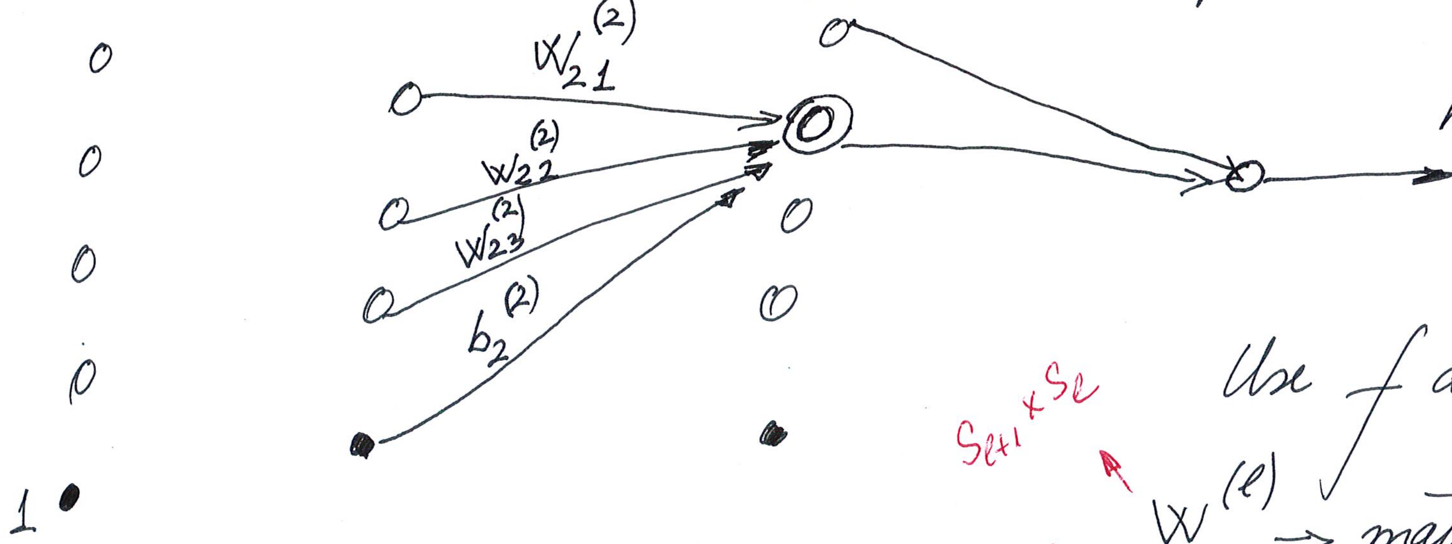
Layer 1 — $x = a^{(1)}$

Layer 2 — $a^{(2)}$

Layer 3 — $a^{(3)}$

Layer 4 — $a^{(4)}$ $= h(x; W, b)$.



$h(x) = h_{W,b}(x)$

Regression

Use f as activation function.

$W^{(l)}$ → matrix of params. connect. layer l to $l+1$ ($s_{l+1} \times s_l$)

$b^{(l)}$ → vector of bias params. for layer $l+1$ ($s_{l+1} \times 1$)

s_l = # of neurons on layer l

$s_1 = 4$ $\quad s_2 = 3$ $\quad s_3 = 4$ $\quad s_4 = 1$

n_l = # layers

$$a^{(3)} = [a_1^{(3)}, a_2^{(3)}, a_3^{(3)}, a_4^{(3)}]$$

$$a^{(3)} = f(\underbrace{W^{(2)}}_{s_3 \times s_2} \cdot \underbrace{a^{(2)}}_{s_2 \times 1} + \underbrace{b^{(2)}}_{s_3 \times 1}) \Longrightarrow$$

$$z^{(3)} = W^{(2)} a^{(2)} + b^{(2)}$$

$$a^{(3)} = f(z^{(3)})$$

$$a_1^{(3)} = f(\ldots)$$

$$a_2^{(3)} = f(W_{21}^{(2)} \cdot a_1^{(2)} + W_{22}^{(2)} a_2^{(2)} + W_{23}^{(2)} a_3^{(2)} + b_2^{(2)}).$$

$$a_3^{(3)} = f(\ldots)$$

$$a_4^{(3)} = f(\ldots).$$

$$\boxed{\begin{aligned} z^{(l+1)} &= W^{(l)} a^{(l)} + b^{(l)} \\ a^{(l+1)} &= f(z^{(l+1)}) \end{aligned}}$$

$$W = \{ W^{(1)}, W^{(2)}, \dots, W^{(n_\ell)} \}$$

$$\|W\|^2 \longrightarrow \|W^{(1)}\|^2 + \|W^{(2)}\|^2 + \dots + \|W^{(n_\ell)}\|^2$$

$$\|A\|_2^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \qquad A \in \mathbb{R}_{m \times n}$$

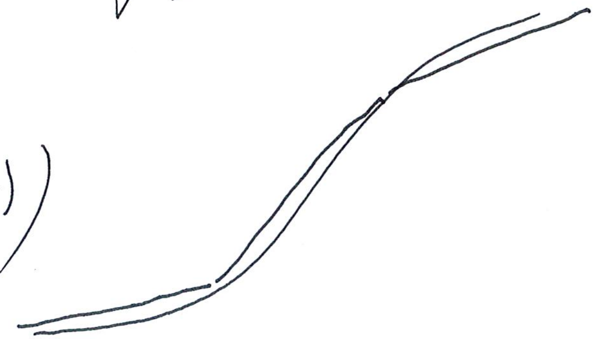
↳ L_2 norm, Frobenius norm.

$$\|W^{(\ell)}\|^2 \neq \mathrm{sum}\left(W^{(\ell)}\cancel{.T} * W^{(\ell)}\right) \neq \sum_{i=1}^{n_{\ell+1}} \sum_{j=1}^{n_\ell} \left(W_{ij}^{(\ell)}\right)^2$$

$$J(W, b) = \frac{1}{2}\left\| a^{(n_\ell+1)} - y \right\|^2 \qquad \frac{\partial J}{\partial a^{(n_\ell+1)}} = \left(a^{(n_\ell+1)} - y\right)$$

$W^{(1)}, W^{(2)}, \dots W^{(n_\ell)}$ $\qquad$ $b^{(1)}, b^{(2)}, \dots, b^{(n_\ell)}$

$$f(z) = \sigma(z) = \frac{1}{1+e^{-z}} \qquad f'(z) = \underbrace{\sigma(z)}_{<1}\underbrace{\left(1-\sigma(z)\right)}_{<1}$$



$\sigma(z) \in [0,1]$, $\sigma(z) \in (0,1)$