

MACHINE LEARNING ANSWERS

1 - D) All of the above

2 - B) Adaboost

3 - A) The regularization will increase

4 - A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.

5 - A) It's an ensemble of weak learners.

6 - C) Both of them

7 - B) Bias will decrease, Variance increase

8 - A) model is underfitting

10 - A random forest is simply a collection of decision trees whose results are aggregated into one final result. Their ability to limit overfitting without substantially increasing error due to bias is why they are such powerful models. One way Random Forests reduce variance is by training on different samples of the data.

11 - 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 - 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 the framework of imbalanced data-sets, accuracy is no longer a proper measure, since it does not distinguish between the numbers of correctly classified examples of different classes. Hence, it may lead to erroneous conclusions

14 - 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.

15 - 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 ANSWERS

1 - B. Candidate keys

2 – B and D

3 - C. Insert

4 - C. ORDERBY

5 - C. SELECT

6 - C. 3NF

7 - C. All of the above can be done by SQ

8 - B. DML

9 - B. Table

10 - A. 1 NF

11 - A JOIN clause is used to combine rows from two or more tables, based on a related column between them

12 - (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.

13 - 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. SQL Server supports ANSI SQL, which is the standard SQL (Structured Query Language) language.

14 - 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 - 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 ANSWERS

1 - b) 0.135

2 - c) 0.45

3 - c) 0.745

4 - b) 0.577

5 - c) 0.6

6 - a) 0.33

7 - c) 0.33

8 - b) 0.22

9 - a) 0.66

10 - a) 0.33

11 - c) 0.5

12 - 0.34

13 - 0.25

14 - d) 0.06

15 - b) $\frac{2}{3}$