Regularization - Ridge Regression & Lasso Regression

In real-life cases, the machine learning models follow the existing patterns in the training data up to different extents. The main purpose of modeling is to make it perform well on unseen data. To do so, the model needs to be free from both overfitting and underfitting.

In machine learning, regularization is the process of making a model **simpler by reducing the importance of variables that are less significant, i.e., adding a penalty in the cost function if more importance is given to such variables**.

A variable plays its role in a model with its **values and the coefficients associated** with it. This indicates that **regularization can be done by regulating this coefficient**. In general, regularization works well when we have a **high number** of features.

Regularization is a technique used to reduce the overfitting of the model. There are two major types of regularization techniques used in regression problems. They are given as follows:

. Ridge regression
. Lasso regression

Let us discuss them one by one:

. **Ridge Regression:** In Ridge regression, as a modification in the cost function, an additional term that is equal to the **product of a regularization parameter and the sum of the square of the coefficients of the model is added**. Ridge regression is also called **L2 regularization** because the **additive term** contains the **second order** of the coefficients.

Mathematically, the loss function for Ridge regression is as follows:

$$Loss\ Function\ =\ min_{\theta}\left[\sum_{i=1}^{n}(Y_i - \theta^T X_i)^2\ +\ \alpha\sum_{j=1}^{m}\theta_j^2\right]$$

Where, α is the regularization parameter, n is the total number of observations, and m is the total number of columns.

The higher the value of the regularization parameter, the higher the penalty, and therefore the magnitude of the coefficients is reduced.

The L2 regularization will **reduce the effect of some coefficients** by minimizing them**,** but it will not completely eradicate the effect of the feature/variable from the model.

. **Lasso Regression:** It is another type of regression that uses regularization but in a different way. In Lasso regression, as a modification in the cost function, **the product of the regularization parameter and the sum of the modulus of the model coefficients is added**. As the model parameters are of the first order, it is also called L1 regularization.

Mathematically, the loss function for Lasso regression is given as follows:

$$Loss\ Function\ =\ min_{\theta}\left[\sum_{i=1}^{n}(Y_i - \theta^T X_i)^2\ +\ \alpha\sum_{j=1}^{m}|\theta_j|\right]$$

Where, α is the regularization parameter, n is the total number of observations, and m is the total number of columns.

The higher the value of the regularization parameter, the bigger the penalty, and therefore the magnitude of the coefficients is reduced.

The **L1 regularization will drive some coefficients to 0** and hence performs **feature selection** because multiplying a variable with 0 removes the effect of that variable from the model.

< Previous

Next >

Help