

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

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

A) Linear regression is sensitive to outliers

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

B) Negative

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

B) Correlation

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

C) Low bias and high variance

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

B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

D) Regularization

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

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

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

B) False

11. Pick the feature extraction from below:

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?

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

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

13. Explain the term regularization?

-- Regularization involves various techniques which are used for the calibration of the ML models, in such a way to minimize the loss of the functions and to avoid overfitting and underfitting.

- Using regularization, the ML models can be fitted appropriately on the given test sets and thus errors can be removed/reduced.

-Regularization, without a substantial increase in its bias, significantly reduces the variance of the model.

-There are 2 types of regularization techniques:

- a. Ridge or L2
- b. LASSO (Least Absolute Shrinkage and Selection Operator)/L1

14. Which particular algorithms are used for regularization?

-There are 2 types of regularization techniques:

- a. Ridge or L2
- b. LASSO (Least Absolute Shrinkage and Selection Operator)/L1

The working of all these algorithms is quite similar to that of Linear Regression, it's just the loss function that keeps on changing!

-**Ridge regularization** is also called as Ridge Regression or L2 Regularization. It modifies the under fitted and overfitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients. It means that the mathematical function representing ML model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Ridge regularization performs regularization by shrinking the coefficients present and reduce model complexity and multi-collinearity.

-**LASSO Regularization** modifies the under and over fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients. It also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients. It means that the coefficient sum can also be 0, because of the presence of negative coefficients. LASSO regression converts coefficients of less important features to zero, shrinks the coefficients of remaining features to reduce the model complexity, thereby avoiding overfitting.

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

-On the scatter plot the liner regression line does not pass through all the data points. We need to find the best fitted line where maximum number of points are passed through the line.

-Linear regression uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by measuring the distance of the observed y-values from the predicted y-values at each value of x; squaring each of these distances; calculating the mean of each of the squared distances.

-The equation for linear regression can be given as $Y = A + Bx_1 + Bx_2 + Bx_3 \dots + E$,

Where, Y= the dependent variable

X = the independent variable

E is the error of the estimate

A is the intercept.

B is the coefficient.

E is the error of the estimate or how much variation there is in our estimate of the regression coefficient.