|  |  |
| --- | --- |
|  |  |
|  | **Machine Learning**  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 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: |
|  | a) 1 only b) 2 and 3 c) 2 and 4 d) 1 and 3 |
|  | Answer : Ignored as the Question was wrong and taken confirmation from SME |
|  |  |
|  | 10. 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 |
|  |  |
|  | 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. |
|  | Options: a) 1 only b) 2 only c) 3 and 4 d) All of the above |
|  | Answer : D |
|  |  |
|  | 12. 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 |
|  |  |
|  | Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly |
|  |  |
|  | 13. 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. |
|  |  |
|  | 14. 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 |
|  |  |
|  | 15. 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.  **SQL**   |  |  | | --- | --- | |  | 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 : C | |  |  | |  | 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 : D | |  |  | |  | 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 : D | |  |  | |  | 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 : C | |  |  | |  | 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 : A | |  |  | |  | 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 : B | |  |  | |  | 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 : B | |  |  | |  | 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 : B | |  |  | |  | 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 | |  | Answer : A | |  |  | |  | Row 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 and 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, C and D  **Statistics**   |  | | --- | |  | |  | 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 : D | |  |  | |  | 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,\_\_\_\_\_\_\_ dependent variables are used | |  | A) 2 or more B) 2 C) 1 D) 1 or more | |  | Answer : A | |  |  | |  | 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 | |  | 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 : C | |  |  | |  | 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 | | |