Ridge

Huber Loss :

Huber loss is a cost function. whenever we want to learn about the outliers as well as ignores them then we can go to this function because it is a Combination of Mean Square Error and Mean Absolute Error. So it balances and Combines good properties of both MAE and MSE.

→ According to MSE, it penalizes the model for making large errors (outliers) by squaring them. So, Mean Squared Error is not robust to outliers.

→ But Mean Absolute Error is robust to outliers. But there we cannot learn about outliers.

So, Huberloss Comes under in that situation.

$$\boxed{\text{Huberloss} = \text{MSE} + \text{MAE}}$$

The Loss function of Huberloss is

Predicted value

$$\text{Loss} = \begin{cases} \underbrace{\frac{1}{2}(y-\hat{y})^2}_{\text{Quadratic Equation}}, & \text{if } |\underset{\text{Actual value}}{y}-\hat{y}| \le \delta \\ \underbrace{\delta|y-\hat{y}|}-\frac{1}{2}\delta^2, & \text{otherwise} \end{cases}$$

Hyperpara-
-meter

↳ Linear Equation

So, Huber loss is Combination of Quadratic Equation and Linear Equation.

if $|y-\hat{y}| > \delta$ then we use $\delta|y-\hat{y}| - \frac{1}{2}\delta^2$

means if $|y-\hat{y}|$ is huge (outlier) then we use the about Linear equation. to calculate the cost of the residuals.

If $|y-\hat{y}| \leq \delta$ then we use Quadratic equation $\frac{1}{2}(y-\hat{y})^2$ to calculate the cost of the residuals.

Advantages:

- It is differentiable at zero.
- Outliers are handled properly due to linearity above δ.
- The hyper parameter, δ can be tuned to maximize model accuracy.

Disadvantages:

- The additional conditionals and comparisons make Huberloss computationally expensive for large datasets.
- In order to maximize model accuracy, δ (hyper parameter) needs to be optimized and it is an iterative process.
- It is differentiable only once.

RMSE [Root Mean Squared Error]:

RMSE is a square root of Mean Square Error [MSE] $= \sqrt{MSE}$.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2} = \sqrt{MSE}$$

→ RMSE is also called the Root Mean Square Deviation.

→ It measures the average magnitude of the residuals/ errors.

- It Concers with the deviation from the actual value.

- If RMSE = 0 means the model has a perfect fit.

- The lower the RMSE value, the better the model and its predictions.

- The higher the RMSE indicates that there is a large devation from the residual to the ground truth.

So simply, RMSE is inversly proportional to the performance of the model.

$$\boxed{RMSE \propto \frac{1}{\text{Performance of the model}}}$$

- RMSE can be used with different features as it helps in figuring out if the feature is improving the model's prediction (or) not.

<u>Advantages:</u>

- RMSE is easy to understand.
- It Serves as a heuristic for training models.
- RMSE does not penalize the errors as much as MSE due to the Square root.

<u>Disadvantages:</u>

- RMSE is sensitive to outliers.
- RMSE increases with an increase in size of the test sample.
- It increases in magnitude if the scale of the error increases.