

## Worksheet 2- Machine Learning

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 sensitive to outliers?

Answer- K-means is an unsupervised algorithm which aims to divide  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest centroid. The purpose of the algorithm is to minimize the Euclidean distances between the observation and the centroid of the cluster to which it belongs.

An outlier is a point which is different from the rest of the data points. Presence of outliers in the data causes the mean to increase and the outliers show up as separate clusters and other clusters tend to merge. Hence, it is evident that the process of clustering is not efficient in the presence of outliers.

13. Why is K-means better?

- K-means algorithm can be used with large datasets conveniently because it has linear time complexity. When the data is unlabelled and big, K-means offers many benefits as an unsupervised clustering algorithm.
- K-means is easy to use. By using the default parameters, it can be initialised. Parameters like number of clusters, maximum iterations, initial centroid initialisation can be adjusted to suit the task goals.
- Even when the data has no labels, column headers K-means will still cluster the data efficiently.

- There is an ease of interpreting the clusters by the K-means. This simplicity makes it useful in cases where there is a requirement of quick assessment of the data.

14. Is K-means a deterministic algorithm?

The basic K-mean is based on non-deterministic algorithm, which signifies running the algorithm several times on the same data will give different results. The non-deterministic nature of the algorithm is due to the random selection of data points as initial centroids. The main purpose of the algorithm is to select data points which belong to dense regions and which are adequately placed among other features.

## **Worksheet 2- SQL**

1. d
2. d
3. a
4. a
5. b
6. c
7. a
8. c
9. a
10. b
11. a
12. c
13. a
14. b,c
15. a,b

## Worksheet 2- Statistics

1. c
2. c
3. d
4. c
5. b
6. b
7. a
8. b
9. d
10. a
11. c
12. d
13. d
14. a
15. d