The Gradient in Gradient Descent

 [(https://github.com/learn-co-curriculum/dsc-the-gradient-in-gradient-descent)](https://github.com/learn-co-curriculum/dsc-the-gradient-in-gradient-descent)
[(https://github.com/learn-co-curriculum/dsc-the-gradient-in-gradient-descent/issues/new/choose)](https://github.com/learn-co-curriculum/dsc-the-gradient-in-gradient-descent/issues/new/choose)

Introduction

As you know, we entered our discussion of derivatives to determine the size and direction of a step with which to move along a cost curve. We first used a derivative in a single variable function to see how the output of our cost curve changed with respect to change a change in one of our regression line's variables. Then we learned about partial derivatives to see how a *three-dimensional cost curve* responded to a change in the regression line.

However, we have not yet explicitly showed how partial derivatives apply to gradient descent.

Well, that's what we hope to show in this lesson: explain how we can use partial derivatives to find the path to minimize our cost function, and thus find our "best fit" regression line.

Objectives

You will be able to:

- Define a gradient in relation to gradient descent

What is the gradient?

Now gradient descent literally means that we are taking the shortest path to *descend* towards our minimum. However, it is somewhat easier to understand gradient *ascent* than descent, and the two are quite related, so that's where we'll begin. Gradient ascent, as you could guess, simply means that we want to move in the direction of steepest ascent.

Now moving in the direction of greatest ascent for a function $f(x, y)$, means that our next step is a step some distance in the x direction and some distance in the y direction which is the steepest upward at that point.

Note how this is a different task from what we have previously worked on for multivariable functions. So far, we have used partial derivatives to calculate the **gain** from moving directly in either the x direction or the y direction.

> Here, in finding gradient ascent, our task is not to calculate the gain from a move in either the x or y direction. Instead, our task is to **find some combination of a change in x,y that brings the largest change in output**.

So if you look at the path our climbers are taking in the picture above, *that* is the direction of gradient ascent. If they tilt their path to the right or left, they will no longer be moving along the steepest upward path.

The direction of the greatest rate of increase of a function is called the gradient. We denote the gradient with the nabla, which comes from the Greek word for harp, which is kind of what it looks like: ∇. So we can denote the gradient of a function, $f(x, y)$, with $\nabla f(x, y)$.

Calculating the gradient

Now how do we find the direction for the greatest rate of increase? We use partial derivatives. Here's why.

As we know, the partial derivative $\frac{df}{dx}$ calculates the change in output from moving a little bit in the x direction, and the partial derivative $\frac{df}{dy}$ calculates the change in output from moving in the y direction. Because with gradient ascent our goal is to make a nudge in x, y that produces the greatest change in output, if $\frac{df}{dy} > \frac{df}{dx}$, we should make that move more in the y direction than the x direction, and vice versa. That is, we want to get the biggest bang for our buck.


Let's relate this again to mountain climbers. Imagine the vertical edge on the left is our y-axis and the horizontal edge is on the bottom is our x-axis. For the climber in the yellow jacket, imagine his step size is three feet. A step straight along the y-axis will move him further upwards than a step along the x-axis. So in taking that step, he should direct himself more towards the y-axis than the x-axis. That will produce a bigger increase per step size.

In fact, the direction of greatest ascent for a function, $\nabla f(x, y)$, is the direction which is a proportion of $\frac{df}{dy}$ steps in the y direction and $\frac{df}{dx}$ in the x direction. So, for example, if $\frac{df}{dy} = 5$ and $\frac{df}{dx} = 1$, the direction of gradient ascent is five times more in the y direction than the x direction. And this seems to be the path, more or less that our climbers are taking - some combination of x and y, but tilted more towards the y direction.

Applying Gradient Descent

Now that we have a better understanding of a gradient, let's apply our understanding to a multivariable function. Here is a plot of a function:

$$f(x, y) = 2x + 3y$$
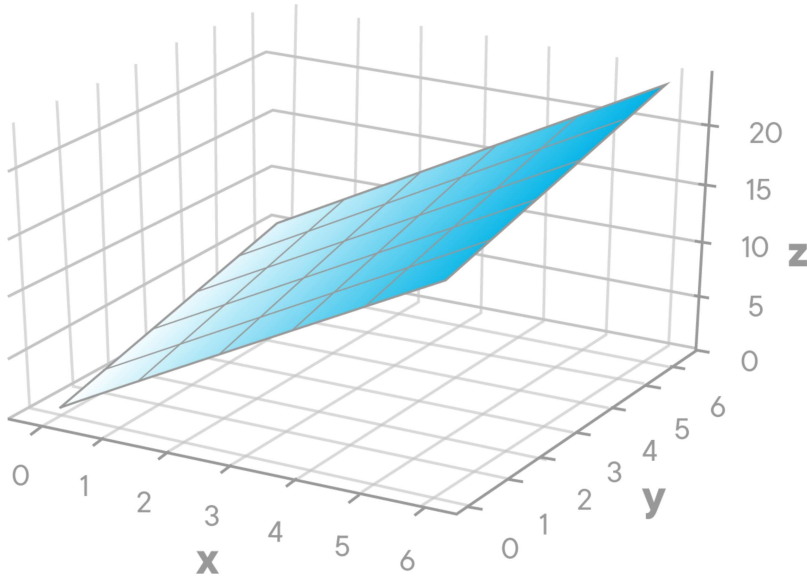
Imagine being at the bottom left of the graph at the point $x = 1, y = 1$. What would be the direction of steepest ascent? It seems, just sizing it up visually, that we should move both in the positive y direction and the positive x direction. Looking more carefully, it seems we should move **more** in the y direction than the x direction. Let's see what our technique of taking the partial derivative indicates.

