Linear Regression

In statistics, linear regression is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables.

Consider a basic linear regression model:

$$y = mx + b$$

Another representation would be

$$y = \beta_0 + \beta_0 x_1$$

Weight Derivation

To derive the weights, or β in the linear regression formula, we calculate $\varepsilon^T \varepsilon$.

$$\hat{\beta}_{LS} = (X^T X)^{-1}(X^T y)$$

$\hat{\beta}_{LS}$ is used to predict β.

Bias Term

If you set all features in a given dataset to zero, what is the response value?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

$$y = \beta_0 + \beta_1(0) + \beta_2(0) + \ldots + \beta_p(0) = \beta_0$$

β_0 represents the bias term. Think about it as some sort of offset. Without a bias term, there will be no offset. If there exists no offset, $y = \beta_1 x_1$. To form a line of better fit, the bias term must be implemented:

$$y = \beta_0 x_0 + \beta_1 x_1$$

We do this so that the residuals are closer to the linear regression line and minimize ε.

Model Prediction

To represent these weights:

$$\hat{y} = x\beta$$

or

$$\hat{y} = X\beta$$

for all n records.

The algorithm goes as follows:

1. Generate weights
2. Multiple the weights $\times$ input

Error

For each entry $x \in X$,

1. predict $\hat{y}$.
2. Calculate residual/error: $y - \hat{y}$
3. Square residual/error and add it to the sum

As a result, the algorithm generates an SSE, or **Sum of Squared Error***.

$$SSE = (y - X\beta)^T (y - x\beta)$$
$$SSE = \sum_i (y_i - \hat{y}_i)^2$$
$$SSE = \sum_i (y_i - \hat{y}\beta)^2$$
$$SSE = \sum_i (y_i - \sum_k (x_{i,k}\beta_k))^2$$

where $x_{i,k}$ is the i^{th} row and k^{th} column and β_k is the k^{th} weight.

Additionally, the **Mean Squared Error** or MSE is computed by:

$$MSE = \frac{1}{n} \cdot SSE$$

which can be implemented in a loss function.