The idea behind the implementation of the Multi-Linear Model

Rohan Nair

June 2024

0.1 Standard Linear Regression (Closed Form)

For some Dataset that has one dependant variable and one target variable, our approach would be to solve it through Simple Linear Regression.

Let us denote our linear regression equation to be,

$$y = B_1 x + B_0$$

Where $B_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ and $B_0 = \bar{y} - B_1 \bar{x}$

0.2 Standard Linear Regression (Optimization)

Just like the regular closed form approach, we have one target variable and one dependant variable. This time we use an iterative approach to update the slope(B_1) and intercept(B_0) to minimize the loss function(L) after every iteration.

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

where $\hat{y} = B_1 * x_i + B_0$

This iterative approach is known as Gradient Descent. Now the question might arise, "How do we update the weight and intercept through the loss function after every iteration?". We use derivatives of the loss function.

$$w = w - \alpha \cdot \frac{\partial L}{\partial w}$$

$$b = b - \alpha \cdot \frac{\partial L}{\partial b}$$

where α symbolizes the learning rate (generally 0.001). It is really important to chose the correct value for the learning rate as choosing a learning rate that is too high or a value that is too low may cause issues. Through this we see that the loss function reduces every iteration and after a point reaches a minimum.

0.3 Multi-Linear Regression (Optimization)

The multilinear optimization approach follows a similar pattern to that of simple linear regression. The key distinction lies in handling multiple dependent variables, which are typically organized into a matrix $\mathbf{y}$. The loss function is computed using matrix operations, specifically through matrix subtraction.
The loss function L for multilinear regression is defined as:

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

Here, $\hat{y}_i$ represents the predicted value given by the multilinear model:

$$\hat{y}_i = B_1 \cdot \mathbf{x}_i + B_0$$

where B_1 is the vector of coefficients for the independent variables $\mathbf{x}_i$, and B_0 is the intercept term. In this context, $\mathbf{x}_i$ denotes one data entry (typically a row) from the matrix of independent variables.

This iterative approach, known as Gradient Descent, operates similarly in multilinear models as it does in simple linear regression. The primary difference lies in handling multiple dependent variables simultaneously. The goal remains to update the weights and intercept iteratively by minimizing the loss function.

In the context of multilinear regression, we update the weight vector $\mathbf{w}$ and intercept b using derivatives of the loss function:

$$w = w - \alpha \cdot \frac{\partial L}{\partial w}$$

$$b = b - \alpha \cdot \frac{\partial L}{\partial b}$$

where α represents the learning rate, typically set to a small value such as 0.001. Selecting an appropriate learning rate is crucial; too high or too low a value can lead to convergence issues. The objective is to observe a reduction in the loss function with each iteration until it reaches a minimum.

This iterative process ensures that the multilinear model converges towards optimal weights and intercept, thereby minimizing prediction errors across multiple dependent variables.