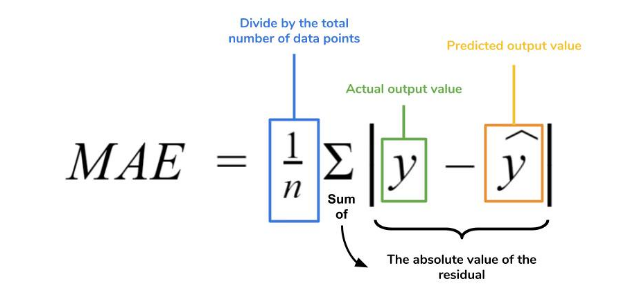
Error Metrics for regression

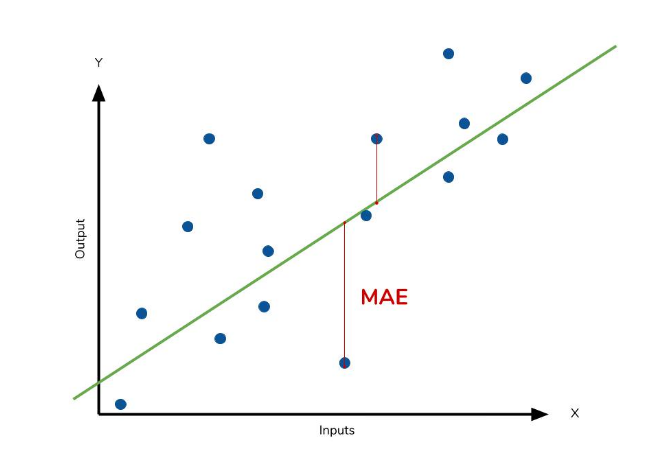
**Mean Absolute Error (MAE)**

The MAE is a simple way to measure error magnitude. It consists on the average of the absolute differences between the predictions and the observed values. The measure goes from 0 to infinite, being 0 the best value you can get.

***How to read****: Values closer to 0 are better than higher values.*



The picture below is a graphical description of the MAE. The green line represents our model’s predictions, and the blue points represent our data.



The MAE is also the most intuitive of the metrics since we’re just looking at the absolute difference between the data and the model’s predictions. Because we use the absolute value of the residual, the MAE does not indicate underperformance or overperformance of the model (whether or not the model under or overshoots actual data). Each residual contributes proportionally to the total amount of error, meaning that larger errors will contribute linearly to the overall error. Like we’ve said above, a small MAE suggests the model is great at prediction, while a large MAE suggests that your model may have trouble in certain areas. A MAE of 0 means that your model is a perfect predictor of the outputs (but this will almost never happen).

While the MAE is easily interpretable, using the absolute value of the residual often is not as desirable as squaring this difference. Depending on how you want your model to treat outliers, or extreme values, in your data, you may want to bring more attention to these outliers or downplay them. The issue of outliers can play a major role in which error metric you use.

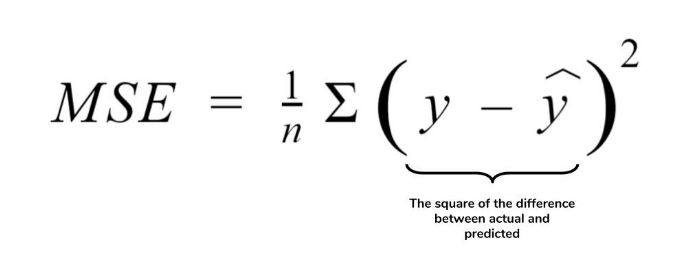
**Usage Tips:**

• Uses a similar scale to input data

• Can be used to compare series of different scale

**Mean Squared Error or MSE**

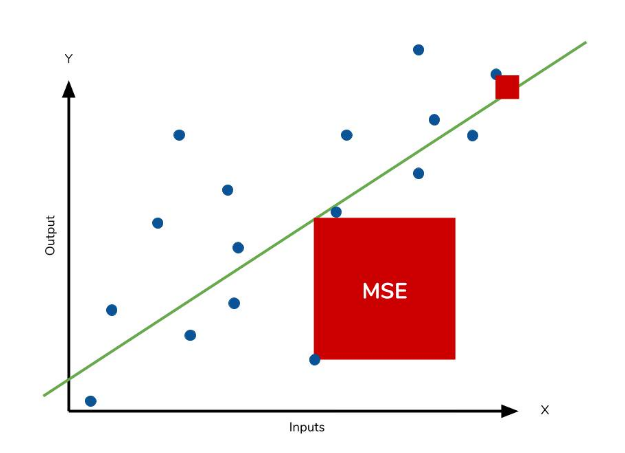
MSE is calculated by taking the average of the square of the difference between the original and predicted values of the data.



Here N is the total number of observations/rows in the dataset. The sigma symbol denotes that the difference between actual and predicted values taken on every i value ranging from 1 to n.

**Consequences of the Square Term**

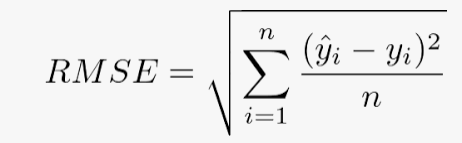
Because we are squaring the difference, the MSE will almost always be bigger than the MAE. For this reason, we cannot directly compare the MAE to the MSE. We can only compare the model’s error metrics to those of a competing model. The effect of the square term in the MSE equation is most apparent with the presence of outliers in the data. While each residual in MAE contributes proportionally to the total error, the error grows quadratically in MSE. This ultimately means that outliers in the data will contribute to much higher total error in the MSE than they would the MAE. Similarly, the model will be penalized more for making predictions that differ greatly from the corresponding actual value. This is to say that large differences between actual and predicted are punished more in MSE than in MAE. The following picture graphically demonstrates what an individual residual in the MSE might look like.

