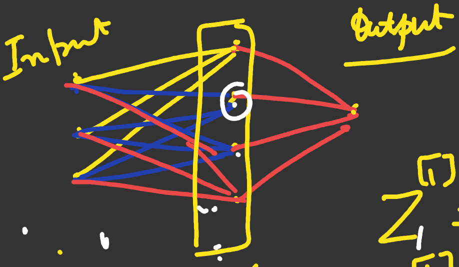
20 Jan, 2023



~~Hidden layer~~

$a\left(\underline{W^T X + b}\right)$

$X = A^{0}$, $A^{[0]}$, $A^{[l]}$ → layer no

$W \longrightarrow W^{[l]}$

$n_x = 3$

$Z_1^{[1]} = W_1^{[1]^T} X^{[0]} + b_1^{[1]} \equiv W^{[1]^T} A^{[0]} + b^{[1]}$

$A^{[1]} = a(Z^{[1]})$

$Z_2^{[1]} = W_2^{[1]^T} X^{[0]} + b_2^{[1]}$

$Z_3^{[1]} = W_3^{[1]^T} X^{[0]} + b_3^{[1]}$

$Z_4^{[1]} = W_4^{[1]^T} X^{[0]} + b_4^{[1]}$

$a_1(Z_1^{[1]})$ $a_3(Z_3^{[1]})$

$a_2(Z_2^{[1]})$ $a_4(Z_4^{[1]})$

$n_x = n^{[0]}$

$$\begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n_x} \\ W_{21} & W_{22} & \text{——} & W_{2n_x} \\ & & & \\ W_{n^{[1]}1} & W_{n^{[1]}2} & & W_{n^{[1]}n^{[0]}} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n_x} \end{bmatrix}$$

$n^{[1]} = 4$

$$z_1^{[1]} = w_1^{[1]} X + b_1^{[1]}$$

$(3,1)$ $(n^{[0]},1)$

$$z_{n^{[1]}}^{[1]} = w_{n^{[1]}}^{[1]} X + b_{n^{[1]}}^{[1]}$$

$$W^{[1]} X + b$$

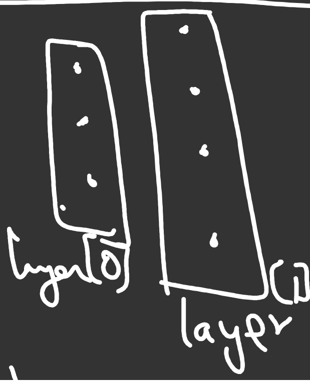
$$Z^{[1]} = W^{[1]} X + b^{[1]}$$

$(m,n)(n,p) \to (m,p)$

$$b^{[1]} = \begin{pmatrix} b_1^{[1]} \\ \vdots \\ b_{n_0}^{[1]} \end{pmatrix} \qquad \begin{bmatrix} b_1^{[1]} & \cdots & b_{n_x}^{[1]} \end{bmatrix}$$

$[n_x, 1]$

$(n^{[1]}, 1)$



$$Z^{[l]} = W^{[l]} A^{[l-1]} + b^{[l]}$$

- Z $(n^{[l]}, 1)$ ✓
- W $(n^{[l]}, n^{[l-1]})$ ✓
- A $(n^{[l-1]}, 1)$ ✓
- b $(n^{[l]}, 1)$ ✓

l – index of layer.

$n^{[l]}$: No of neurons in layer l

$m \times \underline{n}$, $\underline{n} \times p$

$$\underbrace{\begin{bmatrix} \text{———} \\ \text{———} \\ \text{———} \\ \vdots \\ \text{———} \end{bmatrix}}_{(n^{[l]},\, n^{[l-1]})} \begin{bmatrix} A_1^{[l-1]} \\ A_2^{[l-1]} \\ \vdots \\ A_{n^{[l-1]}}^{[l-1]} \end{bmatrix}$$

$$A^{[1]} = \cancel{\sigma}(Z^{[1]})$$

$$A^{[1]} = Z^{[1]} = W^{[1]} A^{[0]} + b^{[1]}$$

$$A^{[2]} = \sigma(Z^{[2]})$$

$$\begin{aligned}
Z^{[2]} &= W^{[2]} A^{[1]} + b^{[2]} \\
&= W^{[2]}\left(W^{[1]} A^{[0]} + b^{[1]}\right) + b^{[2]} \\
&= \left(W^{[2]} W^{[1]}\right) A^{[0]} + W^{[2]} b^{[1]} + b^{[2]} \\
&= W A^{[0]} + W^{[2]} b^{[1]} + b^{[2]} = \underline{W} A^{[0]} + \underline{b}
\end{aligned}$$

