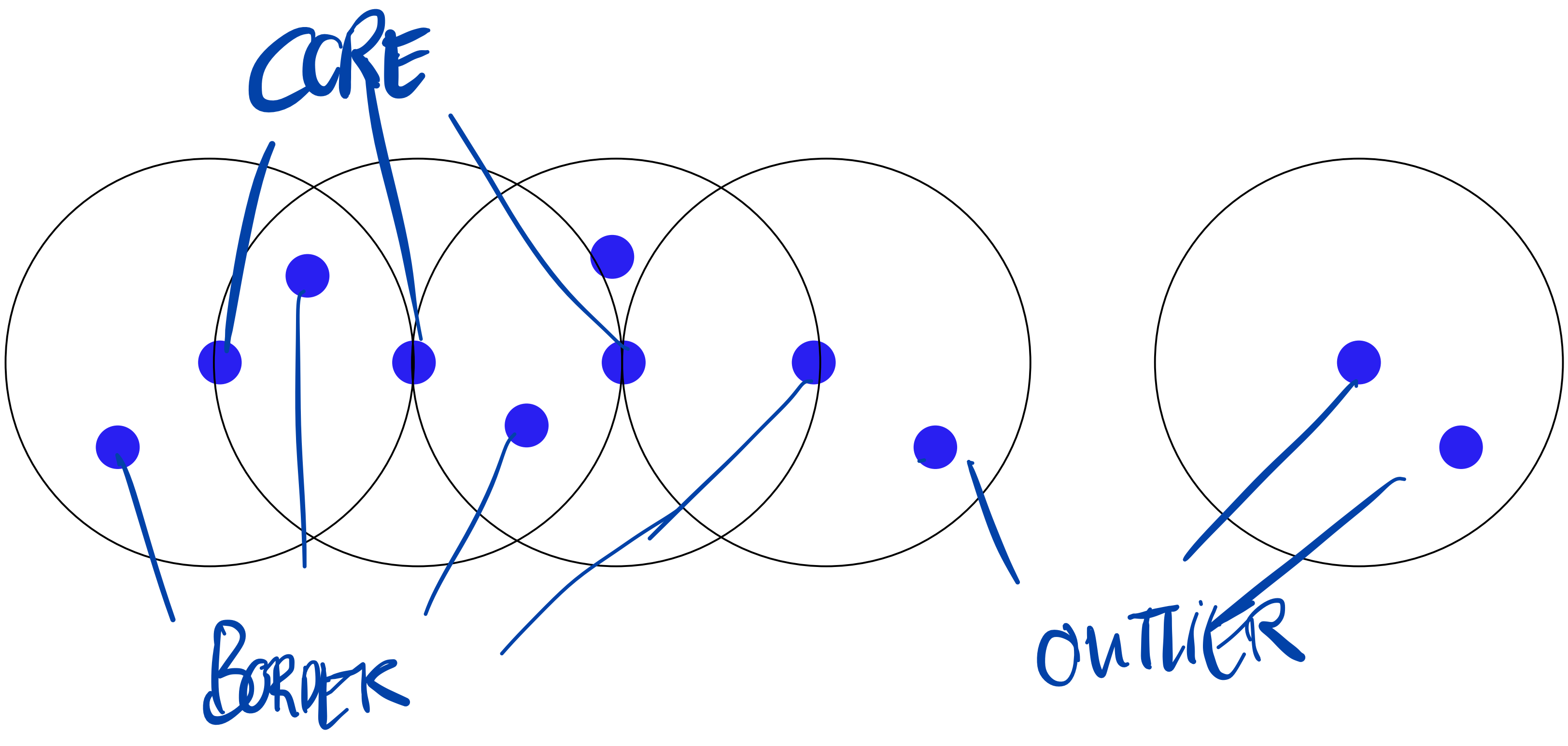
- Three Point-classes
  - Core
  - Border
  - Outlier



