Explanatory Notes for 6.390

Shaunticlair Ruiz (Current TA)

Fall 2022

Many Layers

We are finally ready to build our **complete** neural network. We'll just retrace the steps of the 2-layer case.

Notation 1

The total **number** of **layers** in our **neural network** is notated as L.

Typically we notate an **arbitrary** layer as ℓ (or l).

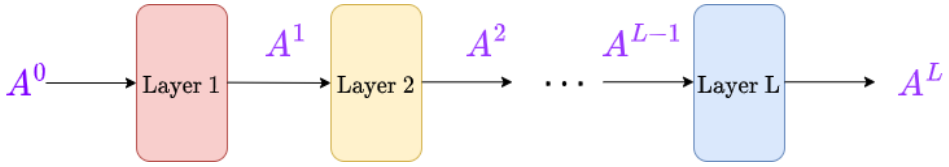
Since x is, for all purposes, **equivalent** to a vector A, we will call it A^0.

Notation 2

Our **neural network**'s input x is used in the **same** way as every term A^ℓ.

So, we will **represent** it as

$$x = A^0$$



Again, we see that the **output** of layer ℓ is the **input** of layer $\ell + 1$.

Concept 3

Each layer **feeds** into the next layer.

A^ℓ is the **output** of layer ℓ, and the **input** of layer $\ell + 1$.

This means that the **output** dimension must **match** the next **input** dimension.

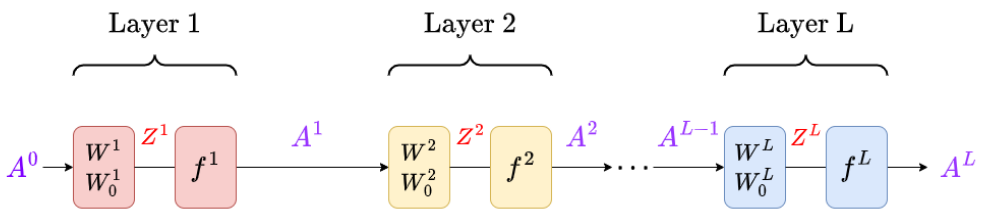
$$\overbrace{n^\ell}^{\text{Output}} = \overbrace{m^{\ell+1}}^{\text{Output}}$$

And the **dimension** of A^ℓ is $(n^\ell \times 1) = (m^{\ell+1} \times 1)$.

Our Complete Neural Network

We can break our layers into components, so we can see the functions involved.

With this, we build our final neural network:



With this, we can see how each layer is **related** to each other: as we **mentioned**, the **output** of one layer is the **input** of the next layer.

Here is the computation we do for layer ℓ:

Key Equation 4

The calculations done by layer ℓ are given by

$$Z^\ell = (W^\ell)^T A^{\ell-1} + W_0^\ell$$

and

$$A^\ell = f(Z^\ell)$$

Which combine into:

$$A^\ell = f(Z^\ell) = f\left((W^\ell)^T A^{\ell-1} + W_0^\ell\right)$$

One more comment: a useful definition.

Definition 5

A **hidden layer** is any layer except for the **last** one.

It is called a "**hidden**" layer because, if you're viewing the whole neural network based on

- **Input** x (first input)
- **Output** A^L (final output)

Then you can't see the **output** of any of the layers except for the **last** one.

Sometimes you'll hear someone say that a hidden layer is any except the "**first** or **last**": by that, they mean you can view the **input** for the **first** layer, as well as the **output** for the **last** layer.

But, when we're talking about **activation** functions (which we often are when we mention hidden layers, see below), we only care about whether the **output** is hidden!