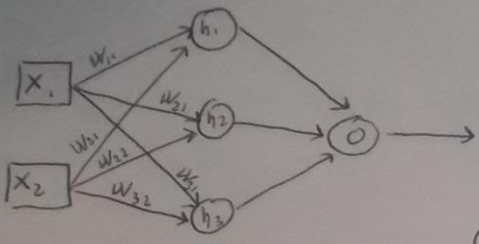
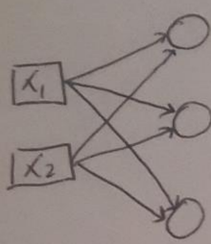


① 1° forward step



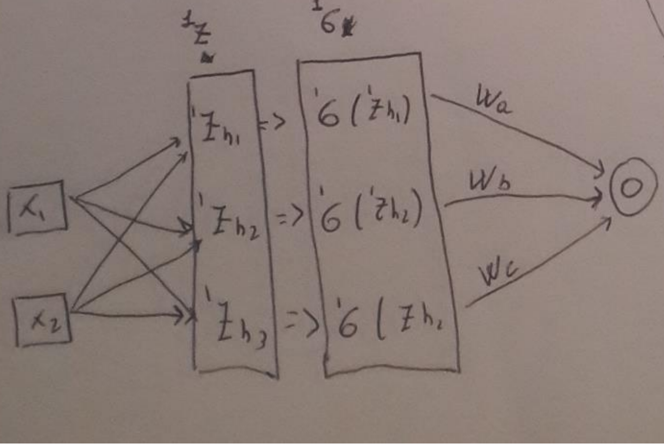
$$
\begin{aligned}
Z^1_{h_1} &= W_{11}\cdot x_1 + W_{12}x_2 + b_1 \\
Z^1_{h_2} &= W_{21}x_1 + W_{22}x_2 + b_2 \\
Z^1_{h_3} &= W_{31}x_1 + W_{32}x_2 + b_3
\end{aligned}
= \underbrace{\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \\ W_{31} & W_{32} \end{bmatrix}}_{\text{"w\_layer1"}} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_{\text{"b\_layer1"}}
$$

$$Z^1 = \text{W\_layer1} \cdot X + \text{b\_layer1} = W \cdot x + b$$

② 1° Activation

$$
\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \\ W_{31} & W_{32} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \longrightarrow \boxed{\frac{1}{1+e^{-z}}} \longrightarrow \sigma^1(z) = \begin{bmatrix} \frac{1}{1+e^{-z_{h_1}}} \\ \frac{1}{1+e^{-z_{h_2}}} \\ \frac{1}{1+e^{-z_{h_3}}} \end{bmatrix} = \begin{bmatrix} \sigma^1(z^1_{h_1}) \\ \sigma^1(z^1_{h_2}) \\ \sigma^1(z^1_{h_3}) \end{bmatrix}
$$

③ 2° forward step



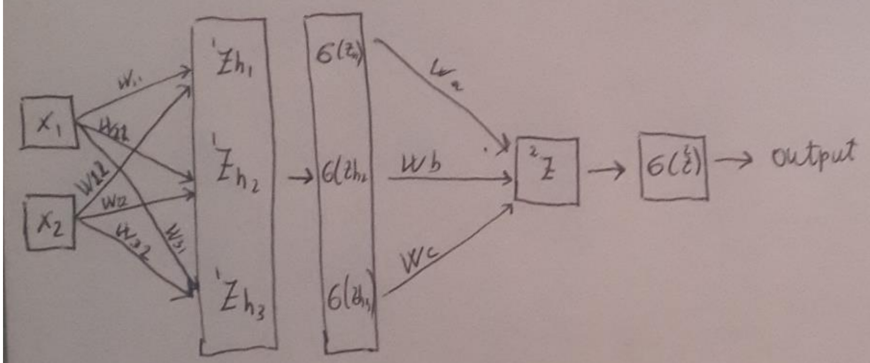
$$^2Z = W_a \cdot {}^1\sigma(Z_{h_1}) + W_b({}^1Z_{h_2}) + W_c \cdot \sigma(Z_{h_3}) + b_4$$

$$^{2}Z = \underbrace{[W_a \quad W_b \quad W_c]}_{\text{w\_layer2}} \cdot \begin{bmatrix} \sigma(Zh_1) \\ \sigma(Zh_2) \\ \sigma(Zh_3) \end{bmatrix} + \underbrace{[b_4]}_{\text{b\_layer 2}}$$

