

Lab Assignment 9

20 marks

CO 3	Develop various classification and clustering techniques.	L4
------	---	----

Roll numbers 1-22

SET 1 Hierarchical Clustering

Use glass.csv

I. Do the Exploratory data analysis on the data (5 marks)

- Reading and Understanding the Data
- Data Cleaning
- Data Preparation

II. Analyze the Clusters in data using Hierarchical Clustering (10 marks)

- How successful has the clustering been in this regard?
- Looking at each class individually, can you spot the particular class that is well identified by the clustering? Classes that are poorly identified?
- Which classes are mostly confused with each other?
- Find the clustering parameters
 - Estimated number of clusters
 - Estimated number of noise points
 - Homogeneity
 - Completeness
 - V-measure
 - Adjusted Rand Index
 - Adjusted Mutual Information
 - Silhouette Coefficient

III. Compare the results with K-means (5 marks)

Sample code for K-means

https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_iris.html

Roll numbers 23-46

Set 2 DBSCAN

Use pima-indians-diabetes.csv

I. Do the Exploratory data analysis on the data (5 marks)

- a. Reading and Understanding the Data
- b. Data Cleaning
- c. Data Preparation

II. Analyze the Clusters in data using DBSCAN (10 marks)

- a. How successful has the clustering been in this regard? $Eps=0.8$, $minPts=19$
- b. Looking at each class individually, can you spot the particular class that is well identified by the clustering? Classes that are poorly identified?
- c. Which classes are mostly confused with each other?
- d. Find the clustering parameters
 - ix. Estimated number of clusters
 - x. Estimated number of noise points
 - xi. Homogeneity
 - xii. Completeness
 - xiii. V-measure
 - xiv. Adjusted Rand Index
 - xv. Adjusted Mutual Information
 - xvi. Silhouette Coefficient

III. Compare the results with K-means (5 marks)

Sample code for K-means

https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_iris.html

Set 3 Hierarchical Rollnumbers 47-69

Use weather.csv

I. Do the Exploratory data analysis on the data (5 marks)

- a. Reading and Understanding the Data
- b. Data Cleaning
- c. Data Preparation

II. Analyze the Clusters in data using Hierarchical (10 marks)

- a. How successful has the clustering been in this regard?
- b. Looking at each class individually, can you spot the particular class that is well identified by the clustering? Classes that are poorly identified?
- c. Which classes are mostly confused with each other?
- d. Find the clustering parameters
 - xvii. Estimated number of clusters
 - xviii. Estimated number of noise points
 - xix. Homogeneity
 - xx. Completeness
 - xxi. V-measure
 - xxii. Adjusted Rand Index
 - xxiii. Adjusted Mutual Information
 - xxiv. Silhouette Coefficient

III. Compare the results with K-means (5 marks)

Sample code for K-means

https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_iris.html