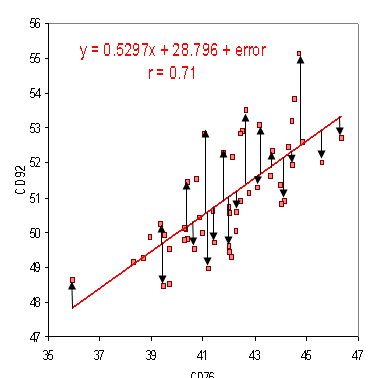


Outliers will produce these exponentially larger differences, and it is our job to judge how we should approach them.

**Root Mean Squared Error**

Root Mean Squared Error (RMSE) measures the average magnitude of the error by taking the square root of the average of squared differences between prediction and actual observation. It tells us how concentrated the data is around the line of best fit. The RMSE is the square root of the variance of the residuals. Lower values of RMSE indicate a better fit. RMSE is a good measure of how accurately the model predicts the response. It is often used to judge the quality of prediction.





**Usage Tips:**

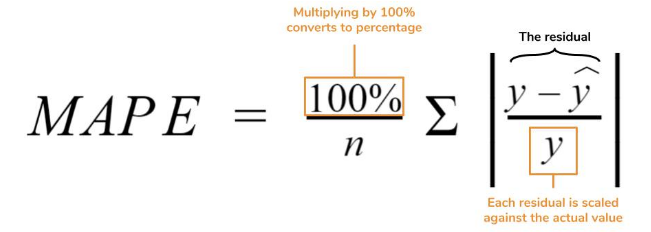
• A lower value for RMSE is favourable.

• Sensitive to outliers.

• Highly dependent on fraction of data used (low reliability)

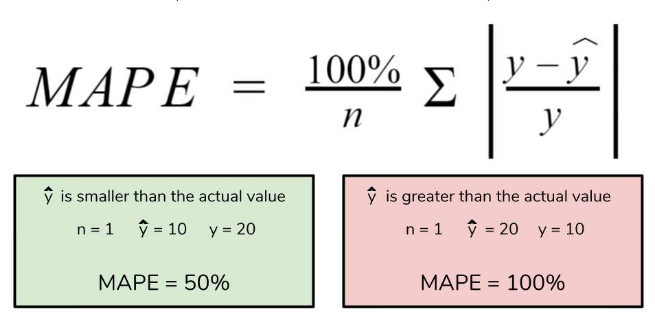
**Mean Absolute Percentage Error**

The **mean absolute percentage error** (MAPE) is the percentage equivalent of MAE. The equation looks just like that of MAE, but with adjustments to convert everything into percentages.



Just as MAE is the average magnitude of error produced by your model, the MAPE is how far the model’s predictions are off from their corresponding outputs on average. Like MAE, MAPE also has a clear interpretation since percentages are easier for people to conceptualize. Both MAPE and MAE are robust to the effects of outliers thanks to the use of absolute value.

However for all of its advantages, we are more limited in using MAPE than we are MAE. Many of MAPE’s weaknesses actually stem from use division operation. Now that we have to scale everything by the actual value, MAPE is undefined for data points where the value is 0. Similarly, the MAPE can grow unexpectedly large if the actual values are exceptionally small themselves. Finally, the MAPE is biased towards predictions that are systematically less than the actual values themselves. That is to say, MAPE will be lower when the prediction is lower than the actual compared to a prediction that is higher by the same amount. The quick calculation below demonstrates this point.



We have a measure similar to MAPE in the form of the **mean percentage error.** While the absolute value in MAPE eliminates any negative values, the mean percentage error incorporates both positive and negative errors into its calculation.

**Usage Tips:**

• Commonly-used as a loss function

• Cannot be used if there are actual zero values.

• Percentage error cannot exceed 1.0 for small predictions.

• There is no upper limit to percentage error in predictions that are too high.

• Non-symmetrical (adversely affected if a predicted value is larger or smaller than the corresponding actual value)

**R Square Error**

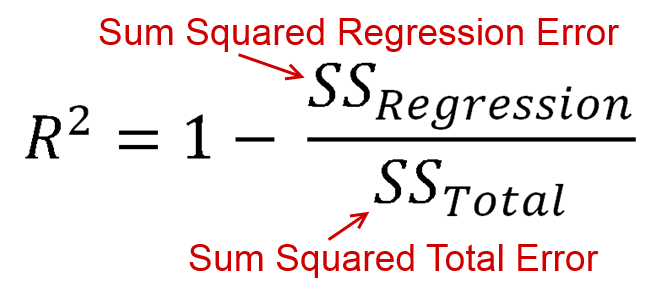
R-Squared (R² or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the [independent variable](https://corporatefinanceinstitute.com/resources/knowledge/modeling/independent-variable/). In other words, r-squared shows how well the data fit the regression model (the goodness of fit).

The most common interpretation of r-squared is how well the regression model fits the observed data. For example, an r-squared of 60% reveals that 60% of the data fit the regression model. Generally, a higher r-squared indicates a better fit for the model.

However, it is not always the case that a high r-squared is good for the regression model. The quality of the statistical measure depends on many factors, such as the nature of the variables employed in the model, the units of measure of the variables, and the applied [data transformation](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3043340/). Thus, sometimes, a high r-squared can indicate the problems with the regression model.

A low r-squared figure is generally a bad sign for predictive models. However, in some cases, a good model may show a small value.

There is no universal rule on how to incorporate the statistical measure in assessing a model. The context of the experiment or [forecast](https://corporatefinanceinstitute.com/resources/knowledge/modeling/forecasting-methods/) is extremely important and, in different scenarios, the insights from the metric can vary.



**Usage Tips:**

A perfect fit is achieved when the numerator equals to zero