Assignment 6: K-Means Algorithm Implementation

Prepared by: Malini K Bhaskaran

Results of Executions

Num of Clusters	Value for SSE	Iterations for Convergence
3	1.9106917060670063	4
5	1.4276654831723032	6
8	1.015360103623189	8
10	0.9282167771160637	9
15	0.7498426920581872	5
20	0.537269604040404	7

Analysis

From the observations above we could see that as we increase the number of clusters we get less squared sum error. Also determining the random data points also have an impact in the accuracy. As per this implementation, I have chosen the initial random numbers to be in the range of 0-1 as the data set have x and y values in between -1 to 1. If the initial random centroids are taken as the points among the data points and similarly in the next iterations then it will require lesser time for convergence and that approach is as part of the K-Medoids algorithm.