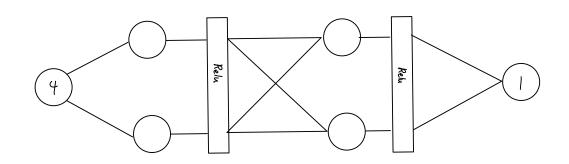
Answer

First update - weight

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

Second update - bias

	Layer 1	Layer 2	Layer 3	
Parameter	$W^{[0]} = egin{bmatrix} 0.08 \ 0.08 \end{bmatrix}$, $b^{[0]} = egin{bmatrix} -0.73 \ -0.73 \end{bmatrix}$	$W^{[1]} = egin{bmatrix} 0.5975 & 0.5975 \ 0.5975 & 0.5975 \end{bmatrix}, b^{[1]} = egin{bmatrix} -0.615 \ -0.615 \end{bmatrix}$	$m{W}^{[2]} = [0.2525 0.2525], b^{[2]} = [-0.615]$	
Gradient	$W^{[0]}=egin{bmatrix} 0 \ 0 \end{bmatrix}$, $b^{[0]}=egin{bmatrix} 0 \ 0 \end{bmatrix}$	$dLdW^{[1]} = egin{bmatrix} 0 & 0 \ 0 & 0 \end{bmatrix}, dLdb^{[1]} = egin{bmatrix} 0 \ 0 \end{bmatrix}$	$dLdW^{[2]} = egin{bmatrix} 0 & 0 \end{bmatrix}, dLdb^{[2]} = egin{bmatrix} -1.615 \end{bmatrix}$	
Updated	$W^{[0]} = egin{bmatrix} 0.08 \ 0.08 \end{bmatrix}, b^{[0]} = egin{bmatrix} -0.73 \ -0.73 \end{bmatrix}$	$m{W}^{[1]} = egin{bmatrix} 0.5975 & 0.5975 \ 0.5975 & 0.5975 \end{bmatrix}, m{b}^{[1]} = egin{bmatrix} -0.615 \ -0.615 \end{bmatrix}$	$m{W}^{[2]} = [0.2525 0.2525], b^{[2]} = [-0.59885]$]
4				•



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

All weights: $| \nabla loss = y^*-y$

All bias: -0.5

Iteration $0:$

Forward:

Reh $M_{(i)}^{(i)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad M_{(i)}^{(i)} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad M_{(i)}^{(i)} = \begin{bmatrix} 1 & 1 \end{bmatrix}$ $P_{(0)} = \begin{bmatrix} -0.2 \\ -0.2 \end{bmatrix}$ $P_{(1)} = \begin{bmatrix} -0.2 \\ -0.2 \end{bmatrix}$ $P_{(1)} = \begin{bmatrix} -0.2 \\ -0.2 \end{bmatrix}$ 7 [0] = W [0] A [0]

 $= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 4 \end{bmatrix} + \begin{bmatrix} -\alpha 5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 3.5 \\ 3.5 \end{bmatrix}$

 $A^{[1]} \cdot Relu\left(\begin{bmatrix} 3.5\\3.5 \end{bmatrix}\right) \cdot \begin{bmatrix} 3.5\\3.5 \end{bmatrix}$ $Z \stackrel{\text{OJ}}{=} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3.5 \\ 3.5 \end{bmatrix} + \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 6.5 \\ 6.5 \end{bmatrix}$

 $Z^{(2)} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 6.5 \\ 6.5 \end{bmatrix} + \begin{bmatrix} -6.5 \end{bmatrix} = \begin{bmatrix} 12.5 \end{bmatrix}$

Update weight and bias:

 $A^{(1)}$: Relu $\left(\begin{bmatrix} 6.5\\6.5 \end{bmatrix}\right)$: $\begin{bmatrix} 6.5\\6.5 \end{bmatrix}$

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

Kelu Backward:

dl dz(2) = 12,5 - 1= 11.5 dldb [= dld [[11.5] dldw [2] = dld7 [1] a [2]T = [11.5] [6.5 6.5] = [74.75 74.75] dldA(E) = W(E)T. dldZ(E)

Relu Forward:

Backward:

125

 $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ [n.t] $\begin{bmatrix} 115 \\ 0.5 \end{bmatrix}$ $d(d|z^{(1)} = R'(z^0) \times d(dR^{(2)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 11.7 \\ 11.7 \end{bmatrix} = \begin{bmatrix} 11.7 \\ 11.7 \end{bmatrix}$ dldb(1) = [11.5]

 $dldW^{(1)} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} [3.5 \ 3.7] = \begin{bmatrix} 40.27 & 40.27 \\ 40.25 & 40.25 \end{bmatrix}$

