Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Data Management for Digital Health
Revision of Exercise IV

Borchert, Dr. Schapranow

Data Management for Digital Health
Winter 2023

Exercise IV
Topics

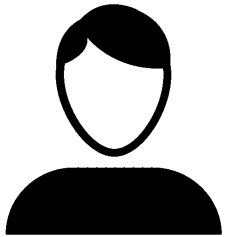
- Unsupervised Learning
- Digital Nephrology (Guest Lecture)
- Infectious Diseases
- Data Management for Epidemiology
- Medical Image Analysis

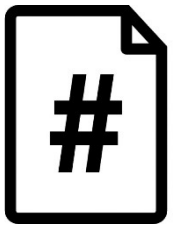
Exercise IV
Key Stats

Hasso Plattner Institut

25 Questions 50 Points	25 Students 25 Passed	Average score 45.0 / 90%	Average time 64 min

<< 3h





Q5 What can you say about the results of applying k-Means clustering to dataset 1 [Jupyter Notebook]?

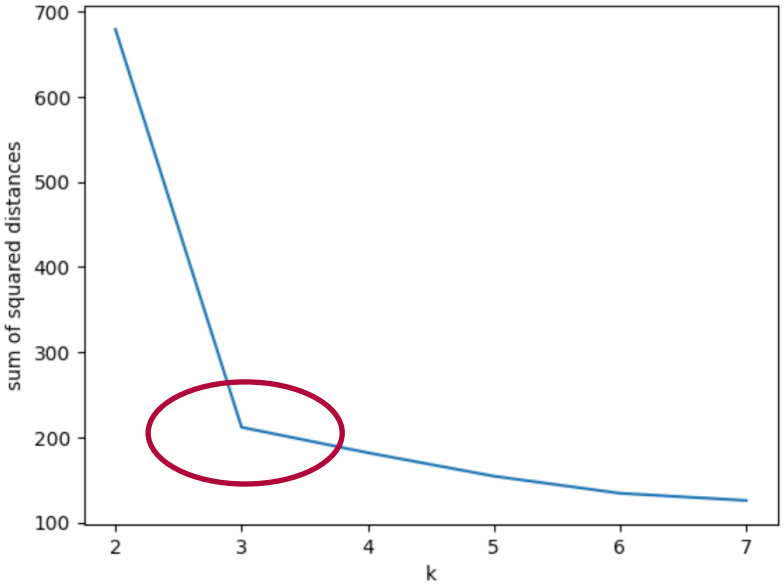
Hasso Plattner Institut

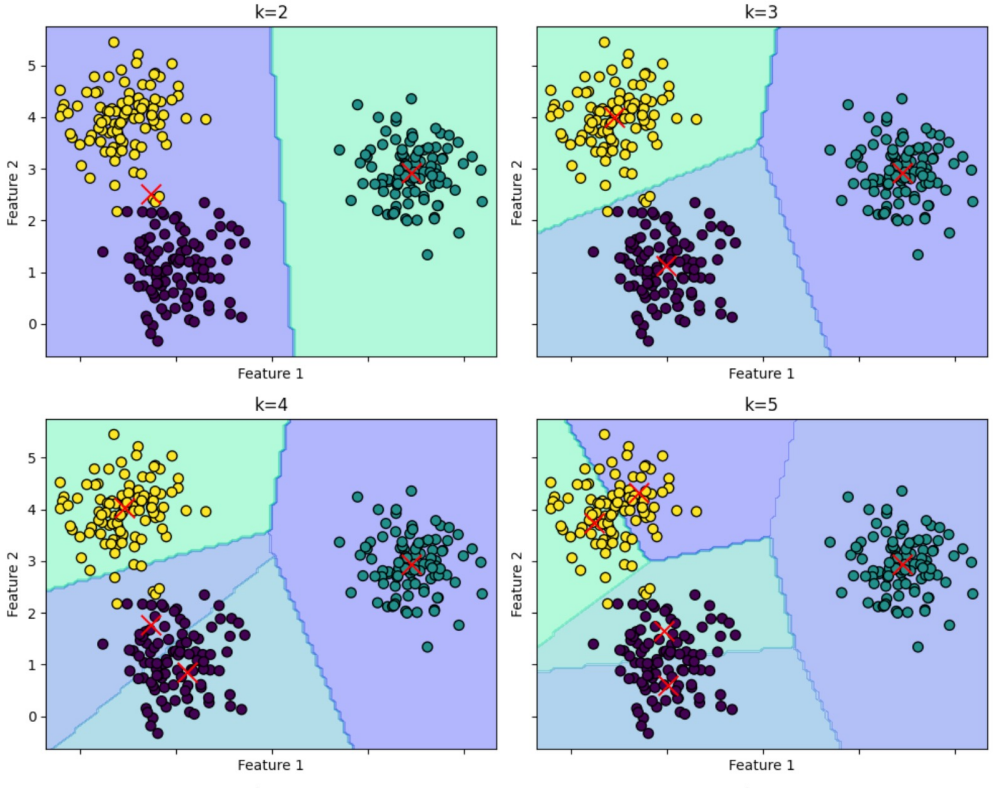
- ✓ The silhouette coefficient has its maximum for k=3, while the second highest value is attained for k=2.
- ✗ The Rand index has its maximum for k=3, while the second highest value is attained for k=2.
- ✗ In the elbow plot, the sum of squared distances decreases until k=3 and increases again for larger values of k.
- ✗ The contingency matrix for k=2 indicates perfect classification accuracy.

