MEAN SQUARE ERROR

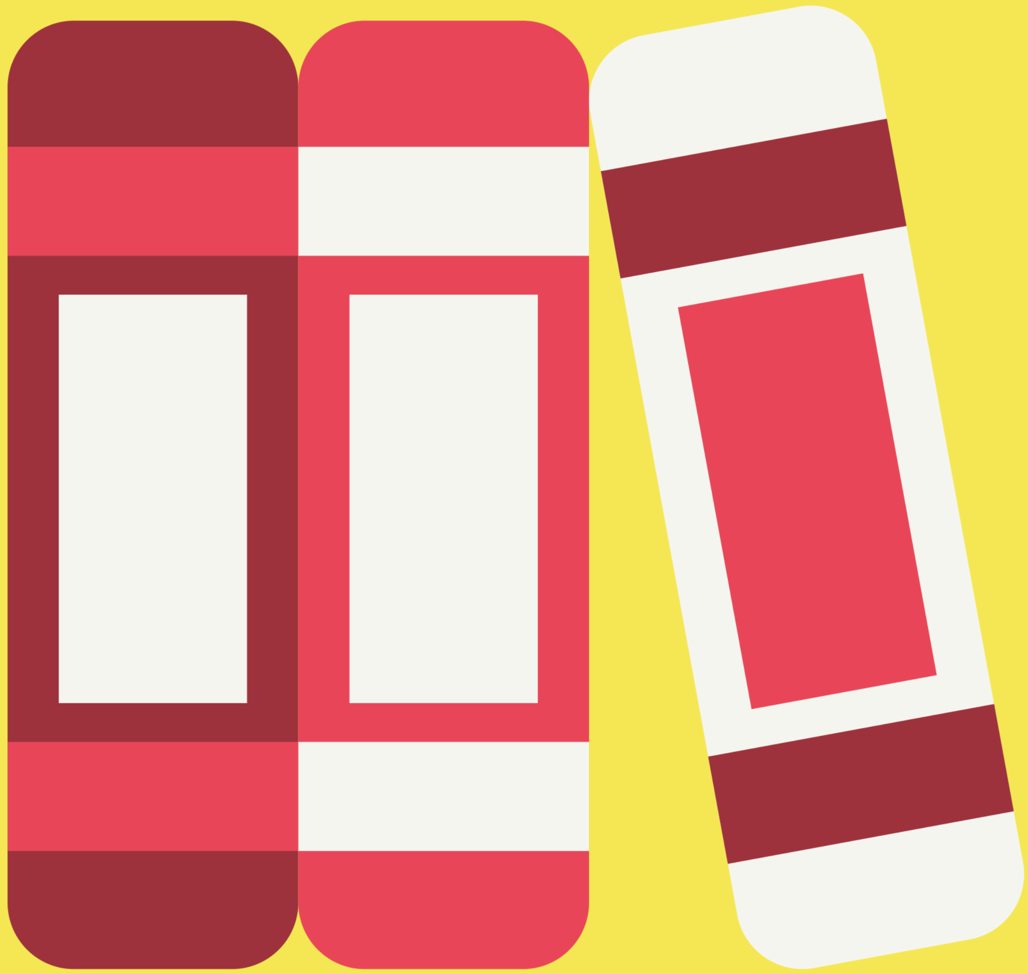
PROBABILITY AND STATISTICS FOR DATA SCIENCE

ABHIN P T

MEAN SQUARE ERROR

THE AVERAGE SQUARED DIFFERENCE BETWEEN THE ESTIMATED VALUES AND THE ACTUAL VALUE.

A FUNCTION MAPPING ARBITRARY INPUTS TO A SAMPLE OF VALUES OF SOME RANDOM VARIABLE.



