Getting started with Machine Learning

Raunak J

What we've covered so far now ?

The Simple Linear Regression Model with

Mathematical orientation and Implementation in Python

So now what next ?

We’re supposed to track and learn the current progress of the model and see to it improvements can be made.

So we'll start with Metrics Evaluation

So what really is Metrics Evaluation ?

Metrics are associated with finding the differences of predicted values and actual values.

So we can perform hypothesis, indicating that our predictions are as good as real values.

So we can relate this to testing ?

In this tutorial…

- We’ll have a look at Mean Squared Error.
- We’ll also learn the implementation.

So what is Mean Squared Error ?

- In statistics, the mean squared error (MSE) or mean squared deviation (MSD) of an estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value.

Source : https://en.wikipedia.org/wiki/Mean_squared_error

So the equation is represented by ?

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

$$n = Number\ of\ samples$$
$$y_i = Actual\ target\ values$$
$$\hat{y}_i = Predicted\ Values$$

Now let us understand with an example we learned earlier

Assume a problem where you have to predict the weight of the person given his height. Here weight is the dependent variable and height is independent variable.

Height of the person in cms (x)	Weight of the person in kgs (y)
160	72
171	76
182	77
180	83
154	76

For the given example we considered Simple Linear Regression

$$\hat{y} = \beta_0 + \beta_{1.} x_i = 41.5648 + (0.208 * x_i)$$

So considering the problem :

Height of the person in cms (x)	Weight of the person in kgs (y)
160	72
171	76
182	77
180	83
154	76

We try to find prediction values for each and every single training data sample.

Height of the person in cms (x)	Weight of the person in kgs (y)	Predicted Values y-hat
160	72	
171	76	
182	77	
180	83	
154	76	

$$\hat{y}_i = \beta_0 + \beta_{1.} x_i = 41.5648 + (0.208 * x_i)$$

Height of the person in cms (x)	Weight of the person in kgs (y)	Predicted Values y-hat
160	72	74.8448
171	76	
182	77	
180	83	
154	76	

$$\widehat{y_1} = \beta_0 + \beta_1.x_1 = 41.5648 + (0.208 * 160) = 74.8448$$

Height of the person in cms (x)	Weight of the person in kgs (y)	Predicted Values y-hat
160	72	74.8448
171	76	77.1328
182	77	
180	83	
154	76	

$$\widehat{y_2} = \beta_0 + \beta_1 . x_2 = 41.5648 + (0.208 * 171) = 77.1328$$

Height of the person in cms (x)	Weight of the person in kgs (y)	Predicted Values y-hat
160	72	74.8448
171	76	77.1328
182	77	79.4208
180	83	
154	76	

$$\widehat{y_3} = \beta_0 + \beta_1 . x_3 = 41.5648 + (0.208 * 182) = 79.4208$$

Height of the person in cms (x)	Weight of the person in kgs (y)	Predicted Values y-hat
160	72	74.8448
171	76	77.1328
182	77	79.4208
180	83	79.0048
154	76	

$$\widehat{y_4} = \beta_0 + \beta_1 . x_4 = 41.5648 + (0.208 * 180) = 79.0048$$

Height of the person in cms (x)	Weight of the person in kgs (y)	Predicted Values y-hat
160	72	74.8448
171	76	77.1328
182	77	79.4208
180	83	79.0048
154	76	73.5968

$$\widehat{y_5} = \beta_0 + \beta_1 . x_5 = 41.5648 + (0.208 * 154) = 73.5968$$

Now comes the calculation of MSE

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

Height of the person in cms (x)	Weight of the person in kgs (y)	Predicted Values y-hat
160	72	74.8448
171	76	77.1328
182	77	79.4208
180	83	79.0048
154	76	73.5968

$$MSE = \frac{1}{5}\{(72-74.8448)^2+(76-77.1328)^2+(77-79.4208)^2+(83-79.0048)^2+(76-73.5968)^2\}$$

$$MSE = \frac{8.092 + 1.283 + 5.860 + 15.962 + 5.775}{5}$$

$$MSE = 7.3944$$