

24. Cost Function

What is Cost Function:

- A cost function is an important parameter that determines how well a machine learning model performs for a given dataset
- Cost function is a measure of how wrong the model is in estimating the relationship b/w $x(\text{input})$ and $y(\text{output})$ parameter.
- With the help of cost function, you draw best fit line
- Cost function and loss functions are both functions of error - to make the error minimum from the best fit line

Types of cost function:

- Regression cost function
- Classification cost function

1) Regression Cost Function:

- Regression models are used to make a prediction for the continuous variables.
 1. MSE (Mean Square Error)
 2. RMSE (Root Mean Square Error)
 3. MAE (Mean Absolute Error)
 4. R^2 Accuracy

2) Binary Classification Cost Function: Classification models are used to make predictions of categorical variables, such as predictions for 0 or 1, cat or dog, etc.

3) Multi-class Classification Cost Function: A multi-class classification cost function is used in the classification problems for which instances are allocated to one of more than two classes. Binary Cross Entropy Cost Function or Log Loss Function

24.1 Regression Cost Function



No description has been provided for this image

- red line represents prediction line
- blue points represent original data
- red triangles represent error

- please note that the error value should be minimum
- For this we use **cost function to make the error minimum from prediction line**

24.2 Mean Square Error

Mean Square Error (MSE) is the mean squared difference b/w the actual and predicted values. MSE penalizes high errors caused by outliers by squaring the error. MSE is also known as **L2 Loss**.

The Mean Squared Error (MSE) is calculated as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Whereas : $-y_i = \text{Original value}$ $-\hat{y}_i = \text{Predicted value}$ $-n = \text{number of rows}$

Advantages of using MSE are:

1. It is differentiable

Disadvantages of using MSE:

1. When outlier is present in data, it will increase exponentially when squaring took place, so it will give wrong predictions
2. The data does not remain in original form, rather available in squared form, and also if original data is in cm. then it will change the unit in cm^2 as well. So the data as well as units will not be in original format.

In []:

The differentiation of $(y = mx + c)$ with respect to (x) is:

$$\frac{d}{dx}(y) = \frac{d}{dx}(mx + c) = m$$

In this differentiation:

- $(\frac{d}{dx}(y))$ represents the derivative of (y) with respect to (x) .
- $(\frac{d}{dx}(mx + c))$ is the derivative of the function $(mx + c)$.
- The result (m) is the slope of the line, which is constant in this linear equation.

The update formula for finding **m(new)** is given by:

$$M_{\text{new}} = M_{\text{old}} - \lambda \left(\frac{dz}{dm} \right)$$

24.3 Mean Absolute Error

Mean Absolute Error (MAE) is the mean absolute difference b/w the actual values and the predicted values. MAE is more robust to outliers. The insensitivity to outliers is b/c it does not penalize high errors caused by outliers.

The Mean Absolute Error (MAE) is calculated as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Advantages of MAE:

1. Error remains in original form
2. It treats outlier well

Disadvantages are:

1. This is not differentiable equation,

24.4 Root Mean Squared Error

Root Mean Squared Error (RMSE) is the root squared mean of the difference b/w actual and predicted values. RMSE can be used in situations where we want to penalize high errors but not as much as MSE does.

The Root Mean Square Error (RMSE) is calculated as:

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

24.5 How to Find Best Fit Line

- For finding best line:
 1. Keep the error (loss) minimum (Which will be calculated through cost function)
 2. Through quadratic equation, gradient descent, we take the minimum value



No description has been provided for this image



No description has been provided for this image

- loss function will be minimum after looking for best m (slope) and c (intercept) values in $y = mx + c$
- m: donates angle
- c: donates intercept at y-axis.
- this whole process is called gradient descent technique

In []: