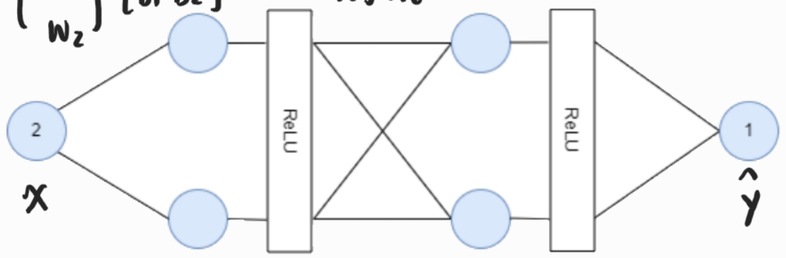
$\bar{W}_1$ $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ B_1 $[b_1\ b_2]$ $\bar{W}_2$ $\begin{bmatrix} w_3 & w_4 \\ w_5 & w_6 \end{bmatrix}$ B_2 $[b_3\ b_4]$ $\bar{W}_3$ $[w_7\ w_8]$ B_3 $[b_5]$



x $\hat{y}$

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & \text{others} \end{cases}$$

$$f'(x) = \begin{cases} 1, & x > 0 \\ 0, & \text{others} \end{cases}$$

Forward

$$x\bar{W}_1^T + B_1 = [xw_1 + b_1 \quad xw_2 + b_2] = \overset{O_1}{[S_1 \quad S_2]} \xrightarrow{f(x)} \overset{f(O_1)}{[f(S_1) \quad f(S_2)]}$$

$$O_1\bar{W}_2^T + B_2 = [w_3 f(S_1) + w_4 f(S_2) + b_3 \quad w_5 f(S_1) + w_6 f(S_2) + b_4]$$

$$= \overset{O_2}{[S_3 \quad S_4]} \xrightarrow{f(x)} \overset{f(O_2)}{[f(S_3) \quad f(S_4)]}$$

$$O_2\bar{W}_3^T + B_3 = w_7 f(S_3) + w_8 f(S_4) + b_5 = \hat{y}$$

$$\begin{cases} S_1 = xw_1 + b_1 \\ S_2 = xw_2 + b_2 \\ S_3 = w_3 f(S_1) + w_4 f(S_2) + b_3 \\ S_4 = w_5 f(S_1) + w_6 f(S_2) + b_4 \end{cases}$$

Gradient.

$$L = \frac{1}{2}(\hat{y} - y)^2 \rightarrow \frac{\partial L}{\partial \hat{y}} = (\hat{y} - y) = \Delta$$

$$f(g(x))$$

$$= g'(x) f'(g(x))$$

$$\frac{\partial L}{\partial B_3} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial B_3} = \Delta [1] = \Delta$$

$$\frac{\partial L}{\partial \bar{W}_3} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \bar{W}_3} = \Delta [f(S_3) \quad f(S_4)] = \Delta f(O_2)$$

$$\frac{\partial L}{\partial B_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial B_2} = \Delta [w_7 \cdot 1 \cdot f'(S_3) \quad w_8 \cdot 1 \cdot f'(S_4)]$$

$$= \Delta \bar{W}_3 \odot f'(O_2)$$

$\odot \rightarrow$ Elementwise multiplication

$$\frac{\partial L}{\partial \bar{W}_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \bar{W}_2} = \Delta \begin{bmatrix} w_7 f(S_1) f'(S_3) & w_7 f(S_2) f'(S_3) \\ w_8 f(S_1) f'(S_4) & w_8 f(S_2) f'(S_4) \end{bmatrix}$$

$$= \Delta \begin{bmatrix} w_7 f'(S_3) \\ w_8 f'(S_4) \end{bmatrix} \begin{bmatrix} f(S_1) & f(S_2) \end{bmatrix}$$

$$= \Delta \cdot (\bar{W}_3 \circ f'(O_2))^T \cdot O_1$$

$$\frac{\partial L}{\partial B_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial B_1}$$

$$= \Delta \begin{bmatrix} w_7 f'(S_3) w_3 f'(S_1) + w_8 f'(S_4) w_5 f'(S_1) & w_7 f'(S_3) w_4 f'(S_2) + w_8 f'(S_4) w_6 f'(S_2) \end{bmatrix}$$

