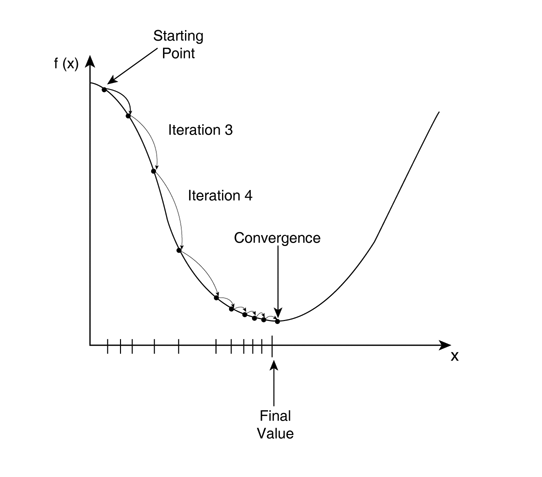
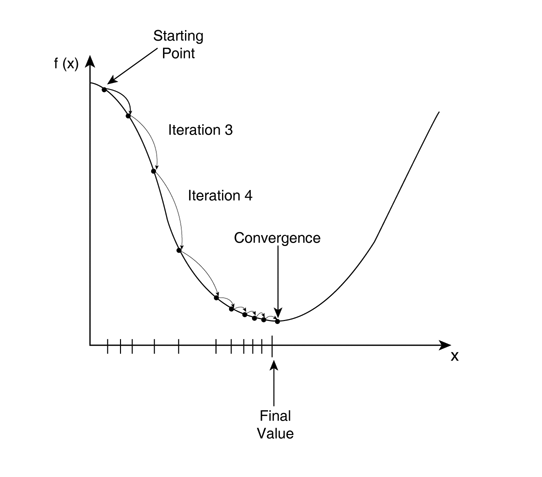
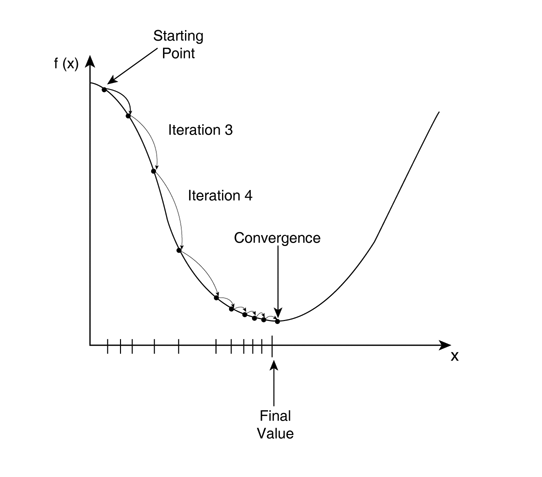
Multiple Linear Regression is a type of machine learning that works with multiple variables, unlike the Simple Linear Regression. The basic process of the Multiple Linear Regression is to split the data (set of x-inputs for each variable and y-inputs) into either a training set or test set. Using the 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.

The linear equation can be defined as . In order to find the equation that best correlates with the data, the values must be manipulated to find the exact value in which a minimum discrepancy between the predicted output and actual output occurs. However, before doing so, a process of data preparation must take place where the x-inputs for each variables must be normalized (adjusting values to a common scale, without distorting differences in the range of values) using formula , so that the estimate y-output is not influenced or biased by certain variables. 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 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 from the cost function. The learning rate is assigned, and is usually a value close to 0. As shown below, the equations defined on the right side shows the process of updating the parameters, and finding the optimizing values that allow minimizes the cost function.

After iterating through the algorithm about depending on the size of the data to optimize all parameters, the cost function will start to converge. This means that the parameters have reached an optimized value, where it is most ideal to predict the y-outputs. By substituting the optimized parameters back into the linear equation or by doing a dot product with the x-inputs and variables, the prediction y-outputs can be calculated.

The accuracy of the predicted y-outputs to the actual y-outputs can be measured using the RMSE.