Input = 28x28 pixel grid

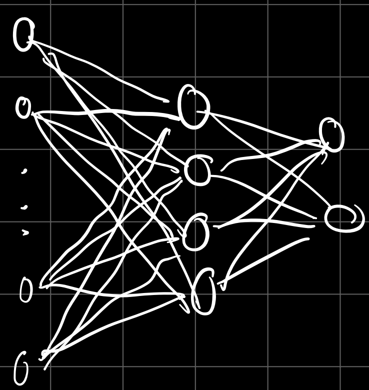
↓

Break apart into 784 individual pixels

↓

Assign each pixel a grayscale value from 0-1

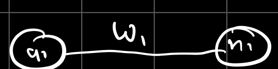
What are Neurons



You have input neurons & Output neurons

In the middle we have the

"Hidden Layer"



$\therefore$ $w_1 + a_1$ = Weighted sum

Sigmoid $= \dfrac{1}{1+e^{-x}} = \sigma(x)$

$$\sigma\left[\begin{bmatrix} w_{11} & w_{12} & \cdots w_{1j} \\ w_{21} & w_{22} & \cdots w_{ji} \end{bmatrix}\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}\right] = n_1$$

Soft Max

$$\text{soft max} = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

$$\begin{array}{cc} & \times 2 \\ \times 3 & \begin{bmatrix} w_1 & w_2 \\ w_3 & w_4 \\ w_5 & w_6 \end{bmatrix} \end{array} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

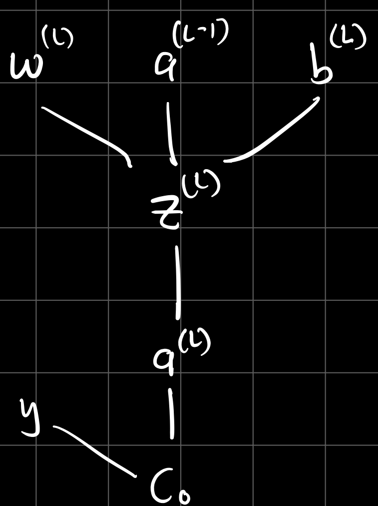
Backpropagation

$$C_0(\ldots) = (a^{(L)} - y)^2$$

$$z^{(L)} = w^{(L)} a^{(L-1)} + b^{(L)}$$

$$a^{(L)} = \sigma(z^{(L)})$$

Desired Ouput



$$\frac{\partial C_0}{\partial w^{(L)}} = \frac{\partial z^{(L)}}{\partial w^{(L)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C_0}{\partial a^{(L)}}$$

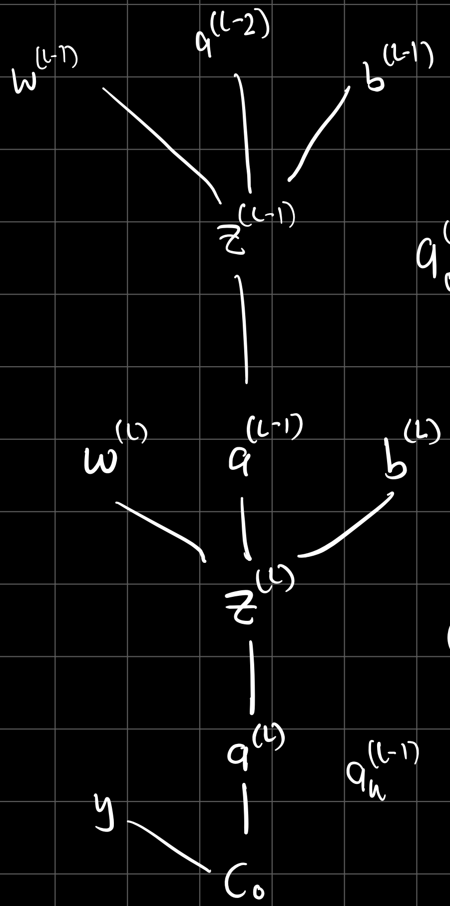
$$\frac{\partial C_0}{\partial w^{(L)}} = a^{(L-1)} \cdot \sigma'(z^{(L)}) \cdot 2(a^{(L)} - y)$$

