MACHINE LEARNING

SUBJECTIVE TYPE QUESTION ANSWER

1. Explain the term regularization?

Ans. Regularization is one of the basic and most important concept in the world of machine learning. Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting.

The commonly used regularization techniques are:

- L1 regularization
- L2 regularization
- Dropout regularization

A regression model which uses L1 Regularization technique is called LASSO(Least Absolute Shrinkage and Selection Operator).

A regression model that uses L2 regularization is called ridge regularization. Lasso Regression adds "absolute value of magnitude" of coefficient as penalty term to the loss function(L).

Ridge regression adds "squared magnitude" of coefficient as penalty term to the loss function(L).

During Regularization the output function does not change. The change is only in the loss function.

2. Which particular algorithms are used for regularization?

Ans. Overfitting is a phenomenon that occurs when a machine learning model is constraint to training set and not able to perform well on unseen data.

There are many Machine Learning algorithms that are used to provide machines with the ability to learn or determine the required information or insights from datasets.

Linear Regression can be declared as being the best ML algorithm in Python as it is extensively used for ML and preferred by many ML developers. This supervised ML algorithm is popular and acclaimed for predicting outcomes while observing features. It is capable of running on single variables or multiple variables. It is also relatively less complex and easier to apply; hence, it is preferred by a lot of beginners as well.

Linear regression is a common method to model the relationship between a dependent variable and one or more independent variables. Linear models are developed using the parameters which are estimated from the data.

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

Ans Considering the Linear Regression model has been given it will give us an expected value for a certain set of features in data. The difference

between the expected and the actual value is defined on some exogenous factor, this exogenous factor is often termed as error term.

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

If there were no error, all the data points would be located on the regression line; to the extent they are not represents error; this is what the error term summarizes.

To calculate the error we have to follow three steps:

- 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