

MACHINE LEARNING

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
- ii) 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

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.

d) All of the above

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

- i) K- Means clustering algorithm
- ii) Agglomerative clustering algorithm
- iii) Expectation-Maximization clustering algorithm
- iv) Diverse clustering algorithm

Options:

d) 1 and 3

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

a) K-means clustering algorithm

11. 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.

d) All of the above

12. 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

13. Is K sensitive to outliers?

Ans: The **K**-means clustering algorithm is **sensitive to outliers**, because a mean is easily influenced by **extreme values**. We can perform standardization of your data using Standard Scaler before applying clustering techniques or we can use **k-mediod** clustering algorithm, also can use z-score analysis to remove your **outliers**.

14. Why is K means better?

Ans: K-Means for Clustering is one of the popular algorithms for unsupervised learning. Where K means the number of clustering and means implies the statistics mean a problem. K means is better than other techniques:

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

15. Is K means a deterministic algorithm?

Ans: 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. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids.

WORKSHEET 2 SQL

1. Which of the following constraint requires that there should not be duplicate entries?

D) Unique

2. Which of the following constraint allows null values in a column?

C) Null

3. Which of the following statements are true regarding Primary Key?

A) Each entry in the primary key uniquely identifies each entry or row in the table

4. Which of the following statements are true regarding Unique Key?

D) All of the above

5. Which of the following is/are example of referential constraint?

D) All of them

6. How many foreign keys are there in the Supplier table?

C) 2

7. The type of relationship between Supplier table and Product table is:

B) many to one

8. The type of relationship between Order table and Headquarter table is:

D) many to many

9. Which of the following is a foreign key in Delivery table?

A) delivery id

10. The number of foreign keys in order details is:

C) 3

11. The type of relationship between Order Detail table and Product table is:

A) one to many

12. DDL statements perform operation on which of the following database objects?

C) Table

13. Which of the following statement is used to enter rows in a table?

A) Insert in to

14. Which of the following is/are entity constraints in SQL?

B) Unique

C) Primary Key

15. Which of the following statements is an example of semantic Constraint?

A) A blood group can contain one of the following values - A, B, AB and O.

STATISTICS WORKSHEET-2

1. What represent a population parameter?

C) both

2. What will be median of following set of scores (18,6,12,10,15)?

C) 12

3. What is standard deviation?

D) All of the above

4. The intervals should be _____ in a grouped frequency distribution

C) Both of these

5. What is the goal of descriptive statistics?

D) All of these

6. A set of data organized in a participant by variables format is called

B) Data set

7. In multiple regression, _____ dependent variables are used

C) 1

8. Which of the following is used when you want to visually examine the relationship between 2 quantitative variables?

B) Scatterplot

9. Two or more groups means are compared by using

D) Analysis of variance

10. _____ is a raw score which has been transformed into standard deviation units?

A) Z-score

11. _____ is the value calculated when you want the arithmetic average?

C) mean

12. Find the mean of these set of number (4,6,7,9,2000000)?

D) 400005.2

13. _____ is a measure of central tendency that takes into account the magnitude of scores?

C) Median

14. _____ focuses on describing or explaining data whereas _____ involves going beyond immediate data and making inferences

A) Descriptive and inferences

15. What is the formula for range?

D) H-L