Performance Metrics: Regression Model

By Sunil Ghimire - "**Be Unique, Be Identifiable, Be You**"

Today I am going to discuss Performance Metrics, and this time it will be Regression model metrics. As in my previous article, I have discussed Classification Metrics, this time its Regression. I am going to talk about the 5 most widely used Regression metrics:

Let's understand one thing first and that is the Difference between Classification and Regression:
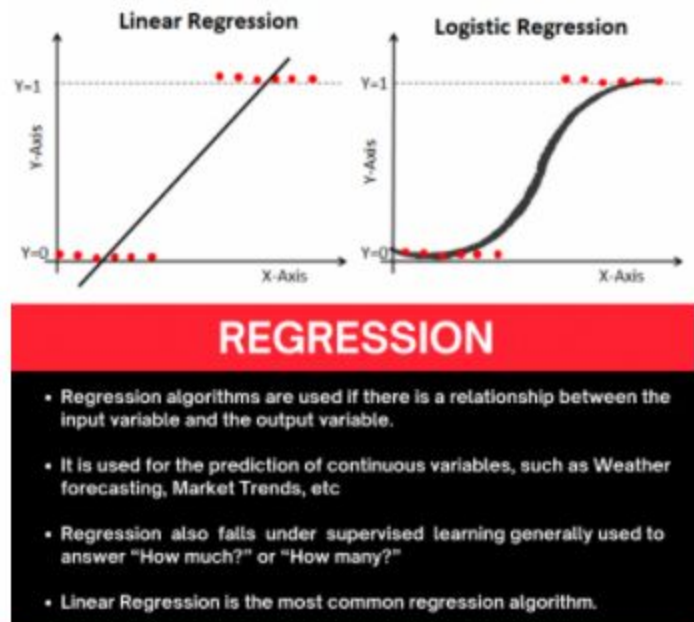
WHAT IS AN ERROR?

Any deviation from the actual value is an error.

Error = $y_{(actual)} - y_{(predicted)}$

So keeping this in mind, we have understood the requirement of the metrics, let's deep dive into the methods we can use to find out ways to understand our model's performance.

MEAN SQUARED ERROR (MSE)

Let's try to breakdown the name, it says Mean, it says Squared, it says Error. We know what Error is, from the above explanation, we know what square is. We square the Error, and then we know what the Mean is. So we take the mean of all the errors which are squared and added.

The first question that should arise is, why are we doing a Square? Why can we not take the error directly?

Let's take the height example, My model predicted 167cm whereas my actual value is 163cm, so the deviation is of +5cm. Let's take another example where my predicted height is 158cm and my actual height is 163cm. Now, my model made a mistake of -5cm.

Now let's find Mean Error for 2 points, so the calculation states $\dfrac{+5 + (-5)}{2} = 0$

This shows that my model has 0 error, but is that true? No right? So to avoid such problems we have to take square to get rid of the Sign of the error.

So let's see the formulation of this Metric:

$$\textbf{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_{(i)} - \overline{y_{(i)}})^2$$

Where,

-> n = total number of data points

-> $y_{(i)}$ = actual value

-> $\overline{y_{(i)}}$ = predicted value

Now the question is, if we already have the **MSE**, why we require **RMSE**?

Let's try to understand it with example. Take the above example of the 2 data points and calculate **MSE** and **RMSE** for them,

$$\textbf{MSE} = \frac{[(+5)^2 + (-5)^2]}{2} = 25$$

$$\textbf{RMSE} = \text{Sqrt}(\textbf{MSE}) = (25)^{0.5} = 5$$

Now, you tell among these values which one is more accurate and relevant to the actual error of the model?

RMSE right? So in actual Squaring off, the values increase them exponentially. While not taking a root might affect our understanding that where my model is actually making mistakes.

MEAN ABSOLUTE ERROR (MAE)

Now, I am sure you might have given this a thought, why squaring? Why not just taking the Absolute value of them, so here we have it. Everything stays the same, the only difference is, we take the Absolute value of our error, this also takes care of the sign issues we had earlier. So let's look into the formula :

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left| y_{(i)} - \overline{y_{(i)}} \right|$$

Where,

-> n = total number of data points

-> $y_{(i)}$ = actual value

-> $\overline{y_{(i)}}$ = predicted value

MAE VS RMSE

Let's understand, **MAE** and **RMSE** can be used together to diagnose the variation in the errors in a set of forecasts. **RMSE** will always be larger or equal to the **MAE**. The greater difference between them, the greater the variance in the individual errors in the sample. If the **RMSE = MAE**, then all the errors are of the same magnitude.

 a. Errors [2, -3, 5, 120, -116, 197]

 b. **RMSE** = 115.5

 c. **MAE** = 88.6

If we see the difference, **RMSE** has a higher value than **MAE**, which states that **RMSE** gives more importance to higher error due to squaring the values.

MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)

It is kind of similar to **MAE** but the difference is: we take the percentage error and not the absolute value. Let's see the formulation:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| y_{(i)} - \overline{y_{(i)}} \right|}{y_{(i)}} * 100\%$$

Where,

-> n = total number of data points

-> $y_{(i)}$ = actual value

-> $\overline{y_{(i)}}$ = predicted value

MAPE represents the error in percentage and therefore it's not relative to the size of the numbers in the data itself, whereas any other performance metrics in the regression model.

$(R)^2$ or COEFFICIENT OF DETERMINATION

It is the Ratio of the **MSE** (Prediction Error) and Baseline Variance of the target Variable, here baseline is the deviation of our Y values from the Mean value.

The metric helps us to compare our current model with constant baseline value (i.e. mean) and tells us how much our model is better, $(R)^2$ is always less than 1, and it doesn't matter how large or small the errors are $(R)^2$ is always less than 1. Let's See the Formulation:

$$(R)^2 = 1- \frac{SS_{(RES)}}{SS_{(TOT)}} = 1 - \frac{\sum\limits_{i=1}^{n} (y_{(i)} - \widehat{y_{(i)}})^2}{\sum\limits_{i=1}^{n} (y_{(i)} - \overline{y_{(i)}})^2}$$

☻ Thanks for your time ☻

What do you think of this "**Performance Metrics: Regression Model**"? (Appreciation, Suggestions, and Questions are highly appreciated).