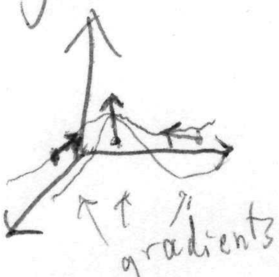
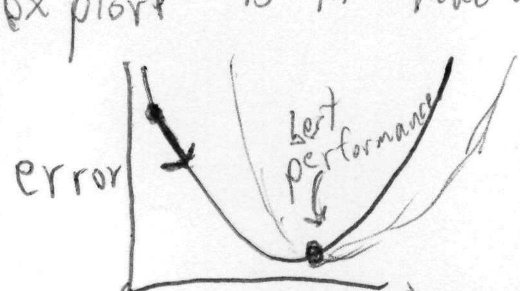
Week 6 | Module 2: Multivariable Calculus

This module is short, because it's mostly about one thing: finding a gradient, which is a derivative when we're dealing with more than one dimension. If we think of a function in 3 dimensions, $f(x,y)$, as a kind of "hilly landscape" the gradient at any point is a vector pointing "uphill."



There are some graphical applications to the gradient, because the way light plays on a surface depends on its exact orientation. But a main application we'll explore is in machine learning. There we are trying

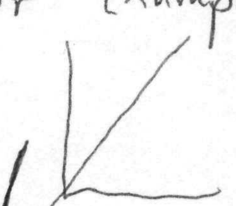


to adjust a variety of parameters to get a system to perform better; if we think of the error as a function of the function parameters, we want to minimize it. But if we only know what the function looks like locally, maybe we can't jump directly to the settings with minimum error. Instead, we go "downhill" on the error surface — moving in a direction directly opposed to the gradient. This is called "gradient descent" and besides being its own method, it's also how neural networks work — as we'll see soon.

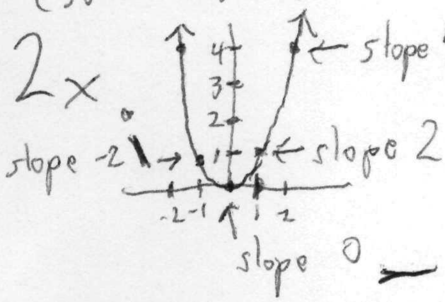
Brief review of calculus

Before going to multivariable we can first review some calculus concepts. (This won't be exhaustive, but it should be good enough for our course.)

The derivative of a function $f(x)$, sometimes denoted $f'(x)$, is a function that gives the slope of a line tangent to $f(x)$ at any point x. For example, if $f(x) = x$, the slope everywhere is 1. (so $f'(x) = 1$)



If $f(x) = x^2$, $f'(x) = 2x$.



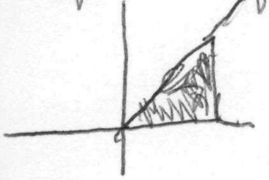
In calculus, you learn a variety of rules for figuring a derivative of a function the most powerful of which (or at least commonly used) is the power rule, $f'(x) = nx^{n-1}$ when $f(x) = x^n$.

Some other useful rules:

- Differentiation is "linear" so the derivative of $f(x) + g(x) = f'(x) + g'(x)$, and the derivative of $cf(x)$ is $cf'(x)$. So if $f(x) = 5x^2 + 3x - 2$, $f'(x) = 10x + 3$.
- The chain rule says if $h(x) = f(g(x))$, $h'(x) = f'(g(x))g'(x)$ so treat $g(x)$ as just another variable, then mult by its derivative.

If $f(x) = \sqrt{1+x^2}$ $f'(x) = \frac{1}{2}(1+x^2)^{-1/2} \cdot 2x = \frac{x}{\sqrt{1+x^2}}$.
$\uparrow (1+x^2)^{1/2}$

In calculus, you also learn to <u>integrate</u>, or perform the reverse of differentiation. Integrating a function gives the area under it. But we won't say much more about this, except in passing when talking about statistics.



$$\int x \, dx = \frac{1}{2}x^2$$

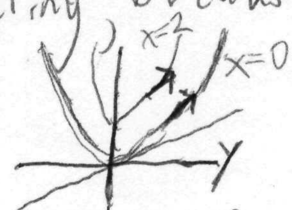
Derivatives with respect to particular variables

So, suppose our function is of two variables, $f(x, y)$.
How do we get a derivative now?
One thing we can do is find the slope along a particular axis, like the y-axis, keeping all other variables fixed.
This is called a partial derivative, denoted $\frac{\partial}{\partial y}$ if we are finding the derivative with respect to y, for example.

The way to generalize our calculus rules to this situation is fairly easy: treat all the other variables as <u>constants</u>. For any particular slice where we vary just y (or whatever), they each effectively are some constant.

So if $f(x,y) = x^2 + y^2$, $\frac{\partial f}{\partial y}(x, y) = 2y$.
The slope along the y-axis here isn't really affected by x.

On the other hand, if $f(x, y) = x^2 + xy + y^2$, then $\frac{\partial f}{\partial y}(x, y) = x + 2y$. The slope depends on x, but it is still okay to think of x as a constant while differentiating because it doesn't change when we change y.



Another way to think of this derivative is a measure of how much impact a particular variable has on the function f. $\frac{\partial f}{\partial y}$ is what you get when you take the limit $\lim_{\Delta y \to 0} \frac{f(x, y+\Delta y) - f(x,y)}{\Delta y} = \lim_{\Delta y \to 0} \frac{\Delta f}{\Delta y}$, a ratio showing how much f changes when you change y. This idea will inform how we think about ML methods like neural networks, where we'll want to know which parameters are most to "blame" for a function's mistakes.

Second derivatives

Once we've taken a derivative, we can do it again. But with multiple variables, we've got options for which derivatives to take.

$f_{xx} = \frac{\partial^2 f}{\partial x^2}$ = 2nd derivative in x direction

$f_{xy} = f_{yx} = \frac{\partial^2 f}{\partial y \partial x}$ = "mixed" derivative in x direction then y or vice versa (the results are the same)

and so on. To find these, we just differentiate, then do it again. $f(x,y) = xy \Rightarrow f_{xx} = \frac{\partial}{\partial x} y = 0$ $f_{xy} = \frac{\partial}{\partial y} y = 1$

You probably recall that a second derivative for the same variable represents how much the "change is changing". f_{xx} represents how much $\frac{\partial f}{\partial x}$ changes in response to changing x. So, the mixed derivatives are similar, but indicate how much one variable affects a different variable's rate of change. $\frac{\partial^2 f}{\partial y x} \; xy = 1$ tells us that changing y by 1, changes the slope in the x direction by 1.

Normal vectors & tangent planes

Once we have the slope of the function in the x direction, and another in the y direction, we could figure out an equation of a plane that passes through a given point and has the relevant slopes. This plane is called the "tangent plane." Its equation at point (x_0, y_0, z_0) is $\leftarrow = f(x_0, y_0)$

$_0$ means value at our point

$$z - z_0 = \left(\frac{\partial f}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial f}{\partial y}\right)_0 (y - y_0)$$

where $\left(\frac{\partial f}{\partial x}\right)_0$ and $\left(\frac{\partial f}{\partial y}\right)_0$ are the values of those partial derivatives at the point. (The equation basically states that as we move away from (x_0, y_0, z_0), we need to adjust z according to $\Delta x \cdot \frac{\partial f}{\partial x} + \Delta y \cdot \frac{\partial f}{\partial y}$.)