Frequently missed

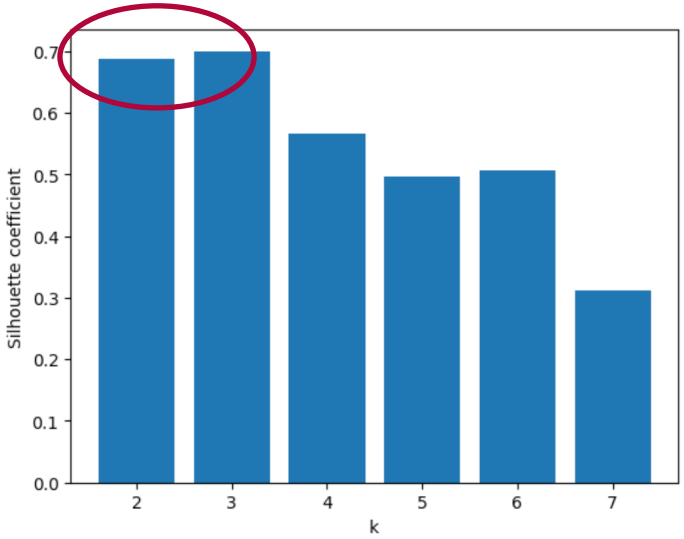
Frequent incorrect answer

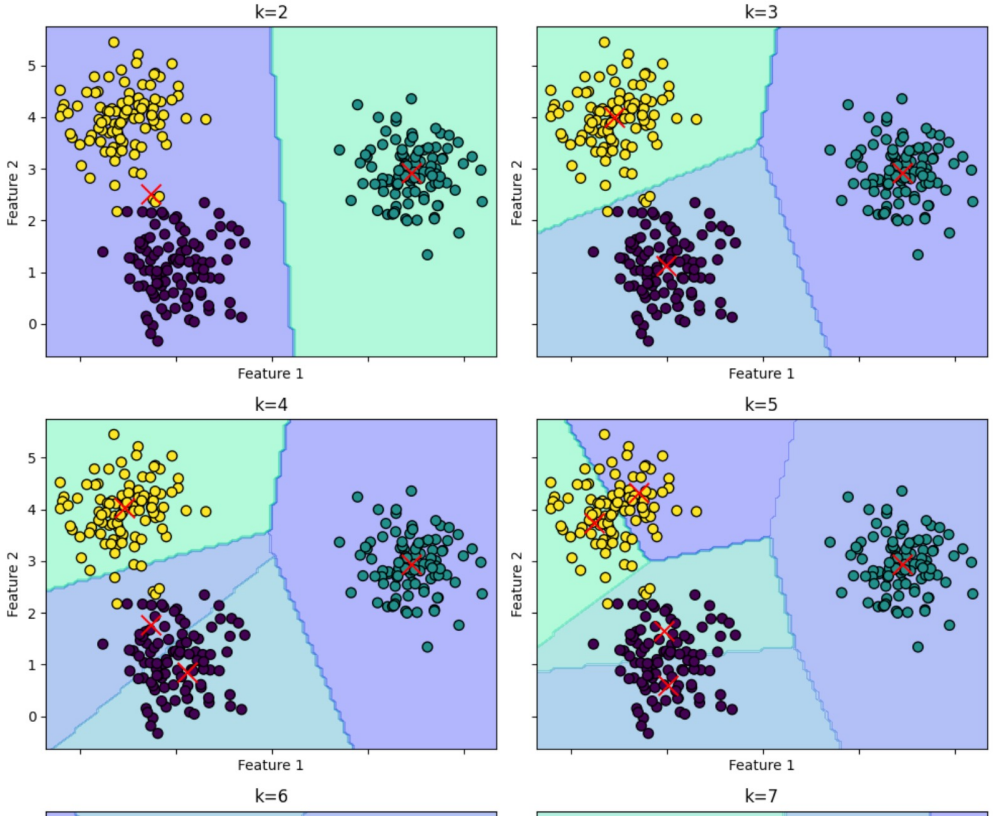
k-Means (Dataset 1)
Elbow Curve



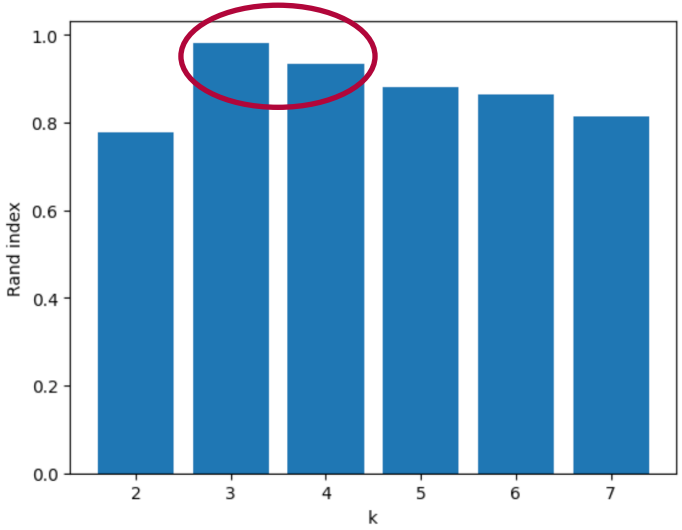


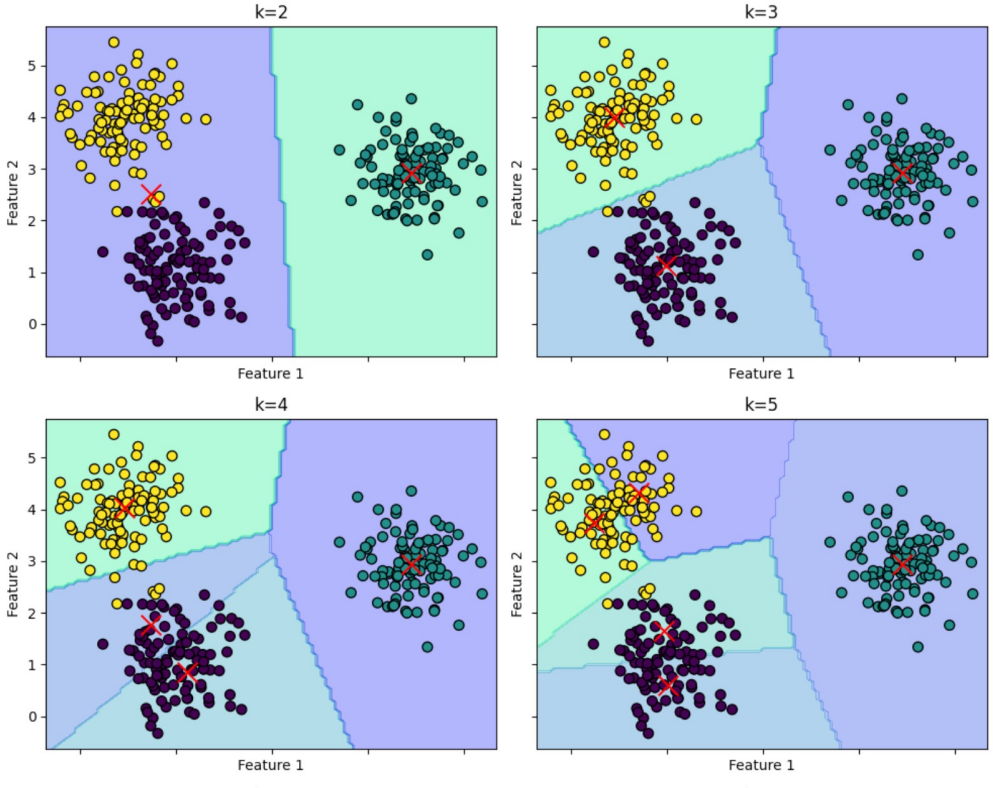
k-Means (Dataset 1)
Silhouette Coefficient





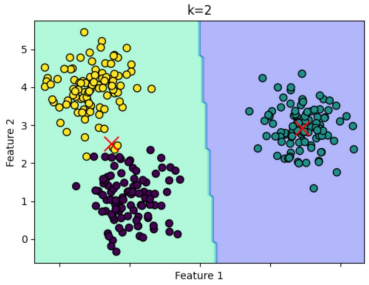
k-Means (Dataset 1)
Rand Index



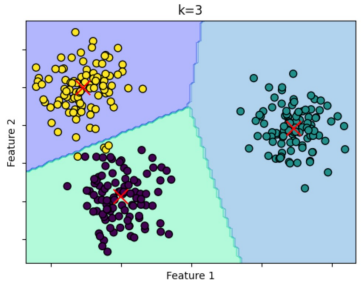


Extrinsic Evaluation

- Evaluation of the ability of clustering algorithms to separate class compared to ground truth
- Contingency Matrix
  - Similar to confusion matrix
  - How often do assignment to cluster and actual class occur together?



	Cluster 1	Cluster 2
Label 1	0	100
Label 2	100	0
Label 3	0	100



	Cluster 1	Cluster 2	Cluster 3
Label 1	100	0	0
Label 2	0	100	0
Label 3	4	0	96

Q6 What can you say about dataset 3? [Jupyter Notebook]

