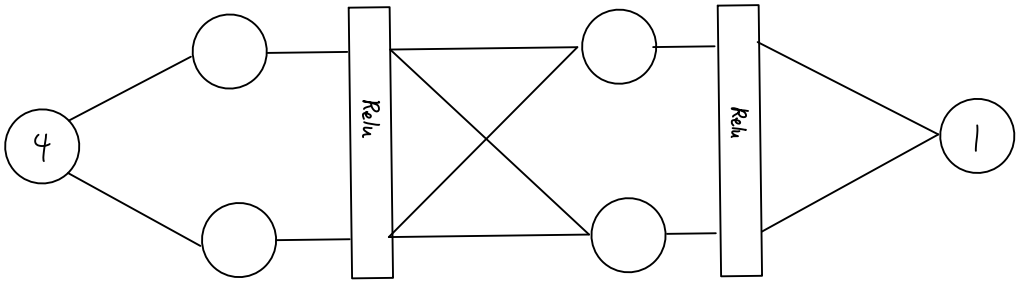
Answer

First update - weight

	Layer 1	Layer 2	Layer 3
Parameter	$W^{[0]} = \begin{bmatrix}1\\1\end{bmatrix}, b^{[0]} = \begin{bmatrix}-0.5\\-0.5\end{bmatrix}$	$W^{[1]} = \begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}, b^{[1]} = \begin{bmatrix}-0.5\\-0.5\end{bmatrix}$	$W^{[2]} = [1 \quad 1], b^{[2]} = [-0.5]$
Gradient	$W^{[0]} = \begin{bmatrix}92\\92\end{bmatrix}, b^{[0]} = \begin{bmatrix}23\\23\end{bmatrix}$	$dLdW^{[1]} = \begin{bmatrix}40.25 & 40.25\\40.25 & 40.25\end{bmatrix}, dLdb^{[1]} = \begin{bmatrix}11.5\\11.5\end{bmatrix}$	$dLdW^{[2]} = [74.75 \quad 74.75], dLdb^{[2]} = [11.5]$
Updated	$W^{[0]} = \begin{bmatrix}0.08\\0.08\end{bmatrix}, b^{[0]} = \begin{bmatrix}-0.73\\-0.73\end{bmatrix}$	$W^{[1]} = \begin{bmatrix}0.5975 & 0.5975\\0.5975 & 0.5975\end{bmatrix}, b^{[1]} = \begin{bmatrix}-0.615\\-0.615\end{bmatrix}$	$W^{[2]} = [0.2525 \quad 0.2525], b^{[2]} = [-0.615]$

Second update - bias

	Layer 1	Layer 2	Layer 3
Parameter	$W^{[0]} = \begin{bmatrix}0.08\\0.08\end{bmatrix}, b^{[0]} = \begin{bmatrix}-0.73\\-0.73\end{bmatrix}$	$W^{[1]} = \begin{bmatrix}0.5975 & 0.5975\\0.5975 & 0.5975\end{bmatrix}, b^{[1]} = \begin{bmatrix}-0.615\\-0.615\end{bmatrix}$	$W^{[2]} = [0.2525 \quad 0.2525], b^{[2]} = [-0.615]$
Gradient	$W^{[0]} = \begin{bmatrix}0\\0\end{bmatrix}, b^{[0]} = \begin{bmatrix}0\\0\end{bmatrix}$	$dLdW^{[1]} = \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}, dLdb^{[1]} = \begin{bmatrix}0\\0\end{bmatrix}$	$dLdW^{[2]} = [0 \quad 0], dLdb^{[2]} = [-1.615]$
Updated	$W^{[0]} = \begin{bmatrix}0.08\\0.08\end{bmatrix}, b^{[0]} = \begin{bmatrix}-0.73\\-0.73\end{bmatrix}$	$W^{[1]} = \begin{bmatrix}0.5975 & 0.5975\\0.5975 & 0.5975\end{bmatrix}, b^{[1]} = \begin{bmatrix}-0.615\\-0.615\end{bmatrix}$	$W^{[2]} = [0.2525 \quad 0.2525], b^{[2]} = [-0.59885]$



learning rate : 0.01 (η) Loss = $\frac{1}{2}(y-y^*)^2$

All weights : 1 ∇ Loss = y^*-y

All bias : -0.5

Relu Forward:

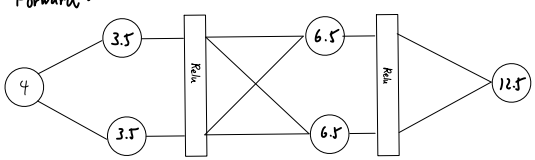
$$R(z)=\begin{cases} z, & \text{if } z>0 \\ 0, & \text{if } z\le 0 \end{cases}$$

