Assignment 6 – Report Part 1

The kmeans clustering algorithm has been implemented in java.

There are 2 java files namely

1. KMeans_Algorithm.java
2. Point.java

To compile the program:

javac KMeans_Algorithm.java

To run the program:

java KMeans_Algorithm <input_file_name> <output_file_name>

We need to have the input file in the same folder as in KMeans_Algorithm.java

Validating the goodness (SSE) of the clusters:

K Value	SSE Value
3	12.127333240496029
5	10.417754298427358
10	7.757273195985593
15	5.915850591885448
20	5.393091781487469

We observe that the value of SSE is inversely proportional to the number of clusters (k).

Elaborating the above statement, the distance between the points and the centroid of the cluster is more when there are fewer number of clusters.