OST
Ostschweizer
Fachhochschule

Lab 12 (solution)

FTP MachLe MSE
HS 2023

Unsupervised Learning: Clustering

Machine Learning
WÜRC

After this unit, ...

Lernziele/Kompetenzen

- you know the *three* clustering algorithms: *k-means*, *dbscan* and a*agglomerative clustering* (using average, complete or Ward linkage).
- you are able to explain the working principle of *k-means*, *dbscan* and *agglomerative clustering*, their advantages and disadvantages and to apply them to data using `scikit-tlearn` in Python.
- you are able to plot the `inertia` and to determine the elbow point of this curve to find an optimum number of clusters as *hyperparameter*.
- you know the way how to *evaluate* a cluster algorithm using *metrics*, namely using `ARI` (*adjusted rand index*), `NMI` (*normalized mutual information*), `SC` (*silhouette score*) and *inertia*.
- you are able to correctly *scale* the data before clustering is applied especially (MinMax, StandardScaler, RobustScaler) or to meaningfully `transform` the data (eg. using `PCA`, `t-SNE` or `NMF`) before a clustering algorithm is applied.
- you know what a *Gaussian Mixture Model* `GMM` is and how the *expectation maximization algorithm* (`EM` algorithm) works. You are able to interpret the `kMeans` algorithm as a form of an `EM` algorithm with an E-step and and M-Step.
- your are able to apply clustering on the faces dataset (agglomerative, k-means and dbscan) to detect and *group* similar faces.
- your are able to apply a *hierarchical cluster analysis* on a voting dataset.

1. Clustering Algorithms [M,I]
This clustering algorithm initially assumes that each data instance represents a single cluster.

Welche der folgenden Aussagen sind wahr und welche falsch?	**wahr**	**falsch**
a) agglomerative clustering	●	○
b) t-SNE	○	●
c) k-means clustering	○	●
d) expectation maximization	○	●

2. Elbow Curve and `sklearn.cluster.KMeans` [A,I]
The solution Juypter notebook can be found on moodle:
`Lab12_A2_MakeBlobs_kMeans.ipynb`

3. k-Means, Gaussian Mixture Models and the EM algorithm [A,II]
The solution Juypter notebook can be found on moodle:
`Lab12_A3_EM_KMeans_MixtureModels.ipynb`

4. Image compression using `kMeans` [A,II]
The solution Juypter notebook can be found on moodle:
`Lab12_A4_ImageCompressionUsingKMeans.ipynb`

5. Detecting similar faces using `DBSCAN` [A,II]
The solution Juypter notebook can be found on moodle:
`Lab12_A5_DBSCAN_DetectSimilarFaces.ipynb`