Lecture 5

→ Training on large datasets, want code to run quickly.

→ What is Vectorization?

$z = w^T x + b$ $\qquad w = \begin{bmatrix} \vdots \end{bmatrix} \qquad x = \begin{bmatrix} \vdots \end{bmatrix}$

Non-vectorized

```
z = 0
for i = 1 to in range(n*x):
    z += w[i] * x[i]
z += b
```

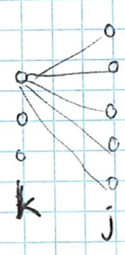
Vectorized

```
z = np.dot(w, x) + b
```

$(z = \underbrace{w @ x}_{w^T x} + b)$ parallelized by using built in functions

$z = \begin{bmatrix} \vdots \end{bmatrix} \qquad \sigma(z) = \begin{bmatrix} \sigma(z_1) \\ \sigma(z_2) \\ \vdots \\ \sigma(z_n) \end{bmatrix}$

$W^l = \begin{array}{c} \\ j \end{array} \left[\begin{array}{c} \xleftarrow{\quad k \quad} \\ \\ \\ \end{array} \right] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} k \qquad b^l = j \begin{bmatrix} \vdots \end{bmatrix}$



$k \times j$ matrix W^l

$= {}_j[\;] + [\;]$

$$\delta^L = \begin{bmatrix} \frac{\partial J}{\partial a_1^{(L)}} \\ \frac{\partial J}{\partial a_2^{(L)}} \\ \vdots \end{bmatrix} \odot \sigma'(z^{(L)})$$

$\boxed{\delta = \text{Delta}}$

$\boxed{\sigma = \text{Sigma}}$

$$\delta^L = \frac{\partial J}{\partial a_k^{(2)}} \times \frac{\partial a_k^{(2)}}{\partial z_k^{(2)}}$$

(red) (blue)

$$[a_k^{(2)} - t_k]$$

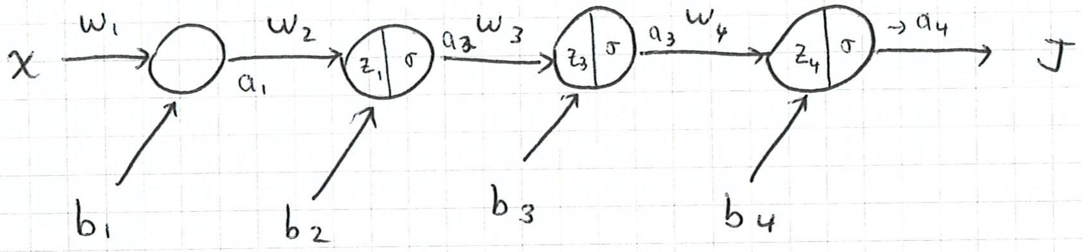
$t_k \to y$

$$\delta^l = \delta^{l+1} \times W^{(l+1)} \odot \sigma'(z^l)$$

$$\delta^{(1)} = \overbrace{\frac{\partial J}{\partial a_k^{(1)}}}^{\text{orange}} \cdot \overbrace{\frac{\partial a_k^{(1)}}{\partial z_k^{(1)}}}^{\text{blue}}$$

$$= \Big(\underbrace{\sum \underbrace{\frac{\partial J}{\partial a_k^{(2)}}}_{\text{red}} \cdot \underbrace{\frac{\partial a_k^{(2)}}{\partial z_k^{(2)}}}_{\text{blue}}}_{\delta^{(2)}} \times \underbrace{\frac{\partial z_k^{(2)}}{\partial a_k^{(1)}}}_{\text{green}} \Big) \times \underbrace{\frac{\partial a_k^{(1)}}{\partial z_k^{(1)}}}_{\text{blue}}$$

$\to W_{kj}^{(2)} \times \sigma'(z^{(1)})$



$$J = \frac{1}{2} \sum (\hat{y} - y)^2$$

$$\hat{y} = \sigma(a_3 w_4 + b_4)$$

$$= \sigma(w_4 \cdot \sigma(w_3 \cdot \sigma(w_2 \cdot \sigma(w_1 x + b_1) + b_2) + b_3) + b_4)$$