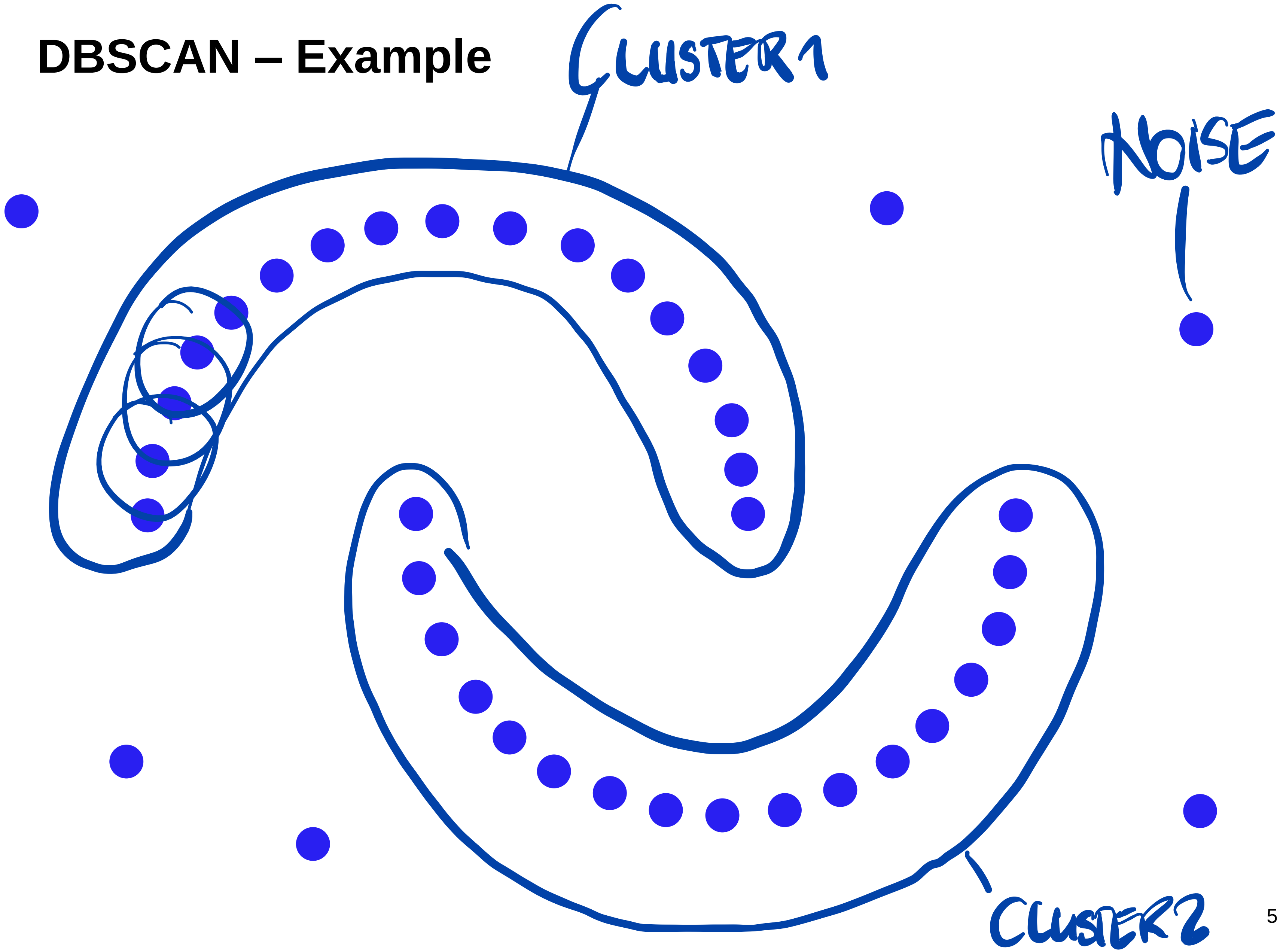


DBSCAN – Density Reachability

- 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 – Example



Discussion DBSCAN

- Pro:
  - **Cluster Form:** Arbitrary 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

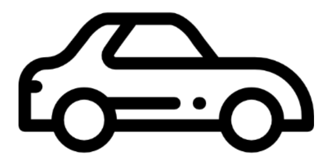
Supervised Learning: Classification

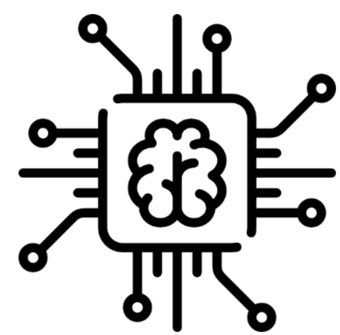
Johannes Betz / Prof. Dr. Markus Lienkamp / Prof. Dr. Boris Lohmann

(Jan Cedric Mertens, M.Sc.)

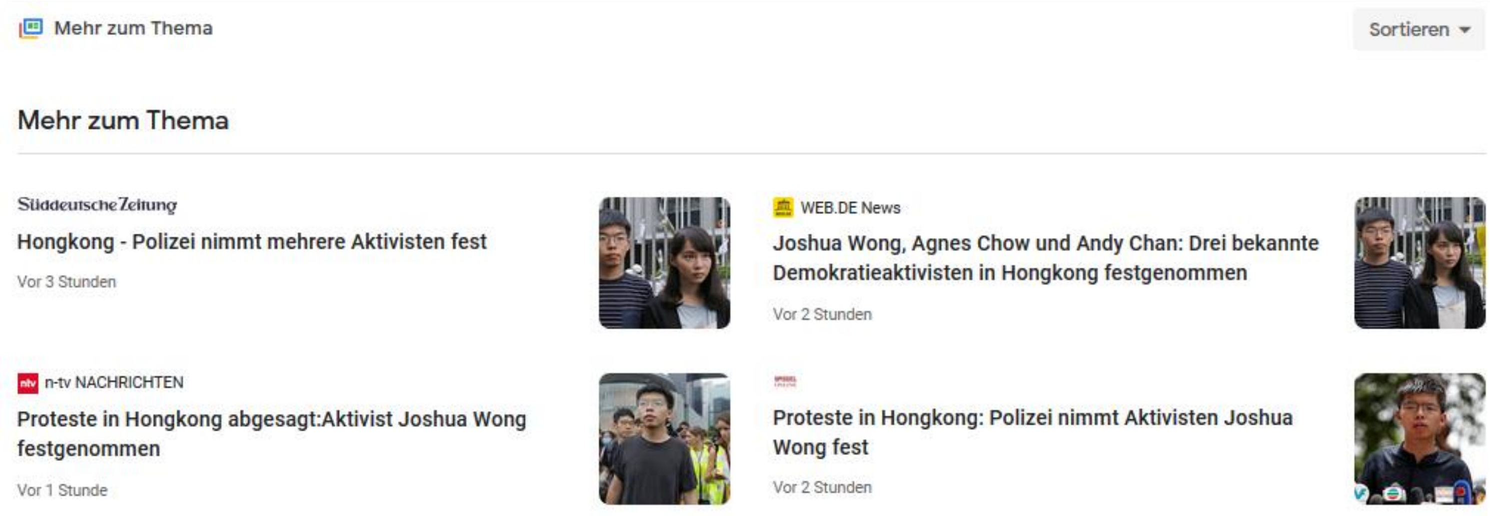
Agenda

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



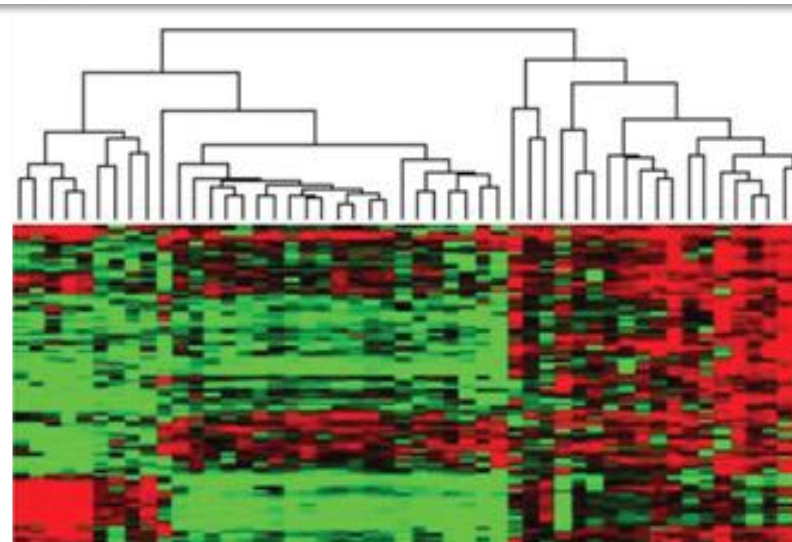


Applications



Google news

