Rail Data Science

AI Methods

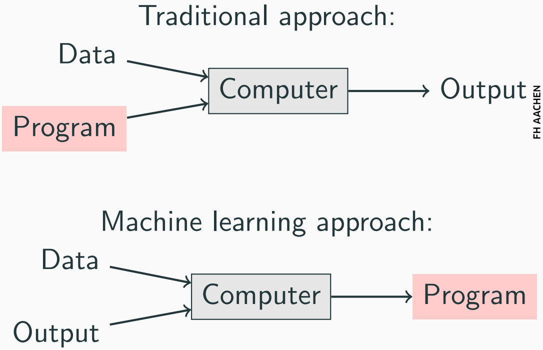
Prof. Dr. Raphael Pfaff

13. Februar 2021

Fachhochschule Aachen

What is machine learning?

- Different paradigm:
  - Derive rules from observations
- Typical procedure:
  - **Train:** Observe set of examples "training data"
  - **Fit:** Infer information of process behind data
  - **Test:** Make prediction on unseen data "Test data"
- Variations:
  - Supervised learning: provide labels with data
  - Unsupervised learning: cluster data based on patterns



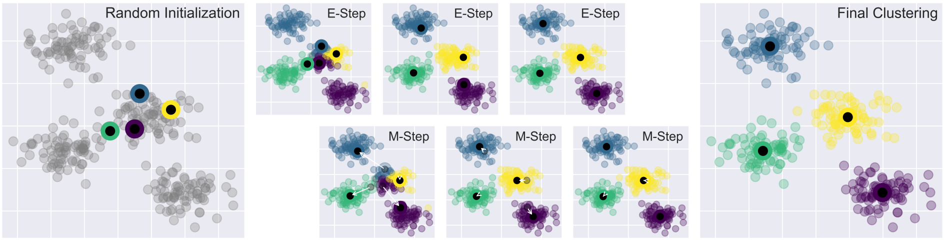
Clustering

Hierarchical clustering

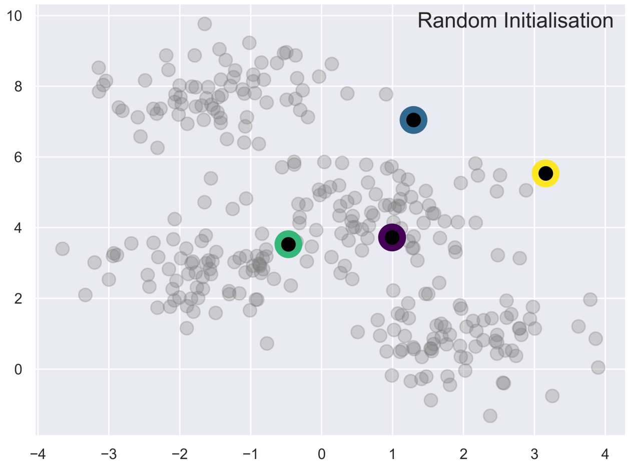
1. Assign each item to one cluster
2. Find closest (most similar cluster): merge them
3. Continue until desired number of clusters k is reached

K-means clustering

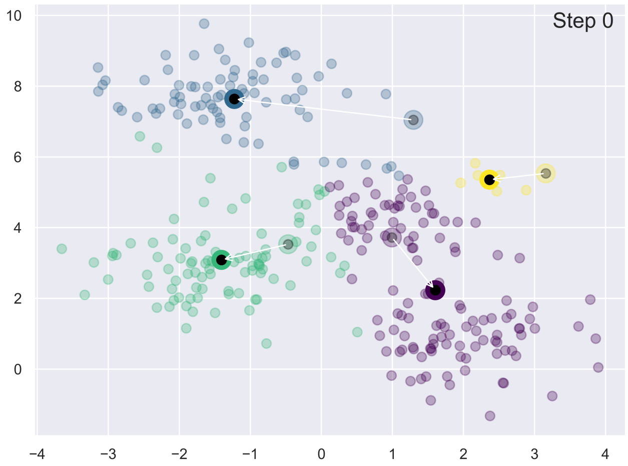
1. Randomly choose k initial centroids
2. While centroids don't change
   - E: Create k clusters by assigning each item to closest centroid
   - M: Compute k new centroids by averaging items in each cluster



