

## 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

**Answer :- 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

**Answer :- d) 1,2 and 4**

3. Can decision trees be used for performing clustering?

a) True

b) False

**Answer :- a) True**

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

**Answer :- a) 1 Only**

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

a) 0

b) 1

c) 2

d) 3

**Answer :- b) 1**

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

a) Yes

b) No

**Answer :- b) No**

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

**Answer :- a) Yes**

## **MACHINE LEARNING**

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

**Answer :- d) All of the above**

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

**Answer :- a) K-means clustering algorithm**

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

**Answer :-d) All of the above**

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

- Proximity function used
- of data points used
- of variables used
- All of the above

**Answer :- 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?

**Answer :- The k-means algorithm is sensitive to the outliers**

13. Why is K means better?

**Answer :- K-means is considered better than other clustering algorithms because it is simple, efficient, and easy to interpret the results. It is also relatively fast and easy to implement, making it a good choice for large datasets.**

14. Is K means a deterministic algorithm?

**Answer : - K-means is a deterministic algorithm, meaning that it will always produce the same results for a given set of data and initial conditions. However, the final cluster assignments may be different if the initial conditions are different. To overcome this problem, we can use the K-means++ initialization method which gives a better result in terms of both quality and time complexity.**