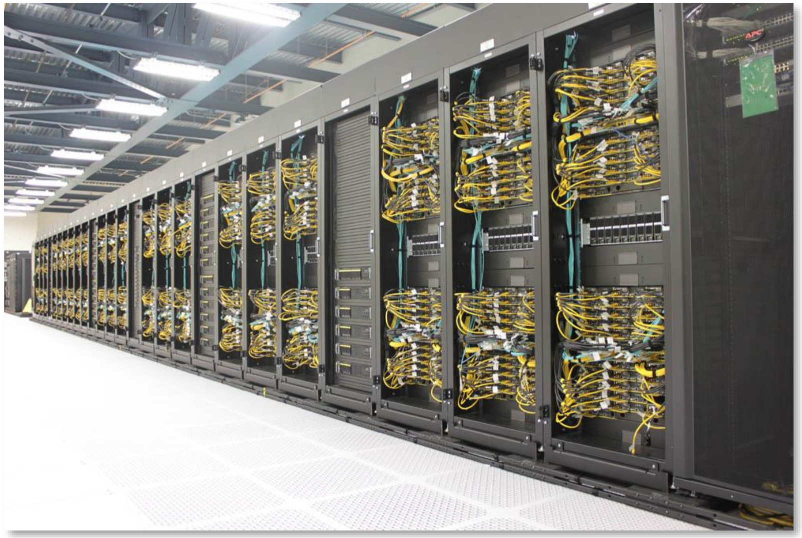
Genome patterns



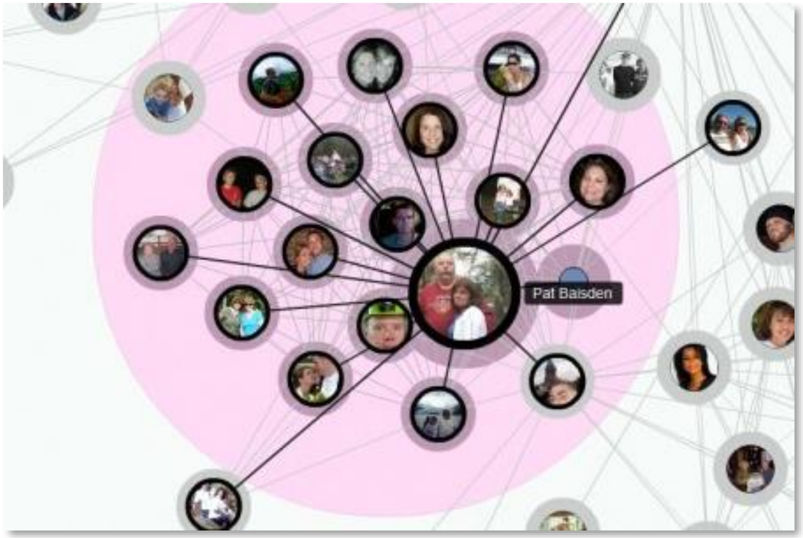
[14]

Applications

[11]



Computing cluster



[12]

Social network

Market segmentation

[13]

