Quiz Solution

1) Need $x_1 > x_2 > x_3 > x_4 > x_5$

We can ensure this by using their difference

$$d_i = x_i - x_{i+1} \quad \forall \ i \in \{1,2,3,4,5\}$$

We have 4 nodes in the hidden layer, and each of them could be fed with one difference d_i

$$d_i = x_i - x_{i+1} > 0 \iff h_i > 1 \ \checkmark \ \text{Ok}$$

$$d_i = x_i - x_{i+1} = 0 \iff h_i = 0 \ \times \ \text{Not ok}$$

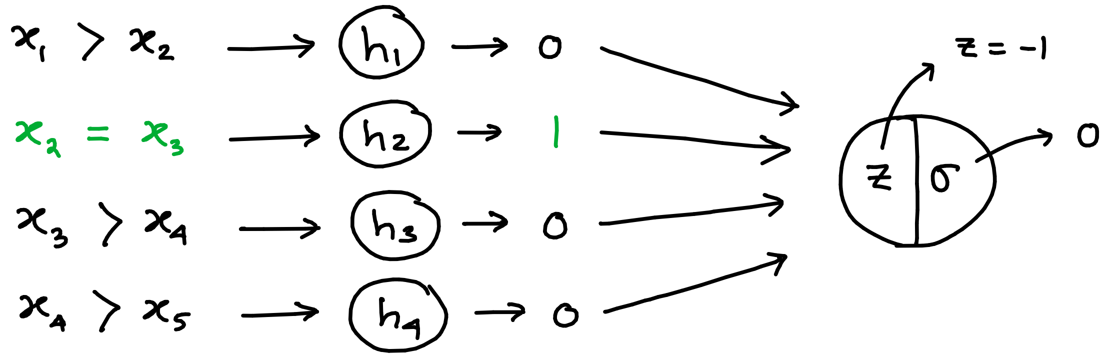
$$d_i = x_i - x_{i+1} < 0 \iff h_i < 0 \ \checkmark \ \text{Ok}$$

Instead, consider $d_i = x_{i+1} - x_i$

$$d_i \geqslant 0 \rightarrow h_i \rightarrow 1$$

$$d_i < 0 \rightarrow h_i \rightarrow 0$$

If all $d_i < 0$, then all $h_i = 0$, and then we would want the output $y = 1$. We could next set $\underline{w}^{(2)} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$



Therefore, one possible solution would be:

(a) $$\underline{\underline{W}}^{(1)} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

(b) $$\underline{b}^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T$$

(c) $$\underline{w}^{(2)} = \begin{bmatrix} -1 & -1 & -1 & -1 \end{bmatrix}^T$$

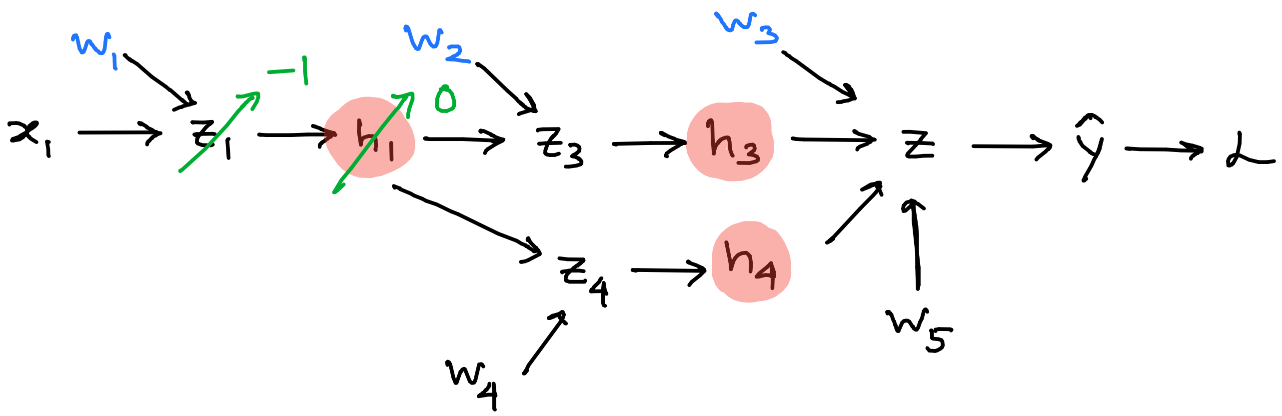
(d) $$b^{(2)} = 0$$

If the model correctly outputs $x_1 > x_2 > x_3 > x_4 > x_5$ for all possible values of x_i's $\rightarrow$ 3

else zero marks

2) Let's draw a partial computational graph

(not necessary to draw)

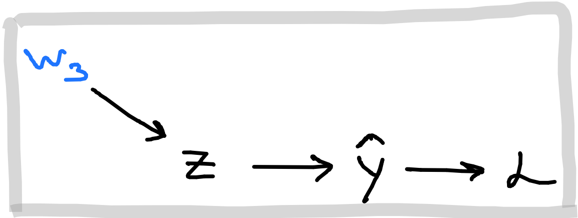


(a) $\dfrac{\partial L}{\partial w_3} = \underbrace{\dfrac{\partial L}{\partial \hat{y}}}_{\Downarrow \text{ Any value}} \underbrace{\dfrac{\partial \hat{y}}{\partial z}}_{\Downarrow \text{ Any value}} \underbrace{\dfrac{\partial z}{\partial w_3}}_{\Downarrow \text{ Any value}}$

