

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
Answer : D) Both A and B
 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
Answer : A) 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
Answer : B) 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
Answer: B) Correlation
 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
Answer : C) 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
Answer : B) Predictive model
 7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation B) Removing outliers
C) SMOTE D) Regularization
Answer : D) Regularization
 8. To overcome with imbalance dataset which technique can be used?
A) Cross validation B) Regularization
C) Kernel D) SMOTE
Answer : D) 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
Answer : A) 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
Answer : B) False
 11. Pick the feature extraction from below:
A) Construction bag of words from a email
B) Apply PCA to project high dimensional data
-

MACHINE LEARNING

C) Removing stop words

D) Forward selection

Answer : A) Construction bag of word from a email

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.

Answer : B) 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?

Answer : Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

The amount of regularization applied to a model is controlled by a hyperparameter called the regularization parameter, which determines the trade-off between fitting the training data well and avoiding overfitting. A higher value of the regularization parameter will lead to more regularization and a simpler model, while a lower value will lead to less regularization and a more complex model.

In linear regression, the term "error" or "residual" refers to the difference between the actual observed values of the dependent variable and the predicted values generated by the linear regression equation.

In linear regression, regularization is achieved by adding a penalty term to the least squares cost function. Two common forms of regularization used in linear regression are Ridge regression and Lasso regression. Ridge regression adds a penalty term proportional to the squared L2 norm of the regression coefficients, while Lasso regression adds a penalty term proportional to the L1 norm of the coefficients.

14. Which particular algorithms are used for regularization?

Answer :

Ridge Regression: This is a linear regression model that uses L2 regularization to prevent overfitting.

Lasso Regression: This is a linear regression model that uses L1 regularization to prevent overfitting.

Elastic Net: This is a linear regression model that uses a combination of L1 and L2 regularization.

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

The linear regression equation attempts to model the relationship between the independent variable(s) and the dependent variable by fitting a straight line through the data points. However, this line may not perfectly capture the true relationship between the variables, and there may be some degree of random variation or noise present in the data.

The error term represents this random variation or noise, and it is often assumed to follow a normal distribution with a mean of zero. In other words, the error term is the difference between the actual values and the predicted values that cannot be explained by the linear regression equation.

Linear regression aims to minimize the sum of the squared errors, or the difference between the actual values and the predicted values squared. This is known as the "sum of squared residuals" or the "residual sum of squares" (RSS), and it is used as a measure of the overall goodness of fit of the linear regression model. The goal of linear regression is to find the coefficients of the equation that minimize this error term, resulting in the best possible fit of the line to the data.
