

1. *Feature Selection*. Given the dataset “features.csv” (available in TeachCenter), which comprises 20 features and one target variable, select the most informative features for the target. Use a feature selection method of your choice.
- (a) Describe preprocessing and feature transformations steps if you made any (e.g., Did you create new features? Did you normalize the data?). Max. three sentences.
  - (b) What is your feature selection method and why? Again max. three sentences.
  - (c) Give ranking of your features together with the scores you got for them.

**Answer (a)** - Preprocessing & feature transformations:

- 

**Answer (b)** - Feature selection method:

- 

**Answer (c)** - Feature ranking:

- Feature #1, Score
- Feature #2, Score
- ...

2. *Dimensionality Reduction.* Given the previous dataset “features.csv” perform dimensionality reduction with PCA on the features (i.e., ignore the target variable in this task).

- (a) Describe preprocessing and feature transformations steps if you made any. Max. three sentences.
- (b) How many dimensions have you selected for this dataset. Explain how you selected them and why. Again max. three sentences.
- (c) Give feature loadings for the dimensions you selected.
- (d) Interpret the dimensions that you selected. Again max. three sentences.

**Answer (a)** - Preprocessing & feature transformations:

- 

**Answer (b)** - Selecting dimensions:

- 

**Answer (c)**

Feature	PCA-1	PCA-2	...
×	×	×	×
×	×	×	×

**Answer (d)**

- PCA-1, Interpretation
- PCA-2, Interpretation
- ...

3. *Classification.* Given the dataset “class.csv” (available in TeachCenter), which comprises 20 features and one binary target variable, implement a classifier of your choice. Evaluate your classifier by a metric of your choice. If your model has hyperparameters cross-validate.

- (a) Describe preprocessing and feature transformations steps if you made any. Max. three sentences.
- (b) What is your model and why? Again max. three sentences.
- (c) Describe your evaluation setup. Max. one sentence.
- (d) Describe hyperparameter optimization if any. Give the final values of hyperparameters. Max. three sentence.
- (e) Give your evaluation results as text or a table.

**Answer (a)** - Preprocessing & feature transformations:

- 

**Answer (b)** - Model choice:

- 

**Answer (c)** - Evaluation setup:

- 

**Answer (d)** - Hyperparameters:

- 

**Answer (e)** - Results:

-

4. *Clustering.* Given the dataset “clustering.csv” (available in TeachCenter), which comprises 10 features, implement a clustering method of your choice.

- (a) Describe preprocessing and feature transformations steps if you made any. Max. three sentences.
- (b) What is your clustering algorithm and why? Again max. three sentences.
- (c) Use an internal evaluation metric to estimate the number of clusters. Plot the evaluation results as `evaluation_metric` vs. `#clusters`.
- (d) How many clusters did you select? Max. one sentence.

**Answer (a)** - Preprocessing & feature transformations:

- 

**Answer (b)** - Clustering algorithm:

- 

**Answer (c)** - Evaluation plot:

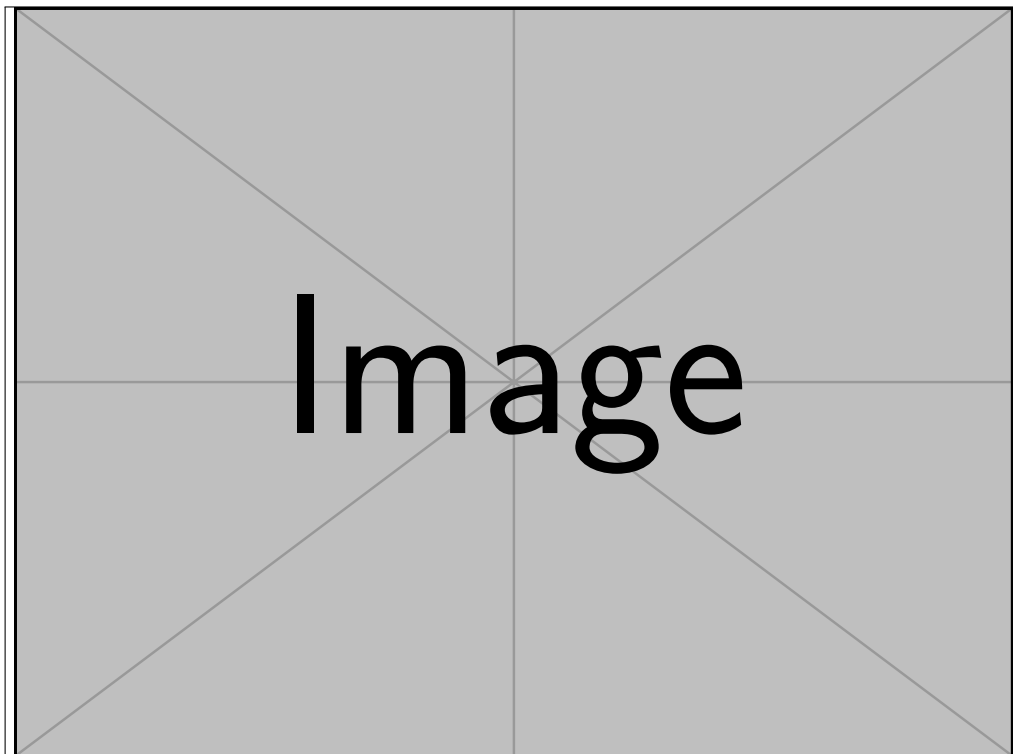


Figure 1: Cluster quality vs. number of clusters

**Answer (d)** - How many clusters:

-