**Name: Deepthi Thalagundamatada**

**Student email ID: dt7637@uncw.edu**

**CSC515: Artificial Intelligence**

**Project Report: K means**

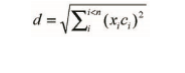
**Project 3**:

Consider the two-dimensional data given in test file [xyz.txt](http://people.uncw.edu/tagliarinig/Courses/415/Problems/xyz.txt) to partition it into three, four, and five clusters, using k-means clustering. Plot the data and the cluster centers for each partition. Repeat the process using different initial cluster centers to construct partitions into three, four, and five groups. Compare the clusters with the first partition

**K-means Algorithm:**

The featured vectors are classified based on the similarities and differences. k number of clusters must be predefined. Based on the centroid the classification of the data takes place. we must classify the new data or if the data from the cluster is deleted as centroid varies. In both condition

In K means clustering the distance between points is calculated using Euclidean distance.

**The flow of K means** **Algorithm**

Step1: In k-means Algorithm, we first assign Centroid to the vector

Step2: for each vector find the nearest centroid using Euclidean distance and assign the centroid

Step3: Recalculate the centroid

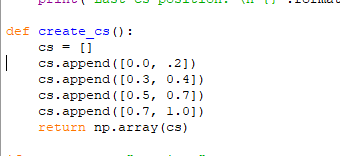
Step4: Look if there is any change in centroid

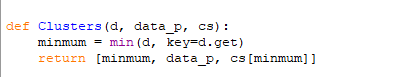
Step 5: IF change in centroid then go to step2 and repeat the process

Else complete the process

Implementation:

**Step1:** In k-means Algorithm, we first assign Centroid to the vector



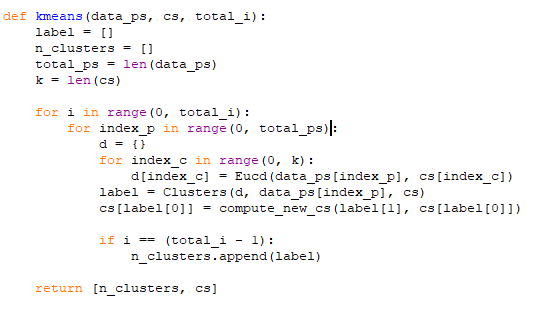


**Step2:** for each vector find the nearest centroid using Euclidean distance and assign the centroid



**Step 3 &4**: Look if there is any change in centroid

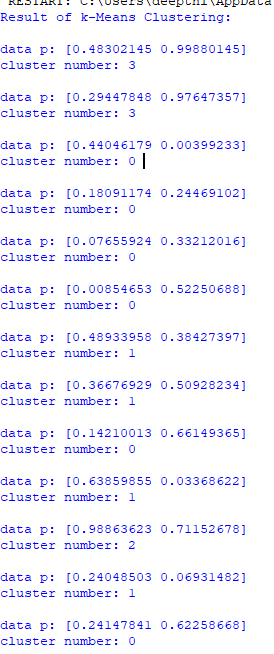
IF change in centroid then go to step2 and repeat the process Else complete the process



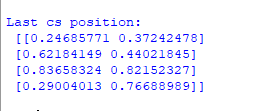
**Code for plotting graph:**

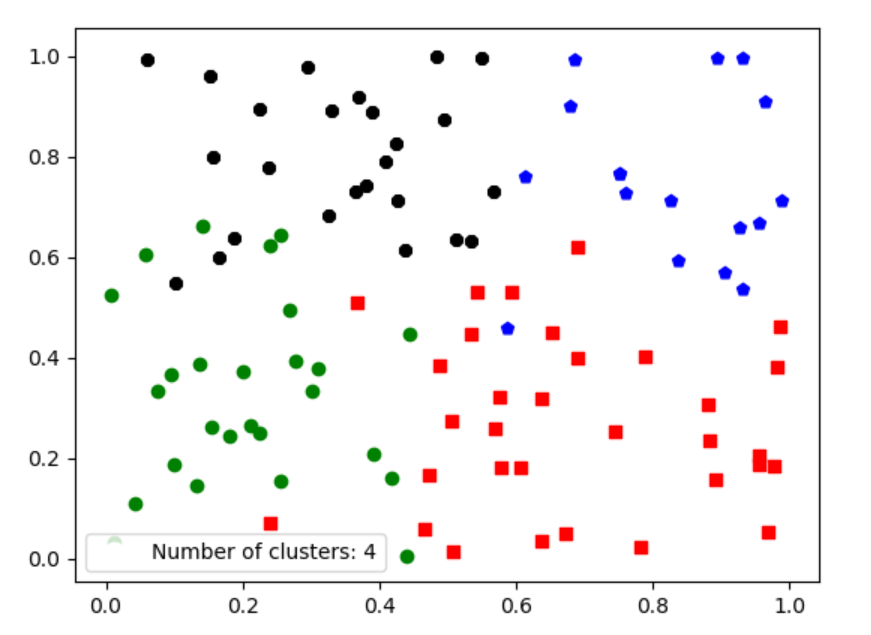
**K means implementation result for the Given XYZ file:**

