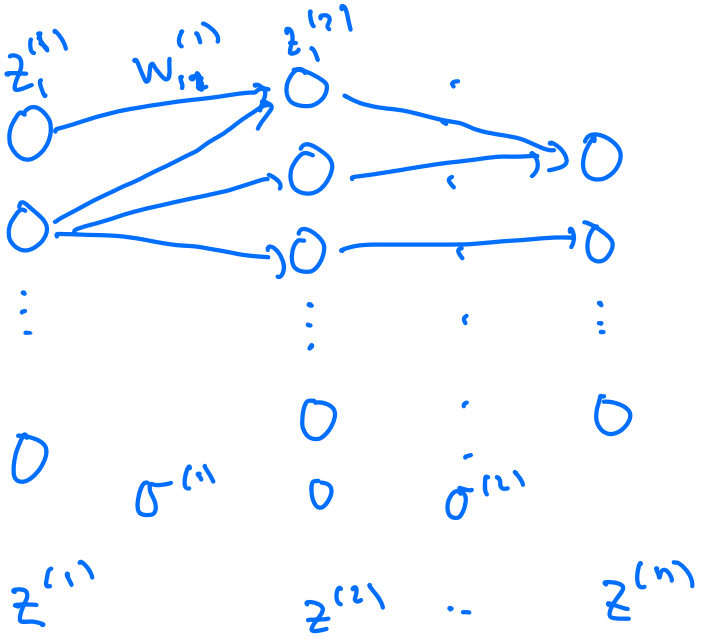
Neural Network Structure



$a_i^{(l)} = \sigma(z_i^{(l)})$

$z_i^{(l)}$

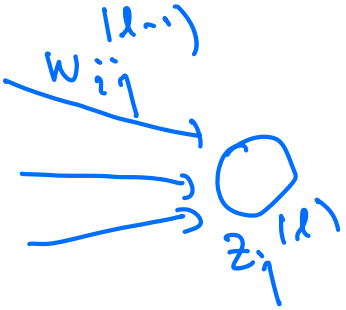
Layer i has n_i entries

Layer $i-1$ has n_{i-1} entries

Layer

σ activation function

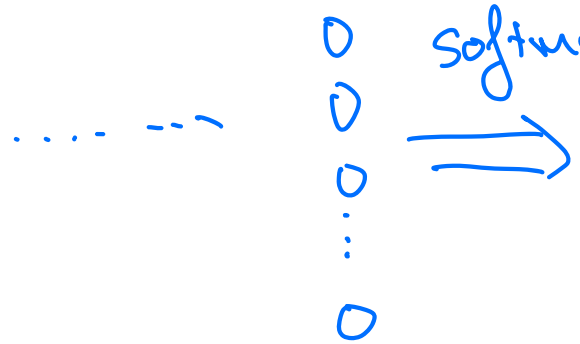
$\sigma(x) = \text{ReLU}(x) = x$ if $x > 0$, 0 otherwise



influence or forward pass

$$z_j^{(l)} = \sum_i a_i^{(l-1)} w_{ij}$$

"output layer"



final output

Loss function

$x^{[i]}$ $y^{[i]}$

$L_W(x^{[i]}, y^{[i]})$

$$\text{Loss}_w = \sum L_w(x^{[i]}, y^{[i]})$$

Goal is to minimize L_w Loss

Find ∇L_w adjust w's by ∇L_w

$\delta^{(j)}$ @ $\delta_i^{(j)} = \frac{\partial L_w}{\partial z_i^{(j)}}$

$\updownarrow$

j^{th} layer $\delta^{(j-1)} = \sigma'(z^{(i-1)}) w^{(i)} \delta^{(i)}$

Layer i has n_i entries

Layer $i-1$ has n_{i-1} entries

$w^{(i)}$ is on

$$\begin{bmatrix} & n_{i-1} & \end{bmatrix} \overset{(i)}{W} \overset{n_i}{=} \begin{bmatrix} \cdots \end{bmatrix}$$

$n_{i-1} \times n_i$ matrix

δ^i is an n_i-entry column vector

$w^i \delta^i$ is an n_{i-1}-column

$\sigma'(z^{(i-1)}) \cdot w^i \delta_i$ entry by entry

Last layer - say layer m:

need $\delta_i^{(m)} = \frac{\partial L}{\partial z_i^m}$

$z_i^{(m)}$

$\begin{cases} 0 \\ 0 \\ 0 \end{cases}$ Loss →

$$\text{Loss} = \| z^{(m)}(x^{(i)}) - y^{[i]} \|^2$$

$$= \sum_K \left(z_K^{(m)} - y_K^{[i]} \right)^2$$

$$\frac{\partial \text{Loss}}{\partial z_K^{(m)}} = 2 \left(z_K^{(m)} - y_K^{(i)} \right)$$

$$\delta^{(m)} = z^{(m)} - y$$

Cross entropy

$$\delta^{(n)} = \underbrace{p^{(m)}}_{=} - y$$

one-hot encoded

$$e^{z_i} / \sum_i e^{z_i}$$

$$\frac{\partial L}{\partial w_{ik}^{(j)}} = \sum_{K=1}^{n^{(j)}} \frac{\partial L}{\partial z_K^{(j)}} \frac{\partial z_K^{(j)}}{\partial w_{ik}^{(j)}}$$

$\frac{\partial L}{\partial z_K^{(j)}} = \delta_K^{(j)}$

$$z_k^{(j)} = \sum_i a_i^{(j-1)} w_{ik}^{(j)}$$

$$a_i^{(j)} = \sigma\left(z_i^{(j-1)}\right)$$

$$\frac{\partial z_k^{(j)}}{\partial w_{it}^{(j-1)}} = \begin{cases} 0 & \text{if } t \neq k \\ a_i^{(j-1)} & \text{if } t = k \end{cases}$$

$$= \sigma\left(z_i^{(j-1)}\right)$$

$$\frac{\partial L}{\partial w_{it}^{(j-1)}} = \sum_k \delta_k^{(j)} \sigma\left(z_i^{(j-1)}\right)$$

$$\frac{\partial L}{\partial w_{it}^{(j-1)}} = \delta^{(j)} \sigma\left(z^{(j-1)}\right)$$

$n_j \times 1$ col vector $\quad 1 \times n_{j-1}$ row vector

$n_j \times n_{j-1}$ matrix

$\rightarrow$ transpose of ∇L