$$= \Delta \begin{bmatrix} w_7 f'(S_3) & w_8 f'(S_4) \end{bmatrix} \begin{bmatrix} w_3 & w_4 \\ w_5 & w_6 \end{bmatrix} \begin{bmatrix} f'(S_1) & f'(S_2) \end{bmatrix}$$

$$= \Delta \, (\bar{W}_3 \circ f'(O_2)) \cdot \bar{W}_2 \circ f'(O_1)$$

$$\frac{\partial L}{\partial \bar{W}_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \bar{W}_1}$$

$$= \Delta \cdot \begin{bmatrix} w_7 f'(S_3) w_3 f'(S_1) \cdot x + w_8 f'(S_4) w_5 f'(S_1) \cdot x \\ w_7 f'(S_3) w_4 f'(S_2) \cdot x + w_8 f'(S_4) w_6 f'(S_2) \cdot x \end{bmatrix}$$

$$= \Delta \cdot x \begin{bmatrix} w_3 & w_5 \\ w_4 & w_6 \end{bmatrix} \begin{bmatrix} w_7 f'(S_3) \\ w_8 f'(S_4) \end{bmatrix} \circ \begin{bmatrix} f'(S_1) \\ f'(S_2) \end{bmatrix}$$

$$= \Delta \cdot x \cdot \bar{W}_2^T \cdot (\bar{W}_3 \circ f'(O_2))^T \circ f'(O_1)^T$$

The calculation processes are in the excel file.

$$\boxed{\text{Iteration } 0}$$

Forward

$$x W_1^T + B_1 = 2\begin{bmatrix}1 & 1\end{bmatrix} + \begin{bmatrix}-0.5 & -0.5\end{bmatrix} = \overset{O_1}{\begin{bmatrix}1.5 & 1.5\end{bmatrix}} \xrightarrow{f(O_1)} \begin{bmatrix}1.5 & 1.5\end{bmatrix}$$

$$f(O_1) W_2^T + B_2 = \begin{bmatrix}1.5 & 1.5\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} + \begin{bmatrix}-0.5 & -0.5\end{bmatrix} = \overset{O_2}{\begin{bmatrix}2.5 & 2.5\end{bmatrix}} \xrightarrow{f(O_2)} \begin{bmatrix}2.5 & 2.5\end{bmatrix}$$

$$f(O_2) W_3^T + B_3 = \begin{bmatrix}2.5 & 2.5\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} + \begin{bmatrix}-0.5\end{bmatrix} = \overset{\hat{y}}{4.5}$$

Backpropagation $(\eta = 0.01)$

$$W_1 \leftarrow W_1 - \eta \frac{\partial L}{\partial W_1} = \begin{bmatrix}1\\ 1\end{bmatrix} - 0.01\begin{bmatrix}14\\ 14\end{bmatrix} = \begin{bmatrix}0.86\\ 0.86\end{bmatrix}$$

$$B_1 \leftarrow B_1 - \eta \frac{\partial L}{\partial B_1} = \begin{bmatrix}-0.5 & -0.5\end{bmatrix} - 0.01\begin{bmatrix}7 & 7\end{bmatrix} = \begin{bmatrix}-0.57 & -0.57\end{bmatrix}$$

$$W_2 \leftarrow W_2 - \eta \frac{\partial L}{\partial W_2} = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} - 0.01\begin{bmatrix}5.25 & 5.25\\ 5.25 & 5.25\end{bmatrix} = \begin{bmatrix}0.9475 & 0.9475\\ 0.9475 & 0.9475\end{bmatrix}$$

$$B_2 \leftarrow B_2 - \eta \frac{\partial L}{\partial B_2} = \begin{bmatrix}-0.5 & -0.5\end{bmatrix} - 0.01\begin{bmatrix}3.5 & 3.5\end{bmatrix} = \begin{bmatrix}-0.535 & -0.535\end{bmatrix}$$

$$W_3 \leftarrow W_3 - \eta \frac{\partial L}{\partial W_3} = \begin{bmatrix}1 & 1\end{bmatrix} - 0.01\begin{bmatrix}8.75 & 8.75\end{bmatrix} = \begin{bmatrix}0.9125 & 0.9125\end{bmatrix}$$

$$B_3 \leftarrow B_3 - \eta \frac{\partial L}{\partial B_3} = -0.5 - 0.01 \cdot 3.5 = -0.535$$

