Gradient Descent Iteration

Let's start with values of 0.0 for both coefficients.

$$B0 = 0.0$$

$$B1 = 0.0$$

$$y = 0.0 + 0.0 * x$$

We can calculate the error for a prediction as follows:

$$error = p(i) - y(i)$$

Where $p(i)$ is the prediction for the i'th instance in our dataset and $y(i)$ is the i'th output variable for the instance in the dataset.

We can now calculate he predicted value for y using our starting point coefficients for the first training instance:

$$x=1, y=1$$

$$p(i) = 0.0 + 0.0 * 1$$

$$p(i) = 0$$

Using the predicted output, we can calculate our error:

$$error = 0 - 1$$

$$error = -1$$

We can now use this error in our equation for gradient descent to update the weights. We will start with updating the intercept first, because it is easier.

We can say that B0 is accountable for all of the error. This is to say that updating the weight will use just the error as the gradient. We can calculate the update for the B0 coefficient as follows:

$$B0(t+1) = B0(t) - alpha * error$$

Where B0(t+1) is the updated version of the coefficient we will use on the next training instance, B0(t) is the current value for B0 alpha is our learning rate and error is the error we calculate for the training instance. Let's use a small learning rate of 0.01 and plug the values into the equation to work out what the new and slightly optimized value of B0 will be:

$$B0(t+1) = 0.0 - 0.01 * -1.0$$

$$B0(t+1) = 0.01$$

Now, let's look at updating the value for B1. We use the same equation with one small change. The error is filtered by the input that caused it. We can update B1 using the equation:

$$B1(t+1) = B1(t) - alpha * error * x$$

Where B1(t+1) is the update coefficient, B1(t) is the current version of the coefficient, alpha is the same learning rate described above, error is the same error calculated above and x is the input value.

We can plug in our numbers into the equation and calculate the updated value for B1:

$$B1(t+1) = 0.0 - 0.01 * -1 * 1$$

$$B1(t+1) = 0.01$$

We have just finished the first iteration of gradient descent and we have updated our weights to be B0=0.01 and B1=0.01. This process must be repeated for the remaining 4 instances from our dataset.

One pass through the training dataset is called an epoch.