Explanatory Notes for 6.390

Shaunticlair Ruiz (Current TA)

Fall 2022

X.8 Vector derivative: a scalar input, vector output

Now, we want to try the flipped version: we swap our vector and our scalar.

$$\frac{\partial(\text{Vector})}{\partial(\text{Scalar})} = \frac{\partial w}{\partial s} \tag{1}$$

We'll take s to be our scalar, and w to be our vector. So, our input is a **scalar**, and our output is a **vector**.

Note that we're using vector w instead of v this time: this will be helpful for our vector/vector derivative: we can use both.

$$\Delta s \longrightarrow \boxed{\text{f}} \longrightarrow \Delta w \tag{2}$$

Written explicitly, like before:

$$\Delta s \longrightarrow \overbrace{\begin{bmatrix} \Delta w_1 \\ \Delta w_2 \\ \vdots \\ \Delta w_n \end{bmatrix}}^{\Delta w} \tag{3}$$

We have 1 **input**, that can affect n different **outputs**. So, our derivative needs to have n elements.

Again, let's look at our **approximation** rule:

$$\Delta w \approx \frac{\partial w}{\partial s} \star \Delta s \qquad \text{or} \qquad \overbrace{\begin{bmatrix} \Delta w_1 \\ \Delta w_2 \\ \vdots \\ \Delta w_n \end{bmatrix}}^{\Delta w} \approx \frac{\partial w}{\partial s} \star \Delta s \tag{4}$$

Here, we can't do a **dot product**: we're multiplying our derivative by a **scalar**. Plus, we'd get the **same shape** as before: we might **mix up** our derivatives.

X.9 Working with the vector derivative

How do we get each of our terms Δw_i?

Well, each term is **separately** affected by Δs: we have our terms $\partial w_i / \partial s$.

So, if we take these terms **individually**, treating it as a scalar derivative, we get:

If you're ever confused with matrix math, thinking about individual elements is often a good way to figure it out!

$$\Delta w_i = \frac{\partial w_i}{\partial s} \Delta s \tag{5}$$

Since we only have **one** input, we don't have to worry about **planar** approximations: we only take one step, in the s direction.

In our matrix, we get:

$$w = \begin{bmatrix} \Delta w_1 \\ \\ \Delta w_2 \\ \\ \vdots \\ \\ \Delta w_n \end{bmatrix} = \begin{bmatrix} \Delta s(\partial w_1/\partial s) \\ \\ \Delta s(\partial w_2/\partial s) \\ \\ \vdots \\ \\ \Delta s(\partial w_n/\partial s) \end{bmatrix} \tag{6}$$

This works out for our equation above!

It could be tempting to think of our derivative $\partial w/\partial s$ as a **column vector**: we just take w and just differentiate each element. Easy!

In fact, this *is* a valid convention. However, this conflicts with our previous derivative: they're both column vectors!

Not only is it **confusing**, but it also will make it harder to do our **vector/vector** derivative.

So, what do we do? We refer back to the equation we used last time:

$$\Delta w = \left(\frac{\partial w}{\partial s}\right)^T \Delta s \tag{7}$$

We take the **transpose**! That way, one derivative is a column vector, and the other is a row vector. And, we know that this equation works out from the work we just did.

$$\Delta w = \begin{bmatrix} \frac{\partial w_1}{\partial s}, & \frac{\partial w_2}{\partial s}, & \cdots & \frac{\partial w_n}{\partial s} \end{bmatrix}^T \Delta s \tag{8}$$

Clarification 1

We mentioned that it is a valid **convention** to have that **vector derivative** be a **column vector**, and have our **gradient** be a **row vector**.

This is **not** the convention we will use in this class - you will be confused if we try!

That means, for whatever **notation** we use here, you might see the **transposed** version elsewhere. They mean exactly the **same** thing!

$$\overbrace{\Delta w}^{(n\times 1)} = \overbrace{\left(\frac{\partial w}{\partial s}\right)^T}^{(n\times 1)} \overbrace{\Delta s}^{(1\times 1)} \tag{9}$$

As we can see, the dimensions check out.

Definition 2

If s is a **scalar** and w is an $(n \times 1)$ **vector**, then we define the **vector derivative** $\partial w / \partial s$ as fulfilling:

$$\Delta w = \left(\frac{\partial w}{\partial s} \right)^T \Delta s$$

Thus, our derivative must be a $(1 \times n)$ vector

$$\frac{\partial w}{\partial s} = \begin{bmatrix} \frac{\partial w_1}{\partial s}, & \frac{\partial w_2}{\partial s}, & \cdots & \frac{\partial w_n}{\partial s} \end{bmatrix}$$