Relu Backward:

$$R'(z)=\begin{cases} 1, & \text{if } z>0 \\ 0, & \text{if } z\le 0 \end{cases}$$

Iteration 0:

Forward:



$$W^{[0]}=\begin{bmatrix}1\\1\end{bmatrix} \quad W^{[1]}=\begin{bmatrix}1&1\\1&1\end{bmatrix} \quad W^{[2]}=(1\ \ 1)$$

$$b^{[0]}=\begin{bmatrix}-0.5\\-0.5\end{bmatrix} \quad b^{[1]}=\begin{bmatrix}-0.5\\-0.5\end{bmatrix} \quad b^{[2]}=[-0.5]$$

$$Z^{[0]}=W^{[0]}A^{[0]}=\begin{bmatrix}1\\1\end{bmatrix}[4]+\begin{bmatrix}-0.5\\-0.5\end{bmatrix}=\begin{bmatrix}3.5\\3.5\end{bmatrix}$$

$$A^{[1]}=\text{Relu}\left(\begin{bmatrix}3.5\\3.5\end{bmatrix}\right)=\begin{bmatrix}3.5\\3.5\end{bmatrix}$$

$$Z^{[1]}=\begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}3.5\\3.5\end{bmatrix}+\begin{bmatrix}-0.5\\-0.5\end{bmatrix}=\begin{bmatrix}6.5\\6.5\end{bmatrix}$$

$$A^{[2]}=\text{Relu}\left(\begin{bmatrix}6.5\\6.5\end{bmatrix}\right)=\begin{bmatrix}6.5\\6.5\end{bmatrix}$$

$$Z^{[2]}=\begin{bmatrix}1&1\end{bmatrix}\begin{bmatrix}6.5\\6.5\end{bmatrix}+[-0.5]=[12.5]$$

Backward:

$$dLdZ^{[2]}=12.5-1=11.5$$

$$dLdb^{[2]}=dLdZ^{[2]}=[11.5]$$

$$dLdW^{[2]}=dLdZ^{[2]}\cdot A^{[2]T}=\begin{bmatrix}11.5\end{bmatrix}\begin{bmatrix}6.5&6.5\end{bmatrix}=\begin{bmatrix}74.75&74.75\end{bmatrix}$$

$$dLdA^{[2]}=W^{[2]T}\cdot dLdZ^{[2]}=\begin{bmatrix}1\\1\end{bmatrix}[11.5]=\begin{bmatrix}11.5\\11.5\end{bmatrix}$$

$$dLdZ^{[1]}=R'(Z^{[1]})\times dLdA^{[2]}=\begin{bmatrix}1\\1\end{bmatrix}\times\begin{bmatrix}11.5\\11.5\end{bmatrix}=\begin{bmatrix}11.5\\11.5\end{bmatrix}$$

$$dLdb^{[1]}=\begin{bmatrix}11.5\\11.5\end{bmatrix}$$

$$dLdW^{[1]}=\begin{bmatrix}11.5\\11.5\end{bmatrix}\begin{bmatrix}3.5&3.5\end{bmatrix}=\begin{bmatrix}40.25&40.25\\40.25&40.25\end{bmatrix}$$

$$dLdA^{[1]}=\begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}11.5\\11.5\end{bmatrix}=\begin{bmatrix}23\\23\end{bmatrix}$$

$$dLdZ^{[0]}=\begin{bmatrix}1\\1\end{bmatrix}\times\begin{bmatrix}23\\23\end{bmatrix}=\begin{bmatrix}23\\23\end{bmatrix}=dLdb^{[0]}$$

$$dLdW^{[0]}=\begin{bmatrix}23\\23\end{bmatrix}[4]=\begin{bmatrix}92\\92\end{bmatrix}$$