**Points are classified according to the centroids: 4 centers defined**

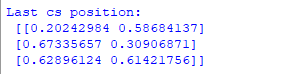


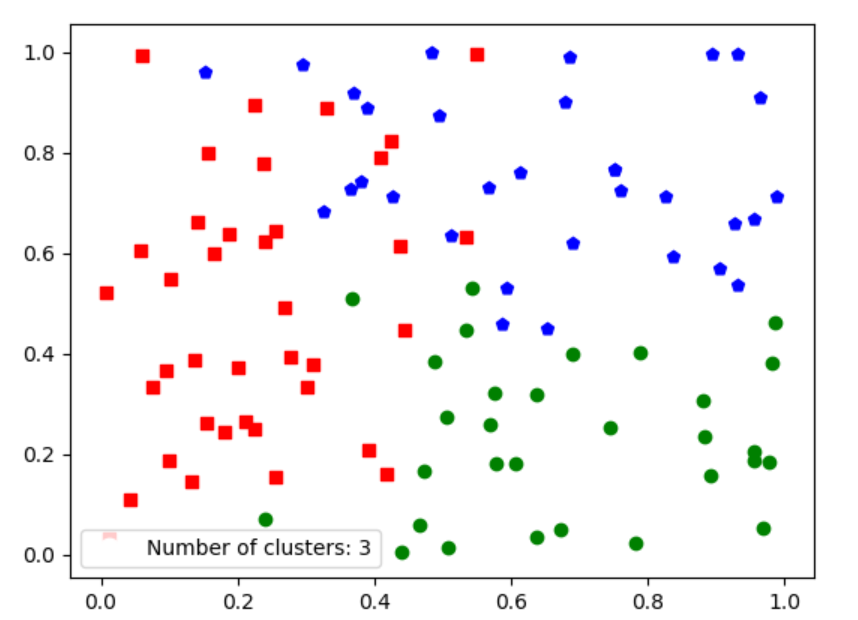
**Clustering position new (4 centers):**



