MACHINE LEARNING

Q1 to Q11 have only one correct answer. Choose the correct option to answer your question.

1. Movie Recommendation systems are an example of:
   i) Classification
   ii) Clustering
   iii) Regression
   Options:
   a) 2 Only
   b) 1 and 2
   c) 1 and 3
   d) 2 and 3

Ans. A) 2 Only

2. Sentiment Analysis is an example of:
   i) Regression
   ii) Classification
   iii) Clustering
   iv) Reinforcement
   Options:
   a) 1 Only
   b) 1 and 2
   c) 1 and 3
   d) 1, 2 and 4

Ans. D) 1, 2 and 4

3. Can decision trees be used for performing clustering?
   a) True
   b) False

Ans. A) True

4. Which of the following is the most appropriate strategy for data cleaning before performing clusteringanalysis, given less than desirable number of data points:
   i) Capping and flooring of variables
   ii) Removal of outliers
   Options:
   a) 1 only
   b) 2 only
   c) 1 and 2
   d) None of the above

Ans. A) 1 Only

5. What is the minimum no. of variables/ features required to perform clustering?
   a) 0
   b) 1
   c) 2
   d) 3

Ans. B) 1

6. For two runs of K-Mean clustering is it expected to get same clustering results?
   a) Yes
   b) No

Ans. B) No

7. Is it possible that Assignment of observations to clusters does not change between successiveiterations in K-Means?

MACHINE LEARNING

a) Yes
b) No
c) Can't say
d) None of these

Ans. A) Yes

MACHINE LEARNING

8. Which of the following can act as possible termination conditions in K-Means?
   i) For a fixed number of iterations.
   ii) Assignment of observations to clusters does not change between iterations. Except for cases witha bad local minimum.
   iii) Centroids do not change between successive iterations.
   iv) Terminate when RSS falls below a threshold.
   Options:
   a) 1, 3 and 4
   b) 1, 2 and 3
   c) 1, 2 and 4
   d) All of the above

Ans. D) All of the above

9. Which of the following algorithms is most sensitive to outliers?
   a) K-means clustering algorithm
   b) K-medians clustering algorithm
   c) K-modes clustering algorithm
   d) K-medoids clustering algorithm

Ans. A) K-means clustering algorithm

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regressionmodel (Supervised Learning):
    i) Creating different models for different cluster groups.
    ii) Creating an input feature for cluster ids as an ordinal variable.
    iii) Creating an input feature for cluster centroids as a continuous variable.
    iv) Creating an input feature for cluster size as a continuous variable.
    Options:
    a) 1 only
    b) 2 only
    c) 3 and 4
    d) All of the above

Ans. D) All of the above

11. What could be the possible reason(s) for producing two different dendrograms using agglomerativeclustering algorithms for the same dataset?
    a) Proximity function used
    b) of data points used
    c) of variables used
    d) All of the above

Ans. D) All of the above

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

12. Is K sensitive to outliers?
13. Why is K means better?
14. Is K means a deterministic algorithm?

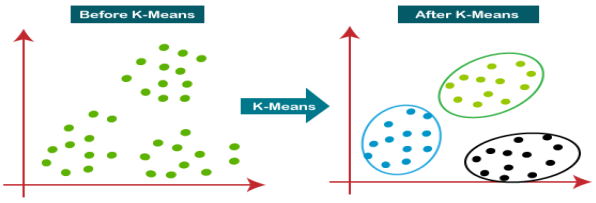
MACHINE LEARNING

12. Is K sensitive to outliers?

Ans. The K-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. K-medoids clustering is a variant of K-means that is more robust to noises and outliers., because a mean is easily influenced by extreme values. K-medoids clustering is a variant of K-means that is more robust to noises and outliers.

13. Why is K means better?

Ans. K means is better because it allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training. It is a centroid-based algorithm, where each cluster is associated with a centroid.



14. Is K means a deterministic algorithm?

Ans. 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.
k-means is a partitioning-based clustering algorithm. k-means method for clustering is an iterative process in which an initial partition of given k clusters is then improved by applying a search algorithm to the data. Simplifying, given a pre-defined number (k) of clusters, the algorithm:

-begins with an initial set of k cluster centers (i.e. the centroids)

-(re)assigns objects to the closest centroids

-recalculates centroids according to new memberships of the data points.

-repeats the last two steps until a consistent result is found or until the maximum number of iterations is reached.