

WORKSHEET-7 MACHINE LEARNING

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

9. Suppose that you have a dataset with two classes, Class 1 and Class 2, and that you are trying to classify an object O using the method and you calculated the distance between O and the centroids of the two classes and obtained the following:

Squared distance to Class 1: 1.3871

Squared distance to Class 2: 3.7342

10. Random forest regression takes mean value of the results from decision trees. Random forests reduce the risk of overfitting and accuracy is much higher than a single decision tree. Furthermore, decision trees in a random forest run in parallel so that the time does not become a bottleneck.

11. In the case of neural networks, an independent variable with a spread of values may result in a large loss in training and testing and cause the learning process to be unstable. Normalization and Standardization are the two main methods for the scaling of the data. Which are widely used in the algorithms where scaling is required.

12. Advantages:

Flexibility: Gradient Descent can be used with various cost functions and can handle non-linear regression problems.

Scalability: Gradient Descent is scalable to large datasets since it updates the parameters for each training example one at a time.

Convergence: Gradient Descent can converge to the global minimum of the cost function,

provided that the learning rate is set appropriately.

13. The very simple metric to measure classification is basic accuracy i.e. ratio of correct predictions to the total number of samples in dataset. However, in the case of imbalanced classes this metric can be misleading, as high metrics doesn't show prediction capacity for the minority class

14. The F-score (also known as the F1 score or F-measure) is a metric used to evaluate the performance of a Machine Learning model. It combines precision and recall into a single score. F-measure formula:
$$F\text{-score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

15. The fit () method helps in fitting the data into a model, transform () method helps in transforming the data into a form that is more suitable for the model. Fit_transform () method, on the other hand, combines the functionalities of both fit () and transform () methods in one step

WORKSHEET- 7 SQL

1. b)

2. b)

3. a)

4. c)

5. b)

6. c)

7. c)

8. b)

9. c)

10. b)

11. SQL join statements allow us to access information from two or more tables at once. They also keep our database normalized. Normalization allows us to keep data redundancy low so that we can decrease the amount of data anomalies in our application when we delete or update a record.

12. Here are the different types of the 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. SQL Server is an application software for Relational Database Management System (RDBMS), from Microsoft, that can be used for creating, maintaining, managing, and implementing the RDBMS systems. It is an extensively used application as it enables multiple users simultaneously to work on the database systems, where users can range from minor office-based machines to huge Internet-based servers. Provisions any variety of SQL

programming extending from ANSI SQL (for traditional SQL) through SQL to T-SQL (Transact-SQL) used for advanced relational databases.

14. SQL - Primary Key. A primary key is a field in a table which uniquely identifies each row/record in a database table. Primary keys must contain unique values. A primary key column cannot have NULL values. A table can have only one primary key, which may consist of single or multiple fields.

15. Extract, Transform and Load

STATISTICS WORKSHEET-7

1. b)
2. d)
3. c)
4. b)
5. c)
6. d)
7. d)
8. b)
9. a)
10. b)
11. d)
12. c)
13. d)
14. a)
15. c)