Module 2

The Hessian

Can be thought of as an ample extension of Jacobian vector

- For the Jacobian we collected together all of the First order derivatives of function into a vector

- Now we going to collect all of the second order derivatives together into a matrix

⇒ function of n variable, will look like this:

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

This is one of these scenarios where abbreviated notation style comes in really handy

$$H = \begin{bmatrix} \partial_{x_1x_1}f & \partial_{x_1x_2}f & \cdots & \partial_{x_1x_n}f \\ \partial_{x_2x_1}f & \partial_{x_2x_2}f & \cdots & \partial_{x_2x_n}f \\ \partial_{x_nx_1}f & \cdots & \cdots & \partial_{x_nx_n}f \end{bmatrix}$$

In previous lesson, we saw how we can keep differentiating a function, using the same method to find higher and higher derivatives.

Similarly for a partial derivative, if you want to find a second derivative w.r.t. x_1 then x_2, it's as simple as just differentiating w.r.t. x_1 assuming all the other variables are constant, THEN differentiating w.r.t. x_2 again assuming all other variables are constant.

As we can see, from this general form, the Hessian matrix will be a $n \times n$ square matrix, where n is number of variables in our function f

Quick example:

Easier to find the Jacobian first, then and then differentiate its terms again to find the Hessian.

$$f(x,y,z) = x^2yz$$

Lets first build the Jacobian for this wrt x, y, z

$$J = \left[2xyz,\ x^2z,\ x^2y\right]$$

Using this, we can then differentiate again w.r.t. each of variables, which will then give us, our Hessian matrix.

$$\begin{array}{ll} & \begin{array}{ccc} \text{wrt } x & \text{wrt } y & \text{wrt } z \end{array} \\ \text{wrt. } x \text{ (again)} \quad H = & \begin{bmatrix} 2yz & 2xz & 2xy \\ 2xz & 0 & x^2 \\ 2xy & x^2 & 0 \end{bmatrix} \end{array}$$

What we notice here is that the Hessian matrix is symmetrical accross the ~~a~~ leading diagonal.

∴ once I have worked out the top right region, I could just ~~have~~ written these directly in ~~for~~ bottom left region

This will always be true if the function is continuous, meaning it has no sudden step changes

We can now pass our Hessian an x,y coordinate, and it will return a matrix of numbers, which hopefully will tell us something about that point in space

In order to visualize this, we going to have to 2Dimensions again

Consider the function,

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

Calculating the Jacobian and Hessian

$$J = [2x, 2y] \quad H = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

are fairly straightforward,
and visualize how it will look in your head.

However if you did not know ~~which~~ what function we dealing with, and calculated the value of the Jacobian at point $(0,0)$, you would have seen that the gradient vector was also zero.

But how could you know whether the thing was a maximum or minimum at that point. You can go and check some other point and see if it was above or below. But this is not very robust.

~~Instead~~, we can look at the Hessian, which in a simple case, no longer even a function of ~~x~~ x or y.

Its determinant is clearly just 2x2, + 0x0 = 4

$$|H| = 4$$

The power of the Hessian is firstly, that ~~its~~ if its determinant is positive, we know we dealing with either a maximum or minimum

Secondly, we then just look at first term, which is sitting at the top left hand corner of the Hessian, ~~if~~

$\Rightarrow$ If this guy is also positive, we know we have a minimum, as is in this case

Whereas if its negative, we have a maximum

Lastly slightly modifying our function to include a minus sign, and recalculating our Jacobian and Hessian, and Hessian determinant, we see third interesting case:

$$f(x,y) = x^2 - y^2$$

$$J = \begin{bmatrix} 2x, -2y \end{bmatrix}$$

$$H = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \quad |H| = -4$$

This time, our Hessian determinant are negative, so we know that we not dealing with a maximum or minimum, but clearly at the point 0,0, the gradient is flat. What happening? Slopes coming down towards one direction, and up towards any other.
as per one mathematician

– we call this feature a saddle point.

– which can cause a lot of confusion when searching for peak...