ERROR METRICS:

This is to represent the predictive ability of our model.The difference between the actual and predicted model is estimated as the residue.If our residue is small,then the model prediction is good.but if it is large,it indicates the model as poor estimator.We technically can inspect all the residuals to judge the accuracy, but when the data points grows,it becomes difficult.

So this is the scenario where we used summary measurement to take all residuals and condense them into single value.Here comes the error metrics.

This is also called as loss functions.

DIFFERENT TYPES OF ERROR METRICS

1. Mean Absolute Error (MAE)
2. Mean Squared Error (MSE)
3. Root Mean Squared Error(RMS)
4. R squared Error
5. Absolute R squared Error

1.Mean Absolute Error(MAE)

As the name indicates,this is used to sum of all values irrespective of negative / positive sign .i.e., It is the sum of absolute difference between actual and predicted values.

MAE = summation of all values |Actual –Predicted Value| /total number.

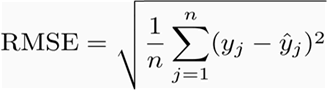
2. Mean Squared Error (MSE)

This is the average of squared values of difference between actual and predicted values. This will always be positive. If this is closer to 0 or lower, Then it is better

MSE= summation of all values (Actual –Predicted Value)^sq. /total number.

3.Root Mean Squared Error(RMSE)

This is similar to MSE but it square root with it.It represent the standard deviation of the difference between actual and predicted value.



|  |  |  |
| --- | --- | --- |
| MAE | MSE | RMSE |
| This take absolute value | This value will be always positive | This value will always be positive |
| This is less biased for higher values. | This is highly biased | This is highly biased . |
| As sample size increases ,value get smaller | The values get increase with sample size | The values get increase with sample size |
| This is more useful when impact is proportionate with increase in Error. | When impact is disproportionate with increase in error. | When impact is disproportionate with increase in error. |
| This is not sensitive to outliers. | Mostly used when outliers present | Avoid taking absolute value and useful in most mathematical calculation |
| It is usually used when the performance is measured on continuous variable data. | It is most commonly used but least useful when dataset have more noises | RMSE is much more useful when large errors are present and they drastically affect the model's performance |

4.R squared

This is to indicate how our model fits good with the given dataset.It indicates how close the regression line is to actual data values. The **value lies between 0 and 1** where 0 indicates that this model doesn't fit the given data and 1 indicates that the model fits perfectly to the dataset provided.