Baranidharan Sridhar 1001113886

**K-means clustering for Seed Categorization:**

**Implemented K –means clustering to the seed dataset and used classification algorithm to classify based on the cluster id.**

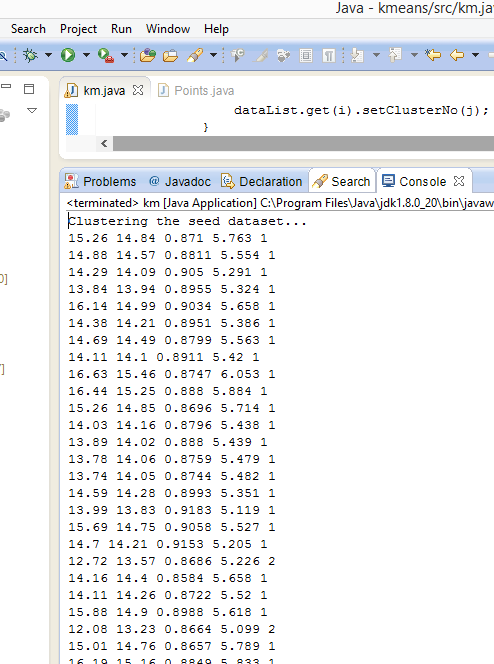
Design:

1. Implemented k-means clustering with k=3 (k-no of clusters)
2. Initialized centroids for each cluster.
3. Finding Euclidian distance between centroids and all data points and assign the data points with the cluster id that is closest.
4. Recalculate the centroid.
5. Step 3 and step 4 is repeated till the convergence.

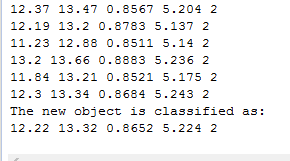
Then the cluster id is added as one of the features to the seed dataset.

1. The dataset is then split into training set and test set. In this project I am taking one random test data and the rest as training data.
2. Applying KNN classification algorithm to classify test data
3. Setting the limit as k=5 (5 nearest neighbors)
4. The distance between each data point with that of test data is calculated.
5. The 5 nearest neighbors of that test data is considered and the majority of those classes is considered as the class of the test data.
6. Each time a random test data is classified in this manner.

Clustering results as below:



The test data is classified as below using the clustered data set.



How to run the program:

Please import the zip file into eclipse and run it.