**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? replaced with mean/mode
  3. How many instances are there in clusters 0, 1, and 2?

0 60 ( 34%)

1 55 ( 31%)

2 63 ( 35%)

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

99.7416 +/-14.2825

* 1. Which cluster has below average Alcohol level? Cluster 2
  2. Find the number of incorrectly classified instances. 10.0 5.618 %
  3. Which classes of wine were misclassified? Cluster 2 & Cluster 3
  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 | 0.88 | 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.

Proline

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