Assignment-2

MACHINE LEARNING

1. Movie Recommendation systems are an example of
   Answer- a) 2 only
2. Sentiment Analysis is an example of:
   Answer- d) 1,2 and 4
3. Can decision trees be used for performing clustering?
   Answer- 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:
   Answer- a) 1 only
5. What is the minimum no. of variables/ features required to perform clustering?
   Answer- b) 1
6. For two runs of K-Mean clustering is it expected to get same clustering results?
   Answer- b) No
7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
   Answer- A) yes
8. Which of the following can act as possible termination conditions in K-Means?
   Answer- d) All of the above
9. Which of the following algorithms is most sensitive to outliers?
   Answer- a) K-means clustering algorithm
10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression
    Answer- 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?
    Answer-d) All of the above
12. Is K sensitive to outliers?

Answer-K-means is one of ten popular clustering algorithms. However, k-means performs poorly due to the presence of outliers in real datasets. Besides, a different distance metric makes a variation in data clustering accuracy.

Yes. K means is sensitive to outliers.

For e.g. Data set point are 1 2 3 7 8 80

Now 80 is outlier.

13. Why is K means better?

    Answer-It's ideal to choose K-means when you have no idea on what basis you are classifying the data. Since k-means is an unsupervised learning algorithm it does't have any attribute based on which it will learn to classify, rather it all group all similar data points and form clusters.

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

STATISTICS WORKSHEET-2

1. What represent a population parameter?
   Answer- c) Both
2. What will be median of following set of scores (18,6,12,10,15)?
   Answer- c)12
3. What is standard deviation?
   Answer- d) All of the above
4. The intervals should be _____ in a grouped frequency distribution
   Answer-c) Both of these
5. What is the goal of descriptive statistics?
   Answer- B) Summarizing and explaining a specific set of data
6. A set of data organized in a participant by variables format is called
   Answer- B) Data set
7. In multiple regression,_____ independent variables are used
   Answer- A)2 or more
8. Which of the following is used when you want to visually examine the relationship between
   Answer- B) Scatterplot
9. Two or more groups means are compared by using
   Answer- D) Analysis of variance
10. _____is a raw score which has been transformed into standard deviation units?
    Answer- A) Z-score
11. . _____is the value calculated when you want the arithmetic average?
    Answer- C) Mean
12. Find the mean of these set of number (4,6,7,9,2000000)?
    Answer- D) 400005.2
13. _____ is a measure of central tendency that takes into account the magnitude of scores?
    Answer- D) Mean
14. _____ focuses on describing or explaining data whereas _____involves going beyond immediate data and making inferences
    Answer-A) Descriptive and inferences
15. What is the formula for range?
    Answer- D) H-L

WORKSHEET 2 SQL

1. Which of the following constraint requires that there should not be duplicate entries?
   Answer- D) Unique
2. Which of the following constraint allows null values in a column?
   Answer- A) Primary key
3. Which of the following statements are true regarding Primary Key?
   Answer- D) None of the above
4. Which of the following statements are true regarding Unique Key?
   Answer- A) There should not be any duplicate entries
5. Which of the following is/are example of referential constraint?
   Answer- B) Foreign Key
6. How many foreign keys are there in the Supplier table?
   Answer- D) 1
7. The type of relationship between Supplier table and Product table is:
   Answer- A) 0
8. The type of relationship between Order table and Headquarter table is:
   Answer- C) one to one
9. Which of the following is a foreign key in Delivery table?
   Answer A) delivered id
10. The number of foreign keys in order details is:
    Answer- D) 2
11. The type of relationship between Order Detail table and Product table is:
    Answer- B) many to one
12. DDL statements perform operation on which of the following database objects?
    Answer- C) Table
13. Which of the following statement is used to enter rows in a table?
    Answer- A) Insert in to
14. Which of the following is/are entity constraints in SQL?
    Answer-B,C,D
15. Which of the following statements is an example of semantic Constraint?
    Answer- A) A blood group can contain one of the following values - A, B, AB and O.
    B) A blood group can only contain characters