

CSE – 575 STATISTICAL MACHINE LEARNING

PROJECT 2 : K-MEANS STRATEGY

Analysis:

In this project, we implement K-Means algorithm with two strategies. The selection of initialization centroids is random initialization in first strategy and in second strategy, we pick first centroid randomly and then for the i -th center ($i > 1$), choose a sample (among all possible samples) such that the average distance of this chosen one to all previous ($i - 1$) centers is maximal.

After centroids calculation, in both strategies, we import KMeans() from sklearn.cluster package, and pass the centroids, number of clusters. After that we fit the data, we calculate the cluster centers and the loss. For the purpose of quiz, we calculate the loss and centroids for given values. But we have a for loop for calculating the centroids and loss for number of clusters = 2 to 10.

Results:

Strategy 1:

K1 = 3

Initialization point 1 = [[3.79752017,0.69134312],[3.81485895,6.91844078],[2.25790845,7.44778003]]

Result:

Centroids: [[5.47740039,2.25498103],[6.49724962,7.52297293],[2.56146449,6.08861338]]

Loss: 1293.7774523911348

K2 = 5

Initialization point 2 = [[9.26998864,9.62492869],[6.05509889,7.23007608],[7.74867074,1.71812324],[7.25412082,2.77862318],[7.57805025,3.82487017]]

Result:

Centroids: [[7.49365367,8.52417952],[3.22202355,7.15937996],[7.55616782,2.23516796],[2.68198633,2.09461587],[5.37514379,4.53101654]]

Loss: 592.0694342732747

Strategy 2:

K1 = 4

Initialization point 1 = [3.81135136,5.98125361]

Result:

Centroids: [array([[3.33995748, 2.59215224],[7.38076264, 2.33245532],

[6.60345839, 7.57042104],[2.85859235, 6.93136525]]])
Loss: [788.2693490065562]

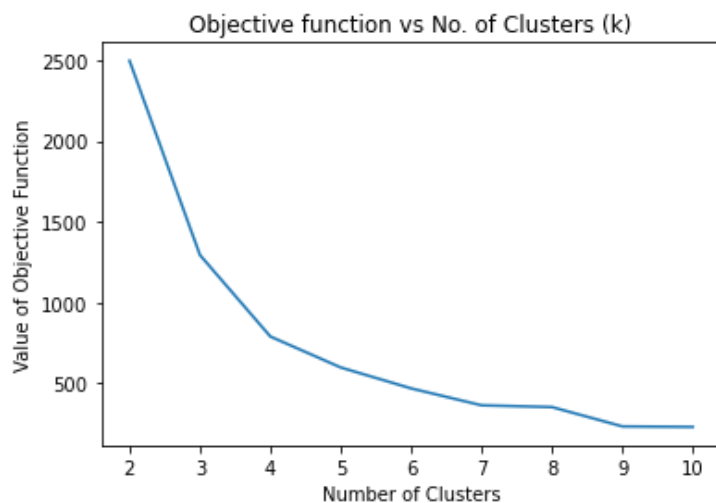
K2 = 6

Initialization point 2 = [2.16641743 2.99414637]

Result:

Centroids: [array([[3.49556658, 3.56611232],[7.75648325, 8.55668928],[3.14506148, 0.90770655],[2.56333815, 6.9782248],[7.41419243, 2.32169114],[5.46427736, 6.83771354]])]
Loss: [476.118751676353]

Results for Strategy 1 (k=2 to 10):



Results for Strategy 2 (k=2 to 10):

