

CS-E4650 - Assignment 2

Arttu Häkkinen - 596077

October 23, 2021

Task 4

a)

The alternative internal validation index, τ , receives values between 0 and 1. The greater the value, the better the clustering according to this metric. Based on the τ computed, the best clustering for both K-means and spectral clustering methods is now $K = 2$, since τ reaches its maximum value with $K = 2$ clusters.

b)

The results in a table below for K-means clustering:

	SI	DB	τ
K=2	0.35	1.17	0.73
K=3	0.36	0.88	0.61
K=4	0.35	0.88	0.51
K=5	0.35	0.90	0.45

The results in a table below for spectral clustering:

	SI	DB	τ
K=2	0.35	1.17	0.58
K=3	0.36	0.88	0.35
K=4	0.35	0.88	0.33
K=5	0.35	0.90	0.31

The best choice for K-means clustering would be SI, since it indicates the correct amount of clusters $K = 3$. For spectral clustering, none of the indexes indicate the correct amount of clusters, so I cannot say that τ index is a better choice in this case.

References

- [1] <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html>
- [2] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html
- [3] <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>
- [4] <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- [5] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.davies_bouldin_score.html
- [6] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.normalized_mutual_info_score.html
- [7] <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.SpectralClustering.html>