Standard Deviation

A measure of dispersion

One of two parameters of a major probability density: The Normal Distribution

$$N(\mu, \sigma)$$

While it's not directly observable, it can be estimated by the sample standard deviation of a group (sample) of observations

$$s = \sqrt{\frac{\sum(y - \bar{y})^2}{n - 1}}$$

Note: Most of the time we do not know µ (the population average) and we estimate it with x (the sample average). The formula for s2 measures the squared deviations from x rather than µ . The xi's tend to be closer to their average x rather than µ , so we compensate for this by using the divisor (n-1) rather than n.

Variance

Standard Deviation is the square root of the Variance. Standard Deviation is often preferred to Variance since it's on the same unit scale as our observations, and more interpretable

$$S^2 = \frac{\sum(y - \bar{y})^2}{n - 1}$$

Total Sum of Squares (Total Variation)

The Numerator of Variance is the Total Sum of Squares, or Total Variation

$$SST = \sum (y - \bar{y})^2$$

Decomposition of Total Variation

A fundamental formula in statistics

$$\sum (y - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2$$

$$SST = SSR + SSE$$

It states that Total Variation can be broken down into Explained Variation (SSR) and Unexplained Variation, and that they're additive

Performance Metrics

MSE

Mean squared error (MSE) is the average of sum of squared difference between actual value and the predicted or estimated value. It is also termed as **mean squared deviation (MSD).** This is how it is represented mathematically:

$$MSE = \frac{1}{n}\Sigma\underbrace{\left(y - \hat{y}\right)}_{\text{The square of the difference between actual and predicted}}{}^{2}$$

Fig 1. Mean Squared Error

The value of MSE is always positive or greater than zero. A value close to zero will represent better quality of the estimator / predictor (regression model). **An MSE of zero (0) represents the fact that the predictor is a perfect predictor**. When you take a **square root of [MSE](MSE)** value, it becomes **root mean squared error (RMSE).** In the above equation, Y represents the actual value and the Y' is predicted value. Here is the diagrammatic representation of MSE:

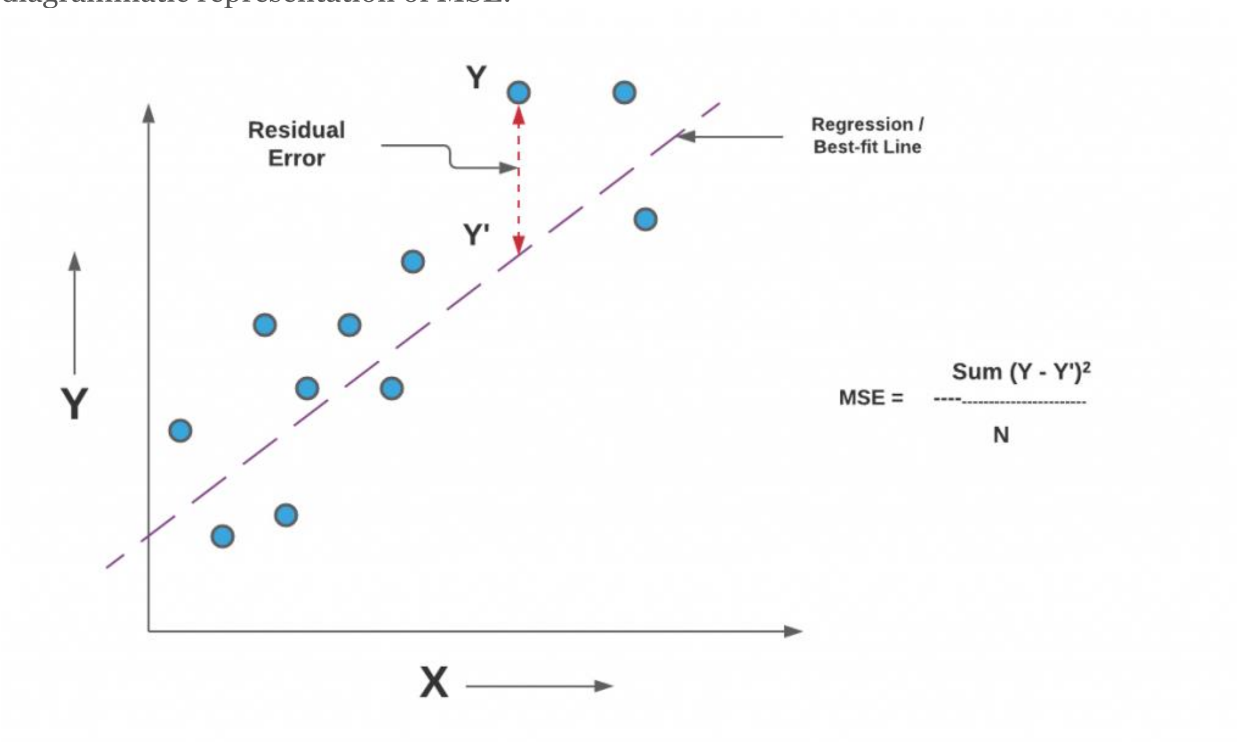

Fig 2. Mean Squared Error Representation

R^2

R squared (a performance metric you're likely familiar with) can be calculated using the components in the Decomposition of Total Variation.

R-Squared is the ratio of Sum of Squares Regression (SSR) and Sum of Squares Total (SST). Sum of Squares Regression is amount of variance explained by the regression line. R-squared value is used to measure the **goodness of fit.** Greater the value of R-Squared, better is the regression model. However, we need to take a caution. This is where **adjusted R-squared** concept comes into picture. This would be discussed in one of the later posts. R-Squared is also termed as the **coefficient of determination.** For the training dataset, the R^2 is bounded between 0 and 1, but it can become negative for the test dataset if the SSE is greater than SST. If the value of R-Squared is 1, the model fits the data perfectly with a corresponding *MSE* = 0.

Here is a visual representation to understand the concepts of R-Squared in a better manner.

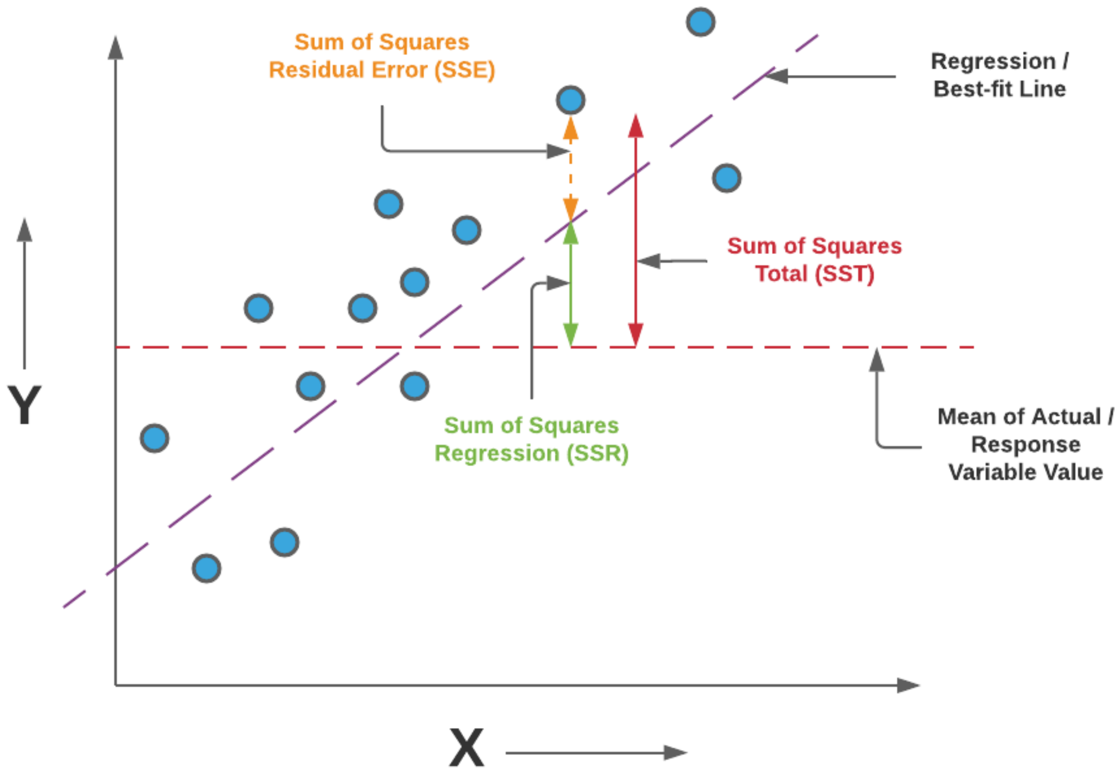

Fig 4. Diagrammatic representation for understanding R-Squared

$$R^2 = \frac{SSR}{SST} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

Four Assumptions of Linear Regression

1) A linear relationship exists between the dependent variable and the independent variable(s)
2) The residuals are Normally distributed
3) The residuals are identically distributed
4) The residuals are independently distributes