Project Part-2

In this part, we implement the k-means algorithm and apply the implementation on the given dataset, which contains a set of 2-D points. We implement two different strategies for choosing the initial cluster centers.

STRATEGY-1

Step-1: We randomly pick the initial centers from the given samples.

Step-2: Then we assign the cluster numbers to the given samples.

Step-3: After forming the clusters we find the mean of the clusters to find the new cluster center.
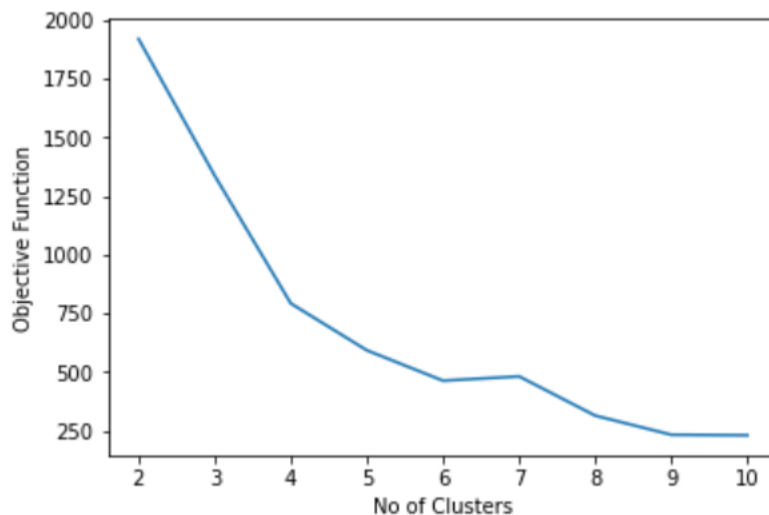
Step-4: We repeat this until the centers don't change.

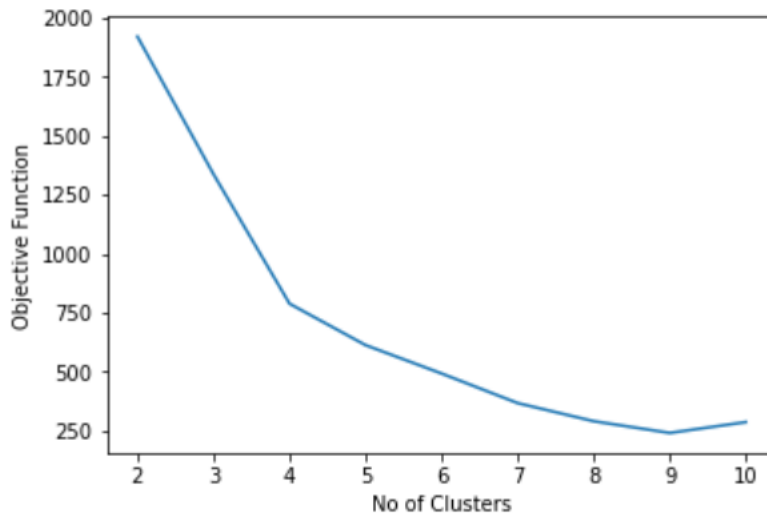Step-5: We compute the objective function as a function of k (k = 2, 3, …, 10).

Step-6: We repeat the above steps with another initialization.

After doing the above steps we plot a graph between k and objective function for two initializations.

First Initialization:

Second Initialization:

We get optimal cluster at K=4 by elbow method.

STRATEGY-2

Step-1: We pick the first center randomly.

Step-2: Then we choose the point which is farthest from the first center and make it as the second center.

Step-3: Based on k value we select the center based on the maximum distance between the average of the centers.

Step-4: After forming the clusters we find the mean of the clusters to find the new cluster center.
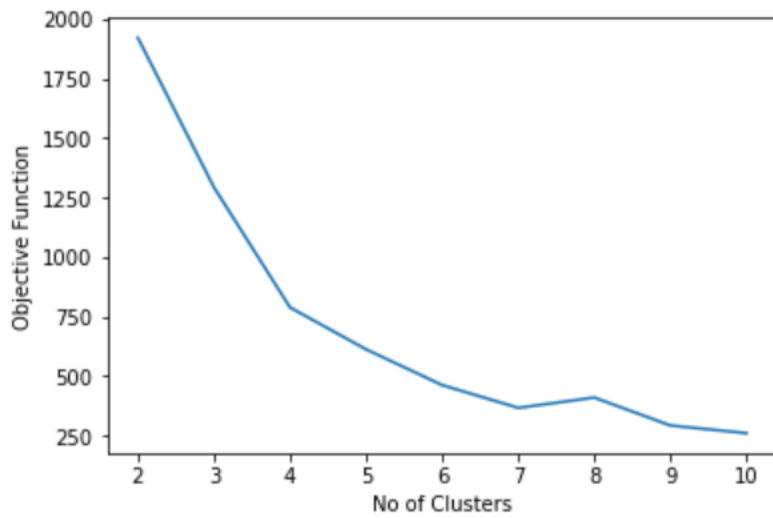
Step-5: We repeat this until the centers do not change.

Step-6: We compute the objective function as a function of k (k = 2, 3, …, 10).
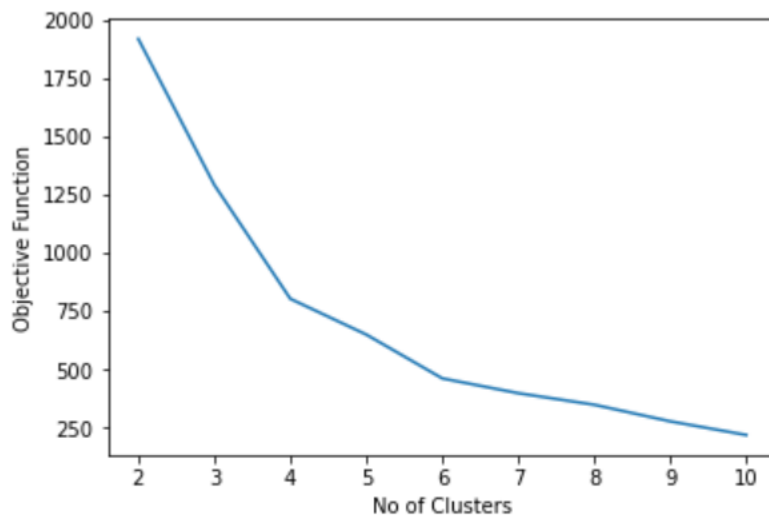
Step-7: We repeat the above steps with another initialization.

After doing the above steps we plot a graph between k and objective function for two initializations.

First Initialization:

Second Initialization:

We get the optimal cluster at k=4 using elbow method.