MACHINE LEARNING

1. D
2. D
3. C
4. C
5. D
6. C
7. B
8. A
9. What are the advantages of Random Forests over Decision Tree?
10. **What are the advantages of Random Forests over Decision Tree?**
    Random forests consist of multiple single trees each based on a random sample of the training data. They are typically more accurate than single decision trees. The following figure shows the decision boundary becomes more accurate and stable as more trees are added.
11. **What is the need of scaling all numerical features in a dataset? Name any two techniques used for scaling.**
    The most common techniques of feature scaling are Normalization and Standardization. Normalization is used when we want to bound our values between two numbers, typically, between [0,1] or [-1,1]. While Standardization transforms the data to have zero mean and a variance of 1, they make our data unitless.
12. **Write down some advantages which scaling provides in optimization using gradient descent algorithm.**

    Faster than Batch version because it goes through a lot less examples than Batch (all examples).
    Randomly selecting examples will help avoid redundant examples or examples that are very similar that don't contribute much to the learning.
13. **In case of a highly imbalanced dataset for a classification problem, is accuracy a good metric to measure the performance of the model. If not, why?**

    This is the most common mistake made by beginners to imbalanced classification.When the class distribution is slightly skewed, accuracy can still be a useful metric. When the skew in the class distributions are severe, accuracy can become an unreliable measure of model performance.The reason for this unreliability is centered around the average machine learning practitioner and the intuitions for classification accuracy.Typically, classification predictive modeling is practiced with small datasets where the class distribution is equal or very close to equal. Therefore, most practitioners develop an

intuition that large accuracy score (or conversely small error rate scores) are good, and values above 90 percent are great.

14. What is “f-score" metric? Write its mathematical formula

The formula for the standard F1-score is the harmonic mean of the precision and recall. A perfect model has an F-score of 1.

F1 score=2*presision*recall/ presision+recall

15. What is the difference between fit(), transform() and fit_transform()?

The fit() function calculates the values of these parameters. The transform function applies the values of the parameters on the actual data and gives the normalized value. The fit_transform() function performs both in the same step. Note that the same value is got whether we perform in 2 steps or in a single step.

SQL

1. B
2. B
3. C
4. C
5. C
6. C
7. C
8. B
9. B
10. B
11. **What are joins in SQL?**
    A JOIN clause is used to combine rows from two or more tables, based on a related column between them.
12. **What are the different types of joins in SQL?**
    (INNER) JOIN: Returns records that have matching values in both tables

    LEFT (OUTER) JOIN: Returns all records from the left table, and the matched records from the right table

    RIGHT (OUTER) JOIN: Returns all records from the right table, and the matched records from the left table

    FULL (OUTER) JOIN: Returns all records when there is a match in either left or right table

13. **What is SQL Server?**

    SQL SERVER is a relational database management system (RDBMS) developed by Microsoft. It is primarily designed and developed to compete with MySQL and Oracle database.

14. **What is primary key in SQL?**

    The PRIMARY KEY constraint uniquely identifies each record in a table. Primary keys must contain UNIQUE values, and cannot contain NULL values. A table can have only ONE primary key; and in the table, this primary key can consist of single or multiple columns (fields).

15. **What is ETL in SQL?**
    ETL stands for Extract, Transform and Load, which is a process used to collect data from various sources, transform the data depending on business rules/needs and load the data into a destination database.

STATISTICS

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