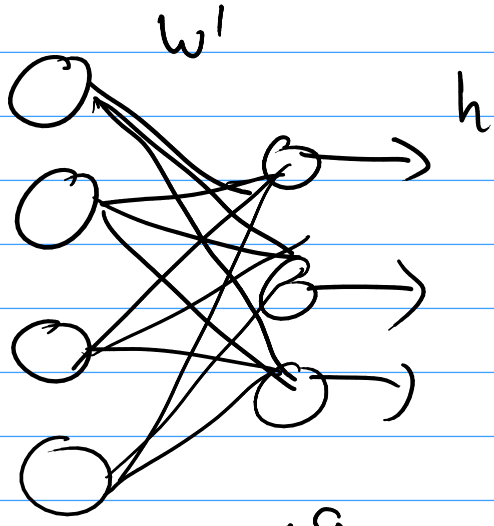
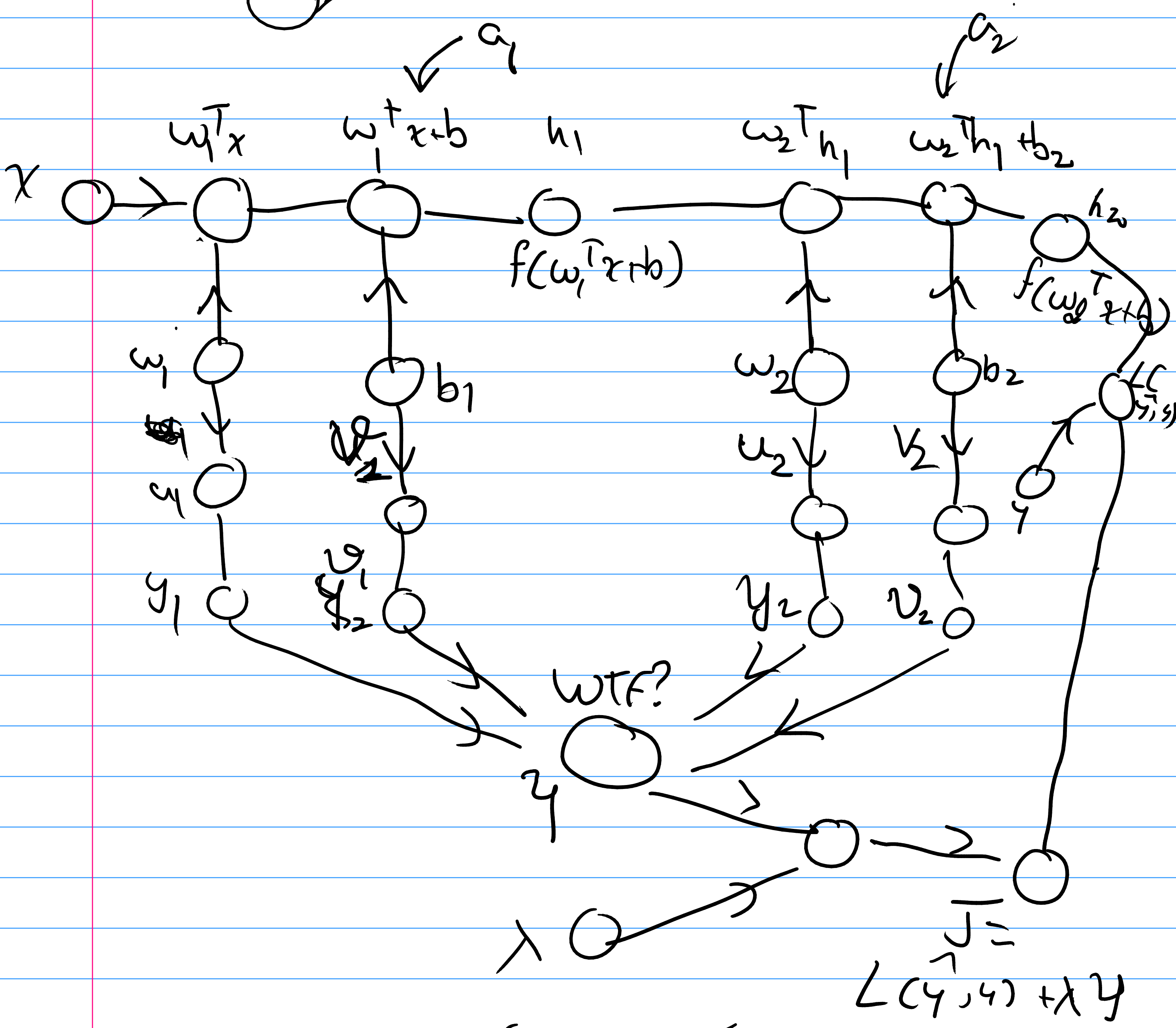


$f(\omega^T$



$y = \Omega(\theta)$ (regularized)

$$\min L(y, \hat{y}) + \lambda \Omega(\Theta)$$

$$\Downarrow$$

w, b

$$\Omega(\Theta) = w_1^2 + w_2^2 + \cdots + w_{10}^2 + w_{\Theta}^2 \cdots_n$$

$\cdots$

$\nabla_{w_1} J$ $\quad$ $\nabla_{b_1} J$

$\nabla_{w_2} J$ $\quad$ $\nabla_{b_2} J$

$\nabla_{w_3} J$ $\quad$ $\nabla_{w_3} J$

a matrix of weights can be treated as a vector of vectors

How can we find the derivative of a scalar with a vector

3 layered NN

$$\nabla_{a_3}(L(\hat{y}, y)) = \nabla_{a_3}(h_3) \nabla_{h_3}(L(\hat{y}, y)) \quad (1)$$

$$= \nabla_{a_3}(f(a_3)) \odot g = f'(a_3) \odot g$$

element-wise mul

$$\frac{\partial L}{\partial a_3} = \frac{\partial L}{\partial h_3} \cdot \frac{\partial h_3}{\partial a_3}$$

$$\nabla_{w_3}(J) = \nabla_3 (L(\hat{y}, y) + \lambda \Omega(\Theta))$$

$$= \nabla_{w_3}(L(\hat{y}, y)) + \lambda \nabla_{w_3}(\Omega(\theta))$$

$$= \nabla_{a_3}(L(\hat{y}, y)) \nabla_{w_3}(a_3)$$

$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial a_3} \cdot \frac{\partial a_3}{\partial w_3} + \lambda \nabla_{w_3}(\Omega(\theta))$$

$$\left[\nabla_{a_3}(L(\hat{y}, y)) = f'(a_3) \odot g \right] \quad \dots \quad (1)$$

$$f'(a_3) \odot g \; \nabla_{w_3}(w_3^T h_2 + b_3)$$

$$\left. \frac{\partial J}{\partial w_3} \right|_{w_3 = w_3^{(k)}}$$

<u>|||ly</u> $\frac{\partial L}{\partial b_3} = \frac{\partial L}{\partial a_3} \cdot \frac{\partial a_3}{\partial b_3} = f'(a_3) \odot g \; \nabla_{b_3}(w_3^T h + b_3)$

keep applying chain rule