Iteration 1

Forward

$$xW_1^T + B_1 = 2\,[0.86\ 0.86] + [-0.57\ \ -0.57] = \overset{O_1}{[1.15\ \ 1.15]}$$

$$\xrightarrow{f(O_1)} [1.15\ \ 1.15]$$

$$f(O_1)W_2^T + B_2 = [1.15\ 1.15]\begin{bmatrix} 0.9475 & 0.9475 \\ 0.9475 & 0.9475 \end{bmatrix} + [-0.535\ -0.535] = \overset{O_2}{[1.644\ \ 1.644]}$$

$$\xrightarrow{f(O_2)} [1.644\ \ 1.644]$$

$$f(O_2)W_3^T + B_3 = [1.644\ 1.644]\begin{bmatrix} 0.9125 \\ 0.9125 \end{bmatrix} + [-0.535] = \overset{\hat{y}}{2.466}$$

Backpropagation $(\eta = 0.01)$

$$W_1 \leftarrow W_1 - \eta\frac{\partial L}{\partial W_1} = \begin{bmatrix} 0.86 \\ 0.86 \end{bmatrix} - 0.01\begin{bmatrix} 5.0691 \\ 5.0691 \end{bmatrix} = \begin{bmatrix} 0.8093 \\ 0.8093 \end{bmatrix}$$

$$B_1 \leftarrow B_1 - \eta\frac{\partial L}{\partial B_1} = [-0.57\ -0.57] - 0.01[2.5436\ 2.5436] = [-0.5953\ \ -0.5953]$$

$$W_2 \leftarrow W_2 - \eta\frac{\partial L}{\partial W_2} = \begin{bmatrix} 0.9475 & 0.9475 \\ 0.9475 & 0.9475 \end{bmatrix} - 0.01\begin{bmatrix} 1.5381 & 1.5381 \\ 1.5381 & 1.5381 \end{bmatrix} = \begin{bmatrix} 0.9321 & 0.9321 \\ 0.9321 & 0.9321 \end{bmatrix}$$

$$B_2 \leftarrow B_2 - \eta\frac{\partial L}{\partial B_2} = [-0.535\ \ -0.535] - 0.01[1.3375\ \ 1.3375] = [-0.5484\ \ -0.5484]$$

$$W_3 \leftarrow W_3 - \eta\frac{\partial L}{\partial W_3} = [0.9125\ 0.9125] - 0.01\,[2.4101\ \ 2.4101] = [0.8884\ \ 0.8884]$$

$$B_3 \leftarrow B_3 - \eta\frac{\partial L}{\partial B_3} = -0.535 - 0.01\cdot 1.4658 = -0.5497$$

Iteration 2

Forward

$$\begin{array}{c} \\ x W_1^T + B_1 = 2[0.8093 \quad 0.8093] + [-0.5953 \quad -0.5953] = \end{array} \begin{array}{c} O_1 \\ [1.0233 \quad 1.0233] \end{array}$$

$$f(O_1)$$

$$\rightarrow [1.0233 \quad 1.0233]$$

$$\begin{array}{c} \\ f(O_1) W_2^T + B_2 = [1.0233 \quad 1.0233] \begin{bmatrix} 0.9321 & 0.9321 \\ 0.9321 & 0.9321 \end{bmatrix} + [-0.5484 \quad -0.5484] = \end{array} \begin{array}{c} O_2 \\ [1.3592 \quad 1.3592] \end{array}$$

$$f(O_2)$$

$$\rightarrow [1.3592 \quad 1.3592]$$

$$\begin{array}{c} \\ f(O_2) W_3^T + B_3 = [1.3592 \quad 1.3592] \begin{bmatrix} 0.8884 \\ 0.8884 \end{bmatrix} + [-0.5499] = \end{array} \begin{array}{c} \hat{y} \\ 1.8654 \end{array}$$