MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

1. **Which of the following methods do we use to find the best fit line for data in Linear Regression?**

   Answer. D) Both A and B

2. **Which of the following statement is true about outliers in linear regression?**

   Answer. A) Linear regression is sensitive to outliers

3. **A line falls from left to right if a slope is _____?**

   Answer. B) Negative

4. **Which of the following will have symmetric relation between dependent variable and independent variable?**

   Answer. B) Correlation

5. **Which of the following is the reason for over fitting condition?**

   Answer. C) Low bias and high variance

6. **If output involves label then that model is called as:**

   Answer. B) Predictive modal

7. **Lasso and Ridge regression techniques belong to _____?**

   Answer. D) Regularization

8. **To overcome with imbalance dataset which technique can be used?**

Answer. D) SMOTE

9. **The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?**

Answer. A) TPR and FPR

10. **In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.**

Answer. B) False

11. **Pick the feature extraction from below:**

Answer. B) Apply PCA to project high dimensional data

In Q12, more than one options are correct, choose all the correct options:

12. **Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?**

Answers.

A) We don't have to choose the learning rate.

B) It becomes slow when number of features is very large.

C) We need to iterate.

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. **Explain the term regularization?**

Answers.

Regularization is a technique used in machine learning to prevent overfitting of models by adding a penalty term to loss function during training. Regularization help in controlling the complexity of a model and discourages it from fitting the training data too closely. It

does this by adding the penalty term based on the magnitude of the model's parameters. There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization.

14. Which particular algorithms are used for regularization?

Answers.

Some commonly used algorithms for regularization are as follows:

1) **L1 Regularization Lasso – This algorithm adds penalty term to the loss functions based on the absolute values of the model's coefficients.**
2) **L2 Regularization Ridge- This algorithm helps in adding penalty term based on the squared values of the model's coefficients.**
3) **Elastic Net- This is a combination of Lasso and Ridge regularization, providing a balance between feature selection L1 and coefficient Shrinkage L2.**
4) **Dropout- This technique primarily used in neural networks, where randomly selected neurons are ignored during training to prevent c-adaptation of neurons. It also decreases overfitting by avoiding training all the neurons on the complete training data in one go.**

15. Explain the term error present in linear regression equation?

Answers.

The term "error" refers to the difference between the predicted value (output) generated by the linear regression model and the actual observed value (target) in the dataset. This error is also known as "residual". Error is represented as "E" which shows the difference between the actual observed value and the value predicted by the model. It qualifies how well the linear regression model fits the data. Error refers to the residual or the difference between the model's predictions and the actual observed values in the dataset. The linear regression algorithm works to minimize this error to create the most accurate linear relationships between the variables.