

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:

Answer- a) 2 Only

2. Sentiment Analysis is an example of:

Answer-d) 1, 2 and 4

3. Can decision trees be used for performing clustering?

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:

Answer- a) 1 only

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

Answer- b) 1

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

Answer-b) No

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

Answer- a) Yes

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

Answer- d) All of the above

9. Which of the following algorithms is most sensitive to outliers?

Answer- a) K-means clustering

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

Answer- d) All of the above

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

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 clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values.

Example-

The mean of 2,3,4,5 is 3.5

If we add a single 50 to that, the mean becomes 12.8, 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.

That's why we have the k-medians algorithm. It just uses the median rather than the mean and is less sensitive to outliers.

13. Why is K means better?

Answer-K means is better because

- Relatively simple to implement.
- Scales to large data sets.
- Guarantees convergence.
- Can warm-start the positions of centroids.
- Easily adapts to new examples.
- Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

14. Is K means a deterministic algorithm?

Answer- No, K-means is based on a non-deterministic algorithm which means running the algorithm several times on the same data, could give different data.