

Project 2 – Unsupervised Learning (K-Means)

Prem Patel – 1217192561

Introduction:

In this project we are implementing Unsupervised Learning algorithm K-means. K-means is an iterative **algorithm** that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (**clusters**) where each data point belongs to only one group.

Implementation:

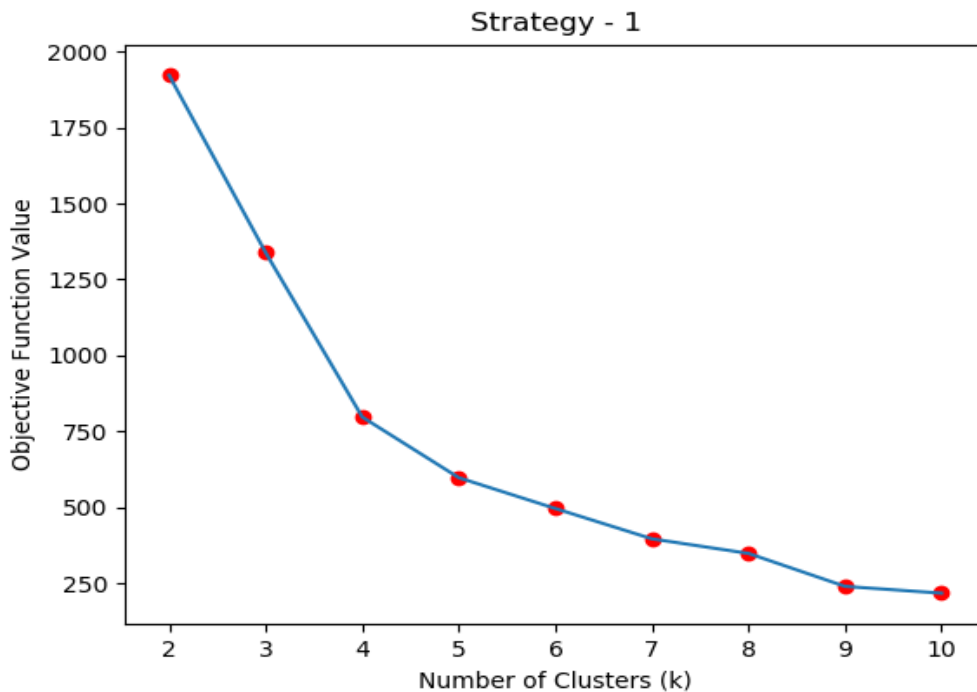
Here we use two strategies to implement the algorithm. These strategies are as follows:

- 1) Strategy_1: In this we randomly pick the initial centroids from the given data and then performs iterative calculations to further optimize the positions of the centroids. This process stops when the centroids have stabilized or the defined number of iterations is achieved.
- 2) Strategy_2: In this we pick the first centre randomly; for the i-th centre ($i > 1$), choose a sample (among all possible samples) such that the average distance of this chosen one to all previous ($i-1$) centres is maximal.

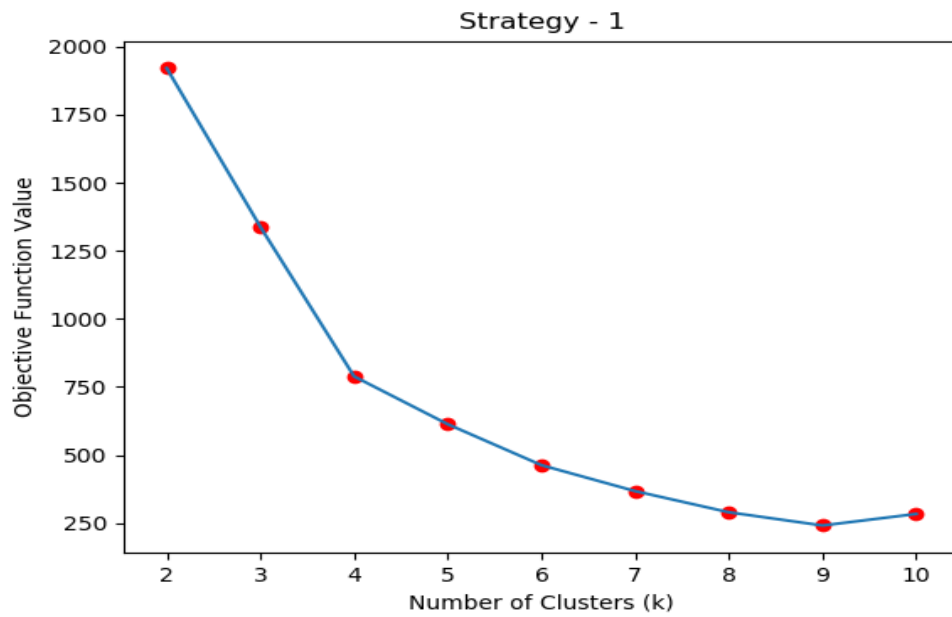
Results:

