A Visual Representation of Linear Regression

To provide more intuition, let us draw a 2-D plot of the first feature x_1 and the target variable y of the dataset with all 50 records. We are using just one feature in this illustration because it is easier to visualize with a 2-D scatter plot (see Figure 19-2).

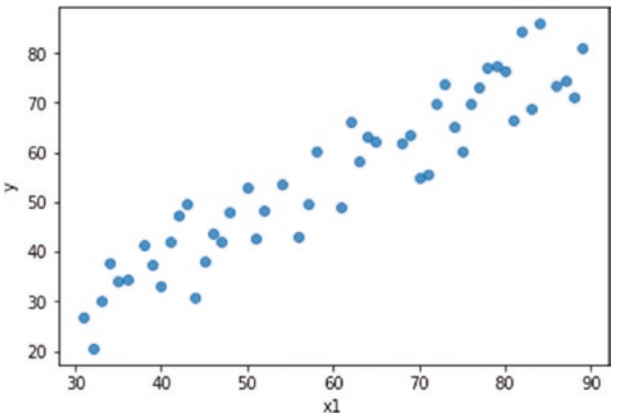

Figure 19-2. *Scatter plot of x_1 (on the x axis) and y (on the y axis)*

The goal of the linear model is to find a line that gives the best approximation or best fit to the data points. When found, this line will look like something in Figure 19-3. The line of best fit is known as the regression line.

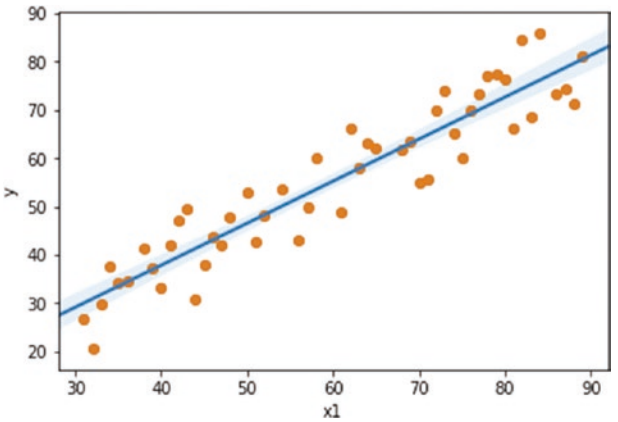

Figure 19-3. *Scatter plot of x_1 (on the x axis) and y (on the y axis) with regression line*

Finding the Regression Line – How Do We Optimize the Parameters of the Linear Model?

To find the regression line, we need to define the cost function, which is also called the loss function. Remember that the cost in machine learning is the error measure that the learning algorithm minimizes. We can also define the cost as the penalty when the model outputs an incorrect prediction.

In the case of the linear regression model, the cost function is defined as half the sum of the squared difference between the predicted value and the actual value. The linear regression cost function is called the ***squared error cost function*** and is written as

$$C(\theta)=\frac{1}{2}\Sigma(\hat{y}-y)^2$$

To put it more simply, the closer the approximate value of the target variable $\hat{y}$ is to the actual variable y, the lower our cost and the better our model.

Having defined the cost function, an optimization algorithm such as gradient descent is used to minimize the cost $C(\theta)$ by updating the weights of the linear regression model.

How Do We Interpret the Linear Regression Model?

In machine learning, the focus of linear regression differs slightly from traditional statistics. In statistics, the goal of a regression model is to understand the relationships between the features and targets by interpreting p-values, whereas in machine learning, the goal of the linear regression model is to predict the targets given new samples.

Figure 19-4 shows a regression model with a line of best fit that optimizes the squared difference between the data features and the targets. This difference is also called the residuals (shown as the purple vertical lines in Figure 19-4). What we care about in a linear regression model is to minimize the error between the predicted labels and the actual labels in the dataset.