NAME: ASHUTOSH UPADHYE

ID:1001581542

PROJECT III

KMEANS-CLUSTERING FOR IRIS DATASET

The clustering method generally doesn’t require any sort of training or testing since k-Means is a unsupervised clustering method.

Generally we take the features such as petal width, petal length, sepal\_width, sepal\_length as features and we calculate mean of each and every distinct feature.

With each observation being very distinct, we partition the data into distinct partition and then we evaluate the distance of a specific point from the calculated centroid. Assign each observation to the cluster whose mean has the least squared Euclidean distance, this is intuitively the "nearest" mean.

Generally we obtain the resultant observations in the form of clustered points which are nearest to the calculated centroid. For this dataset, k=3 has been taken which generates proper clusters and most of the points are clustered accurately. If k=4 or k=5 is taken the performance of the algorithm degrades rapidly. k = 4 is the elbow point for this algorithm because this is where the performance starts to degrade. A plot of cluster points is shown in the program which displays all the clustered points.

For the graph plotted, there was a specific issue with legend. So red points are iris-setosa, green points are iris-virginica, blue points are iris-versicolor