24 Jan 23

$a = \tanh(z)$

$$= \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

Rectified linear unit

$a(z) = z$, if $z \geq 0$

$= 0$ otherwise

Relu RELU

$$\frac{da}{dz} = \frac{(e^{z}+e^{-z})(e^{z}+e^{-z}) - (e^{z}-e^{-z})(e^{z}-e^{-z})}{(e^{z}+e^{-z})^2}$$

$1-\tanh^2(z)$

Leaky Relu $a(z) = 0.001z$

$$Z^{[1]} = W^{[1]} X + b^{[1]}$$
$$A^{[1]} = g^{[1]}(Z^{[1]})$$
$$Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$$
$$A^{[2]} = g^{[2]}(Z^{[2]}) = \sigma(Z^{[2]})$$

$$L = -(y \log a + (1-y)\log(1-a))$$

$$dz = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \; : \; a = \frac{1}{1+e^{-z}}$$
$$= (a-y)$$

$$dz^{[2]} = A^{[2]} - Y$$

$$dW^{[2]} = dz^{[1]} X_1^{(i)} + dz^{[1]} X_2^{(i)}$$

$$Z^{[2]\cancel{(i)}} = W^{[2]} A^{[1]\cancel{(i)}} + b^{[2]}$$

vector

W $\quad$ $A^{[1]}$

$$\begin{bmatrix} \quad \end{bmatrix} \begin{bmatrix} A_1^{[1](1)} & \cdots & A^{[1](m)} \\ \vdots & & \\ A^{[1](1)} & & \\ \vdots & & \\ A_{n^{[1]}}^{[1](1)} & & \end{bmatrix}$$

$$Z^{[1]} = W^{[1]} X + b^{[1]}$$
$$A^{[1]} = g^{[1]}(Z^{[1]})$$
$$Z^{[2]} = W^{[2]} A^{[1]} + b^{[2]}$$
$$A^{[2]} = g^{[2]}(Z^{[2]}) = \sigma(Z^{[2]})$$

$$L = -(y \log a + (1-y)\log(1-a))$$
$$dz = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \quad ; \quad a = \frac{1}{1+e^{-z}}$$
$$= (a-y)$$

$w \quad A^{[1]}$

$$dz^{[2]} = A^{[2]} - Y$$
$$dW^{[2]} = \frac{1}{m} dz^{[2]} A^{[1]T} \qquad z = w_1 x_1 + w_2 x_2 + \cdots$$

$$Z = \_\_\_\_ \qquad \frac{\partial Z}{\partial W}$$

$$db^{[2]} = \frac{1}{m} np.sum(dz^{[2]}, axis=1, keepdims=True)$$

$$\frac{\partial L}{\partial W^{[1]}} = \frac{\partial L}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial Z^{[2]}} \cdot \frac{\partial Z^{[2]}}{\partial a^{[1]}} \cdot \frac{\partial a^{[1]}}{\partial Z^{[1]}} \cdot \frac{\partial Z^{[1]}}{\partial W^{[1]}}$$
$$= \underbrace{\underbrace{\frac{\partial L}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial Z^{[2]}}}_{dZ^{[2]}} \; W^{[2]T} \; g^{[1]'}(Z^{[1]})}_{dZ^{[1]}} X^T$$