 $\operatorname{dldA}^{(r)} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 10.5 \\ 10.7 \end{bmatrix} = \begin{bmatrix} 23 \\ 23 \end{bmatrix}$ $\operatorname{did}_{(0)}: \begin{bmatrix} 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} n \\ n \end{bmatrix} = \begin{bmatrix} n \\ n \end{bmatrix} = \operatorname{did}_{(0)}$ d(dw (0) = [23][4] = [92]

 $b^{(1)} = b^{(1)} - \gamma \cdot d(a)b^{(0)}, \begin{bmatrix} -0x \\ -0x \end{bmatrix} - 0.01 \begin{bmatrix} 21 \\ 21 \end{bmatrix} = \begin{bmatrix} -0.13 \\ -0.73 \end{bmatrix} \qquad b^{(1)} \cdot \begin{bmatrix} -0x \\ -0x \end{bmatrix} - 0.01 \begin{bmatrix} 11.5 \\ 11.7 \end{bmatrix} = \begin{bmatrix} -0.017 \\ -0.017 \end{bmatrix}$

b (1) [- 0.617]

Iteration 1:

Forward: æ -0.613 $W^{(s)} = \begin{bmatrix} 0.08 \\ 0.08 \end{bmatrix}$ $W^{(1)} = \begin{bmatrix} 0.5717 & 0.5717 \\ 0.5717 & 0.7777 \end{bmatrix}$ $W^{(1)} = \begin{bmatrix} 0.2727 & 0.2727 \end{bmatrix}$ $P_{(0)} = \begin{bmatrix} -0.41 \\ -0.41 \end{bmatrix}$ $P_{(1)} = \begin{bmatrix} -0.912 \\ -0.912 \end{bmatrix}$ $P_{(2)} = \begin{bmatrix} -0.912 \\ -0.912 \end{bmatrix}$ 7 [0] = W [0] A [0] $= \begin{bmatrix} 0.08 \\ 0.08 \end{bmatrix} \begin{bmatrix} 4 \end{bmatrix} + \begin{bmatrix} -0.73 \\ -0.71 \end{bmatrix} = \begin{bmatrix} -0.41 \\ -0.41 \end{bmatrix}$ A [1] = Relu ([-0.4]) = [0] $Z = \begin{bmatrix} 0.5995 & 0.5997 \\ 0.5977 & 0.5977 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -0.617 \\ -0.617 \end{bmatrix} = \begin{bmatrix} -0.617 \\ -0.617 \end{bmatrix}$ A [2] = Relu ([-0.617]) = [0] Z (2) = [0.2727 0.2727] [0] + [-0.617] = [-0.617]

Backward: $dl dZ^{(s)} = -0.617 - 1 = -1.617$ $dl db^{(s)} = dl dZ^{(s)} [-1.617]$ $dl dw^{(s)} = dl dZ^{(s)} a^{(s)T}$ = [-1.617] [0 0] = [0 0] $dl dA^{(s)} = w^{(s)T} \cdot dl dZ^{(s)}$ = [0.2727] [-1.615] = [-0.4071777] $dl dZ^{(s)} = R^{s}(z^{0}) \times dl dA^{(s)} = [0]$ $dl db^{(s)} = [0]$ $dl dw^{(s)} = [0]$ $dl dw^{(s)} = [0]$ $dl dw^{(s)} = [0.3777 \cdot 0.5777] [0] = [0 0]$ $dl dd^{(s)} = [0.3777 \cdot 0.3777] [0] = [0 0]$ $dl dd^{(s)} = [0]$ $dl dd^{(s)} = [0]$

Update weight and bias:

$$W^{(a)} = \begin{bmatrix} 0.08 \\ 0.08 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.08 \\ 0.08 \end{bmatrix} \qquad W^{(a)} = \begin{bmatrix} 0.5917 & 0.5917 \\ 0.5917 & 0.7917 \end{bmatrix} - 0.01 \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5917 & 0.5917 \end{bmatrix}$$

$$W^{(a)} = \begin{bmatrix} 0.0917 & 0.5917 \\ 0.5917 & 0.7917 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.713 \\ -0.713 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.617 \\ -0.617 \end{bmatrix} - 0.01 \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.617 \\ -0.617 \end{bmatrix}$$

$$W^{(a)} = \begin{bmatrix} -0.713 \\ 0.5917 & 0.7917 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.713 \\ -0.713 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.713 \\ -0.713 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.713 \\ -0.713 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.713 \\ -0.713 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.713 \\ -0.713 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.713 \\ -0.713 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.713 \\ -0.713 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.713 \\ -0.713 \end{bmatrix} - V \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.617 \\ -0.617 \end{bmatrix} - 0.01 \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.617 \\ -0.617 \end{bmatrix}$$