MACHINE LEARNING-2

Q1. Movie Recommendation systems are an example of:

a) 2 Only

Q2. Sentiment Analysis is an example of:

d) 1,2and4

Q3. Can decision trees be used for performing clustering?

a.) True

Q4. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:

a) 1 only

Q5. What is the minimum no. of variables/ features required to perform clustering?

b) 1

Q6. For two runs of K-Mean clustering is it expected to get same clustering results?

b) No

Q7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?

a) Yes

Q8. Which of the following can act as possible termination conditions in K-Means?

d) All of the above

Q9. Which of the following algorithm is most sensitive to outliers?

a) K-means clustering algorithm

Q10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):

d) All of the above

Q11 .What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithm for the same dataset?

d) All of the above

Q12. Is K sensitive to outliers?

K-Means clustering algorithm is most sensitive to outliers as it uses the mean of cluster data points to find the cluster center. As a statistic, is generally sensitive to outliers. The mean of 2,2,2,3,3,3,4,4,4 is 3 . 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, you'll be taking the mean a lot, you wind up with a lot of outlier-sensitive calculations.

Q13. Why is K means better?

K means algorithm is good in capturing structure of the data if clusters have a spherical-like shape. K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters.

Q14. Is K means a deterministic algorithm?

k-means clustering is based on a non-deterministic algorithm. K-means is one of the popular algorithms for gene data clustering due to its simplicity and computational efficiency. K-means algorithm is highly sensitive to the choice of initial cluster centers. The key idea of the algorithm is to select data points which belong to big regions and which are separated in feature space as the initial centroids.