Linear Regression Metrics

R2

- R2 score or the coefficient of determination measures how much the explanatory variables explain the variance of the dependent variable.
- It indicates if the fitted model is a good one, and if it could be used to predict the unseen values.
- The best value of R2 is 1, meaning that the model is a perfect fit of our dataset.
- It could be 0, meaning that the model is a constant and it will always predict the expected average value of the dependent variable y, regardless of the input variables.
- It could also be negative, the model could be arbitrarily worse.

R2…

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

With y*i* the predicted value for the observation i, and y is the average value of the dependent variable:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

Let's introduce another measure RSS "Residual Sum of Square" which as its name indicates, compute the square of the residual error (difference between the real and the predicted value):

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

We can then, write the formula of the R2 as:

$$R^2 = 1 - \frac{RSS}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$

MSE: Mean Squared Error

- Mean square Error compute the average squared error between the true value and the predicted one
- It gives more importance to the highest errors, thus it’s more sensitive to outliers.

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

We can recall here that the RSS (Residual Sum of Square):

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

Then:

$$MSE = \frac{RSS}{N}$$

RMSE: Root Mean Squared Error

- RMSE is the root squared of MSE

$$RMSE = \sqrt{MSE}$$

MAE: Mean Absolute Error

- MAE is the mean absolute error, is more robust to outliers than the MSE:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \widehat{y_i}|$$