(4) 2° activation

$$^{2}Z = W_a\, \sigma(^{1}Zh_1) + W_b \cdot \sigma(Zh_2) + W_c\, \sigma(Zh_3) + b_4 \Rightarrow \boxed{\frac{1}{1+e^{-z}}} \rightarrow {}^{2}\sigma(^{2}z) = \text{output}$$

Summary



$\boxed{\text{Error}}$

$$\text{Loss} = \frac{1}{2}\left(\text{output} - y_{TRUE}\right)^2 = \frac{1}{2}\left(\sigma({}^{2}z) - y_T\right)^2$$

Parameters update

$$W_{11} = W_{11} - \eta \cdot \frac{\partial Loss}{\partial W_{11}}$$

η : learning rate

$$W_{12} = W_{12} - \eta \frac{\partial Loss}{\partial W_{12}}$$

$\vdots$

$$W_{32} = W_{32} - \eta \frac{\partial Loss}{\partial W_{32}}$$

$$b_1 = b_1 - \eta \frac{\partial Loss}{\partial b_1}$$

$\vdots$

$$b_4 = b_4 - \eta \frac{\partial Loss}{\partial b_4}$$

Derivatives computation

$$\frac{\partial L}{\partial W_{11}} = \frac{\partial L}{\partial \sigma({}^{2}z)} \cdot \frac{\partial \sigma({}^{2}z)}{\partial {}^{2}z} \cdot \frac{\partial {}^{2}z}{\partial \sigma(z_{h_1})} \cdot \frac{\partial \sigma(z_{h_1})}{\partial z_{h_1}} \cdot \frac{\partial z_{h_1}}{\partial W_{11}}$$

$$\frac{\partial L}{\partial \sigma({}^{2}z)} = \frac{1}{2} \cdot 2\left(\sigma({}^{2}z) - y_T\right) \cdot 1 = \sigma({}^{2}z) - y_T$$

$$\frac{\partial \sigma({}^{2}z)}{\partial {}^{2}z} = \sigma({}^{2}z)\left(1 - \sigma({}^{2}z)\right)$$

$$\frac{\partial {}^{2}z}{\partial \sigma(z_{h_1})} = W_a$$

$$\frac{\partial \sigma(z_{h_1})}{\partial z_{h_1}} = \sigma(z_{h_1})\left[1 - \sigma(z_{h_1})\right]$$

$$\frac{\partial z_{h_1}}{\partial W_{11}} = x_1$$

$$zh_1 = W_{11}x_1 + W_{12}x_2 + b_1 \rightarrow \boxed{\ } \rightarrow \sigma(zh_1)$$

$$zh_2 = W_{21}x_1 + W_{22}x_2 + b_2 \rightarrow \boxed{\ } \rightarrow \sigma(zh_2)$$

$$zh_3 = W_{31}x_1 + W_{32}x_2 + b_3 \rightarrow \boxed{\ } \rightarrow \sigma(zh_3)$$

$${}^{2}z = W_a\,\sigma(zh_1) + W_b\,\sigma(zh_2) + W_c\,\sigma(zh_3) + b_4 \rightarrow \boxed{\ } \rightarrow \sigma({}^{2}z) = \text{output}$$

$$\text{Loss} = \frac{1}{2}\left(\sigma({}^{2}z) - y\right)^2$$

$$\frac{\partial L}{\partial W_{11}} = \frac{\partial L}{\partial \sigma({}^{2}z)} \cdot \frac{\partial \sigma({}^{2}z)}{\partial {}^{2}z} \cdot \frac{\partial {}^{2}z}{\partial \sigma(zh_1)} \cdot \frac{\partial \sigma(zh_1)}{\partial zh_1} \cdot \frac{\partial zh_1}{\partial W_{11}} = \left[\sigma({}^{2}z) - y\right] \cdot \left[\sigma({}^{2}z)(1-\sigma({}^{2}z))\right] \cdot \left[W_a\right] \cdot \left[\sigma(zh_1)\cdot(1-\sigma(zh_1))\right] \cdot \left[x_1\right]$$

$$\frac{\partial L}{\partial W_{12}} = \frac{\partial L}{\partial \sigma({}^{2}z)} \cdot \frac{\partial \sigma({}^{2}z)}{\partial {}^{2}z} \cdot \frac{\partial {}^{2}z}{\partial \sigma(zh_1)} \cdot \frac{\partial \sigma(zh_1)}{\partial zh_1} \cdot \frac{\partial zh_1}{\partial W_{12}} = '' \cdot '' \cdot '' \cdot '' \cdot \left[x_2\right]$$