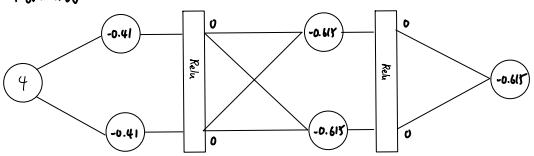
Update weight and bias:

$$W^{[0]}=W^{[0]}-\eta\cdot dLdW^{[0]}=\begin{bmatrix}1\\1\end{bmatrix}-0.01\begin{bmatrix}92\\92\end{bmatrix}=\begin{bmatrix}0.08\\0.08\end{bmatrix} \quad W^{[1]}=\begin{bmatrix}1&1\\1&1\end{bmatrix}-0.01\begin{bmatrix}40.25&40.25\\40.25&40.25\end{bmatrix}=\begin{bmatrix}0.5975&0.5975\\0.5975&0.5975\end{bmatrix} \quad W^{[2]}=[0.2525\ \ 0.2525]$$

$$b^{[0]}=b^{[0]}-\eta\cdot dLdb^{[0]}=\begin{bmatrix}-0.5\\-0.5\end{bmatrix}-0.01\begin{bmatrix}23\\23\end{bmatrix}=\begin{bmatrix}-0.73\\-0.73\end{bmatrix} \quad b^{[1]}=\begin{bmatrix}-0.5\\-0.5\end{bmatrix}-0.01\begin{bmatrix}11.5\\11.5\end{bmatrix}=\begin{bmatrix}-0.615\\-0.615\end{bmatrix} \quad b^{[2]}=[-0.615]$$

Iteration 1:

Forward:



$W^{[0]} = \begin{bmatrix} 0.08 \\ 0.08 \end{bmatrix}$ $W^{[1]} = \begin{bmatrix} 0.5995 & 0.5995 \\ 0.5995 & 0.5995 \end{bmatrix}$ $W^{[2]} = [0.2525 \ 0.2525]$

$b^{[0]} = \begin{bmatrix} -0.73 \\ -0.73 \end{bmatrix}$ $b^{[1]} = \begin{bmatrix} -0.615 \\ -0.615 \end{bmatrix}$ $b^{[2]} = [-0.615]$

$$Z^{[0]} = W^{[0]} A^{[0]} = \begin{bmatrix} 0.08 \\ 0.08 \end{bmatrix} [4] + \begin{bmatrix} -0.73 \\ -0.73 \end{bmatrix} = \begin{bmatrix} -0.41 \\ -0.41 \end{bmatrix}$$

$$A^{[1]} = Relu\left(\begin{bmatrix} -0.41 \\ -0.41 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Z^{[1]} = \begin{bmatrix} 0.5995 & 0.5995 \\ 0.5995 & 0.5995 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -0.615 \\ -0.615 \end{bmatrix} = \begin{bmatrix} -0.615 \\ -0.615 \end{bmatrix}$$

$$A^{[2]} = Relu\left(\begin{bmatrix} -0.615 \\ -0.615 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Z^{[2]} = [0.2525 \ 0.2525] \begin{bmatrix} 0 \\ 0 \end{bmatrix} + [-0.615] = [-0.615]$$

Backward:

$$dL/dZ^{[2]} = -0.615 - 1 = -1.615$$

$$dL/db^{[2]} = dL/dZ^{[2]} = [-1.615]$$

$$dL/dW^{[2]} = dL/dZ^{[2]} \cdot A^{[1]T} = [-1.615] \begin{bmatrix} 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

$$dL/dA^{[2]} = W^{[2]T} \cdot dL/dZ^{[2]} = \begin{bmatrix} 0.2525 \\ 0.2525 \end{bmatrix} [-1.615] = \begin{bmatrix} -0.40798875 \\ -0.40798875 \end{bmatrix}$$

$$dL/dZ^{[1]} = R'(Z^{[1]}) \times dL/dA^{[2]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$dL/db^{[1]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$dL/dW^{[1]} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$dL/dA^{[1]} = \begin{bmatrix} 0.5995 & 0.5995 \\ 0.5995 & 0.5995 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$dL/dZ^{[0]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = dL/db^{[0]}$$

$$dL/dW^{[0]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Update weight and bias:

$$W^{[0]} = \begin{bmatrix} 0.08 \\ 0.08 \end{bmatrix} - \eta \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.08 \\ 0.08 \end{bmatrix} \qquad W^{[1]} = \begin{bmatrix} 0.5995 & 0.5995 \\ 0.5995 & 0.5995 \end{bmatrix} - 0.01 \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.5995 & 0.5995 \\ 0.5995 & 0.5995 \end{bmatrix} \qquad W^{[2]} = [0.2525 \ 0.2525]$$

$$b^{[0]} = \begin{bmatrix} -0.73 \\ -0.73 \end{bmatrix} - \eta \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.73 \\ -0.73 \end{bmatrix} \qquad b^{[1]} = \begin{bmatrix} -0.615 \\ -0.615 \end{bmatrix} - 0.01 \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.615 \\ -0.615 \end{bmatrix} \qquad b^{[2]} = [-0.59885]$$