Applications

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

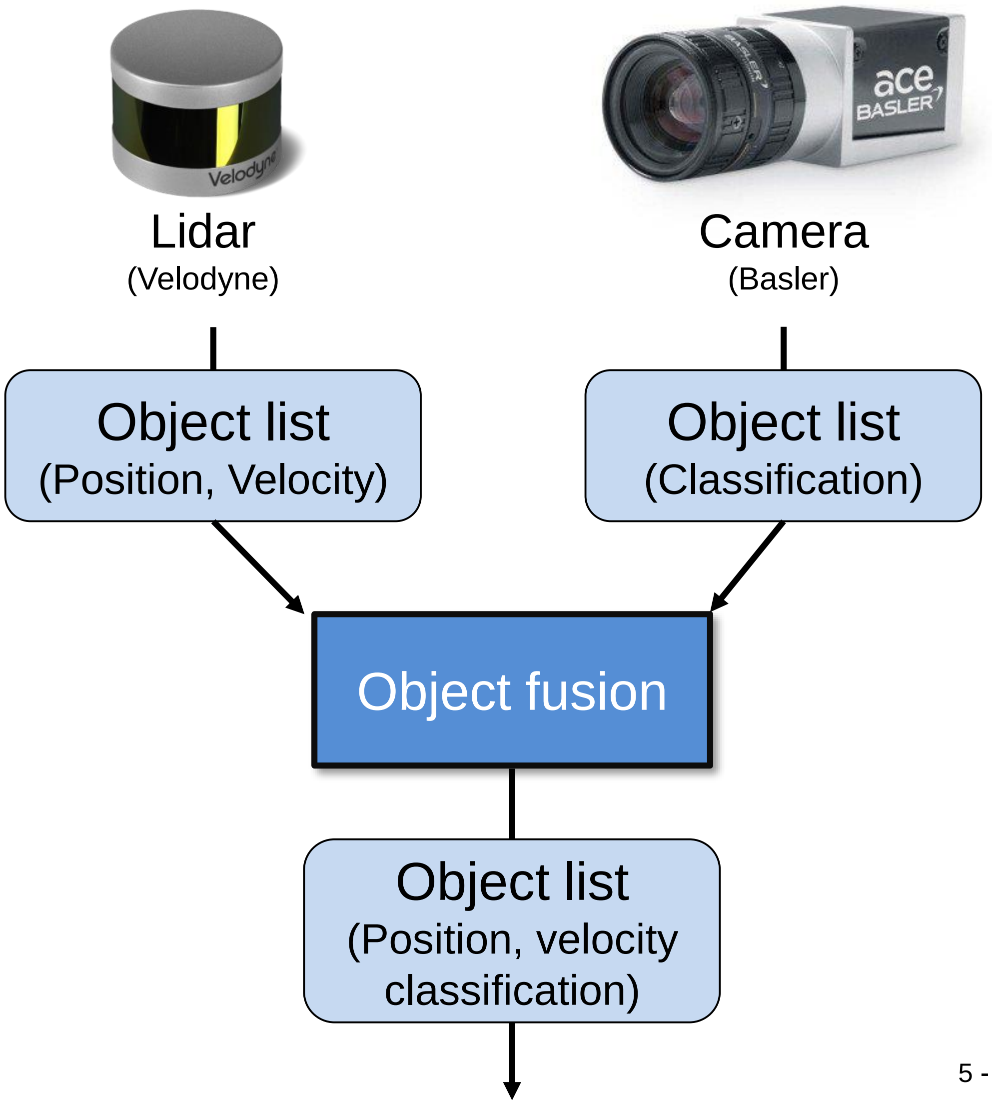
  - Netflix 1,000,000 $ challenge from 2006

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 generalisation of trajectories and traffic flow
  - Use knowledge for city planing and to identify bottlenecks