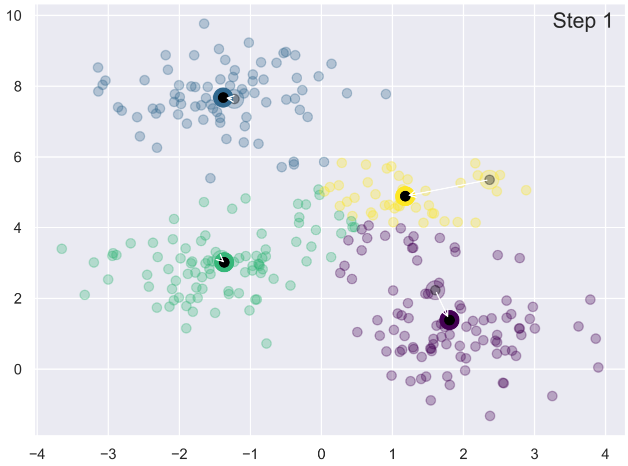
K-means clustering procedure



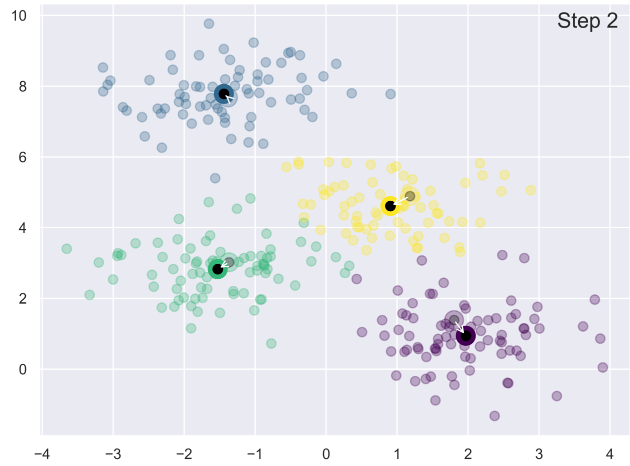
K-means clustering procedure



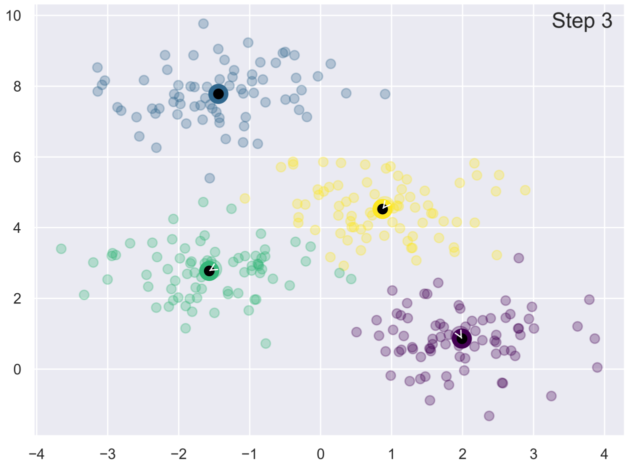
K-means clustering procedure



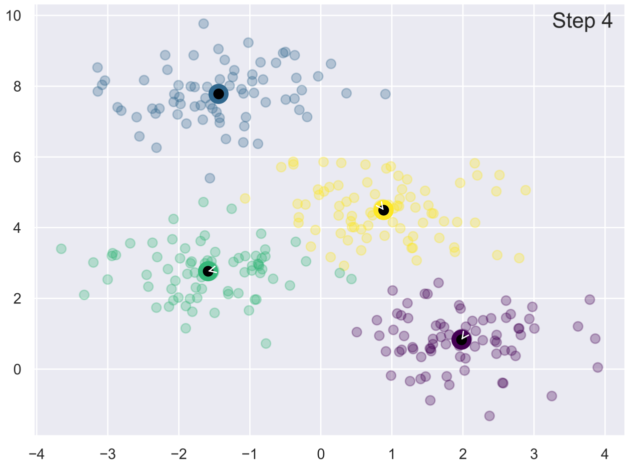
K-means clustering procedure



K-means clustering procedure

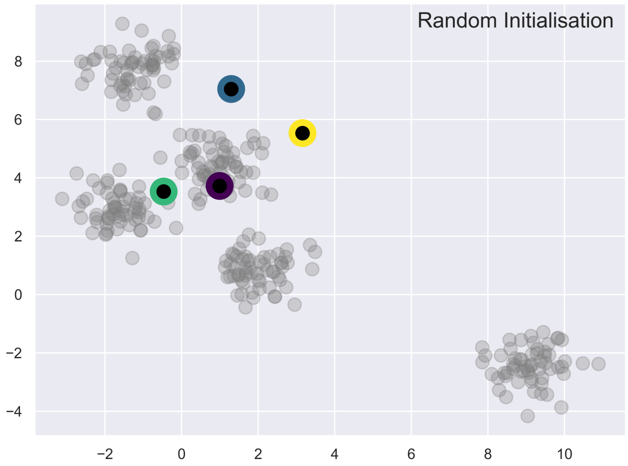


K-means clustering procedure



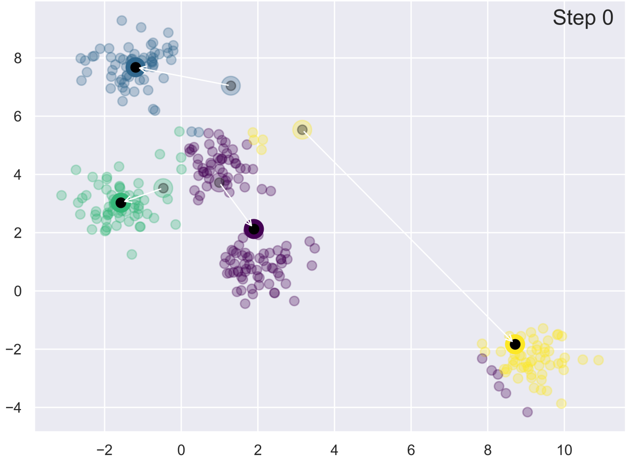
FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