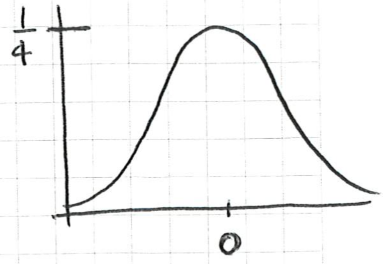
$$\boxed{\begin{array}{l} \text{define} \\ z_4 = a_3 w_4 + b_4 \\ a_4 = \sigma(z_4) \end{array}}$$

$$\frac{\partial J}{\partial b_1} = \frac{\partial J}{\partial a_4} \times \underbrace{\frac{\partial a_4}{\partial z_4}}_{\sigma'(z_4)} \times \underbrace{\frac{\partial z_4}{\partial a_3}}_{w_4} \times \underbrace{\frac{\partial a_3}{\partial z_3}}_{\sigma'(z_3)} \times \underbrace{\frac{\partial z_3}{\partial a_2}}_{w_3} \times \underbrace{\frac{\partial a_2}{\partial z_2}}_{\sigma'(z_2)}$$

$$\times \underbrace{\frac{\partial z_2}{\partial a_1}}_{w_2} \times \underbrace{\frac{\partial a_1}{\partial z_1}}_{\sigma'(z_1)} \times \underbrace{\frac{\partial z_1}{\partial b_1}}_{1}$$

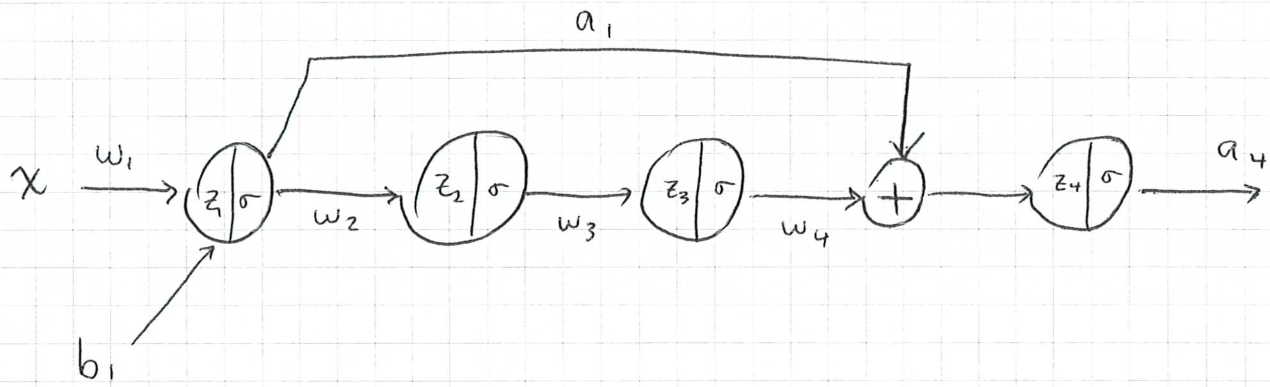
$$= \frac{\partial J}{\partial a_4} \times \sigma'(z_4) \times w_4 \, \sigma'(z_3) \, w_3 \, \sigma'(z_2) \, w_2 \, \sigma'(z_1) \times 1$$

Now assume σ = sigmoid σ'



$$(0.3)^4 = 0.0081$$

$$(0.3)^{20} = 0.00000000000348 \rightarrow 3.48 \times 10^{-11} \rightarrow \text{ReLu's}$$



$$z_4 = w_4 a_3 + a_1$$

$$\frac{\partial J}{\partial b_1} = \frac{\partial J}{\partial a_4} \times \frac{\partial a_4}{\partial z_4} \left[\underbrace{\frac{\partial z_4}{\partial a_3} \times \frac{\partial a_3}{\partial z_3} \times \frac{\partial z_3}{\partial a_2} \times \frac{\partial a_2}{\partial z_2} \times \frac{\partial z_2}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial b_1}}_{\text{very very small}} + \underbrace{\frac{\partial z_4}{\partial a_1} \times \frac{\partial a_1}{\partial z_1} \times \frac{\partial z_1}{\partial b_1}}_{\text{much less small}} \right]$$