Data Mining Assignment-2

Name – Ritu Kumari
Roll No. - 2016078
Branch – CSE, IIIrd Year

4.Answer – a) Yes. The output is:

```
Cluster 0 :
1.0
2.0
3.0
4.0
5.0
Cluster 1 :
8.0
9.0
10.0
11.0
12.0
Cluster 2 :
24.0
28.0
32.0
36.0
40.0
New Reps 0 : 3.0
New Reps 1 : 10.0
New Reps 2 : 32.0
Cluster 0 :
1.0
2.0
3.0
4.0
5.0
Cluster 1 :
8.0
9.0
10.0
11.0
12.0
Cluster 2 :
24.0
28.0
32.0
36.0
40.0
New Reps 0 : 3.0
New Reps 1 : 10.0
New Reps 2 : 32.0
```

b) Yes. The output is:

```
Cluster 0 :
1.0
Cluster 1 :
```

```
2.0
Cluster 2 :
8.0
3.0
9.0
24.0
28.0
10.0
4.0
32.0
11.0
5.0
12.0
36.0
40.0
New Reps 0 : 1.0
New Reps 1 : 2.0
New Reps 2 : 17.076923
Cluster 0 :
1.0
Cluster 1 :
8.0
2.0
3.0
9.0
4.0
5.0
Cluster 2 :
24.0
28.0
10.0
32.0
11.0
12.0
36.0
40.0
New Reps 0 : 1.0
New Reps 1 : 5.1666665
New Reps 2 : 24.125
Cluster 0 :
1.0
2.0
3.0
Cluster 1 :
8.0
9.0
10.0
4.0
11.0
5.0
12.0
Cluster 2 :
24.0
28.0
32.0
36.0
40.0
New Reps 0 : 2.0
New Reps 1 : 8.428572
New Reps 2 : 32.0
Cluster 0 :
1.0
```

```
2.0
3.0
4.0
5.0
Cluster 1 :
8.0
9.0
10.0
11.0
12.0
Cluster 2 :
24.0
28.0
32.0
36.0
40.0
New Reps 0 : 3.0
New Reps 1 : 10.0
New Reps 2 : 32.0
Cluster 0 :
1.0
2.0
3.0
4.0
5.0
Cluster 1 :
8.0
9.0
10.0
11.0
12.0
Cluster 2 :
24.0
28.0
32.0
36.0
40.0
New Reps 0 : 3.0
New Reps 1 : 10.0
New Reps 2 : 32.0
```

c) Observation for initial seed cluster {1, 11, 28}:

1. The K-Means algorithms converges in just 2 iterations.
2. The actual clusters and representatives are found in the 1st iteration itself.
3. The difference between actual representatives and the initial seed representatives is close.

Observation for initial seed cluster {1, 2, 3}:

1. The K-Means algorithms converges in 5 iterations.
2. The actual clusters and representatives are found in the 4th iteration.
3. The difference between actual representatives and the initial seed representatives is large.
4. Most of the datapoints get clustered in the Cluster 2 (initial seed representative = 3) which then gradually move to other clusters as new

representatives are calculated.

Based on the observations above, we can infer that the choice of initial seed clusters determines the no. of iterations it takes for the K-Means algorithm to converge and produce optimal clusters.