Loss Function

Loss function is the measure of wrongness of the model. In general it is calculated by the difference between actual and predicted values. The goal of any model is to minimize the loss function as much as possible.
There are various types of loss function based on regression and classification problems.

Regression:

1. Mean Squared Error
2. Mean Absolute Error
3. Huber loss

Classification:

1. Binary cross-entropy
2. Categorical cross-entropy

<u>Mean Squared Error:</u>

This is used for the regression problem where the output/target variable is continuous.
Formula:

MSE = 1/n $\sum_{i=1}^{n}(y_i - \widehat{y})$^2

Advantage:

- ➔ As the equation is in the form of quadratic ie, when expanding $(y_i - \widehat{y})$^2
    - Plotting the quadratic equation, we get a gradient descent with only one global minima
    - We don't get any local minima.
- ➔ This penalizes the model for making large errors by squaring them.

Disadvantage:

- ➔ It is not robust to outliers.

<u>Mean Absolute Error:</u>

This is used for the regression problem where the output/target variable is continuous.
Formula:

MSE = 1/n $\sum_{i=1}^{n}|y_i - \widehat{y}|$

Advantage:

- ➔ It is robust to outliers

Disadvantage:

- ➔ Computation is difficult because of modulus operator
- ➔ It may have local minima

Huber loss:
This is used for the regression problem where the output/target variable is continuous.
This combines the advantages of MSE and MAE
Formula:
Loss = 1/2 (y - $\widehat{y}$)^2 , if |y- $\widehat{y}$| ≤ δ ; otherwise $\delta|y_i$ - $\widehat{y}$| - 1/2δ^2
Advantage:

- Robust to outliers
- Range of value lies between MSE and MAE

Disadvantage:

- This is complex to implement.

Binary cross-entropy:
This is used for the classification problem where the output/target variable is binary classification. For finding $\widehat{y}$, we use sigmoid activation function at the last neuron of the neural network.
Formula:
Loss = - 1/n $\sum_{i=1}^{n}$ (y log($\widehat{y}$) + (1-y) log(1-$\widehat{y}$))
Advantage:

➔ Cost function is differentiable

Disadvantage:

➔ Multiple local minima

Categorical cross-entropy:
This is used for the classification problem where the output/target variable is multiclass classification. For finding $\widehat{y}$, we use softmax activation function at the last neuron of the neural network.
Formula:
Loss = - $\sum_{j=1}^{c} y_{ij}$ log($\widehat{y}$)
c - number of class in the target variable
Advantage:

➔ This function is differentiable

Disadvantage:

➔ Impact of imbalance data as it prioritize the most occurring class as output
➔ Requires one hot encoding