Worksheet 2 Machine Learning

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

Worksheet 2 Machine Learning

13. Why is K means better?

    Ans. K-means is better because when the number of variables is higher or huge it computes more faster than other. It produces tighter cluster especially when if the clusters are globular.

14. Is K means a deterministic algorithm?

    Ans. K means is a Non-deterministic algorithm it means that running the algorithm several times on the same data, could gives the different results.

Worksheet 2 SQL

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

Worksheet 2 Statistics

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