Linear Regression

Equation of Linear Regression

The equation of Linear Regression:

$$y = b_0 + b_1\, x \qquad \text{(Eq. 1)}$$

Where:
$y =$ Dependent Variable

$x =$ independent Variable

$b_0 =$ Intercept

$b_1 =$ Slope

We can redefine Eq. 1 as:

$$\hat{y} = b_0 + b_1\, x \qquad \text{(Eq. 2)}$$

Where:
$\hat{y} =$ Estimated Linear Regression Dependent Variable.

The best model will have the least error; therefore, the minimum Mean Squared Error is required. The MSE can be defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad \text{(Eq. 3)}$$

Where:

$n =$ Number of samples

The slope can be found using the average of x and y ($\overline{x}$ and $\overline{y}$). We can define the first equation in the algorithm to find the slope as:

$$b_1 = \frac{\sum_{i=1}^{s}(x_i - \overline{x}\,)(y_i - \overline{y})}{\sum_{i=1}^{s}(x_i - \overline{x}\,)^2} \qquad \text{(Eq. 4)}$$

The Intercept can therefore be calculated rearranging Eq. 2, solving to find b_0, however using the averages of x and y ($\overline{x}$ and $\overline{y}$):

$$b_0 = \overline{y} - b_1\overline{x} \qquad \text{(Eq. 5)}$$