$A^{[0]} \quad W^{[1]} \quad W^{[2]}$

$$\frac{\partial L}{\partial W^{[1]}} = \underbrace{\frac{\partial L}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial z^{[2]}}}_{dz^{[2]}} \cdot \underbrace{\frac{\partial z^{[2]}}{\partial a^{[1]}}}_{W^{[2]T}} \underbrace{\frac{\partial a^{[1]}}{\partial z^{[1]}}}_{g'^{[1]}(z^{[1]})} \underbrace{\frac{\partial z^{[1]}}{\partial W^{[1]}}}_{X = A^{[0]}}$$

$$z_1^{[2]} = W_{11}^{[2]} a_1^{[1]} + W_{12}^{[2]} a_2^{[1]} + \dots + W_{1n^{[1]}}^{[2]} a_{n^{[1]}}^{[1]} + b$$

$$\begin{bmatrix} z_1^{[2]} \\ \vdots \\ z_{n^{[2]}}^{[2]} \end{bmatrix} = \begin{bmatrix} W_{11}^{[2]} & W_{12}^{[2]} & \dots & W_{1n^{[1]}}^{[2]} \\ W_{21}^{[2]} & W_{22}^{[2]} & & W_{2n^{[1]}}^{[2]} \\ W_{n^{[2]}1}^{[2]} & W_{n^{[2]}2}^{[2]} & & W_{n^{[2]}n^{[1]}}^{[2]} \end{bmatrix} \begin{bmatrix} a_1^{[1]} \\ \vdots \\ a_{n^{[1]}}^{[1]} \end{bmatrix}$$

$$\begin{bmatrix} W_{11}^{[2]} & W_{21}^{[2]} & W_{n^{[2]}1}^{[2]} \\ W_{1n^{[1]}}^{[2]} & W_{2n^{[1]}}^{[2]} & W_{n^{[2]}n^{[1]}}^{[2]} \end{bmatrix}$$

$x = (x, y, z)$

$W^T A + b$

$$\frac{\partial L}{\partial W^{[1]}} = \underbrace{\frac{\partial L}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial z^{[2]}}}_{dz^{[2]}} \cdot \underbrace{\frac{\partial z^{[2]}}{\partial a^{[1]}}}_{W^{[2]T}} \underbrace{\frac{\partial a^{[1]}}{\partial z^{[1]}}}_{g'^{[1]}(z^{[1]})} \underbrace{\frac{\partial z^{[1]}}{\partial W^{[1]}}}_{X = A^{[0]}}$$

$$Z_1^{[2]} = W_{11}^{[2]} a_1^{[1]} + W_{12}^{[2]} a_2^{[1]} + \dots + W_{1n^{[1]}}^{[2]} a_{n^{[1]}}^{[1]} + b$$

