Machine Learning Lab, B.Tech 5th Semester

Instructions

1. You are required to submit your assignment responses by 10:30 AM today through the Google Form that has been emailed to you.
2. There will be evaluation for this assignment.
3. Plagiarism checking will be performed on all the submissions for this assignment. If plagiarism is detected, your assignment will not be evaluated.

Midsem

1. Implement the K-means clustering algorithm (without using any built-in library functions) on the IRIS dataset (https://www.kaggle.com/datasets/saurabh00007/iriscsv), excluding the label column.
   (a) Evaluate your algorithm's performance and determine the number of clusters formed.
   (b) For each cluster, select the 25 samples that are closest to the centroid. These selected samples, along with their class labels, will form the training and validation set for your logistic regression model. The remaining data will serve as the test set. Apply Logistic Regression to the selected samples and evaluate the model.

   For example, if K-means clustering results in 3 clusters, you would select 25 samples from each cluster closest to the centroid, yielding 75 samples in total. These 75 samples will be used as the training and validation set (divided in any ratio), while the remaining data will form the test set. Perform Logistic Regression on these samples and report the model's performance outcomes.

Note: The marks will be awarded based on the design of your algorithm.