

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. D)

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. B)

3. Can decision trees be used for performing clustering?

- a) True
- b) False

Ans. B)

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. D)

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

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

Ans. B)

 FLIP ROBO

MACHINE LEARNING

6. For two runs of K-Mean clustering is it expected to get same clustering results?
- a) Yes
 - b) No
- Ans. B)
7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
- a) Yes
 - b) No
 - c) Can't say
 - d) None of these
- Ans. A)
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 with a 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. C)
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)
10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (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)

MACHINE LEARNING

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

- a) Proximity function used
- b) of data points used
- c) of variables used
- d) All of the above

Ans. A)

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

12. Is K sensitive to outliers?

Ans. Yes, K-means clustering algorithm is sensitive to outliers in the data. K-means algorithm works by iteratively assigning data points to the nearest centroid and updating the centroid based on the mean of the points in the cluster.

13. Why is K means better?

Ans. K-means clustering algorithm is a popular and widely used clustering algorithm due to advantages:

Simplicity: K-means is a simple and easy-to-understand algorithm that can be implemented quickly and efficiently.

Scalability: K-means can handle large datasets and can be easily parallelized, making it a good choice for big data applications.

Flexibility: K-means can work with various types of data and distance metrics, and can be modified to incorporate domain-specific knowledge.

Interpretability: K-means produces clusters that are easy to interpret and can be visualized in a simple scatter plot.

Performance: K-means can converge to a local optimum quickly and efficiently, making it suitable for many real-world applications.

14. Is K means a deterministic algorithm?

Ans. Yes, K-means is a deterministic algorithm. Given the same initial conditions

(i.e., same starting points for the K centroids), K-means will always converge to the same clustering solution, meaning that it will produce the same clusters for the same data set and the same value of K
