

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?
A) Least Square Error B) Maximum Likelihood
C) Logarithmic Loss D) Both A and B

Ans:- A) Least square error

2. Which of the following statement is true about outliers in linear regression?
A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers
C) Can't say D) none of these

Ans:- Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?
A) Positive B) Negative C) Zero D) Undefined

Ans:- Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?
A) Regression B) Correlation C) Both of them D) None of these

Ans:- Both of them

5. Which of the following is the reason for over fitting condition?
A) High bias and high variance B) Low bias and low variance
C) Low bias and high variance D) none of these

Ans:- Low bias and high variance

6. If output involves label then that model is called as:
A) Descriptive model B) Predictive modal
C) Reinforcement learning D) All of the above

Ans:- Predictive modal

7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation B) Removing outliers
C) SMOTE D) Regularization

Ans:- Regularization

8. To overcome with imbalance dataset which technique can be used?
A) Cross validation B) Regularization
C) Kernel D) SMOTE

Ans:- SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?
- A) TPR and FPR B) Sensitivity and precision
C) Sensitivity and Specificity D) Recall and precision

Ans:- TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
- A) True B) False

Ans:- False

11. Pick the feature extraction from below: A) Construction bag of words from a email
B) Apply PCA to project high dimensional data
C) Removing stop words
D) Forward selection

Ans:- 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?
- 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.
D) It does not make use of dependent variable.

Ans:- We don't have to choose the learning rate.

It becomes slow when number of features is very large

MACHINE LEARNING

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

13. Explain the term regularization?

Ans:-

Regularization is a technique used to reduce errors by fitting the function appropriately on the given training set and avoiding overfitting.

Regularization is like a guardrail for machine learning models.

It helps prevent them from memorizing the training data too much and instead encourages them to understand the underlying patterns.

It's like finding a sweet spot between fitting the known data well and being ready for new, unfamiliar data.

Lasso and Ridge are tools within regularization that ensure this balance, making our models smarter and more reliable.

14. Which particular algorithms are used for regularization?

Ans:-

Two common types of regularization used in linear regression are Lasso (L1 regularization) and Ridge (L2 regularization). These methods add a regularization term to the cost function, which is a function of the model's coefficients:

1. Lasso (L1 regularization): • Adds the absolute values of the coefficients as a penalty term. • Encourages sparsity in the model, meaning some coefficients may become exactly zero, effectively eliminating certain features.
2. Ridge (L2 regularization): • Adds the squared values of the coefficients as a penalty term. • Discourages extreme values in the coefficients, leading to a more stable model.

The strength of regularization is controlled by a hyperparameter (often denoted as λ or alpha), and tuning this parameter allows for a balance between fitting the training data and preventing overfitting

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

Ans:-

In the context of linear regression, the term "error" refers to the difference between the predicted values and the actual values of the dependent variable. The linear regression equation models the relationship between the independent variable(s) and the dependent variable by estimating the coefficients that define the slope and intercept of the best-fit line. However, due to various factors such as random variation, unaccounted influences, or inherent noise in the data, the model may not perfectly predict the dependent variable.

The linear regression equation is typically represented as:

$$[Y = \text{beta_0} + \text{beta_1}X + \text{epsilon}]$$

Here:

- (Y) is the dependent variable (the variable we are trying to predict),
- (X) is the independent variable,
- (beta_0) is the intercept,
- (beta_1) is the slope,
- (epsilon) represents the error term.

The error term (epsilon) captures the discrepancies between the predicted ($\text{beta_0} + \text{beta_1}X$) and actual (Y) values. In a well-fitted model, the goal is to minimize the sum of squared errors, which means reducing the squared differences between the predicted and actual values for all data points.