

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
 2. Sentiment Analysis is an example of:
i) Regression
ii) Classification iii) Clustering iv) Reinforcement Options:
D) 1, 2 and 4
 3. Can decision trees be used for performing clustering?
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
iii) Removal of outliers
Options:
a) 1 only
 5. What is the minimum no. of variables/ features required to perform clustering?
b) 1
 6. For two runs of K-Mean clustering is it expected to get same clustering results?
b) No
 7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
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 with a bad local minimum.
 - iii) Centroids do not change between successive iterations.
 - iv) Terminate when RSS falls below a threshold. Options:

d. All of the above

9

Which of the following algorithms is most sensitive to outliers?

a. K-means clustering algorithm

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

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?

d. All of the above

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

2 Is K sensitive to outliers?

We observe that the outlier increases the mean of data by about 10 units. This is a significant increase considering the fact that all data points range from 0 to 1. This shows that the mean is influenced by outliers. Since K-Means algorithm is about finding mean of clusters, the algorithm is influenced by outliers.

3 Why is K means better?

K-means has been around since the 1970s and fares better than other clustering algorithms like density-based, expectation-maximisation. It is one of the most robust methods, especially for image segmentation and image annotation projects. According to some users, K-means is very simple and easy to implement.

4 Is K means a Deterministic algorithm?

Hierarchical Agglomerative Clustering is deterministic except for tied distances when not using single-linkage. DBSCAN is deterministic, except for permutation of the data set in rare cases. k-means is deterministic except for initialization.

