

MACHINE LEARNING

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 B) Correlation

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 modal

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 A) TPR and FPR

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

A) True B) False

Q11. Pick the feature extraction from below:

Ans

- A) Construction bag of words from an email
- B) Apply PCA to project high dimensional data
- C) Removing stop words
- D) Forward selection

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.

Q13. Explain the term regularization?

Ans.

In general, regularization means to make things regular or acceptable.

In the context of machine learning, regularization is the process which regularizes or shrinks the coefficient towards zero.

Regularization discourages learning. This technique converts a complex model into a simpler one, so as to avoid the risk of overfitting and shrinks the coefficients, for lesser computational cost.

Using regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

➤ ***There are two main types of regularization techniques:***

Ridge regularization and lasso regularization.

Q14. Which particular algorithms are used for regularization?

Ans

Regularization algorithms are:

- Ridge regularization
- Lasso regularization

➤ Ridge Regularization:

Also known as regression. It modifies the overfitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients. This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Regression performs regularization by shrinking the coefficients present. The function depicted below shows the cost function of Ridge regression.

➤ Lasso Regularization:

It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients. Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients. This means that the coefficient sum can also be 0, because of the presence of negative coefficients. We can control the coefficient values by controlling the penalty terms, just like we did in Ridge Regression.

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

Ans

Linear regression most often uses mean-square error (MSE) to calculate the error of the model.

MSE is calculated by:

1. measuring the distance of the observed y-values from the predicted y-values at each value of x;
2. squaring each of these distances;
3. calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.