

MACHINE LEARNING

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

Ans :- A) Least Square Error

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

Ans :- A) Linear regression is sensitive to outliers

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

Ans :- B) Negative

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

Ans :- C) Both of them

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

Ans :- C) Low bias and high variance

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

Ans :- D) All of the above

7) Lasso and Ridge regression techniques belong to _____?

Ans :- D) Regularization

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

Ans :- A) Cross validation

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

Ans :- C) Sensitivity and Specificity

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

Ans :- A) True

11) Pick the feature extraction from below:

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

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

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

C) We need to iterate.

13) Explain the term regularization?

Ans :- Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it. Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique.

Regularization works by adding a penalty or complexity term to the complex model

There are mainly two types of regularization techniques

- Ridge Regression :- Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.
- Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization.
- In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called ridge regression penalty . We can calculate it by multiplying with the lambda to the squared weight of each individual feature
- Lasso Regression :- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for least absolute and selection operator.
- It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.
- Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.

- It is also called as L1 regression The equation for the cost function of Lasso regression.

14) Which particular algorithms are used for regularization?

Ans :- The most useful and popular ML regression algorithm is Linear regression algorithm which further divided into two types :

- Simple Linear Regression algorithm :- Linear regression may be defined as the statistical model that analylies the linear relationship between a dependent variable with given set of independent variables. Linear relationship between variables means that when the value of one or more independent variables will change, the value of dependent variable will also change accordingly .

Mathematically the relationship can be represented with the help of following equation

$$Y = mX + b$$

Here, Y is the dependent variable we are trying to predict X is the dependent variable we are using to make predictions. m is the slop of the regression line which represents the effect X has on Y b is a constant, known as the Y -intercept. If $X = 0$, Y would be equal to b .

- Multiple Linear Regression algorithm :- It is the extension of simple linear regression that predicts a response using two or more features. Mathematically we can explain it as follows: Consider a dataset having n observations, p features i.e. independent variables and y as one response i.e. dependent variable the regression line for p features can be calculate

$$h(x_i) = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_px_i$$

Here, $h(x_i)$ is the predicted response value and $b_0, b_1, b_2, \dots, b_p$ are the regression coefficients. Multiple Linear Regression models always includes the errors in the data known as residual error which changes the calculation as follows

$$h(x_i) = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_px_i + e$$

Hence its algorithm used to machine learning in python on regression

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

Ans :- First of all error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables.

For example given Linear regression is a form of analysis that relates to current trends experienced by a particular security or index by providing a relationship between a dependent and independent variables, such as the price of a security and the passage of time, resulting in a trend line that can be used. regression line always has an error term because, in real life, independent variables are never perfect predictors of the dependent variables. Rather the line is an estimate based on the available data.