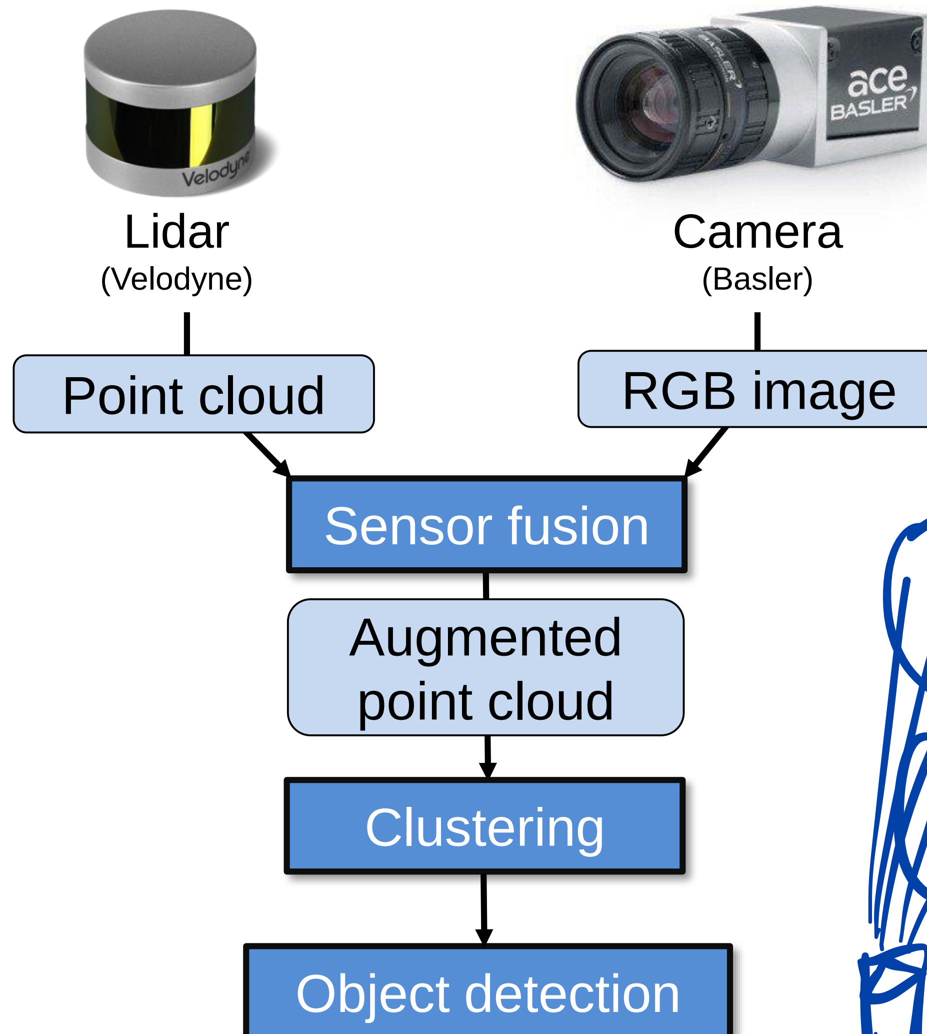
Clustering for Automotive Technology

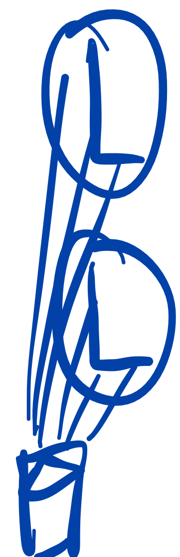
- High level object fusion
  - Object detection is based on limited data (only from one sensor)
  - Object fusion is based on processed Object-List (already information loss)



