CS 435 FALL 2011 ASSIGNMENT-3 / UMUT CAN GENLIK/ 11.17.2011

For the first question of assignment i have created a modified k-means algorithm without initial k value, program takes different values of k and takes data points from given file and puts them in an array list according to dimensions. In program firstly clusters are settled after that iteration happens and checks if centroids of data points change or not, in a scenario of centroids changes new distance of centroids is calculated by algorithm and also new clusters are settled to activate our main goal which is to calculate radius of clusters by taking the average of . In a scenario of k is in the range of 3-9 output would be like

radius of cluster 3= 61.9219862184

radius of cluster 4= 51.455735836

radius of cluster 5= 52.4314519607

radius of cluster 6= 50.1942638958

radius of cluster 7= 49.5369738436

radius of cluster 8= 47.2639293507

radius of cluster 9= 46.4907561203

Clustering of samples with k-means algorithm is crucial in the case of huge data or high dimensions. However, my algorithm to modified k-means help me to converge both local minima and global minima, my algorithm calculates clusters incrementally and executes k pieces of data set it uses k-1 cluster centers from previous iteration. As observed on above output when k=6 we need to have minimum value than previous k values, after k=6 value of k either can increase or decrease.