

MACHINE LEARNING

In Question 1 to Question 11:

Q1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

Ans: - (A) Least Square Error

Q2. Which of the following statement is true about outliers in linear regression?

Ans: - (A) Linear regression is sensitive to outliers

Q3. A line falls from left to right if a slope is _____?

Ans: - (B) Negative

Q4. Which of the following will have symmetric relation between dependent variable and independent variable?

Ans: - (C) Both of them

Q5. Which of the following is the reason for over fitting condition?

Ans: - (C) Low bias and high variance

Q6. If output involves label then that model is called as:

Ans: - (B) Predictive model

Q7. Lasso and Ridge regression techniques belong to _____?

Ans: - (D) Regularization

Q8. To overcome with imbalance dataset which technique can be used?

Ans: - (D) SMOTE

Q9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

Ans: - (D) Recall and precision

Q10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

Ans: - (B) False

Q11. Pick the feature extraction from below:

Ans: - (B) Apply PCA to project high dimensional data

In Question 12:

Q12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

Ans: - (A) We don't have to choose the learning rate.
(B) It becomes slow when number of features is very large

In Question 13, 14, 15:

13. Explain the term regularization?

Ans: -

- **Regularization** is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting. It reduces the parameters and simplifies the model.
- The model we are training will have low accuracy if it is overfitting. This happens because the model is trying too hard to capture the data points (data points don't represent the true properties of the data).
- So, it will shrink/simplify the model by tuning the function by adding an additional penalty term in the error function.
- Regularization makes slight modifications to the learning algorithms such that the model generalizes better. This in turn improves the model's performance on the unseen data as well.

Q14. Which particular algorithms are used for regularization?

Ans: - The different algorithm we use in regularization are as follows-

- **Ridge Regression** – Ridge regression is also known as **L2 Regularization**. Ridge regression shrinks the coefficients as it helps to reduce the model complexity and multi-collinearity.
Expression= Loss
- **LASSO (Least Absolute Shrinkage and Selection Operator) Regression** – The LASSO regression shrinks the less important feature's coefficient to zero thus, removing some feature altogether. So, this works well for feature selection in case we have a huge number of features. LASSO regression is also known as **L1 Regularization**.
- **Elastic-Net Regression** – Elastic-Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods respectively.

Q15. Explain the term error present in linear regression equation?

Ans: -

- The vertical distance between the data point and the regression line is known as **error** or **residual**.
- Error is the difference between the Actual value and Predicted value. The objective is to reduce the difference between the Actual value and Predicted value.
- Mathematical Approach:

Residual/Error = Actual values – Predicted Values

Sum of Residuals/Errors = Sum (Actual- Predicted Values)

Square of Sum of Residuals/Errors = (Sum (Actual- Predicted Values))²

----- ❖ ----- ❖ ----- ❖ -----