$$\frac{\partial C}{\partial w^{(L)}} = \overbrace{\frac{1}{n} \sum_{k=0}^{n-1}}^{\text{avg of all training examples}} \frac{\partial C_k}{\partial w^{(L)}}$$

$$\nabla C = \begin{bmatrix} \frac{\partial C}{\partial w^{(1)}} \\ \frac{\partial C}{\partial b^{(1)}} \\ \vdots \\ \frac{\partial C}{\partial w^{(L)}} \\ \frac{\partial C}{\partial b^{(L)}} \end{bmatrix}$$

$$\frac{\partial C_0}{\partial b^{(L)}} = \frac{\partial z^{(L)}}{\partial b^{(L)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C_0}{\partial a^{(L)}} = 1 \cdot \sigma'(z^{(L)}) \cdot 2(a^{(L)} - y)$$

$$\frac{\partial C_0}{\partial a^{(L-1)}} = \frac{\partial z^{(L)}}{\partial a^{(L-1)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C_0}{\partial a^{(L)}} = w^{(L)} \sigma'(z^{(L)}) 2(a^{(L)} - y)$$

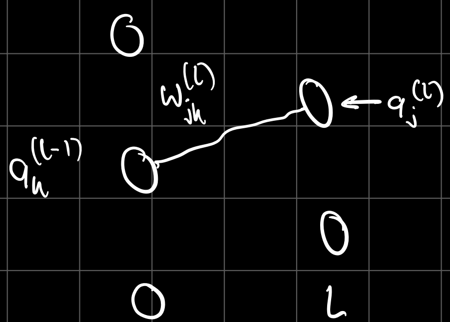


$$C_0 = \sum_{j=0}^{n_L - 1} (a_j^{(L)} - y_j)^2$$

$$a_0^{(1)} = \sigma(w_{0,0} a_0^{(0)} + w_{0,1} a_1^{(0)} + \dots + w_{0,n} a_n^{(0)} + b_0)$$

$$z_j^{(L)} = w_{j0}^{(L)} a_0^{(L-1)} + w_{j1}^{(L)} a_1^{(L-1)} + w_{j2}^{(L)} a_2^{(L-1)} + b_j^{(L)}$$

$$a_j^{(L)} = \sigma(z^{(L)})$$



$$\frac{\partial C_0}{\partial a^{(L-1)}} = \frac{\partial z^{(L)}}{\partial a^{(L-1)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C_0}{\partial a^{(L)}} = w^{(L)} \sigma'(z^{(L)}) 2(a^{(L)} - y)$$

$$\frac{\partial C_0}{\partial w_{jk}^{(L)}} = \frac{\partial z_j^{(L)}}{\partial w_{jk}^{(L)}} \cdot \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \cdot \frac{\partial C_0}{\partial a_j^{(L)}}$$

$$\frac{\partial C_0}{\partial a_k^{(L-1)}} = \underbrace{\sum_{j=0}^{n_L - 1} \frac{\partial z_j^{(L)}}{\partial a_k^{(L-1)}} \cdot \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \cdot \frac{\partial C_0}{\partial a_j^{(L)}}}_{\text{Sum over layer } L}$$

$$\nabla C = \sum_{j=0}^{n_{L+1} - 1} w_{jk}^{(L+1)} \sigma'(z_j^{(L+1)}) \frac{\partial C}{\partial a_j^{(L+1)}}$$

or

$$2(a_j^{(L)} - y_j)$$

$$\frac{\partial C_0}{\partial w^{(L)}} = \frac{\partial z^{(L)}}{\partial w^{(L)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C_0}{\partial a^{(L)}}$$

