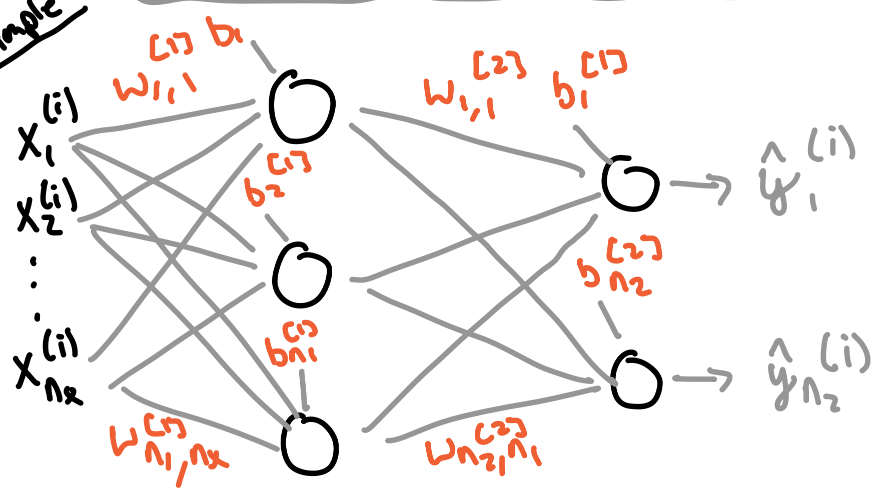
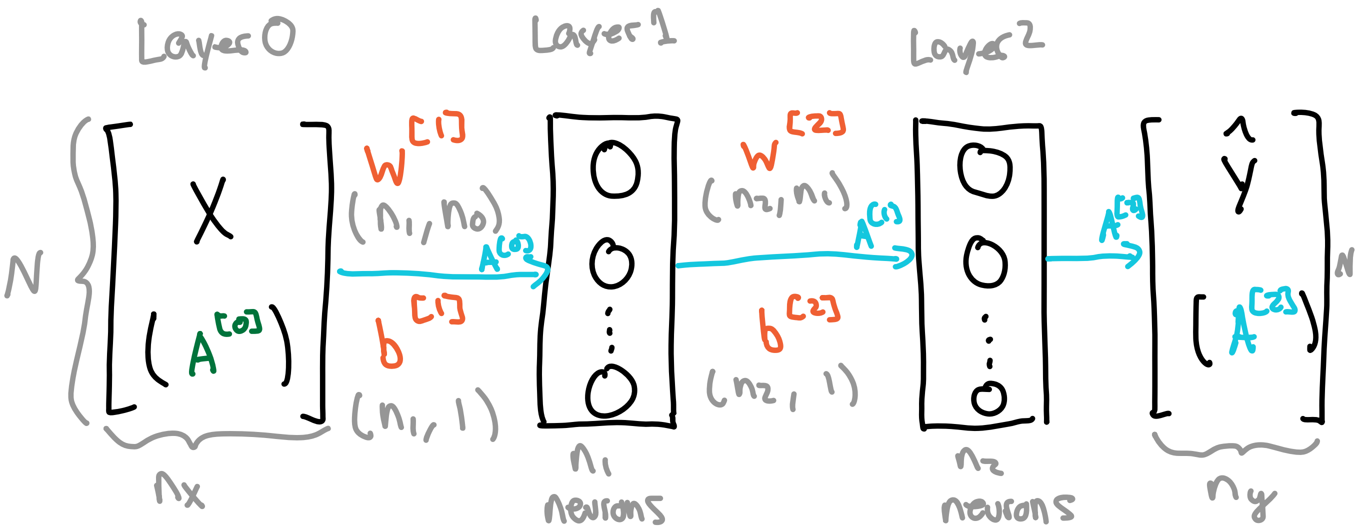
Two-Layer Neural Network

Single Input Example



Vectorized



Vectorized Equations

Layer 1

Input

Linear $$Z^{[1]} = A^{[0]} W^{[1]T} + b^{[1]}$$

Non-Linear $$A^{[1]} = g^{[1]}(Z^{[1]})$$

Activation Function

Parameters

Layer 2

Linear $$Z^{[2]} = A^{[1]} W^{[2]T} + b^{[2]}$$

Non-Linear $$A^{[2]} = g^{[2]}(Z^{[2]})$$

Output

Activation Function

Loss (Binary Cross Entropy)

$$\mathcal{L}(A^{[2]}, Y) = -\| Y \log A^{[2]} + (1-Y)\log(1-A^{[2]}) \|_1$$

Backpropagation

Assume $g^{[1]}(\cdot)$ + $g^{[2]}(\cdot)$ are the sigmoid function.

