SQL WORKSHEET

Q1 and Q2 have one or more correct answer. Choose all the correct option to answer your question.

1) a, c, d

2) a, c

Q3 to Q10 have only one correct answer. Choose the correct option to answer your question.

3) b

4) c

5) b

6) b

7) c

8) c

9) d

10) a

Q11 to Q15 are subjective answer type questions, Answer them briefly.

11)ANS: Denormalization is the process of adding precomputed redundant data to an otherwise normalized relational database to improve read performance of the database.

12)ANS: Database cursor is an identifier associated with a group of rows. It is, in a sense, a pointer to the current row in a buffer.

13)ANS: Navigational search queries

      Informational search queries

      Transactional search queries.

14)ANS: Something that controls what you do by keeping you within particular limits.

15)ANS: Auto-increment allows a unique number to be generated automatically when a new record is inserted into a table.

STATISTICS WORKSHEET

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1) D
2) A
3) D
4) C
5) C
6) D

7) C

8) B

9) B

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10)ANS: Histograms are a special kind of bar graph that shows a bar for a range of data values instead of a single value. A box plot is a data display that draws a box over a number line to show the interquartile range of the data.

11)ANS: Good metrics are important to your company growth and objectives. Your key metrics should always be closely tied to your primary objective.

Good metrics can be improved. Good metrics measure progress, which means there needs to be room for improvement.

Good metrics inspire action.

12)ANS: Statistical significance is often calculated with statistical hypothesis testing, which tests the validity of a hypothesis by figuring out the probability that your results have happened by chance.

13)ANS: Distributions of income; distributions of house prices; distributions of bets placed on a sporting even.

14)ANS: Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.

15)ANS: Probability that a particular outcome is observed when the true value of the parameter is , equivalent to the probability mass on.

MACHINE LEARNING

Q1 to Q5, only one option is correct, Choose the correct option:

1) D

2) B

3) C

4) A

5) B

In Q6 to Q9, more than one options are correct, Choose all the correct options:

6) A , D

7) B , C

8) A , C

9) D

Q10 to Q15 are subjective answer type questions, Answer them briefly.

10)ANS: Compared to a model with additional input variables, a lower adjusted R-squared indicates that the additional input variables are not adding value to the model. Compared to a model with additional input variables, a higher adjusted R-squared indicates that the additional input variables are adding value to the model.

11)ANS: while lasso regression takes the magnitude of the coefficients, ridge regression takes the square.

12)ANS: Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables .VIF of three or below is not a cause for concern.

13)ANS: To ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features.

14)ANS: R-squared, the overall F-test, and the Root Mean Square Error (RMSE).

15)ANS: Accuracy = TP/TP+TN+FP+FN = 0.88

Sensitivity = TP/TP+FN = 0.45

Specificity =   TN/TN+FP = 0.96

Precision = TP/TP+FP = 0.95238