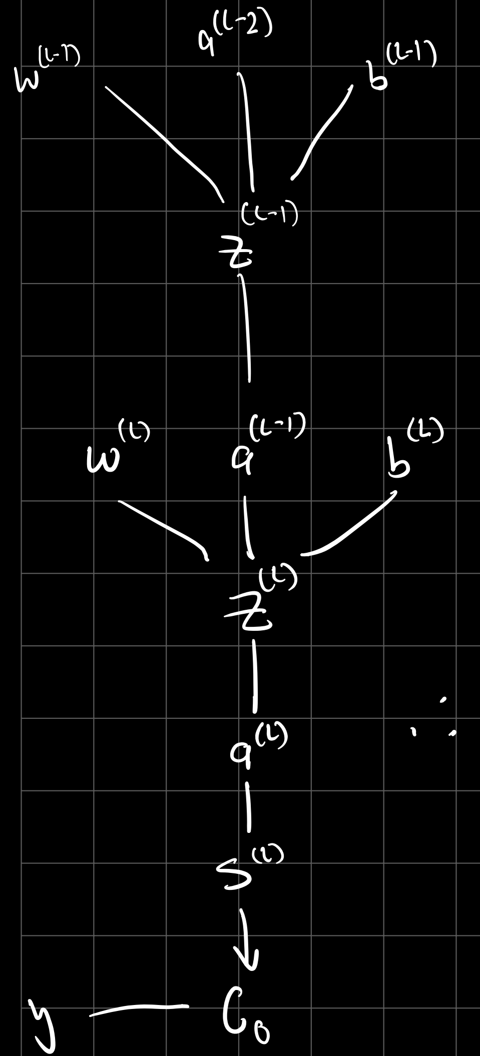
$$\frac{\partial C_0}{\partial w^{(L)}} = a^{(L-1)} \cdot \sigma'(z^{(L)}) \cdot 2(a^{(L)} - y)$$

$$\frac{\partial C_0}{\partial a^{(L-1)}} = \frac{\partial z^{(L)}}{\partial a^{(L-1)}} \cdot \frac{\partial a^{(L)}}{\partial z^{(L)}} \cdot \frac{\partial C_0}{\partial a^{(L)}}$$

$$= w^{(L)} \sigma'(z^{(L)}) 2(a^{(L)} - y)$$

$a^L = sm(z)$

z = sigmoid s = softmax



$$C_0 = -y\ln(s^{(L)})$$

$y =$ labels

$$z^{(L)} = w^{(L)}a^{(L-1)} + b^{(L)}$$

$$a^{(L)} = \sigma(z^{(L)})$$

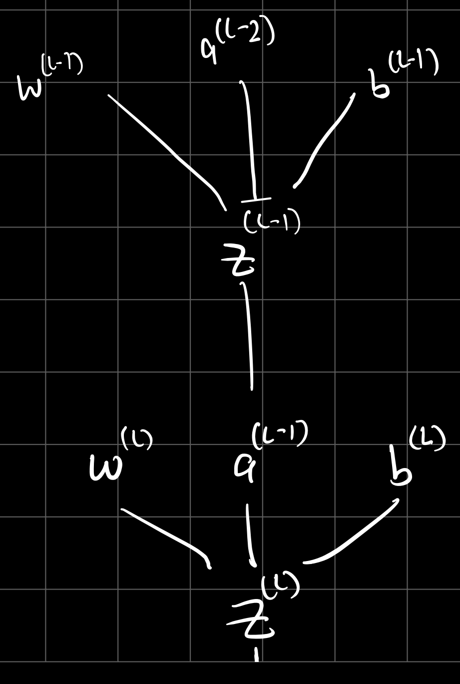
$$s^{(L)} = S_m(a^{(L)}) = \frac{e^{a_i^{(L)}}}{\sum_j e^{a_j^{(L)}}}$$

$$\frac{\partial C_0}{\partial w^{(L)}} = \frac{\partial C_0}{\partial s^{(L)}} \frac{\partial s^{(L)}}{\partial a^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}} \frac{\partial z^{(L)}}{\partial w^{(L)}}$$

$$\frac{\partial C_0}{\partial a^{(L)}} = \frac{\partial C_0}{\partial s^{(L)}} \frac{\partial s^{(L)}}{\partial a^{(L)}} = s^{(L)} - y$$

$$\therefore \frac{\partial C_0}{\partial w^{(L)}} = (s^{(L)} - y)\,\sigma'(z^{(L)}) \cdot a^{(L-1)}$$

-1



$$\frac{\partial C_0}{\partial w^{(L)}} = \frac{\partial C_0}{\partial s} \cdot \frac{\partial s}{\partial z} \cdot \frac{\partial z}{\partial w}$$

$$= (s^{(L)} - y) \cdot a^{(L-1)}$$

$$\begin{tikzcd} & S^{(1)} \arrow[d] \\ y \arrow[r, dash] & C_0 \end{tikzcd}$$