Assignment 8: K means Clustering

Steps:

(i) Load the required libraries

(ii) Import the dataset

(iii) Take any two attributes of Dataset

(iv) Store the Actual target labels in a separate list

(v) Encode the target labels in a separate list

(vi) Scikit-learn Method:

(a) Load the kMeans() Model with maximum iterations and no. of clusters as desired.

(b) Train the model over training dataset

(c) Get the centroids of each cluster.

(d) To get the optimum value of k, calculate intertia.

(vii) Manual Method:

(a) For each iteration, Initialize by taking k random centroids.

(A) Assign each datapoint to their closest centroid

(B) Calculate the variance and assign a new centroid for each cluster.

(C) Reassign each datapoint to the new closest

(D) If any reassignment occurs, repeat step(B) else final centroids are obtained.

(b) To get the optimum value of k, perform Elbow Method.

(viii) Compare results for both methods.