As Luis said, with gradient descent, we take multiple small steps towards our goal. In this case, we want to change the weights in steps that reduce the error. Continuing the analogy, the error is our mountain and we want to get to the bottom. Since the fastest way down a mountain is in the steepest direction, the steps taken should be in the direction that minimizes the error the most. We can find this direction by calculating the *gradient* of the squared error.

Gradient is another term for rate of change or slope. If you need to brush up on this concept, check out Khan Academy's great lectures on the topic.

To calculate a rate of change, we turn to calculus, specifically derivatives. A derivative of a function $f(x)$ gives you another function $f'(x)$ that returns the slope of $f(x)$ at point x. For example, consider $f(x) = x^2$. The derivative of x^2 is $f'(x) = 2x$. So, at $x = 2$, the slope is $f'(2) = 4$. Plotting this out, it looks like: