

Labwork 2 - Clustering

1 K-means

Select two datasets from UCI, and for each of them:

1. Run experiments with k-means. Explain the experimental protocol.
2. Explain the centroid initialization of the k-means used in the first question.
3. Analyze and compare the results with different values of k.
4. Calculate the clustering quality (any criteria in slides).

2 Subspace clustering

1. Select one high-dimensional dataset (>100 features)
2. Use PCA or SVD to visualize the data distribution in 2D/3D.
3. Apply any clustering method (DBSCAN, K-means, SOM, GNG, etc.) for the selected datasets in 2D/3D after using PCA/SVD. Compare the performance before and after the dimensionality reduction.
4. Select a random subspace of the dataset. Re-apply any clustering method on the subspace. Discuss and compare the performance with the previous results.