MACHINE LEARNING ASSIGNMENT – 2

1. b

2. d

3. a

4. a

5. b

6. b

7. a

8. d

9. a

10. d

11. d

12. Is K means sensitive to outliers?

Despite being used widely, the k-means algorithm has several drawbacks. One drawback is that

it is sensitive to noisy data and outliers because the mean, as a statistic, is generally sensitive to

outliers.

e.g. The mean of 2,2,2,3,3,3,4,4,4 is 3 and If we add a single 23 to that, the mean becomes 5, which is larger than any of the other values.

Since in k-means, we'll be taking the mean a lot, we wind up with a lot of outlier-sensitive calculations. That's why we have the k-medians algorithm. It just uses the median rather than the mean and is less sensitive to outliers.

13. Why is K means better?

- It is relatively simple to implement.

- Scales to large data sets.

- Guarantees convergence.

- Can warm-start the positions of centroids.

- Easily adapts to new examples.

- Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

14. Is K means a deterministic algorithm?

The k-means clustering is based on a non-deterministic algorithm. This means that running the algorithm several times on the same data, could give different results.