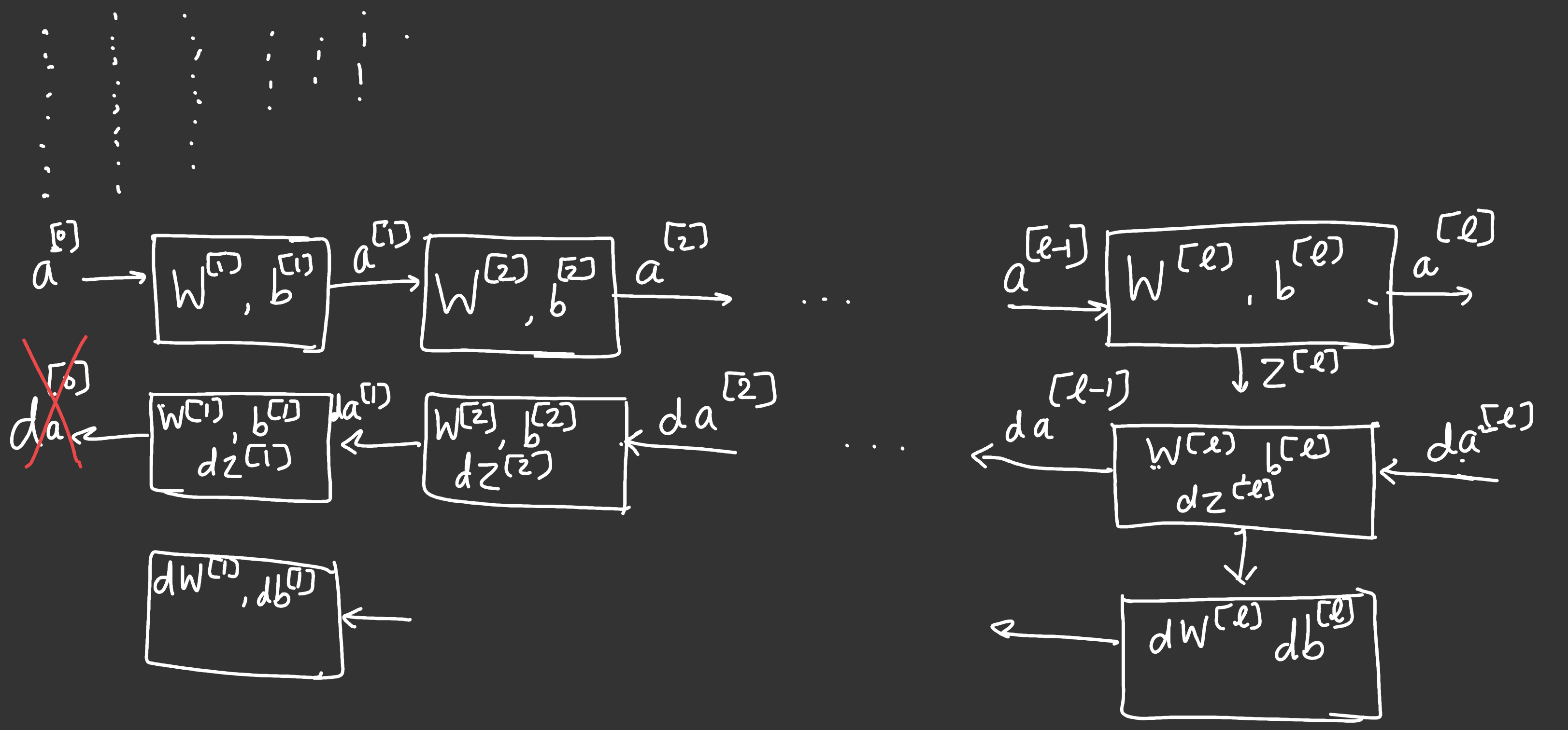
$$dz^{[1]} \quad \begin{bmatrix} Z_1^{[1]} \\ Z_2^{[1]} \\ \vdots \\ Z_{n^{[1]}}^{[1]} \end{bmatrix} = \begin{bmatrix} W_{11}^{[2]} & W_{21}^{[2]} & \dots & W_{n^{[2]}1}^{[2]} \\ \vdots & & & \vdots \\ W_{1n^{[1]}}^{[2]} & W_{2n^{[1]}}^{[2]} & \dots & W_{n^{[2]}n^{[1]}}^{[2]} \end{bmatrix} \begin{bmatrix} dZ_1^{[2]} \\ \vdots \\ dZ_{n^{[2]}}^{[2]} \end{bmatrix} * \begin{bmatrix} g'^{[1]}(Z_1^{[1]}) \\ g'^{[1]}(Z_2^{[1]}) \\ \vdots \\ g'^{[1]}(Z_{n^{[1]}}^{[1]}) \end{bmatrix}$$

matrix multiplication

$$(n^{[1]}, n^{[2]}) \quad (n^{[2]}, 1) \rightarrow (n^{[1]}, 1)$$

elementwise multiplication

03 Feb 2023


$a^{[0]}$
$W^{[1]}, b^{[1]}$
$a^{[1]}$
$W^{[2]}, b^{[2]}$
$a^{[2]}$
$a^{[l-1]}$
$W^{[l]}, b^{[l]}$
$a^{[l]}$
$z^{[l]}$
$da^{[0]}$
$W^{[1]}, b^{[1]}$
$dz^{[1]}$
$da^{[1]}$
$W^{[2]}, b^{[2]}$
$dz^{[2]}$
$da^{[2]}$
$da^{[l-1]}$
$W^{[l]}, b^{[l]}$
$dz^{[l]}$
$da^{[l]}$
$dW^{[1]}, db^{[1]}$
$dW^{[l]}$ $db^{[l]}$

$$dZ^{[l]} = dA^{[l]} * g'^{[l]}(Z^{[l]})$$

$$dW^{[l]} = \frac{1}{m} dZ^{[l]} A^{[l-1]^T}$$

$$db^{[l]} = \frac{1}{m} np.sum(dZ^{[l]}, axis=1, keepdims=True)$$

$$dA^{[l-1]} = W^{[l]^T} dZ^{[l]}$$