

Assignment 4

1. Cluster Analysis:

(a) Use K-Means, DBSCAN, Infomap clustering Algorithms and one obtained in assignment 1 to cluster the datasets (consider 3 datasets) available in UCI Machine Learning Repository.

(b) Compute the following cluster validation indices:
Silhouette index, Dunne Index, Davies Bouldin Index

Note: Infomap is a graph based clustering algorithm, where each object is a node in the graph and similarity between a pair of nodes is the weight of the edge between them.

2. Classification Analysis:

(a) Use Decision Tree, Naïve Bayes, K-Nearest Neighbor, Support Vector Machine and Bagging classification Algorithms to classify the datasets (consider 3 datasets) available in UCI Machine Learning Repository.

(b) Compute the following performance measure metrics using 10-fold cross validation:
Accuracy, Specificity, Sensitivity, Precession, Recall, and F-Measure

(c) Draw the ROC curve for each of the considered classifiers for each dataset.