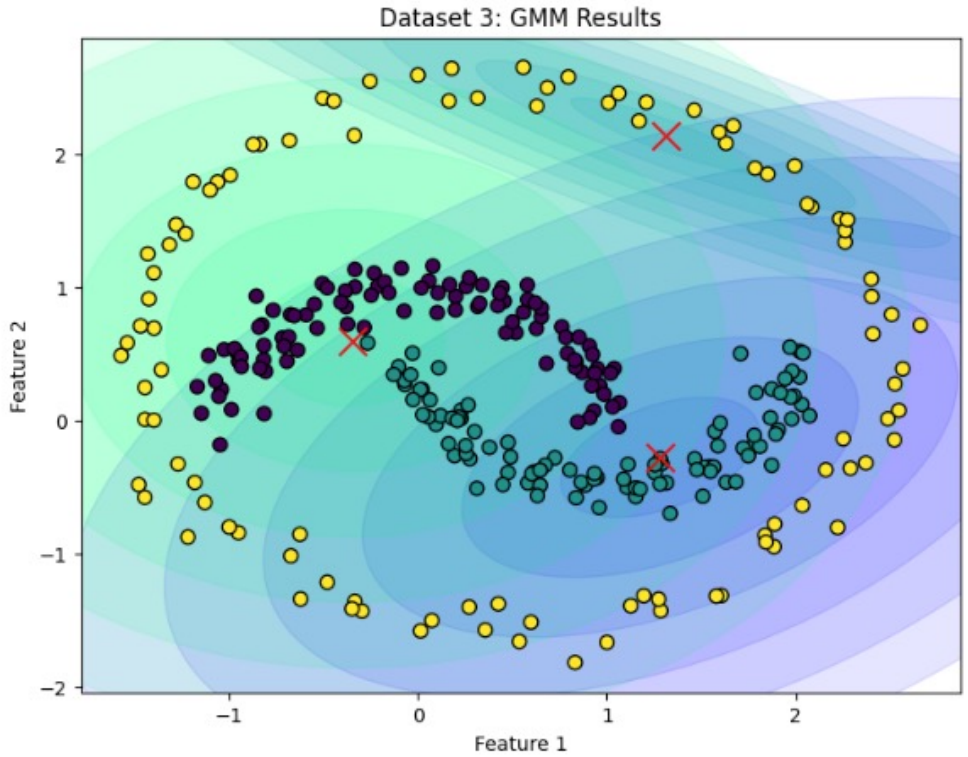
Hasso Plattner Institut

- ✓ The Gaussian Mixture model fails to find a meaningful clustering of the data.
- ✓ DBSCAN with epsilon = 0.4 and a minimum number of samples of 20 identifies 3 clusters.
- ✓ DBSCAN with epsilon = 0.4 and a minimum number of samples of 20 results in a Rand score of > 0.95, indicating a good fit of the data with respect to the ground truth.
- ✗ The Rand index score of DBSCAN increases linearly with larger values of epsilon.

Frequently missed

Frequent incorrect answer

Q6 What can you say about dataset 3? [Jupyter Notebook]



Q9 Which statements are valid for kidney transplantations in the context of the Eurotransplant Senior Program (ESP)?

Hasso Plattner Institut

- ✓ ESP is designed to address the transplantation requirements of a special group of recipients, i.e., age 65 and older.
- ✓ The ESP prioritizes short transport time of the donor organ over immunological matching of recipient and donor.
- ✗ Older transplanted patients have better survival rates compared to younger patients.
- ✗ More than 90% of patients undergoing kidney transplantation in Germany are older than 65 years.

Frequently missed

Frequent incorrect answer

ESP (Guest Lecture Marcel Naik)

	European Senior Programme
Age Recipient	≥ 65
Age donor	≥ 65
HLA	**No Matching**
Cold ischemia	**As short as possible**
Option „Urgency“	Yes
Waiting time	Yes
	Started 1999

Q16 Please select all appropriate statements about the reproduction number R as discussed in class.

- ✓ The effective reproduction number describes the average number of new infections caused by a single case at a given point in time.
- ✓ The basic reproduction number is hard to measure, therefore it is typically estimated.
- ✗ The effective reproduction number is hard to measure, therefore it is typically estimated.
- ✗ The basic reproduction number describes the total number of cases during a disease outbreak.

Frequently missed

Frequent incorrect answer

Reproduction number R

- Basic reproduction R_0 (typically estimated) := Expected number of new cases caused by a single case at t_0 when all individuals were in compartment S
- Effective reproduction R_t (observed) := Avg. number of new cases caused by a single case at time point t (this is what you find in situational reports)
- Linking R_t and R_0: Let s be the proportion of people in compartment S, who can get infected (e.g. no immunity): $R_t = R_0 * s$
- Herd immunity := Indirect protection against infectious diseases once a specific percentage p_{immune} of the population has become immune so that $R_t < 1$.
  - $R_t < 1 \Leftrightarrow R_0 * s < 1$
  - $R_0 * (1-p_{immune}) < 1$
  - $p_{immune} > 1-R_0^{-1}$