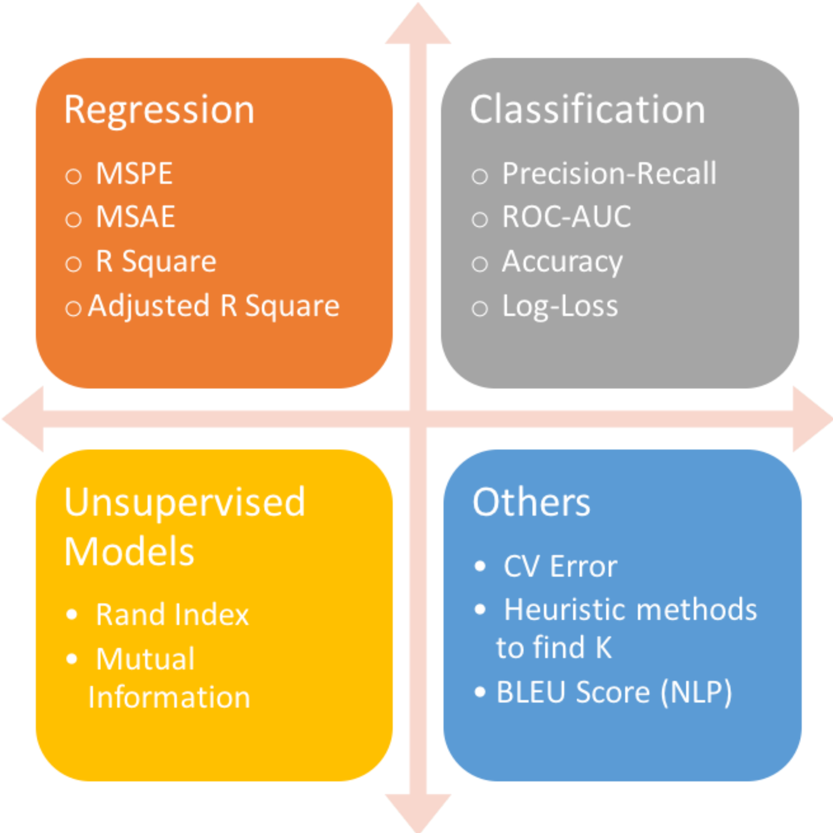
Performance Metrics for Regression



Regression analysis is a subfield of supervised machine learning.

It aims to model the relationship between a certain number of features and a continuous target variable. Following are the performance metrics used for evaluating a regression model:

1.Mean Absolute Error (MAE)

2.Mean Squared Error (MSE)

3.Root Mean Squared Error (RMSE)

4.R-Squared

5.Adjusted R-squared

Mean Squared Error

Mean Squared Error, or MSE for short, is a **popular error metric for regression problems.**

It is also an **important loss function** for **algorithms fit or optimized using the least squares framing of a regression problem**.

Here "**least squares**" refers to **minimizing** the **mean squared error between predictions and expected values.**

The MSE is calculated as the mean or average of the squared differences between predicted and expected target values in a dataset.

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

Where **yi is the i'th expected value** in the dataset and **yhat_i is the i'th predicted value**.

The **difference between these two values is squared, which has the effect of removing the sign, resulting in a positive error value.**

The **squaring** also has the **effect of inflating or magnifying** large errors.

That is, the **larger the difference between the predicted and expected values**, the **larger the resulting squared positive error**.

This has the effect of "punishing" models more for **larger errors when MSE is used as a loss function.**

It also has the **effect of "punishing" models** by **inflating the average error score when used as a metric.**

Here, **the error term is squared and thus more sensitive to outliers** as compared to **Mean Absolute Error (MAE)**.

Root Mean Squared Error

The Root Mean Squared Error, or RMSE, is an extension of the mean squared error.

Importantly, the square root of the error is calculated, which means that the units of the RMSE are the same as the original units of the target value that is being predicted.

As such, it may be common to use MSE loss to train a regression predictive model, and to use RMSE to evaluate and report its performance.

The RMSE can be calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

Where y_i is the i'th expected value in the dataset, yhat_i is the i'th predicted value, and sqrt() is the square root function.

Note that the RMSE cannot be calculated as the average of the square root of the mean squared error values.

You may recall that the square root is the inverse of the square operation.

*MSE uses the square operation to remove the sign of each error value and to punish large errors. *

The square root reverses this operation, although it ensures that the result remains positive.

Mean Absolute Error (MAE)

Mean Absolute Error, or MAE, is a popular metric because, like RMSE, the units of the error score match the units of the target value that is being predicted.

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

Unlike the RMSE, the **changes in MAE are linear** and therefore intuitive.

That is, **MSE and RMSE punish larger errors** more than **smaller errors, inflating or magnifying the mean error score.**

This is due to the **square of the error value**.

The **MAE does not give more or less weight** to **different types of errors** and instead the scores **increase linearly** with **increases in error**.

As its **name suggests**, the **MAE score is calculated as the average of the absolute error values.**

Absolute or abs() is a mathematical function that simply makes a number positive.

R-Squared

SS_{RES} $\sum_i (y_i - \hat{y}_i)^2$

R-squared is calculated by **dividing the sum of squares of residuals (SSres) from the regression model by the total sum of squares (SStot)** of **errors from the average model** and **then subtract it from 1**.

R-squared is also known as the Coefficient of Determination. It explains the **degree to which the input variables explain the variation of the output / predicted variable.**

A **R-squared value of 0.81**, **tells** that **the input variables explains 81 % of the variation in the output variable.**

The higher the R squared, the **more variation is explained by the input variables and better is the model.**

Although, there exists a limitation in this metric, which is solved by the Adjusted R-squared.

Adjusted R-squared

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Here, N- total sample size (number of rows) and p- number of predictors (number of columns)

The limitation of R-squared is that it will **either stay the same or increases with the addition of more variables, even if they do not have any relationship with the output variables.**

To overcome this limitation, Adjusted R-square comes into the picture as it penalizes you for adding the variables which do not improve your existing model.

Hence, if you are **building Linear regression on multiple variables, it is always suggested that you use Adjusted R-squared to judge the goodness of the model.**

If there exists only one input variable, R-square and Adjusted R squared are same.