K-means inappropriate cluster number

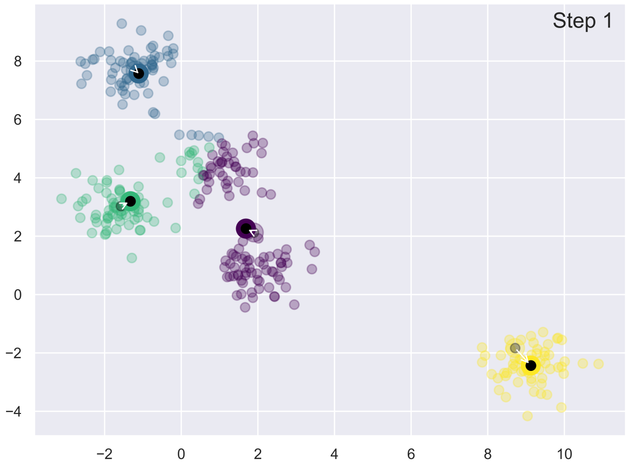


FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

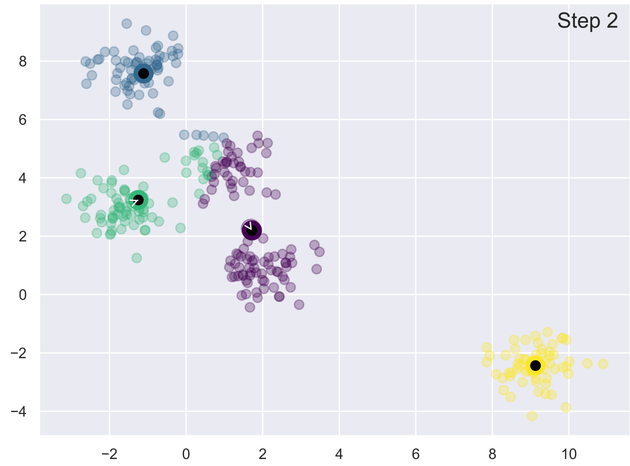
K-means inappropriate cluster number



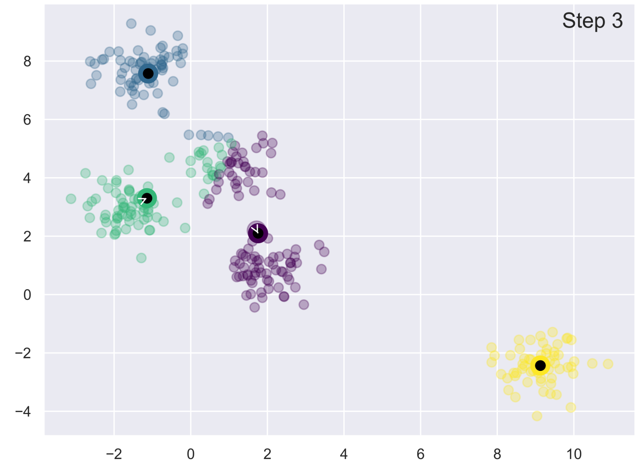
K-means inappropriate cluster number



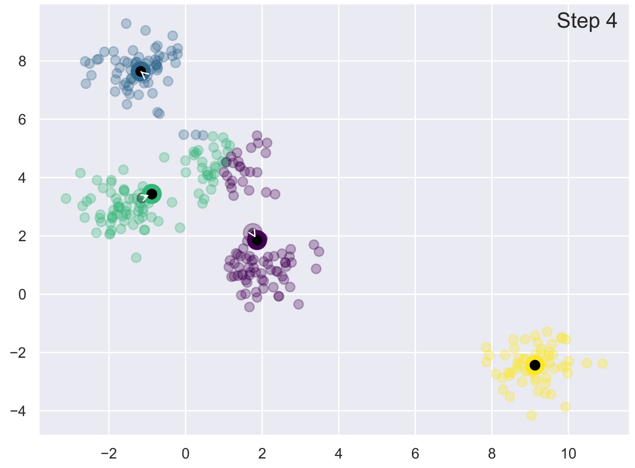
K-means inappropriate cluster number



K-means inappropriate cluster number

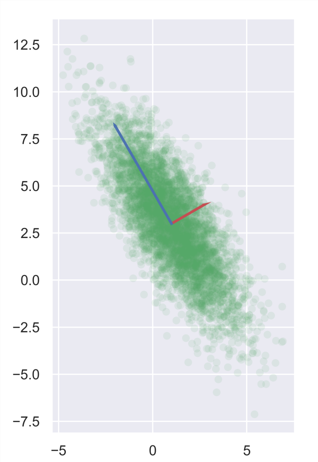


K-means inappropriate cluster number



Principal Component Analysis (PCA)

- Sample $X_1, \ldots, X_n$ is a cloud of points in $\mathbb{R}^d$
- Typically, d is large
- **Question:** can we project the point cloud onto a linear subspace of dimension $d' < d$ while keeping as much of the information as possible?

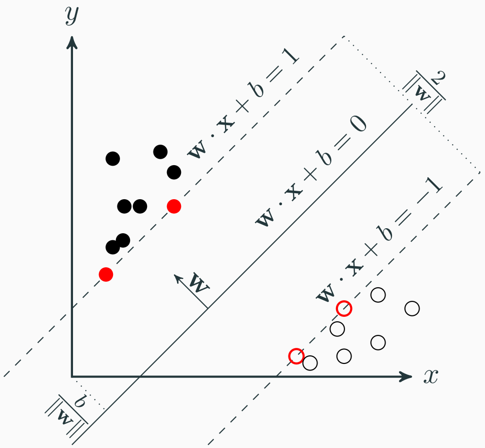


Principal Component Analysis (PCA)

- Sample $X_1, \ldots, X_n$ is a cloud of points in $\mathbb{R}^d$
- Typically, d is large
- **Question:** can we project the point cloud onto a linear subspace of dimension $d' < d$ while keeping as much of the information as possible?
- **Answer:** PCA achieves this by keeping important aspects of the covariance structure in orthogonal directions.