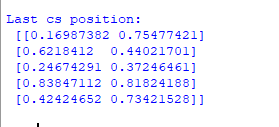


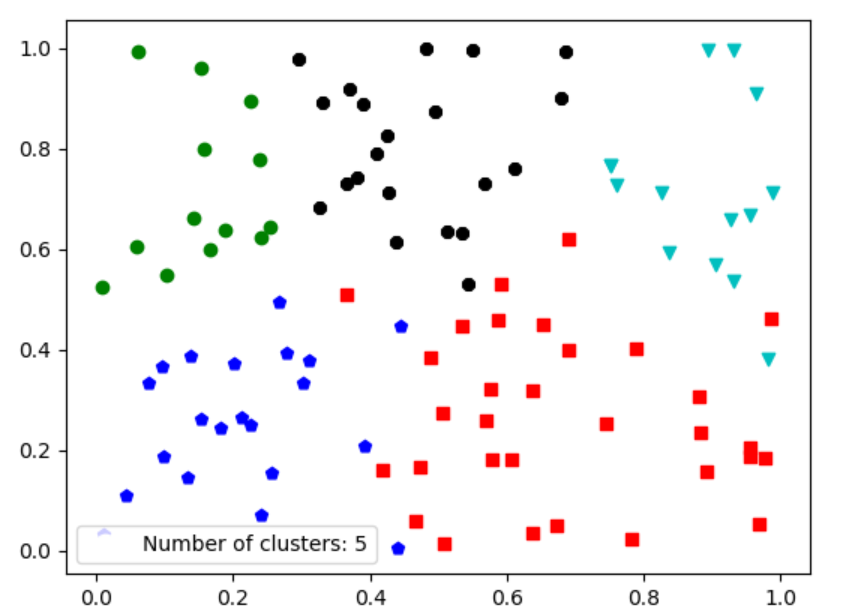
**For 3 centers:**





**For 5 centers:**





**Analyzing the centroid points and comparing the centroid change:**

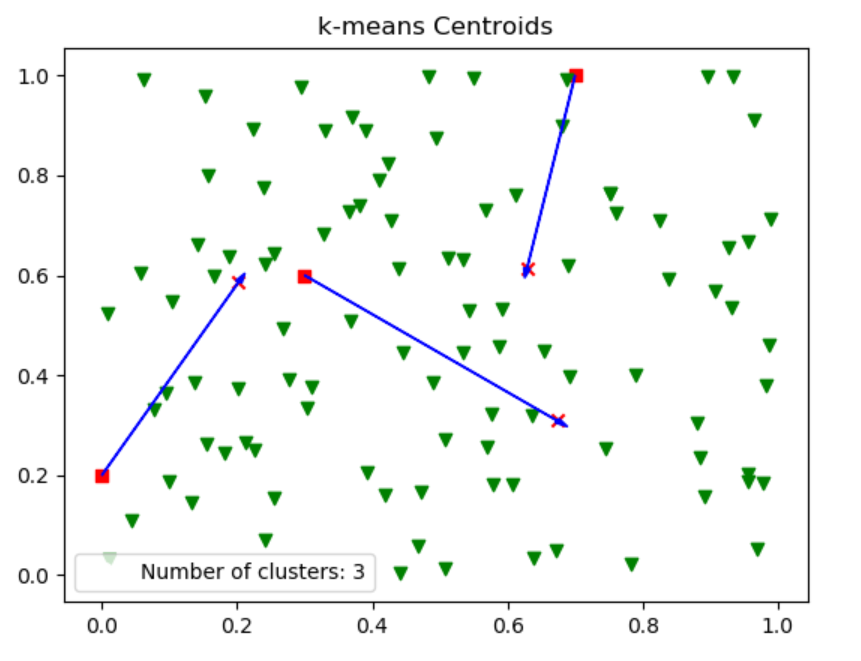
3 centroids

Last centroids position:

 [[0.20242984 0.58684137]

 [0.67335657 0.30906871]

 [0.62896124 0.61421756]]



4 centroids

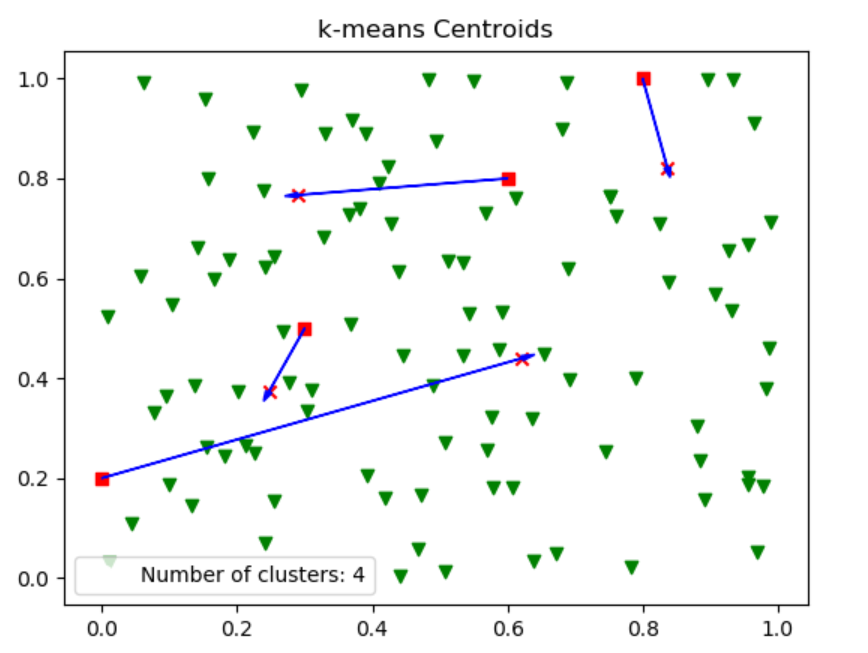
Last centroids position:

 [[0.62184149 0.44021845]

 [0.24685771 0.37242478]

 [0.29004013 0.76688989]

 [0.83658324 0.82152327]]



5 centroids

Last centroids position:

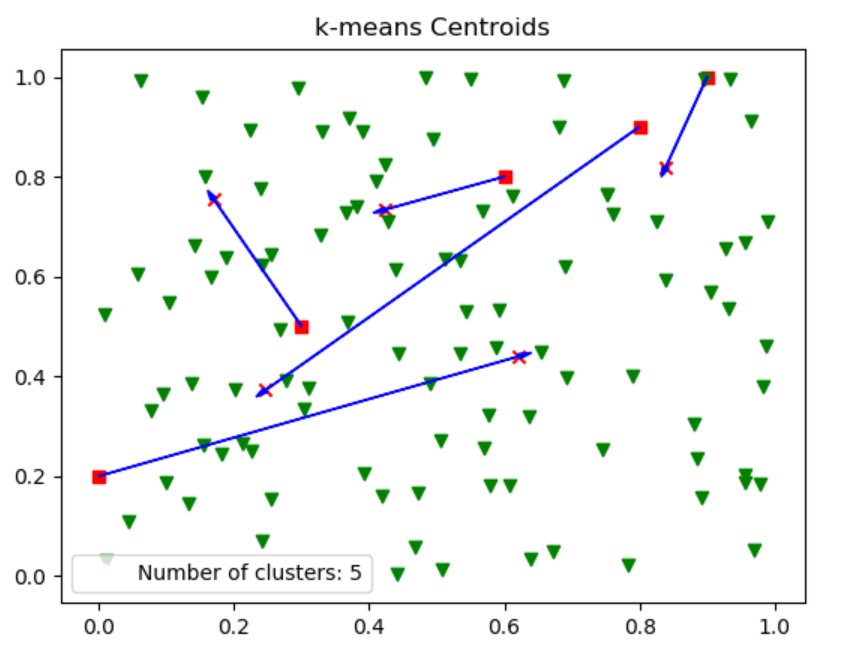
 [[0.6218412  0.44021701]

 [0.16987382 0.75477421]

 [0.42424652 0.73421528]

 [0.24674291 0.37246461]

 [0.83847112 0.81824188]]



**Graph plotting code snippet:**