Strategy_1 Results:

Instance_1:- Cluster List $k = [2, 3, 4, 5, 6, 7, 8, 9, 10]$ Objective Function List = [1921.03, 1338.17, 797.96, 598.55, 497.26, 397.62, 349.75, 241.43, 219.01]

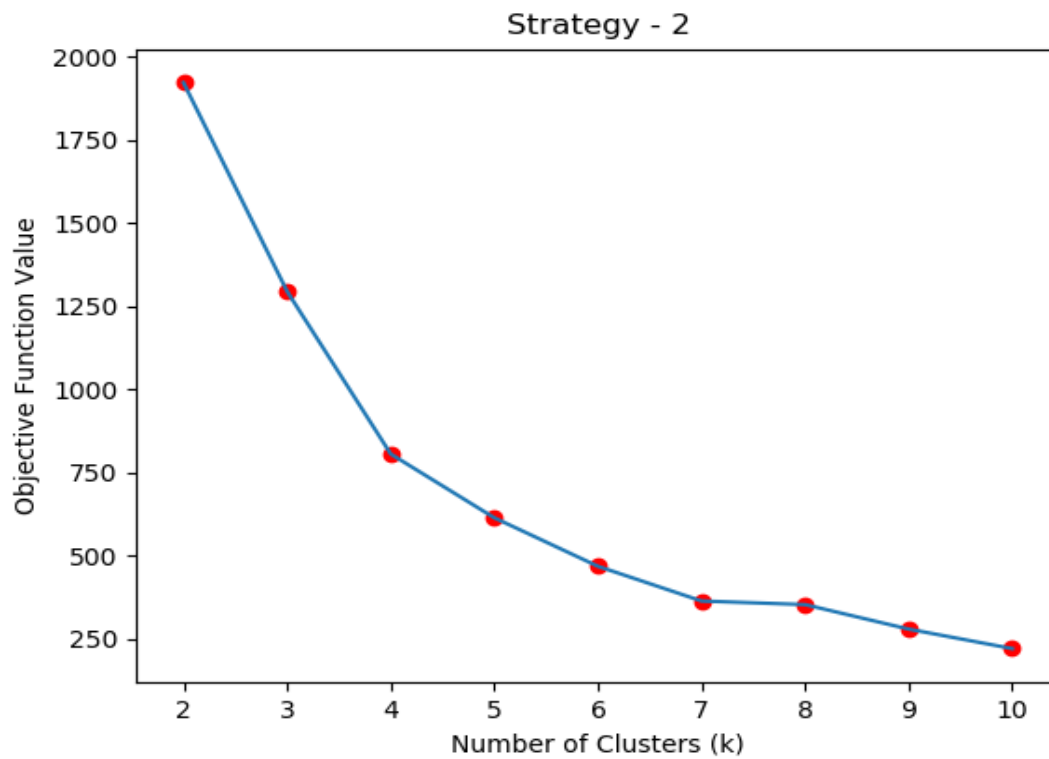


Instance_2:- Cluster List k = [2, 3, 4, 5, 6, 7, 8, 9, 10] Objective Function List = [1921.03, 1338.10, 788.96, 613.28, 462.92, 367.59, 289.93, 241.37, 283.76]