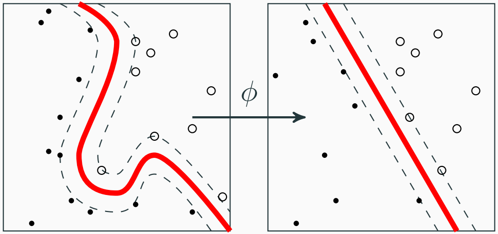
Support Vector Machine

- Idea:
  - Find separating hyperplane maximising the distance between borderline instances
  - For data mixed in feature space, slack parameter (and penalty C) is used
  - If no separating hyperplane can be found in feature space, dimension is increased "Kernel trick"
- Robust:
  - High dimensionality
  - Small datasets
- Simple to complex models



Support Vector Machine

- Idea:
  - Find separating hyperplane maximising the distance between borderline instances
  - For data mixed in feature space, slack parameter (and penalty C) is used
  - If no separating hyperplane can be found in feature space, dimension is increased "Kernel trick"
- Robust:
  - High dimensionality
  - Small datasets
- Simple to complex models



Practical hints for application of data science methods

- Get to know your data
- Use training and validation sets
- For clustering and ML:
  - Regularise
  - Remove outliers, NaN
- When building models:
  - Don't expect perfect fit
  - Inspect confusion matrix

Prediction outcome

		p	**n**	**total**
actual value	**p′**	True Positive	False Negative	P′
	n′	False Positive	True Negative	N′
	total	P	N	

Get to work!