

Worksheet set 2(Machine Learning) Name: Prashant Pathak(Internship33)

Note:- 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 : b

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

3. Can decision trees be used for performing clustering?

- a) True
- b) False

Answer : 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

Answer : a

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

a) 0

b) 1

c) 2

d) 3

Answer : b

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

a) Yes

b) No

Answer : 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

Answer : 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

Answer : d

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

Answer : 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

Answer : d

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

Answer : d

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

12. Is K sensitive to outliers?

Answer :

K-Means clustering algorithm is most sensitive to outliers as it uses the mean of cluster data points to find the cluster center.

13. Why is K means better?

Answer :

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

14. Is K means a deterministic algorithm?

Answer :

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.