Machine Learning Worksheet-2

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

a) 2 Only

b) 1 and 2

c) 1 and 3

d) 2 and 3

Answer : a

2. Sentiment Analysis is an example of:

i) Regression ii) Classification iii) Clustering iv) Reinforcement

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

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.

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.

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

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

12. Is K sensitive to outliers?

Ans: It is known that k-means clustering is highly sensitive to the isolated points (called outliers). Such outliers can significantly influence the final cluster configuration and should be removed to obtain quality solution.

13. Why is K means better?

Ans: k- Means is better because:

- It is relatively simple to implement.
- It can also scale the large dataset.
- It guarantees convergence.(Convergence is the movement in the price of a futures contract toward the spot or cash price of the underlying commodity over time).
- It also easily adapts to new examples.
- Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

14. Is K means a deterministic algorithm

Ans: No, k-mean is a non-deterministic algorithm. And this nature of K-Means is due to its random selection of data points as initial centroids. The key idea of the algorithm is to select data points which belong to dense regions and which are adequately separated in feature space as the initial centroids.

STATISTICS WORKSHEET-2

Q1 to Q15 have only one correct answer. Choose the correct option to answer your question.

1. What represent a population parameter?

A) SD B) mean C) both D) none

Answer :- C

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

A) 14 B) 18 C) 12 D) 10

Answer :- C

3. What is standard deviation?

A) An approximate indicator of how number vary from the mean

B) A measure of variability

C) The square root of the variance

D) All of the above

Answer :- D

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

A) Exhaustive B) Mutually exclusive

C) Both of these D) None

Answer :- C

5. What is the goal of descriptive statistics?

A) Monitoring and manipulating a specific data

B) Summarizing and explaining a specific set of data

C) Analyzing and interpreting a set of data

D) All of these

Answer :- B

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

A) Data junk B) Data set

C) Data view D) Data dodging

Answer :- B

7. In multiple regression,_____ independent variables are used

A) 2 or more B) 2

C) 1 D) 1 or more

Answer :- C

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

A) Line graph B) Scatterplot

C) Bar graph D) Pie graph

Answer :- B

9. Two or more groups means are compared by using

A) analysis B) Data analysis

C) Varied Variance analysis D) Analysis of variance

Answer :- D

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

A) Z-score B) t-score

C) e-score D) SDU score WORKSHEET

Answer :- A

11. _____is the value calculated when you want the arithmetic average?

A) Median B) mode

C) mean D) All

Answer :- C

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

A) 4 B) 7

C) 7.5 D) 400005.2

Answer :- D

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

A) Range B) Mode

C) Median D) Mean

Answer :- D

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

A) Descriptive and inferences

B) Mutually exclusive and mutually exhaustive properties

C) Positive skew and negative skew

D) Central tendency

Answer :- A

15. What is the formula for range?

A) H+L B) L-H

C) LXH D) H-L

Answer :- D

SQL WORKSHEET-2

Q1 to Q13 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following constraint requires that there should not be duplicate entries?
A) No Duplicity B) Different
C) Null D) Unique
Answer: - D

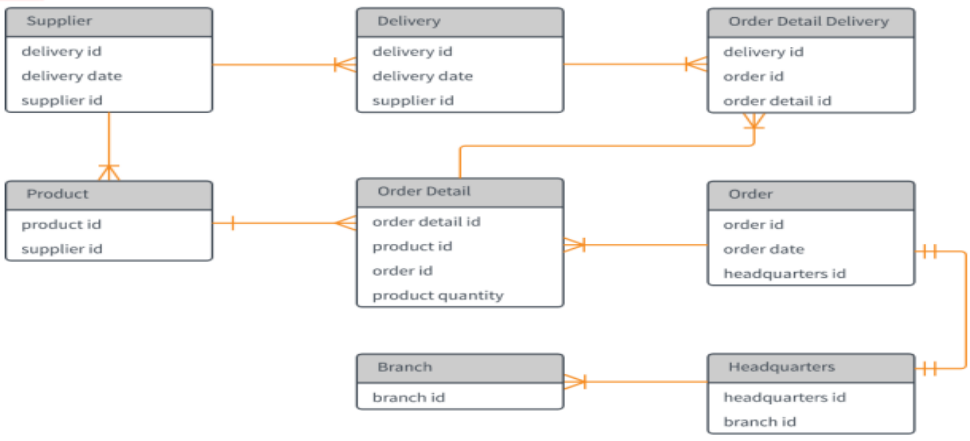
2. Which of the following constraint allows null values in a column?
A) Primary key B) Empty Value
C) Null D) None of them
Answer: - D

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
B) There can be duplicate values in a primary key column
C) There can be null values in Primary key
D) None of the above.
Answer: - A

4. Which of the following statements are true regarding Unique Key?
A) There should not be any duplicate entries
B) Null values are not allowed
C) Multiple columns can make a single unique key together
D) All of the above
Answer: - A

5. Which of the following is/are example of referential constraint?
A) Not Null B) Foreign Key
C) Referential key D) All of them
Answer: - B

For Questions 6-13 refer to the below diagram and answer the questions:



6. How many foreign keys are there in the Supplier table?
A) 0 B) 3
C) 2 D) 1

Answer: - D

7. The type of relationship between Supplier table and Product table is:
A) one to many B) many to one
C) one to one D) many to many

Answer: - D

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

A) one to many B) many to one
C) one to one D) many to many

Answer: - C

9. Which of the following is a foreign key in Delivery table?
A) delivery id B) supplier id
C) delivery date D) None of them

Answer: - A

10. The number of foreign keys in order details is:
A) 0 B) 1
C) 3 D) 2

Answer: - D

11. The type of relationship between Order Detail table and Product table is:
A) one to many B) many to one
C) one to one D) many to many

Answer: - A

12. DDL statements perform operation on which of the following database objects?
A) Rows of table B) Columns of table
C) Table D) None of them

Answer: - C

13. Which of the following statement is used to enter rows in a table?
A) Insert in to B) Update
C) Enter into D) Set Row

Answer: - A

Q14 and Q15 have one or more correct answer. Choose all the correct option to answer your question

14. Which of the following is/are entity constraints in SQL?
A) Duplicate B) Unique C) Primary Key D) Null

Answer: - B,C.

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.
B) A blood group can only contain characters
C) A blood group cannot have null values
D) Two or more donors can have same blood group

Answer: - A, B.