Logistic Regression is another type of machine learning algorithm very much similar to the Simple Linear Regression in its process. However, the main difference is that the Logistic Regression is used to classify predicted values into its appropriate categories depending on their probability score and decision boundaries.

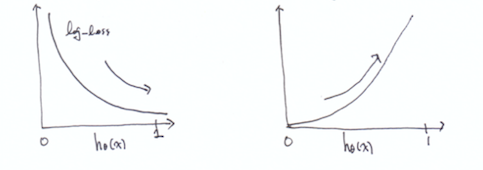
The Logistic Regression also depends on a linear equation , where are the weights and are the x-inputs used to predict the y value. This is also called the weighted sum, where it is calculated by taking the dot product of and . However, this linear equation predicts an output that is within the range of . In order to squeeze the predicted outputs into the range of which allows the predicted output to be represented as a probability, the sigmoid function is utilized.



Decision Boundary

The equation shown above is the sigmoid function, where the predicted outputs have been remodeled into appropriate probability scores. This allows the predicted outputs to be classified into different categories depending on the decision boundary, which in this particular example is set as 0.5. Therefore, if it would be classified into class = 1, and if it would be classified into class = 0.

In order to find the most optimal predicted output that best correlates with the actual output, the cost function must be used. The cost function is a measure of the discrepancy between the actual output and predicted output. However, in Logistic Regression, the optimal predicted output cannot be calculated by immediately using gradient descent. This is because the prediction function for the Logistic Regression is not a linear equation, thus making it difficult to find right the global minima. Instead, logarithmic loss is utilized first to divide the cost function into separate equations for each classification: first for class = 1, and second for class = 0. This creates two equations that are monotonic (always increasing or decreasing), making it easy to use gradient descent. Afterwards, the cost function can be rewritten as one equation.



1. if class = 1
2. if class = 0

Using the cost function above, the gradients of each parameters can be found by taking partial derivates of the cost function with respect to (. The gradient will be used to update the values as the cost function converges to an optimal value.

After finding the optimal values, it is plugged back into the weighted sum to be used in the sigmoid function. This will then produce predicted y values that best represents the actual y values.