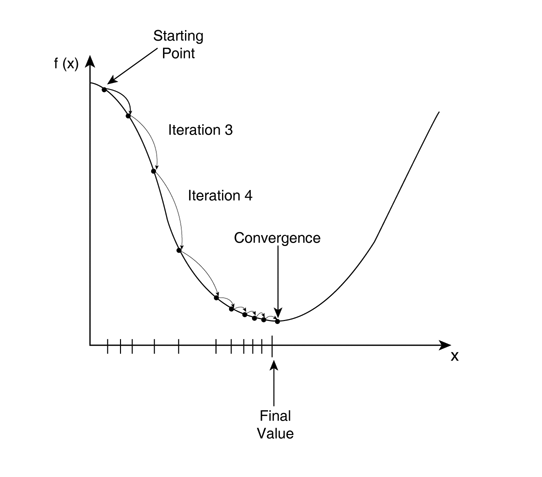
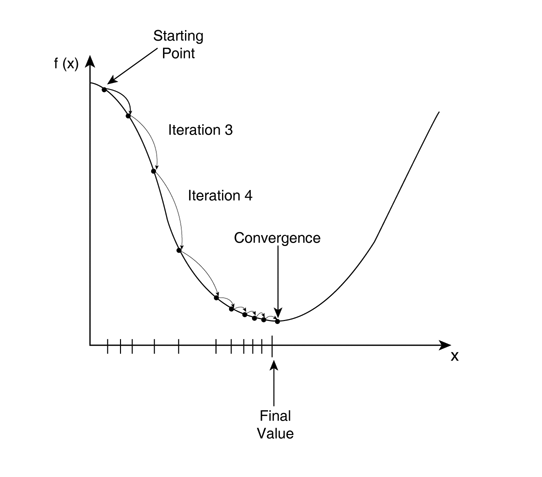
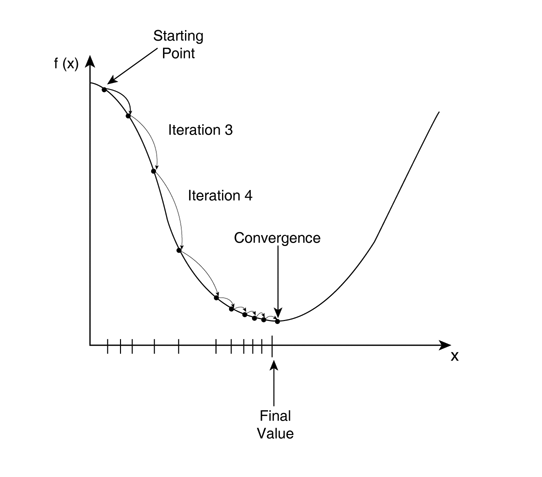
Simple Linear Regression is type of machine learning that aims to map a linear mathematical function through given x-inputs and y-inputs. The basic process of the Simple Linear Regression is to split the data (set of x-inputs and y-inputs) into either a training set or test set. Using the training set, the algorithm calculates the best fit equation. The program optimizes the parameters that will be used to create the model. Then, using the test set as a comparison to the predicted output derived from the equation, the accuracy is checked.

A linear mathematical function can be written as . In order to find the equation that best correlates with the data, the and values must be manipulated to find the exact value in which a minimum discrepancy between the predicted output and actual output occurs. The cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between x and y. The cost function is given as , where is the predicted output and is the actual output. To create the most accurate model of the data, the cost function must be minimized. This is done through a process called “gradient descent”.

Gradient descent is an efficient optimization algorithm that attempts to find a local or global minimum of a cost function. As the model iterates, it gradually converges towards a minimum where the parameters and are optimized to produced little to zero discrepancy. The gradient of the cost function at a specific point is crucial for the model to understand in which direction the parameters must be adjusted to get a lower discrepancy on the next iteration. The learning rate controls the size of the steps taken by each gradient, and must be specified beforehand. If this is too big, the model might miss the local minimum of the cost function. If it is too small, the model will take a long time to converge. This process can be illustrated with the diagram below.



The mathematical process of gradient descent is done by taking the gradients (partial derivatives with respect to and from the cost function. The cost function can be written as by replacing as . The learning rate is assigned, and is usually between . As shown below, the equations defined on the right side shows the process of updating the and parameters, and finding the optimizing values that allow minimizes the cost function.

The equation above is the standard method utilized to find the optimized parameters and in which the gradient of the cost function is minimized. In other words, the mean squared error (MSE) is being minimized by iterating through the model times, allowing it to update the parameters based on the gradient calculated in the previous iteration and the learning rate .

When the derivate of MSE, which is basically the cost function, is set as 0, this is denoted as least squares estimates. In Simple Linear Regression, the mathematics is much simpler and easier to do by hand compared to other types of linear regressions.

The predicted values of the optimized parameters and can be plugged back into the linear function to find the best fit line.