Layer 2

$$\frac{\partial \mathcal{L}}{\partial W^{[2]}} = \frac{\partial \mathcal{L}}{\partial A^{[2]}} \cdot \frac{\partial A^{[2]}}{\partial Z^{[2]}} \cdot \frac{\partial Z^{[2]}}{\partial W^{[2]}}$$

① $$\frac{\partial \mathcal{L}}{\partial A^{[2]}} = \frac{\partial}{\partial A^{[2]}} - \| Y \log A^{[2]} + (1-Y)\log(1-A^{[2]}) \|_1$$

$$= -\left(\frac{Y}{A^{[2]}} - \frac{(1-Y)}{1-A^{[2]}} \right)$$

② $$\frac{\partial A^{[2]}}{\partial Z^{[2]}} = \frac{\partial}{\partial Z^{[2]}} \sigma(Z^{[2]})$$

$$= \sigma(Z^{[2]})(1-\sigma(Z^{[2]}))$$

$$= A^{[2]}(1-A^{[2]})$$

③ $$\frac{\partial Z^{[2]}}{\partial W^{[2]}} = \frac{\partial}{\partial W^{[2]}} A^{[1]} W^{[2]T} + b^{[2]}$$

$$= A^{[1]}$$

$$\frac{\partial \mathcal{L}}{\partial W^{[2]}} = -\left(\frac{Y}{A^{[2]}} - \frac{(1-Y)}{1-A^{[2]}}\right) A^{[2]}(1-A^{[2]}) A^{[1]}$$

$$= -\left[Y(1-A^{[2]}) - (1-Y)A^{[2]}\right] A^{[1]}$$

$$= -\left[Y - \cancel{YA^{[2]}} - A^{[2]} + \cancel{YA^{[2]}}\right] A^{[1]}$$

$$= -(Y - A^{[2]}) A^{[1]}$$

$$\boxed{\frac{\partial \mathcal{L}}{\partial W^{[2]}} = (A^{[2]} - Y) A^{[1]}} = \partial Z^{[2]} A^{[1]}$$

$$\frac{\partial \mathcal{L}}{\partial b^{[2]}} = \underbrace{\frac{\partial \mathcal{L}}{\partial A^{[2]}} \cdot \frac{\partial A^{[2]}}{\partial Z^{[2]}}}_{\text{Call this } \partial Z^{[2]} \text{ since we use it in several computations}} \cdot \overset{①}{\frac{\partial Z^{[2]}}{\partial b^{[2]}}}$$

① $$\frac{\partial Z^{[2]}}{\partial b^{[2]}} = \frac{\partial}{\partial b^{[2]}} A^{[1]} W^{[2]T} + b^{[2]}$$

$$= \underline{1}$$

$$\boxed{\frac{\partial Z^{[2]}}{\partial b^{[2]}} = (A^{[2]} - Y) = \partial Z^{[2]}}$$

Layer 2 Update Equations

$$W^{[2]} := W^{[2]} - \alpha \frac{\partial \mathcal{L}}{\partial W^{[2]}}$$

$$b^{[2]} := b^{[2]} - \alpha \frac{\partial \mathcal{L}}{\partial b^{[2]}}$$

Layer 1

$$\frac{\partial \mathcal{L}}{\partial W^{[1]}} = \underbrace{\frac{\partial \mathcal{L}}{\partial A^{[2]}} \cdot \frac{\partial A^{[2]}}{\partial Z^{[2]}}}_{\partial Z^{[2]} \leftarrow \text{Backward from layer 2}} \cdot \overset{①}{\frac{\partial Z^{[2]}}{\partial A^{[1]}}} \cdot \overset{②}{\frac{\partial A^{[1]}}{\partial Z^{[1]}}} \cdot \overset{③}{\frac{\partial Z^{[1]}}{\partial W^{[1]}}}$$

① $$\frac{\partial Z^{[2]}}{\partial A^{[1]}} = \frac{\partial}{\partial A^{[1]}} A^{[1]} W^{[2]T} + b^{[2]}$$

$$= W^{[2]}$$

② $$\frac{\partial A^{[1]}}{\partial Z^{[1]}} = \frac{\partial}{\partial Z^{[1]}} \sigma\left(Z^{[1]}\right)$$

$$= \sigma\left(Z^{[1]}\right)\left(1 - \sigma\left(Z^{[1]}\right)\right)$$

$$= A^{[1]}\left(1 - A^{[1]}\right)$$

③ $$\frac{\partial Z^{[1]}}{\partial W^{[1]}} = \frac{\partial}{\partial W^{[1]}} A^{[0]} W^{[1]T} + b^{[1]}$$

$$= A^{[0]}$$

$$\boxed{\frac{\partial \mathcal{L}}{\partial W^{[1]}} = \underbrace{\partial Z^{[2]}\, W^{[2]} A^{[1]}\left(1 - A^{[1]}\right)}_{\text{Call this } \partial Z^{[1]}} A^{[0]}}$$

$$\frac{\partial \mathcal{L}}{\partial b^{[1]}} = \underbrace{\frac{\partial \mathcal{L}}{\partial A^{[2]}} \cdot \frac{\partial A^{[2]}}{\partial Z^{[2]}} \cdot \frac{\partial Z^{[2]}}{\partial A^{[1]}} \cdot \frac{\partial A^{[1]}}{\partial Z^{[1]}}}_{\partial Z^{[1]}} \cdot \overset{①}{\frac{\partial Z^{[1]}}{\partial b^{[1]}}}$$

①

$$\frac{\partial Z^{[1]}}{\partial b^{[1]}} = \frac{\partial}{\partial b^{[1]}} A^{[0]} W^{[1]T} + b^{[1]}$$

$$= \underline{1}$$

$$\boxed{\frac{\partial Z^{[1]}}{\partial b^{[1]}} = \partial Z^{[1]}}$$

Layer 1 Update Equations

$$W^{[1]} := W^{[1]} - \alpha \frac{\partial \mathcal{L}}{\partial W^{[1]}}$$

$$b^{[1]} := b^{[1]} - \alpha \frac{\partial \mathcal{L}}{\partial b^{[1]}}$$