

Jacobian

Previously we saw we can differentiate functions of multiple variables and it isn't much harder than the univariate calculus at the start

Here, we going to introduce the Jacobian which brings in some of the ideas of LA. to build these partial derivatives into something particularly useful

The concept of the Jacobian can be applied to variety of different problems

But in the context of getting started with optimization of ML, there's a particular scenario that comes up a lot, which is the Jacobian of a single function of many variables.

(2)

In short, if you have a function with many variables : $\Rightarrow f(x_1, x_2, \dots)$

$$f(x_1, x_2, x_3, \dots)$$

Then the Jacobian is simply a vector, where each entry is the partial derivative of f w.r.t each one of those variables in turn.

$$J = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}, \dots \right]$$

By convention, we write this as a row vector, rather than a column vector for reasons that will become clear later

Let's start by looking at single functions, to see how straightforward building a Jacobian can be.

Consider the function:

(3)

$$f(x,y,z) = x^2y + 3z$$

To build a Jacobian, we just find each of the partial derivatives of the function one by one

wrt x

$$\frac{\partial f}{\partial x} = 2xy \quad \left[\text{Everything else is constants} \right]$$

wrt y

$$\frac{\partial f}{\partial y} = x^2 \quad \left[J = (2xy, x^2, 3) \text{ } \textcircled{A} \right]$$

wrt z

$$\frac{\partial f}{\partial z} = 3.$$

Bringing all of it together \textcircled{A}, But what does this tell us?

→ we have an algebraic expression for a vector, which when we give it a specific x,y,z coordinate, will

Q

return a vector pointing in direction
of steepest slope of this function.

The vector for this particular function has
a constant contribution in the Z direction,
which does not depend on the location
selected.

Eg at point $(0,0)$, we can see that our
Jacobian will just be

$$J(0,0) = [0, 0, 3]$$

∴ Our Jacobian is vector of length $\sqrt{3}$,
pointing directly in the Z direction.

Some of the numerical methods that will be
discussed later in course, will require us to
calculate the Jacobian in 100's of dimensions.

However even for the 3D example we just
solved, graphical representation is quite
difficult.

(5)

So we will revert to 2 dimensions, to see what's going on.

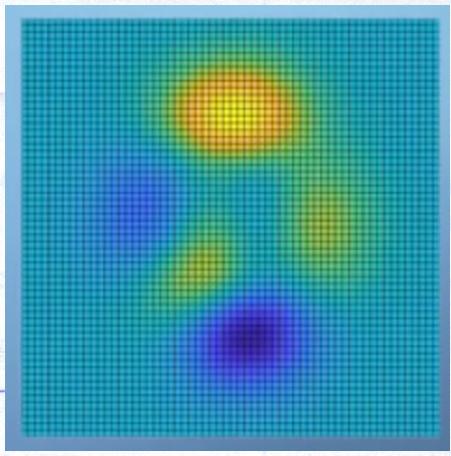
But to keep things interesting, we will lose a lot of ~~per~~ particularly complicated, attractive function.

Here's the equation we're going to plot, which is shown briefly, just so we can see, even though it looks huge, we will agree that just with tools we covered, we really could calculate the partial derivative, but we would not really learn anything new from grinding through it.

Instead we will be shown the results graphically:

$$z(x,y) = 3(1-x)^2 e^{-x^2-(y+1)^2} - 10\left(\frac{x}{5} - x^3 - y^5\right) e^{-x^2-y^2} - \frac{1}{5} e^{-(x+1)^2-y^2}$$

[See ----]



⑥

Because of the nature of this particular function, we know that nothing interesting happens outside of the region shown [so we can forget about it for now]. Although the blocks are fairly clear, our intuition about the gradient is a bit lacking in this format, so let's briefly make things ~~3D~~^{2D}, where the values of 2 are also represented by the height of 2 ~~points~~^{squares}.

So, as indicated by start of ~~the~~^{second} picture, the Jacobian is simply a vector that we can calculate for each location of this plot, that points in direction of steepest uphill slope.

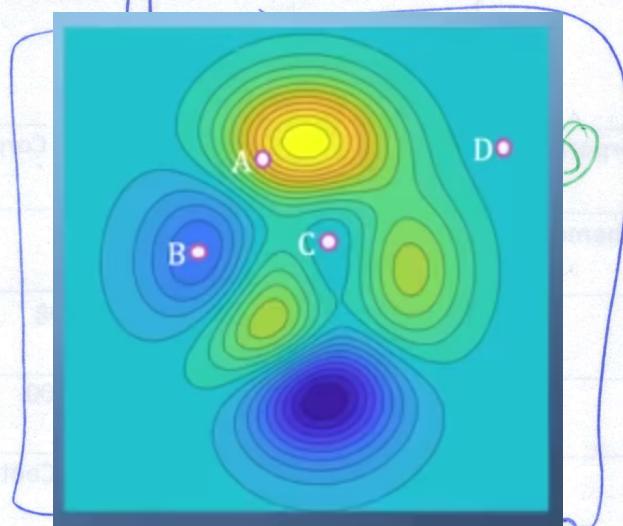
Further more, the steeper the slope the greater the magnitude of the Jacobian at that point.

Hold the image of this 3D space in your mind, as we now go back to 2D, and rather than showing all of the grid points, we use to plot the graph. Let's instead convert to Contour plot, we're just like a map of actual mountains, we will draw a line along the regions of the same height, which means the same value of Z .

This removes all the clutter from plot, which is useful for the final step which will be adding lots of Jacobian vectors on top of our Contour plot.

However, before doing it, let's look at a few points and see if my intuition is correct, by guessing

which one will have the Jacobian with the largest magnitude?



Overlays the Jacobian vector field, we can see that they are clearly all pointing up hill; away from the low value regions, and towards the high bright regions.

Also we see that where the contour lines are tightly packed this is where we find our largest Jacobian vectors such as at Point A.

Whereas at the peak of mountain and bottom of valleys, or even out on the side of flat planes, our gradients are therefore our Jacobians are small.

Hope, that's over 2D example will
give us confidence to treat maths,
when we come up against much
higher dimensions and problems later in
life.

⑨