

SQL Questions

1. A,c,d
2. A,c,d
3. B
4. C
5. C
6. C
7. A
8. C
9. D
10. A

11. Denormalization is a database optimization technique in which we add redundant data to one or more tables. This can help us avoid costly joins in a relational database. Note that denormalization does not mean not doing normalization. It is an optimization technique that is applied after doing normalization.

12. Cursor is a Temporary Memory or Temporary Work Station. It is Allocated by Database Server at the Time of Performing DML operations on Table by User. Cursors are used to store Database Tables.

13. -- DDL (Data Defining Language)
-- DML (Data Manipulation Language)
-- TCL (Transition Control Language)
-- DCL (Data Control Language)
-- DQL (Data Query Language)

14. Constraints are used to limit the type of data that can go into a table. This ensures the accuracy and reliability of the data in the table. If there is any violation between the constraint and the data action, the action is aborted.

15. Auto increment is a method of increasing the count of the particular row even when no entry is provided.

Machine Learning Questions

1. B
2. B
3. C
4. C
5. B
6. A,D
7. B,C
8. A,C
9. A,B

10. **Adjusted R-Squared** tells you the percentage of variation explained by only the independent variables that actually affect the dependent variable. It does this by penalizing the presence of unnecessary predictors and thus gives only the relation percentage where the factors actually affect the outcome.

11. The difference between ridge and lasso regression is that it tends to make coefficients to absolute zero as compared to Ridge which never sets the value of coefficient to absolute zero.

12. Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables. Mathematically, the VIF for a regression model variable is equal to the ratio of the overall model variance to the variance of a model that includes only that single independent variable.

VIF(variance inflation factor) should be less than 10.

13. The machine learning algorithm works on numbers and does not know what that number represents. A weight of 10 grams and a price of 10 dollars represents completely two different things — which is a no brainer for humans, but for a model as a feature, it treats both as same.

14. - RMSE
- MSE
 - R-squared
 - Adjusted R-squared
 - Normalised RMSE
 - MAPE (mean absolute percentage error)

15. Sensitivity-- 0.45, Specificity=0.83, Precision=0.95, Recall = 0.8

Statistics Questions

1. d
2. a
3. a
4. c
5. c
6. a
7. c
8. b
9. a

11. **Boxplot** is the plotting technique that shows all the outliers outside the interquartile range.

Histogram is the plot of frequency of occurrence of an event. It tells us about the probability distribution.

12. Metrics are chosen based on the type of problem.

13. We check that by doing a hypothesis test and knowing about the relationship in between.

14. Sometimes we only get binary values in the data. In such cases, we can see that no gaussian or log-normal distribution is formed.

15. The state or fact of something being likely is called likelihood.