CST8390 - Lab 6 – Answer Document
Clustering by k-Means

Onur Onel 041074824

a. How many iterations were needed for the centroid convergence?

Number of iterations: 8 **After 8 iterations algorithm has been stopped**

b. What method was used to replace missing values globally?

Missing values globally replaced with mean/mode

c. How many instances are there in clusters 0, 1, and 2?

Cluster 0 has 60 instances, which makes up 34% of the total instances.

Cluster 1 has 55 instances, which makes up 31% of the total instances.

Cluster 2 has 63 instances, which makes up 35% of the total instances.

d. Average Magnesium levels and the corresponding standard deviations for all the clusters?

Cluster 0 - *107.8667 +/- 13.2811*

Cluster 1 - *98.6545 +/- 11.0825*

Cluster 2 - *92.9524 +/- 13.9913*

e. Which cluster has below average *(13.0006)* Alcohol level?

Cluster 2 with value of *12.2295* **has a below-average Alcohol level.**

f. Find the number of incorrectly classified instances.

Incorrectly clustered instances: *10.0 5.618 %*

g. Which classes of wine were misclassified?

Classes to Clusters:

0 1 2 ← assigned to cluster

58 0 1 | 1 **Only *1* misclassified**

2 7 62 | 2 ***9*** **misclassified**

0 48 0 | 3 ***0*** **misclassified**

h. Which classes represented by clusters 0, 1 and 2?

Class 1 is represented by cluster 0.

Class 2 is represented by cluster 2.

Class 3 is represented by cluster 1.

6. Record the **initial centroids** of all clusters for attributes Alcohol and Color Intensity in the following table. Repeat clustering for seeds 5, 10, 15, 20, and 25.

Attribute		**Seed = 5**	**Seed = 10**	**Seed = 15**	**Seed = 20**	**Seed = 25**
Alcohol	Cluster 0	13.7115	13.7193	13.0998	13.1074	13.1341
	Cluster 1	13.0998	13.0998	13.7115	13.6569	13.6569
	Cluster 2	12.2132	12.2295	12.2132	12.1798	12.2027
Color Intensity	Cluster 0	5.4446	5.4627	6.9365	7.0085	7.2347
	Cluster 1	6.9365	6.9365	5.4446	5.3617	5.3617
	Cluster 2	3.0115	3.0329	3.0115	2.9385	2.9494

7. Record the **initial and final centroids** for Proline in the following table. Repeat clustering for seeds 5, 10, 15, 20, and 25.

Proline *Cluster 0*

Seed	Initial	Final
5	680	1110.6393
10	1285	1117.8167
15	510	624.8545
20	600	627.2593
25	740	619.0588

Proline *Cluster 1*

Seed	Initial	Final
5	880	624.8545
10	312	624.8545
15	885	1110.6393
20	450	1080.1077
25	770	1080.1077

Proline *Cluster 2*

Seed	Initial	Final
5	630	497.2742
10	680	500.1746
15	502	497.2742
20	378	489.2881
25	515	502.7097