PROPERTIES

PREDICTOR

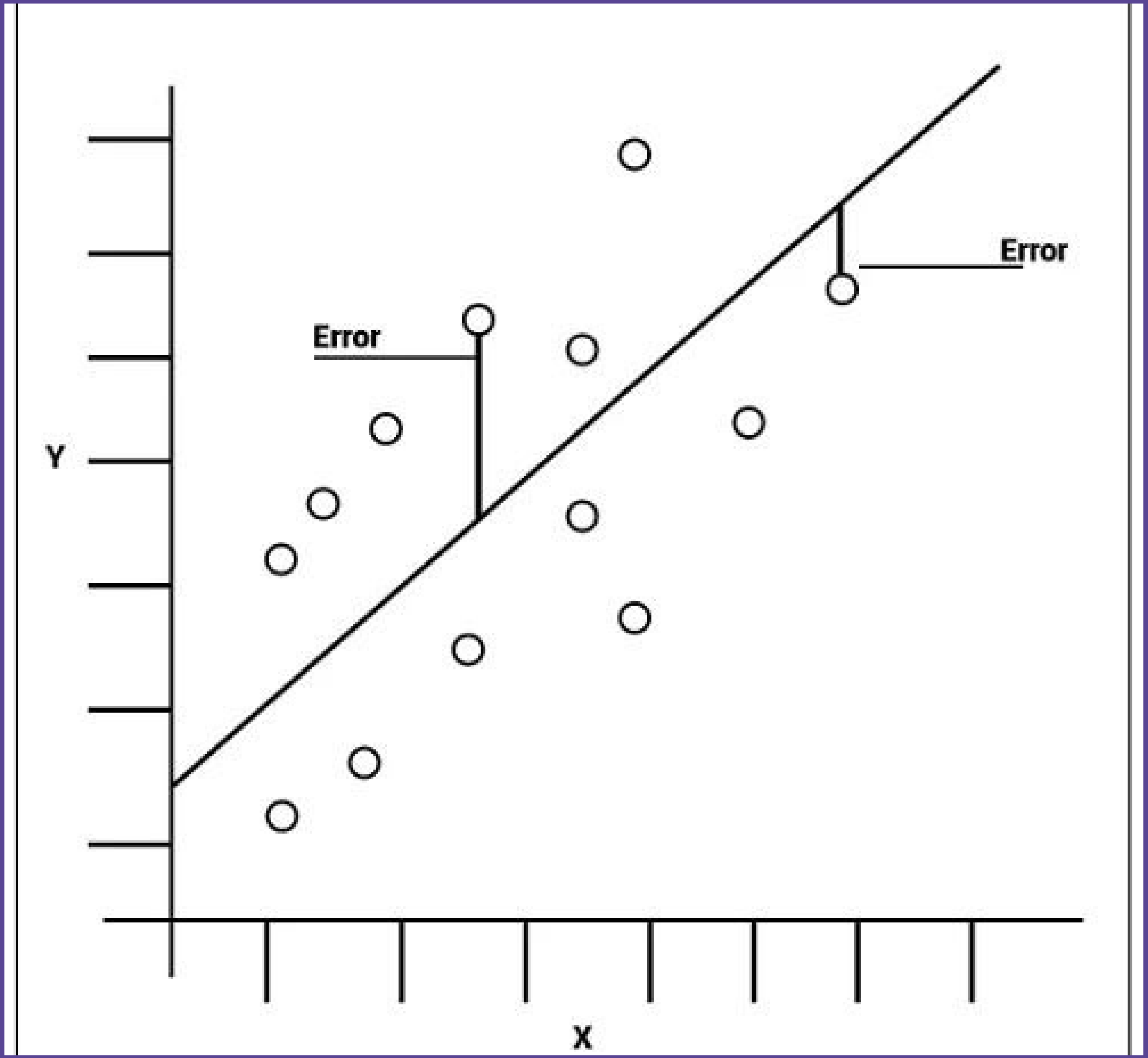
If a vector of n predictions is generated from a sample of n data points on all variables, and displaystyle Y is the vector of observed values of the variable being predicted, with Y^ being the predicted values , then the within-sample MSE of the predictor is computed as

$$\operatorname{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y_i} \right)^2 .$$

IN REGRESSION

PROPERTIES

In regression analysis, plotting is a more natural way to view the overall trend of the whole data. The mean of the distance from each point to the predicted regression model can be calculated, and shown as the mean squared error. The squaring is critical to reduce the complexity with negative signs. To minimize MSE, the model could be more accurate, which would mean the model is closer to actual data.



IN REGRESSION

LOSS FUNCTION

- THE MATHEMATICAL BENEFITS OF MEAN SQUARED ERROR ARE PARTICULARLY EVIDENT IN ITS USE AT ANALYZING THE PERFORMANCE OF LINEAR REGRESSION, AS IT ALLOWS ONE TO PARTITION THE VARIATION IN A DATASET INTO VARIATION EXPLAINED BY THE MODEL AND VARIATION EXPLAINED BY RANDOMNESS.

EXAMPLE

Month	Actual	Forecasted	Squared Error	
January	67	70	9	
February	50	49	1	
March	36	38	4	
April	74	76	4	
May	84	83	1	
June	84	80	16	
July	64	67	9	
August	34	30	16	
September	23	20	9	
October	72	75	9	
November	62	60	4	
December	42	38	16	
			8.166666667	**MSE**

MSE = (1/12) * (98) = 8.166

The MSE for this model is 8.17.