

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 – (c)

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)

3. Can decision trees be used for performing clustering?

- a) True
- b) False

Ans – (a)

4. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, 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)

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

- a) 0
- b) 1
- c) 2
- d) 3

Ans- (b)

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

- a) Yes
- b) No

Ans- (b)

MACHINE LEARNING

7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
- Yes
 - No
 - Can't say
 - None of these

Ans (a)

8. Which of the following can act as possible termination conditions in K-Means?
- For a fixed number of iterations.
 - Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum.
 - Centroids do not change between successive iterations.
 - Terminate when RSS falls below a threshold.

Options:

- 1, 3 and 4
- 1, 2 and 3
- 1, 2 and 4
- All of the above

Ans (d)

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

Ans(a)

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):
- Creating different models for different cluster groups.
 - Creating an input feature for cluster ids as an ordinal variable.
 - Creating an input feature for cluster centroids as a continuous variable.
 - Creating an input feature for cluster size as a continuous variable.

Options:

- 1 only
- 2 only
- 3 and 4
- All of the above

Ans (d)

11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?
- Proximity function used
 - of data points used
 - of variables used
 - All of the above

Ans (d)

MACHINE LEARNING

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

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. Instead of using the mean point as the centre of a cluster, K-medoids uses an actual point in the cluster to represent it. Medoid is the most centrally located object of the cluster, with minimum sum of distances to other points

13. Why is K means better?

Ans - The beauty of K-Means lies in its simplicity: all it really does is **compute the distances between points and group centers, resulting in a linear complexity** . This works perfectly fine with most datasets where you aren't processing millions of data points.

K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K . The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity.

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. However, to ensure consistent results, FCS Express performs k-means clustering using a deterministic method.
