**CST8390 - Lab 6 – Answer Document**

**Clustering by k-Means**

5

* 1. How many iterations were needed for the centroid convergence?   
     8
  2. What method was used to replace missing values globally?   
     mean/mode
  3. How many instances are there in clusters 0, 1, and 2?  
     60, 55, and 63, respectively
  4. Average Magnesium levels and the corresponding standard deviations for all the clusters?

Magnesium:

Cluster 0 – 107.8667+/-13.2811

Cluster 1 – 98.6545+/-11.0825

Cluster 2 – 92.9524+/-13.9913

* 1. Which cluster has below average Alcohol level?   
     2
  2. Find the number of incorrectly classified instances.   
     10
  3. Which classes of wine were misclassified?   
     1 and 2
  4. Which classes represented by clusters 0, 1 and 2?  
     Class 1 = Cluster 0  
     Class 2 = Cluster 2  
     Class 3 – Cluster 1

1. 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 | 13.69 | 13.3 | 12.37 | 12.7 | 13.71 |
| Cluster 1 | 12.45 | 12.22 | 13.05 | 12.67 | 12.93 |
| Cluster 2 | 12.86 | 11.61 | 13.11 | 12.07 | 13.05 |
| Color Intensity | Cluster 0 | 5.88 | 3.95 | 4.68 | 5 | 7.7 |
| Cluster 1 | 7.5 | 2.7 | 5.04 | 2.62 | 4.5 |
| Cluster 2 | 4.1 | 2.65 | 5.3 | 2.76 | 4.8 |

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Attribute** | | **Seed = 5 Initial/Final** | **Seed = 10**  **Initial/Final** | **Seed = 15**  **Initial/Final** | **Seed = 20**  **Initial/Final** | **Seed = 25**  **Initial/Final** |
| Proline | Cluster 0 | 680 / 1110.6393 | 1285 /  1117.8167 | 510 / 624.8545 | 600 / 627.2593 | 740 / 619.0588 |
| Cluster 1 | 880 / 624.8545 | 312 / 624.8545 | 885 / 1110.6393 | 450 / 1080.1077 | 770 / 1080.1077 |
| Cluster 2 | 630 / 497.2742 | 680 / 500.1746 | 502 / 497.2742 | 378 / 489.2881 | 515 / 502.7097 |