The gradient of the function $f(x, y)$, that is $\nabla f(x, y) = 2x + 3y$ is the following:

$\frac{df}{dx}(2x + 3y) = 2$ and $\frac{df}{dy}(2x + 3y) = 3$.

So what this tells us is to move in the direction of greatest ascent for the function $f(x, y) = 2x + 3y$, is to move up three and to the right two. So we would expect our path of greatest ascent to look like the following.
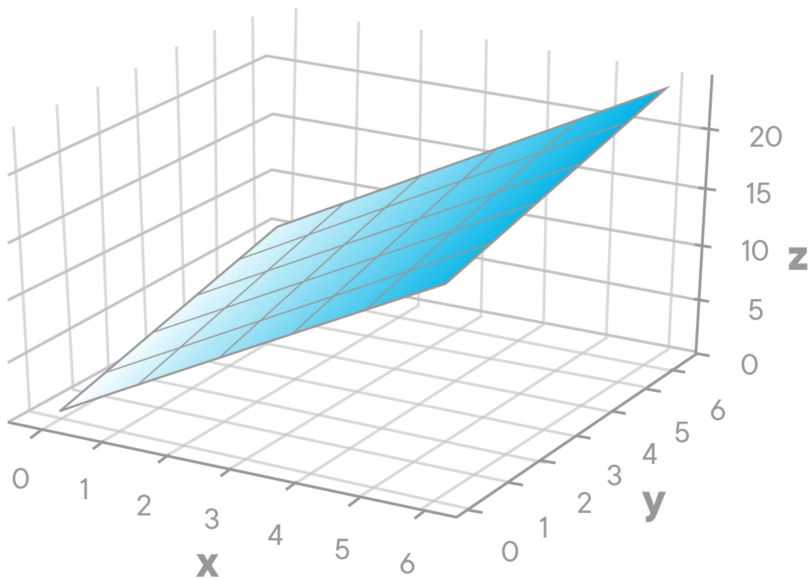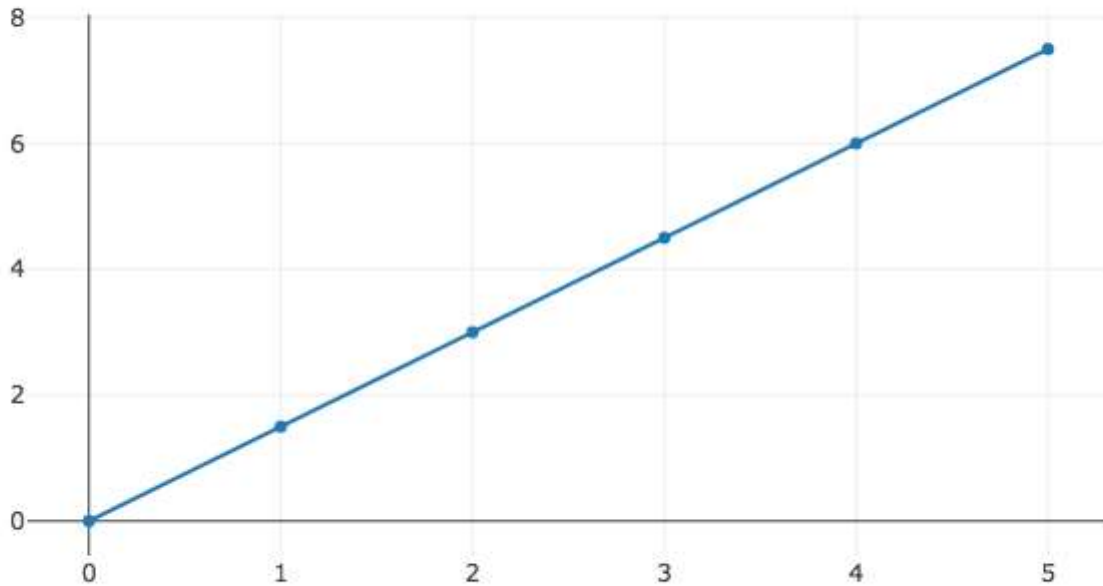
So this path maps up well to what we see visually. That is the idea behind gradient descent. The gradient is the partial derivative with respect to each type of variable of a multivariable function, in this case x and y. And the importance of the gradient is that its direction is the direction of steepest ascent. The negative gradient, that is the negative of each of the partial derivatives, is the direction of steepest descent. So our direction of gradient descent for the graph above is $x = -2, y = -3$. And looking at the two graphs above, it seems that the steepest downward direction is just the opposite of the steepest upward direction. We get that by mathematically by simply taking the multiplying our partial derivatives by negative one.

Summary

In this lesson, you saw how to use gradient descent to find the direction of steepest descent. You saw that the direction of steepest descent is generally some combination of a change in your variables to produce the greatest negative rate of change.

You first how saw how to calculate the gradient **ascent**, or the gradient ∇, by calculating the partial derivative of a function with respect to the variables of the function. So $\nabla f(x, y) = \frac{\delta f}{\delta y}, \frac{\delta f}{\delta x}$. This means that to take the path of greatest ascent, you should move $\frac{\delta f}{\delta y}$ divided by $\frac{\delta f}{\delta x}$. So for example, when $\frac{\delta f}{\delta y} f(x, y) = 3$, and $\frac{\delta f}{\delta x} f(x, y) = 2$, you traveled in line with a slope of 3/2.

For gradient descent, that is to find the direction of greatest decrease, you simply reverse the direction of your partial derivatives and move in $-\frac{\delta f}{\delta y}, -\frac{\delta f}{\delta x}$.

How do you feel about this lesson?

Have specific feedback?

Tell us here! (https://github.com/learn-co-curriculum/dsc-the-gradient-in-gradient-descent/issues/new/choose)