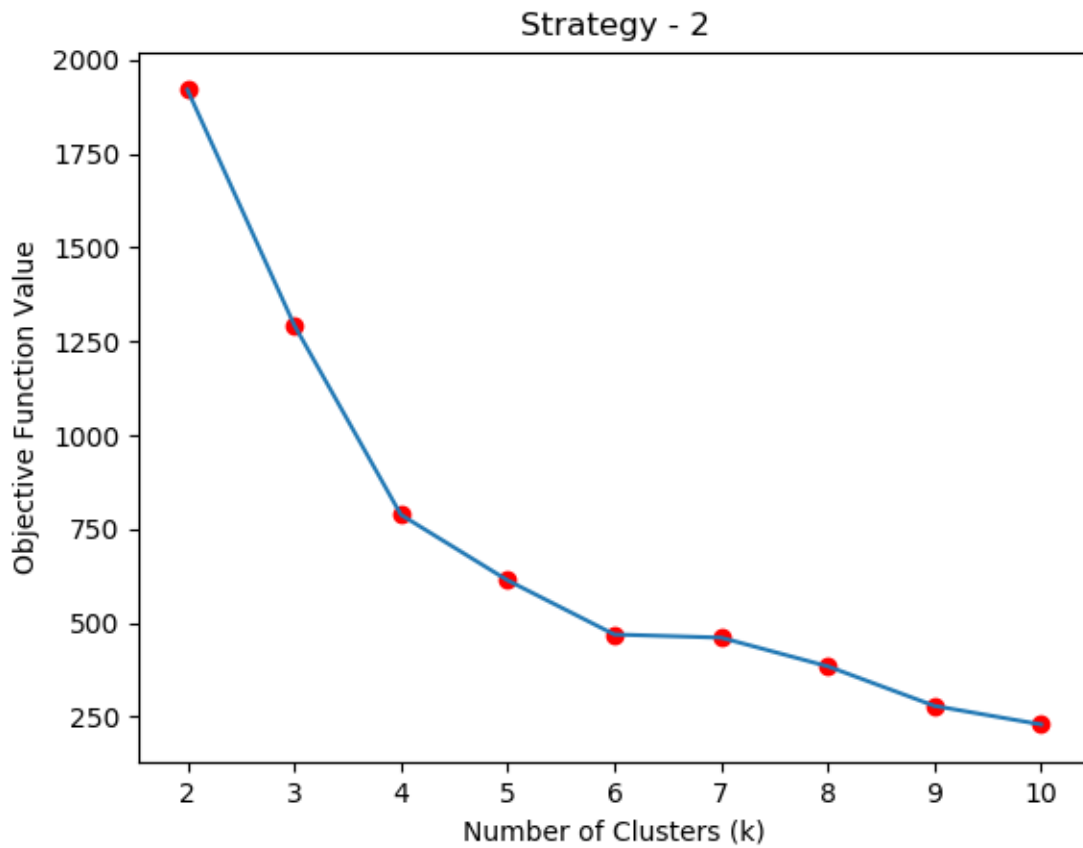


Strategy_2 Results:

Instance_1:- Cluster List k = [2, 3, 4, 5, 6, 7, 8, 9, 10] Objective Function List = [1921.03, 1294.29, 792.58, 653.95, 469.13, 397.44, 290.92, 277.74, 261.38]



Instance_2:- Cluster List k = [2, 3, 4, 5, 6, 7, 8, 9, 10] Objective Function List = [1921.03, 1293.77, 789.23, 613.42, 469.10, 460.67, 290.92, 277.74, 220.15]



Conclusion:

By comparing two strategies, there is not much difference achieved in the result as shown in the graph there is some difference if we go further than two decimal points. So whichever strategy we use the output achieved will be (approximately) similar.