$$\frac{\partial L}{\partial W_{21}} = \frac{\partial L}{\partial \sigma({}^{2}z)} \cdot \frac{\partial \sigma({}^{2}z)}{\partial {}^{2}z} \cdot \frac{\partial {}^{2}z}{\partial \sigma(zh_2)} \cdot \frac{\partial \sigma(zh_2)}{\partial zh_2} \cdot \frac{\partial zh_2}{\partial W_{21}} = '' \cdot '' \cdot \left[W_b\right] \cdot \left[\sigma(zh_2)\cdot(1-\sigma(zh_2))\right] \cdot \left[x_1\right]$$

$$\frac{\partial L}{\partial W_{22}} = '' \cdot '' \cdot '' \cdot '' \cdot \frac{\partial zh_2}{\partial W_{22}} = '' \cdot '' \cdot '' \cdot '' \cdot \left[x_2\right]$$

$$\frac{\partial L}{\partial W_{31}} = '' \cdot '' \cdot \frac{\partial {}^{2}z}{\partial \sigma(zh_3)} \cdot \frac{\partial \sigma(zh_3)}{\partial zh_3} \cdot \frac{\partial zh_3}{\partial W_{31}} = '' \cdot '' \cdot \left[W_c\right] \cdot \left[\sigma(zh_3)(1-\sigma(zh_3))\right] \cdot \left[x_1\right]$$

$$\frac{\partial L}{\partial W_{32}} = '' \cdot '' \cdot '' \cdot '' \cdot \frac{\partial zh_3}{\partial W_{32}} = '' \cdot '' \cdot '' \cdot '' \cdot \left[x_2\right]$$

$$\frac{\partial L}{\partial b_1} = '' \cdot '' \cdot \frac{\partial {}^{2}z}{\partial \sigma(zh_1)} \cdot \frac{\partial \sigma(zh_1)}{\partial zh_1} \cdot \frac{\partial zh_1}{\partial b_1} = '' \cdot '' \cdot \left[W_a\right] \cdot \left[\sigma(zh_1)(1-\sigma(zh_1))\right] \cdot \left[1\right]$$

$$\frac{\partial L}{\partial b_2} = '' \cdot '' \cdot \frac{\partial {}^{2}z}{\partial \sigma(zh_2)} \cdot \frac{\partial \sigma(zh_2)}{\partial zh_2} \cdot \frac{\partial zh_2}{\partial b_2} = '' \cdot '' \cdot \left[W_b\right] \cdot \left[\sigma(zh_2)(1-\sigma(zh_2))\right] \cdot \left[1\right]$$

$$\frac{\partial L}{\partial b_3} = '' \cdot '' \cdot \frac{\partial {}^{2}z}{\partial \sigma(zh_3)} \cdot \frac{\partial \sigma(zh_3)}{\partial zh_3} \cdot \frac{\partial zh_3}{\partial b_3} = '' \cdot '' \cdot \left[W_c\right] \cdot \left[\sigma(zh_3)(1-\sigma(zh_3))\right] \cdot \left[1\right]$$

$$\frac{\partial L}{\partial W_a} = \frac{\partial L}{\partial \sigma({}^{2}z)} \cdot \frac{\partial \sigma({}^{2}z)}{\partial {}^{2}z} \cdot \frac{\partial {}^{2}z}{\partial W_a} = \left[\sigma({}^{2}z) - y\right] \cdot \left[\sigma({}^{2}z)(1-\sigma({}^{2}z))\right] \cdot \left[\sigma(zh_1)\right]$$

$$\frac{\partial L}{\partial W_b} = '' \cdot '' \cdot \frac{\partial {}^{2}z}{\partial W_b} = '' \cdot '' \cdot \left[\sigma(zh_2)\right]$$

$$\frac{\partial L}{\partial W_c} = '' \cdot '' \cdot \frac{\partial {}^{2}z}{\partial W_c} = '' \cdot '' \cdot \left[\sigma(zh_3)\right]$$

$$\frac{\partial L}{\partial b_4} = '' \cdot '' \cdot \frac{\partial {}^{2}z}{\partial b_4} = '' \cdot '' \cdot \left[1\right]$$