Explanatory Notes for 6.390

Shaunticlair Ruiz (Current TA)

Fall 2022

What do these derivatives equal?

Let's look at each of these derivatives and see if we can't simplify them a bit.

First, every gradient needs

- The **loss derivative**:

$$\frac{\partial \mathcal{L}}{\partial a^L} \tag{1}$$

This **depends** on on our loss function, so we're **stuck** with that one.

Next, within each layer, we have

- The **activation function** - between our activation a and preactivation z:

$$\frac{\partial a^\ell}{\partial z^\ell} \tag{2}$$

What does the function between these **look** like?

$$a = f(z) \tag{3}$$

Well, that's not super interesting: we **don't know** our function. But, at least we can **write** it using f: that way, we know that this term only depends on our **activation** function.

$$\frac{\partial a^\ell}{\partial z^\ell} = \Big(\overbrace{f^\ell}^{\text{func for layer } \ell} \Big)'(z^\ell) \tag{4}$$

This expression is a bit visually clunky, but it works.

Between layers, we have

- We can also think about the derivative of the **linear function** that **connects two layers**:

Be careful not to get this mixed up with the last one!
They look similar, but one is within the layer, and the other is between layers.

$$\frac{\partial z^\ell}{\partial a^{\ell-1}} \tag{5}$$

So, we want the function of these two:

$$z^\ell = w^\ell a^{\ell-1} + w_0^\ell \tag{6}$$

This one is pretty simple! We just take the derivative manually:

$$\frac{\partial z^\ell}{\partial a^{\ell-1}} = w^\ell \tag{7}$$

Finally, every gradient will end with

- The derivative that directly connects to a **weight**, again using the **linear function**:

$$\frac{\partial z^\ell}{\partial w^\ell} \tag{8}$$

The linear function is the same:

$$z^\ell = w^\ell a^{\ell-1} + w_0^\ell \tag{9}$$

But with a different **variable**, the **derivative** comes out different:

$$\frac{\partial z^\ell}{\partial w^\ell} = a^{\ell-1} \tag{10}$$

Notation 1

Our **derivatives** for the **chain rule** in a **1-D neural network** take the form:

$$\frac{\partial \mathcal{L}}{\partial a^L} \tag{11}$$

$$\frac{\partial a^\ell}{\partial z^\ell} = (f^\ell)'(z^\ell) \tag{12}$$

$$\frac{\partial z^\ell}{\partial a^{\ell-1}} = w^\ell \tag{13}$$

$$\frac{\partial z^\ell}{\partial w^\ell} = a^{\ell-1} \tag{14}$$

Now, we can rewrite our generalized expression for gradient:

$$\frac{\partial \mathcal{L}}{\partial w^\ell} = \overbrace{\left(\frac{\partial \mathcal{L}}{\partial a^L}\right)}^{\text{Loss unit}} \cdot \overbrace{\left((f^L)'(z^L) \;\cdot\; w^L\right)}^{\text{Layer } L} \cdot \overbrace{\left((f^{L-1})'(z^{L-1}) \;\cdot\; w^L\right)}^{\text{Layer } L-1} \cdot \left(\cdots\right) \cdot \overbrace{\left((f^\ell)'(z^\ell) \;\cdot\; a^{\ell-1}\right)}^{\text{Layer } \ell} \tag{15}$$

Our expressions are more concrete now. It's still pretty visually messy, though.