

SQL Questions

1. A,B
2. B,C
3. C
4. C
5. C
6. C
7. C
8. A
9. B
10. B
11. A Join clause is used to combine rows from two or more tables, based on a related column between them.
12. Left join, Right join, inner join, full join, cross join
13. SQL Server is a software needed to execute the SQL commands and queries.
14. The primary key constraint uniquely identifies each record in a table.
15. ETL is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. Full form of ETL is Extract, Transform and Load.

Statistics Question

1. B
2. D
3. C
4. B
5. C
6. A
7. C
8. B
9. A
10. A
11. C
12. A
13. A
14. C
15. B

Machine Learning Question

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

9. Gini Index = $1 - (0.6 * 0.6) - (0.4 * 0.4) = 0.48$
Entropy = $-(0.4 * \ln(0.4) + 0.6 * \ln(0.6)) = 0.65$

10. Random Forest is not susceptible to over-fitting like Decision Tree. Random Forest can generalize over data more easily than Decision Tree.

11. To normalize the distance between the data we need to scale the data as machine learning algorithms calculate the distance between the data points. Two scaling techniques are:

- Normalization
- Standardization

12. In gradient descent, feature scaling can sometimes improve the convergence speed of the algorithm.

13. In highly imbalanced dataset in classification problem, the model gets trained to predict whatever class' data is in abundance where the data class in scarce is difficult. So when we get the accuracy it will be higher only by predicting one class.

14. F-score = $2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$