ASSIGNMENT-1

2 Question

The Program is executed on iris.txt dataset. The program is run for 10 Iterations with different or random initializations of centroids. Purity is calculated from the obtained cluster assignment. The obtained cluster assignment is compared with the original assignment. The maximum purity score obtained is 0.8933 with mean centroids ([5.01,3.42,1.46,0.24],[5.88,2.74,4.39,1.43],[6.85,3.08,5.72,2.05]).From the observations it is quite clear that the best purity score took more iterations than the normal one to converge and is mainly dependent on the initial centroid selected. The cluster assignments are presented in the code.

ITERATIONS:

[[5.5, 2.5, 4.0, 1.3], [5.7, 3.8, 1.7, 0.3], [6.0, 2.2, 4.0, 1.0]]	0.4853
[[6.4, 3.1, 5.5, 1.8], [7.7, 3.0, 6.1, 2.3], [5.5, 2.6, 4.4, 1.2]]	0.3233
[[4.6, 3.6, 1.0, 0.2], [5.0, 3.2, 1.2, 0.2], [5.8, 2.7, 3.9, 1.2]]	0.5333
[[5.6, 3.0, 4.1, 1.3], [6.7, 3.1, 5.6, 2.4], [7.7, 2.6, 6.9, 2.3]]	0.8744
[[7.1, 3.0, 5.9, 2.1], [5.8, 4.0, 1.2, 0.2], [5.8, 2.7, 5.1, 1.9]]	0.8833
[[5.1, 3.8, 1.5, 0.3], [5.1, 3.7, 1.5, 0.4], [6.3, 3.3, 6.0, 2.5]]	0.5733
[[4.9, 2.4, 3.3, 1.0], [6.7, 3.0, 5.2, 2.3], [6.3, 2.5, 4.9, 1.5]]	0.6588
[[6.7, 3.1, 5.6, 2.4], [6.2, 3.4, 5.4, 2.3], [6.1, 2.6, 5.6, 1.4]]	0.8933
[[5.4, 3.7, 1.5, 0.2], [4.6, 3.6, 1.0, 0.2], [6.2, 2.9, 4.3, 1.3]]	0.8771
[[6.9, 3.1, 5.1, 2.3], [6.6, 3.0, 4.4, 1.4], [5.8, 2.6, 4.0, 1.2]]	0.8933

Thus from the iterations we see the purity scores obtained for each random centroid taken and the clustering is evaluated by means of PURITY.