

MACHINE LEARNING ASSIGNMENT-2 ANSWERS

1) OPTION A

2) OPTION D

3) OPTION A

4) OPTION A

5) OPTION B

6) OPTION B

7) OPTION A

8) OPTION D

9) OPTION A

10) OPTION D

11) OPTION D

Q12 to Q14 are subjective answers type questions, Answers them in their own words briefly

12. Is K sensitive to outliers?

Yes, K-means is sensitive to outliers. K-means aims to minimize the squared Euclidean distances between the observation and the centroid of cluster to which it belongs. But sometimes K-Means algorithm does not give best results. It is sensitive to outliers. An outlier is a point which is different from the rest of data points. The k-means algorithm updates the cluster centers by taking the average of all the data points that are closer to each cluster center. When all the points are packed together, the average makes sense. However, when you have outliers, this can affect the average calculation of the whole cluster. As a result, this will push your cluster center closer to the outlier.

13. Why is K means better?

K-means is the simplest. To implement and to run. All you need to do is choose "k" and run it a number of times. Most cleverer algorithms (in particular the good ones) are much harder to implement efficiently (you'll see factors of 100x in runtime differences) and have much more parameters to set.

Moreover, most people don't need quality clusters. They actually are happy with anything remotely working for them. Plus, they don't really know what to do when they had more complex clusters. K-means, which models clusters using the simplest model ever - a centroid - is exactly what they need: massive data reduction *to centroids*.

14. Is K means a deterministic algorithm?

The basic 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. This nature of K-Means is due to its random selection of data points as initial centroids. The key idea of the algorithm is to select data points which belong to dense regions and which are adequately separated in feature space as the initial centroids.