Module 3

Multivariate Chain Rule Extra

The last session we defined an expression for the multivariate chain rule

Here, we going to start by picking up one more little detail, and follow this up, by adding another link to our chain

If we have function f of x, where x is a vector, and in which each term in x depends on t, which we can compactly write as

$$f(x(t))$$

We also have our compact form of our multivariate chain rule to go with it:

$$\frac{df}{dt} = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt}$$

②

From last time, we notice that our vector of partial derivative, df by dx is just the same as the Jacobian vector which we saw in last module, except we wrote it as a column instead of row vector.

$$\frac{df}{dx} = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \partial f/\partial x_3 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix} = (Jf)^T$$

So from our knowledge of LA, we can say that df by dx is the transpose of the Jacobian of f

The last thing to realize is that taking the dot product of two column vectors is the same operation as multiplying the row vector by column vector. (3)

So finally we can see that our old friend, the Jacobian offers us perhaps the most convenient representation of the multivariate chain rule

$$\frac{df}{dt} = J_f \frac{dx}{dt}$$

Next we going to see that the chain rule still works for more than 2 links.

Start off: we will work through a quick univariant example, where we going to add in another function separating f and t.

$f(x) = 5x$

$x(y) = 1 - y$

$y(t) = t^2$

So we have 3 functions, and we separating f from t, by extra step.

Offcourse we can just sub in each step into each other and find an expression for

$f(t) = 5(1-t^2) = 5 - 5t^2$

whats x
whats y } in one go
whats t

$\frac{df}{dt} = -10t$

or we can apply a 2 step chainrule.

$$\left.\begin{aligned} f(x) &= 5x \\ x(u) &= 1-u \\ u(t) &= t^2 \end{aligned}\right\} \quad \frac{df}{dt} = \frac{df}{dx}\frac{dx}{du}\frac{du}{dt}$$

subbing in for each of our terms

$$\frac{df}{dt} = (5)(-1)(2t)$$

$$= -10t \text{ (same answer)}$$

We can see the approach works for chain of arbitrary functions.

And we can extend it out to as many intermediary functions between f and t as we like.

But what about the multivariate case?

The chain rule does work here too

But we have to pay attention to a few extra details

Lets start by considering:

$$\mathbf{f}(\mathbf{x}(\mathbf{u}(t))) \quad \left(\text{'read out'}: f \text{ of } x \text{ of } u \text{ of } t\right)$$

where function f takes takes vector **x** as input, but this time **x** is a vector valued function, which also takes vector **u** as its input:

$$f(\mathbf{x}) = f(x_1, x_2) \qquad (A)$$

$$\mathbf{x}(\mathbf{u}) = \begin{bmatrix} x_1(u_1, u_2) \\ x_2(u_1, u_2) \end{bmatrix} \qquad (B)$$

Remember, from last session, the bold symbols indicate vectors

and **u** is again a vector valued function, and it takes the scalar t as its input

$$\mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} \qquad (C)$$

Ultimately we still relating the scalar input t, to a scalar output f,
But through two intermediary vector valued functions x and y.
If once again we would like to know the derivative of f w.r.t t, we can essentially ~~the~~ write the same expression as the univariant case, except now several of our terms are in bold

A
B
C

$$\frac{df}{dt} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial y} \frac{dy}{dt}$$

we have already seen that differentiating the scalar valued f w.r.t its input vector x, gives us the Jacobian row vector

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

we've also seen differentiating, a vector valued function u wrt. scalar variable t gives us a column vector of derivatives

$$? \quad \begin{bmatrix} \frac{du_1}{dt} \\ \frac{du_2}{dt} \end{bmatrix}$$

But what about the middle term (?) dx by du

well for the function x, we need to find the derivative of each of the two output variables wrt each of the input variables.

So we end up with 4 terms in total, which as we saw in last module, can be conveniently arranged as a matrix

$$? \text{ (above)} \quad \begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{bmatrix}$$

We still refer to this object as the Jacobian.

So we can now say that the derivative of f wrt vector t is the product of the Jacobian of f, with the Jacobian of x, ~~with~~ and the derivative vector of u

Notice that the dimensions of ~~these~~ those vectors and matrices as shown here, are such that this operation are possible, and that they will return a scalar just as we expected.

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u}\frac{du}{dt} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{bmatrix} \begin{bmatrix} \frac{du_1}{dt} \\ \frac{du_2}{dt} \end{bmatrix}$$

$$(1 \times 1) = (1 \times 2)(2 \times 2)(2 \times 1)$$

scalars

We now see the various threads of LA, in multivariant Calculus weave together.