LAB-Experiment-8(Unsupervised Learning)

ML-LAB (Dr.Trilok Nath pandey)

Lab Question: K-means and K-medoid Clustering in Python

You are provided with a dataset

About Dataset

Measurements of geometrical properties of kernels belonging to three different varieties of wheat.

A soft X-ray technique and GRAINS package were used to construct all seven, real-valued attributes.

Dataset Information:

The examined group comprised kernels belonging to three different varieties of wheat: Kama, Rosa and Canadian, 70 elements each, randomly selected for the experiment. High quality visualization of the internal kernel structure was detected using a soft X-ray technique. It is non-destructive and considerably cheaper than other more sophisticated imaging techniques like scanning microscopy or laser technology. The images were recorded on 13x18 cm X-ray KODAK plates. Studies were conducted using combine harvested wheat grain originating from experimental fields, explored at the Institute of Agrophysics of the Polish Academy of Sciences in Lublin.

The data set can be used for the tasks of classification and cluster analysis.

About this file

To construct the data, seven geometric parameters of wheat kernels were measured:

area A,

perimeter P,

compactness $C = 4piA/P^2$,

length of kernel,

width of kernel,

asymmetry coefficient

length of kernel groove.

Perform K-means and K-medoid clustering on this dataset using Python. You are required to accomplish the following tasks for k values 3, 5 and 7:

1. Load the dataset and preprocess it if necessary.

2. Implement K-means clustering algorithm to segment the customers into K clusters.

3. Implement K-medoid clustering algorithm to segment the customers into K clusters.

4. Visualize the clusters obtained from both K-means and K-medoid algorithms.

5. Compare the results of K-means and K-medoid clustering based on their efficiency, robustness, and cluster interpretability. (For Each Case)

Dataset Description:

The dataset is provided in a CSV file named " Seed_Data_Clustering.csv".

Note: You are encouraged to use appropriate Python libraries for data manipulation, clustering algorithms, and visualization. Ensure to include comments in your code to explain each step clearly.

Marks Distribution:

1. Data Preprocessing and Loading: 5 marks

2. Implementation of K-means Clustering: 10 marks

3. Implementation of K-medoid Clustering: 10 marks

4. Visualization of Clusters: 10 marks

5. Comparison and Analysis: 15 marks

Total Marks: 50

Submission Instructions:

Submit your Python code along with a brief report discussing your observations, insights, and comparisons between K-means and K-medoid clustering algorithms. Ensure your code is well-commented and organized.

Note to Students:

Ensure you understand the concepts of K-means and K-medoid clustering thoroughly before attempting this lab. Experiment with different values of K and observe how it affects the clustering results. Additionally, pay attention to the evaluation metrics for clustering algorithms to compare their performance accurately.