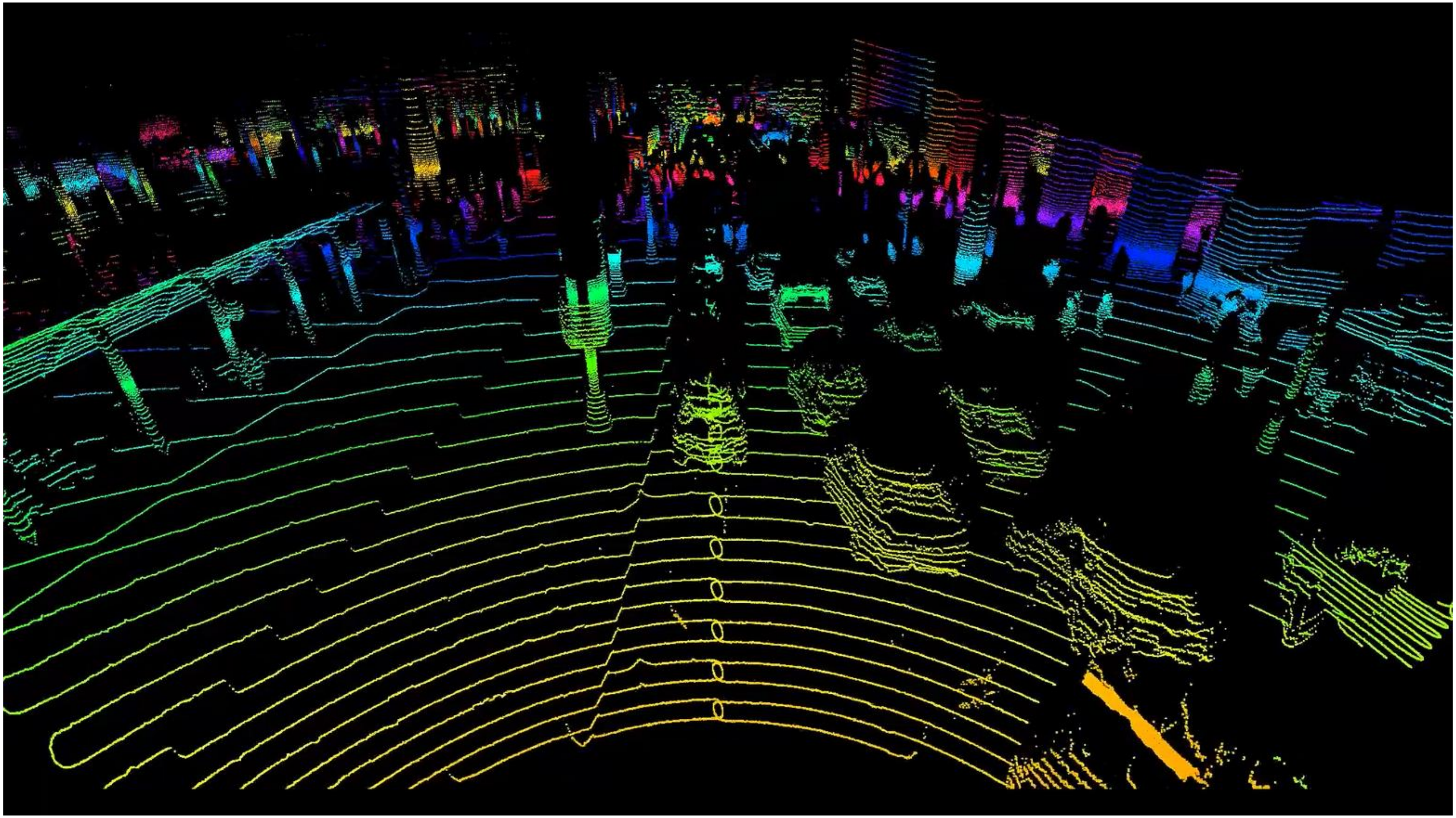
Clustering for Automotive Technology

- Low level object fusion
  - Overlay lidar point cloud with camera image
  - Find cluster in augmented point cloud
  - Object detection based on fused raw data



