**Report**

**Project Part 2: Unsupervised Learning (K-means)**

Submitted by Shashank Davalgi 1219510734 [sdavalgi@asu.edu](mailto:sdavalgi@asu.edu)

Unsupervised Learning: K-means clustering.

Strategy 1: Randomly picking the initial centers from the sample.

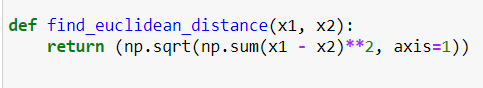
Implementation:

1. Iterating the logic 2 times to get the generalized result.
2. K values ranges from 2 to 10, where K means the number of clusters.
3. Below code is used to get the random data point from the given sample:



*randomGeneratorIndices* will pick k number of centers from the given data sample.

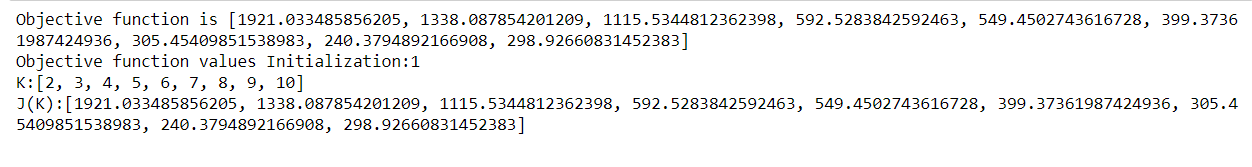
1. For each data sample, calculate the Euclidean distance with the centers generated from the previous step.
2. Euclidean distance is calculated using:



1. Classify the data points to the clusters which has the minimum Euclidean distance.
2. Calculate the average of the 2 data points to find the new centers in each cluster found from the previous step.
3. Repeat step-4 to step-7, until the new center remains the same with the previous center.

Output:

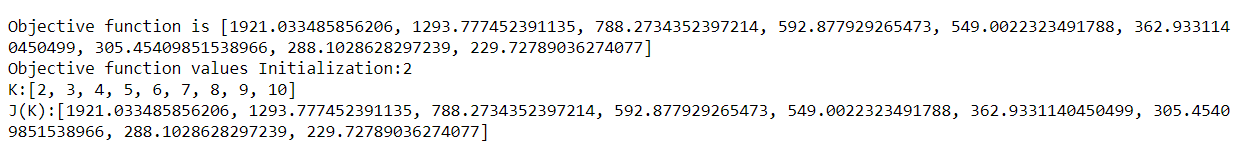
Objective function for Iteration 1:



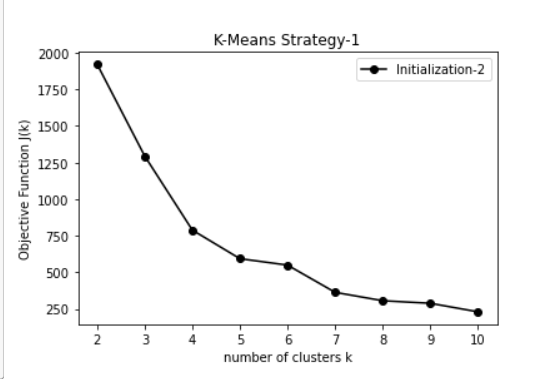
Plot for Iteration 1:



Objective function for Iteration 2:



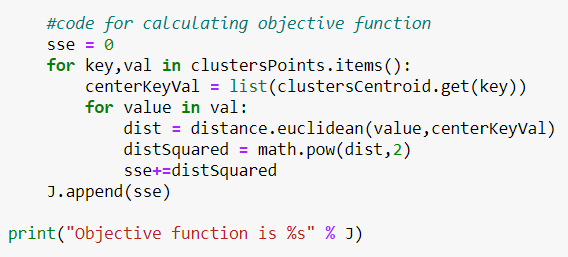
Plot for Iteration 2:



Strategy 2:

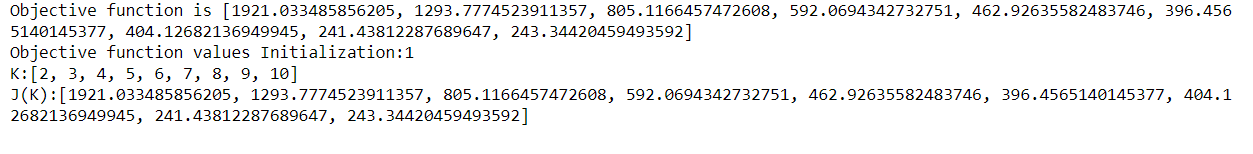
Implementation:

1. Randomly picking the first center.
2. For the i-th center, choose a sample (among all possible samples) such that the average distance of this chosen one to all previous (i-1) centers is maximal.
3. Steps 3 to 8 are similar to strategy 1.
4. Code for calculating objective function:

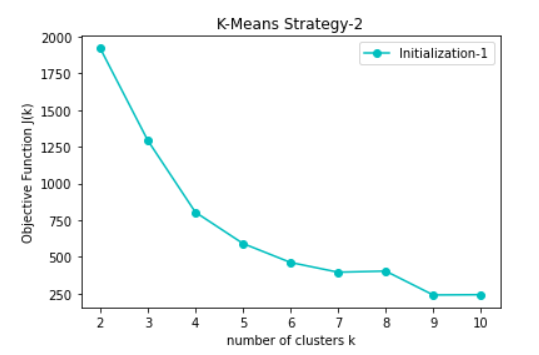


Output:

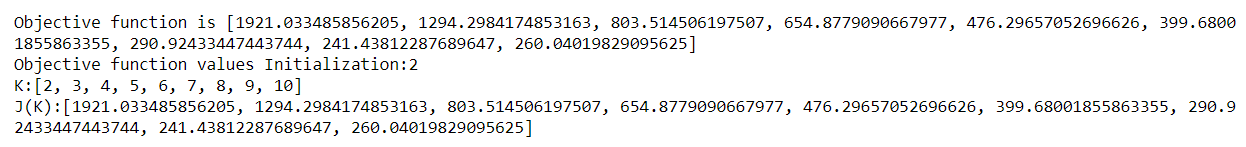
Objective function for Iteration 1:



Plot for Iteration 1:



Objective function for Iteration 2:



Plot for Iteration 2:

