**Regularization-**

Regularization techniques are used to calibrate the ML models in order to minimize the adjusted loss function and prevent overfitting.

In case of the data points classified correctly, if there are outliers in the dataset the model will classify the outliers also correctly which leads to overfitting the model classification of the training data will be perfect in this case but not on the test data.

Use case- To get rid of the problem of overfitting, we use regularization technique.

Our aim, in regularization the aim is to minimize the term.

L2 regularization, also known as weight decay, is a technique used in machine learning and statistical modeling to prevent overfitting and improve the generalization ability of a model. It's particularly common in linear and logistic regression as well as in neural networks.

In the context of L2 regularization, the standard loss function (e.g., mean squared error for regression problems) is augmented with a regularization term that penalizes large weights. The L2 regularization term is proportional to the squared magnitude of the weights, encouraging the model to have smaller weights and preventing any particular weight from becoming too dominant.

The regularized loss function with L2 regularization can be written as follows:

Regularized Loss = Standard Loss +

Where:

- is the regularization parameter (also known as the regularization strength) that controls the amount of regularization applied. A higher \(\lambda\) leads to stronger regularization.

are the weights of the model that are being regularized.

- n is the total number of weights in the model.

The term is the L2 regularization term, where we sum the squares of all weights and scale the sum by

If is infinity then becomes infinity. We are trying to minimize the term and is going to help us do that.

If then and becomes large. It shows that regularization term and loss terms are moving in opposite directions. is a hyper parameter, which can be obtained by cross validation.

If , the model overfits, in this case the influence of regularization term will be negligible.

If is large then the influence of loss term is negligible when compared to the regularization term and doesn’t show any impact on the model, leading to underfitting the model.

**Regularization-**

As we know that regularizer overfits the model, so we have an alternative . In we have instead of .

will always be positive as we are using its absolute values. is almost same as but there is one advantage of sparsity in .

Sparsity- in regularizer all the unimportant features will become 0. In they will convey to a very small value. So if we want the features to be 0 we will use otherwise.