CSE 575: Statistical Machine Learning
Project Part 2 – Unsupervised Learning (K-means)

Submitted by:
Abhishek Mohabe
1219468432

Strategy 1

The strategy 1 involved randomly picking the initial centers from the sample.

Implementation:
1. Iterating the logic twice to evaluate the result.
2. K values ranges from 2 to 10 where the K means the number of clusters.
3. Generate the random indices with k number of centers from the given data samples.
4. Calculating Euclidean distance between each data sample and the randomly generated center.
5. Classification the data points to the clusters which has the minimum Euclidean distance.
6. Calculating the average of the 2 data points to evaluate the new center points.
7. We repeat the previous steps until new center arrives as the previous result.

Results:

```
Abhisheks-MacBook-Pro:Project 2 abhishek$ python3 strategy1.py
K-Means Strategy1

Object Function = [1921.033485856206, 1293.7774523911348, 792.7234630260808, 613.2824392056043, 498.49437735591755, 419.57835150204346, 386.2034470350115, 232.34550801111072, 182.7922711829635]
Initial Objective Function value = 1
K: [2, 3, 4, 5, 6, 7, 8, 9, 10]

J(K): [1921.033485856206, 1293.7774523911348, 792.7234630260808, 613.2824392056043, 498.49437735591755, 419.57835150204346, 386.2034470350115, 232.34550801111072, 182.7922711829635]

Object Function = [1921.033485856205, 1293.7774523911346, 1286.9350250974871, 653.9567908403943, 504.88277080229534, 401.1811397938829, 290.8573092438312, 301.23378534950126, 245.59635641345253]
Initial Objective Function value = 2
K: [2, 3, 4, 5, 6, 7, 8, 9, 10]

J(K): [1921.033485856205, 1293.7774523911346, 1286.9350250974871, 653.9567908403943, 504.88277080229534, 401.1811397938829, 290.8573092438312, 301.23378534950126, 245.59635641345253]
```
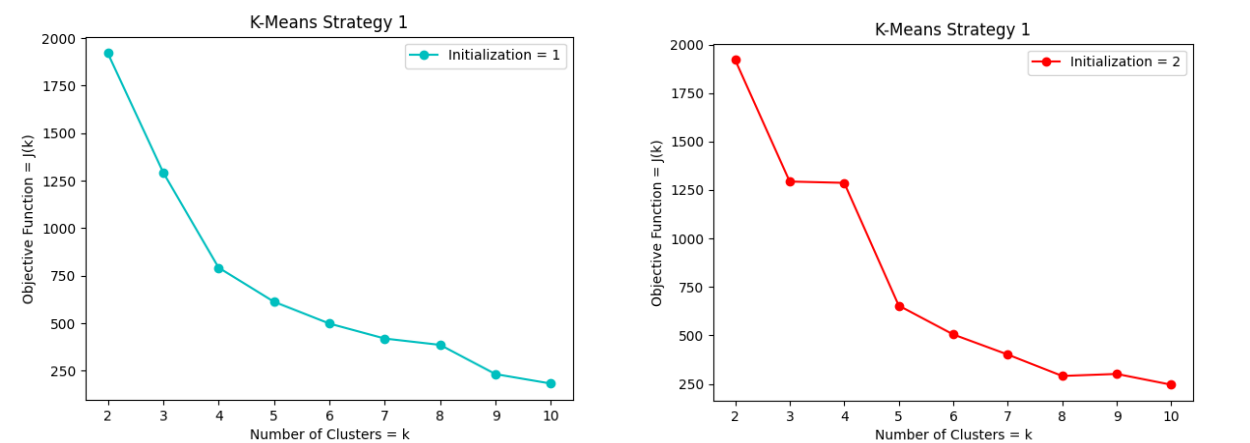
Strategy 2

Implementation:
1. Randomly choose the first center.
2. For the ith center, choose a sample such that the average distance of the chosen one to all the previous (i-1) centers is maximal.
3. Generate the random indices with k number of centers from the given data samples.
4. Calculating Euclidean distance between each data sample and the randomly generated center.
5. Classification the data points to the clusters which has the minimum Euclidean distance.
6. Calculating the average of the 2 data points to evaluate the new center points.
7. We repeat the previous steps until new center arrives as the previous result.

Results:

```
Abhisheks-MacBook-Pro:Project 2 abhishek$ python3 strategy2.py
Objective function = [1921.033485856205, 1293.7774523911346, 805.1166457472607, 592.5283842592472, 476.1187516763529, 400.725799658413, 290.9243344744375, 277.39143397662195, 220.1119983514042]
Initial Objective Function value = 1
K:[2, 3, 4, 5, 6, 7, 8, 9, 10]

J(K):[1921.033485856205, 1293.7774523911346, 805.1166457472607, 592.5283842592472, 476.1187516763529, 400.725799658413, 290.9243344744375, 277.39143397662195, 220.1119983514042]

Objective function = [1921.033485856206, 1293.7774523911346, 805.116645747261, 592.5283842592472, 462.9263558248376, 396.4565140145374, 289.9327260448387, 241.37192245730452, 182.78850958040246]
Initial Objective Function value = 2
K:[2, 3, 4, 5, 6, 7, 8, 9, 10]

J(K):[1921.033485856206, 1293.7774523911346, 805.116645747261, 592.5283842592472, 462.9263558248376, 396.4565140145374, 289.9327260448387, 241.37192245730452, 182.78850958040246]
```


K-Means Strategy 2 — Initialization = 1

K-Means Strategy 2 — Initialization = 2