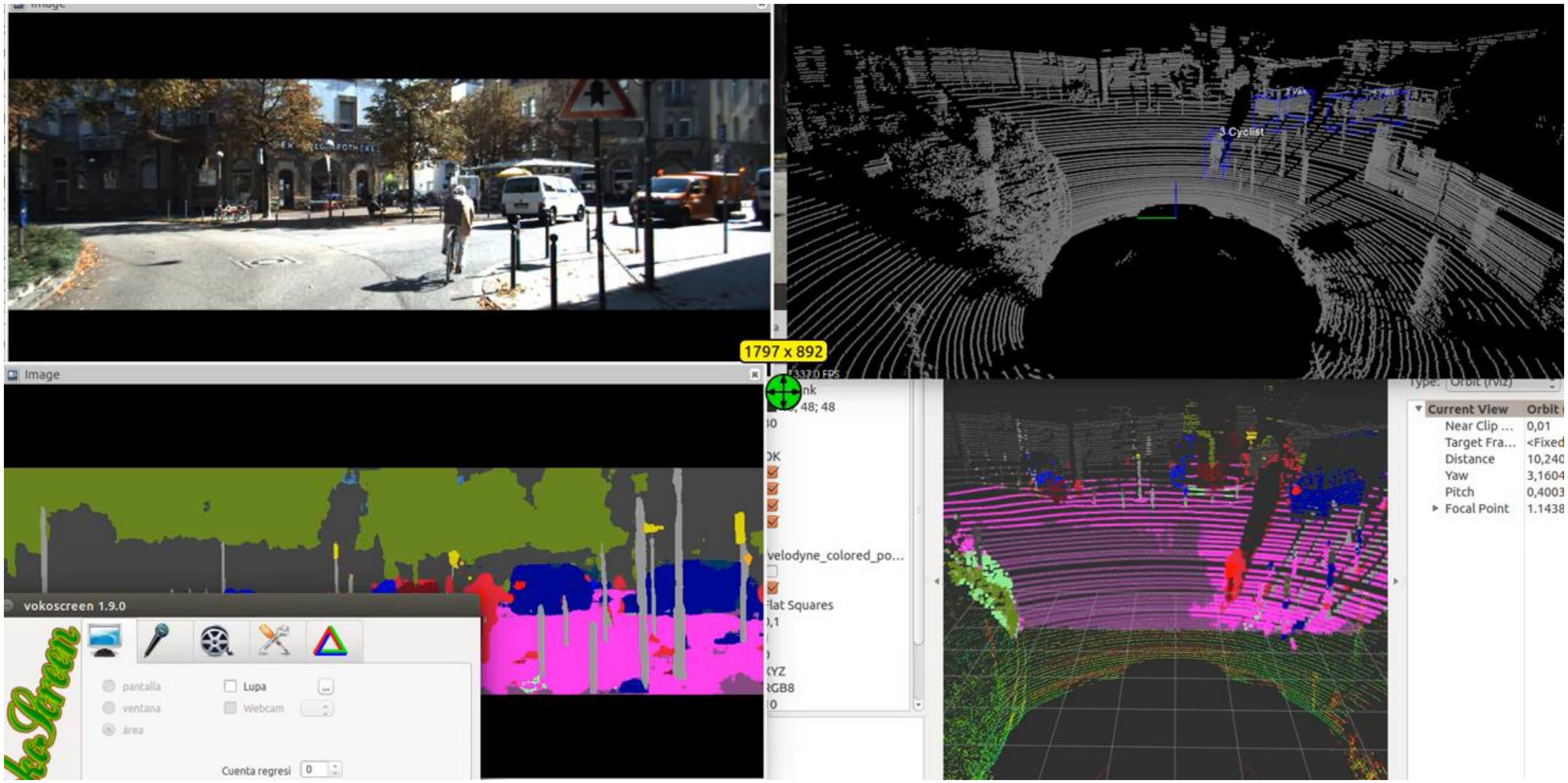


Clustering for Automotive Technology



[15]

Clustering for Automotive Technology



[4]

Supervised Learning: Classification

Johannes Betz / Prof. Dr. Markus Lienkamp / Prof. Dr. Boris Lohmann

(Jan Cedric Mertens, M.Sc.)

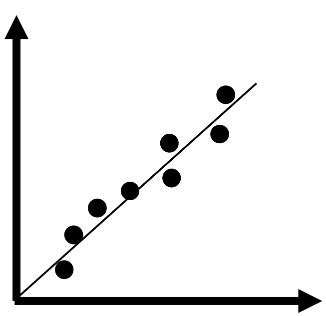
Agenda

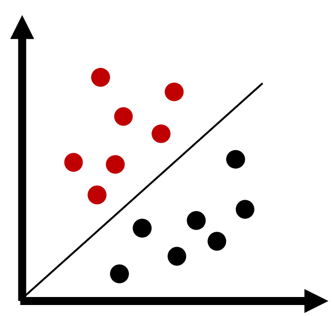
1. Chapter: Introduction
    - 1.1 Overview
    - 1.2 Training and Validation
2. Chapter: Methods
    - 2.1 Hierarchical Clustering
    - 2.2 k-means
    - 2.3 DBSCAN
3. Chapter: Application
4. **Chapter: Summary**

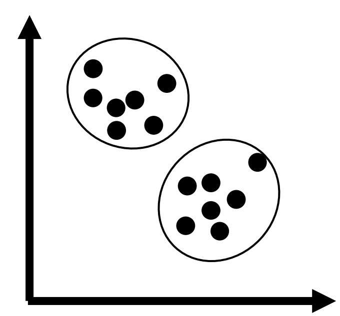
Summary

Pattern Recognition

Regression	Classification	Clustering
• Predict **continuous** valued output • Supervised	• Predict **discrete** valued output • Supervised	• Predict discrete valued output • **Unsupervised**







Summary

What did we learn today:

- **Clustering** is about finding groups in a dataset.
- Clustering is an **optimisation problem**.
- Elements **within** a cluster are **similar**.
- Elements form **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 a cluster.
- 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**.
- **Complex forms** can be grouped as clusters.
- Clustering is applied as **preprocessing** or to find **coherences**.
- Wide range of **clustering applications**, but rarely as stand alone.
- Experts or classification methods **give clusters afterwards meaning**.

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