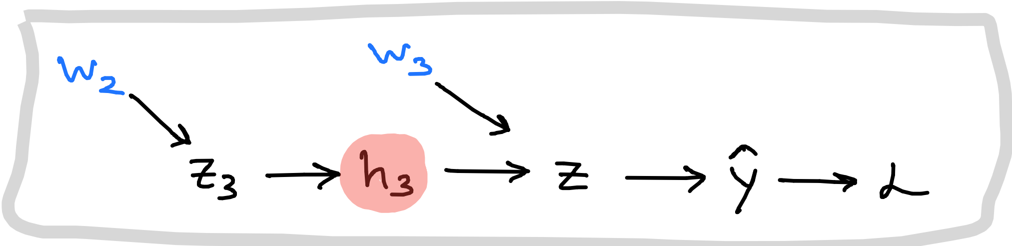
(NO)

Not necessarily zero

0.5



(b)



$$\frac{\partial L}{\partial w_2} = \underbrace{\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z}}_{\Downarrow \text{ Any value}} \underbrace{\frac{\partial z}{\partial h_3}}_{\Downarrow \text{ Any value}} \underbrace{\frac{\partial h_3}{\partial z_3}}_{\Downarrow \text{ Any value}} \cancel{\frac{\partial z_3}{\partial w_2}}^{\,0}$$

Zero (YES)

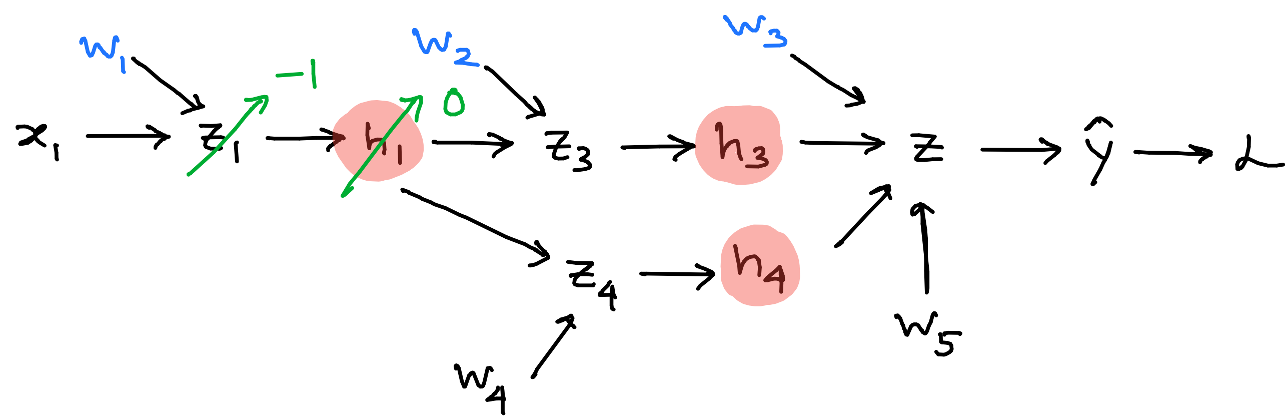
0.75

$\dfrac{\partial z}{\partial h_3} = w_3 \Rightarrow$ Any value

$\dfrac{\partial z_3}{\partial w_2} = h_1 = 0$

$\dfrac{\partial h_3}{\partial z_3} = \dfrac{\partial}{\partial z_3} \max(0, z_3) \Rightarrow$ Any value

(c)



$$\frac{\partial L}{\partial w_1} = \underbrace{\frac{\partial L}{\partial z}}_{A}\left(\underbrace{\frac{\partial z}{\partial h_3}\;\frac{\partial h_3}{\partial z_3}\;\frac{\partial z_3}{\partial h_1}}_{B_1} + \underbrace{\frac{\partial z}{\partial h_4}\;\frac{\partial h_4}{\partial z_4}\;\frac{\partial z_4}{\partial h_1}}_{B_2}\right)\underbrace{\frac{\partial h}{\partial z_1}}_{C}\;\underbrace{\frac{\partial z_1}{\partial w_1}}_{D}$$

A $\quad \dfrac{\partial L}{\partial z} \leftarrow$ need not be zero

B₁

- $\dfrac{\partial z}{\partial h_3} = w_3 \leftarrow$ need not be zero
- $\dfrac{\partial h_3}{\partial z_3} = \dfrac{\partial}{\partial z_3}\,\mathrm{ReLU}(z_3) = \dfrac{\partial}{\partial z_3}\max(0, z_3) \leftarrow$ need not be zero
- $\dfrac{\partial z_3}{\partial h_1} = w_1 \leftarrow$ need not be zero

B₂ Similarly, you can check that $\dfrac{\partial z}{\partial h_4}$, $\dfrac{\partial h_4}{\partial z_4}$, $\dfrac{\partial z_4}{\partial h_1}$ need not be zero

C $\quad \dfrac{\partial h_1}{\partial z_1} = \dfrac{\partial}{\partial z_1}\max(0, \overset{-1}{\cancel{z_1}}) = \dfrac{\partial}{\partial z_1}(0) = 0$

0.75

So the entire product turns out to be zero because of C

$$\therefore \quad \frac{\partial L}{\partial w_1} = 0 \quad \text{(YES)}$$

3) $20 \times 3 \times 16 \times 16$

Batch size, Channels (depth), Height, Width

Kernel size $= 3 \times 5 \times 5$

Output size $= 16 - 5 + 1 = 12$

Output of code: $20 \times 7 \times 12 \times 12$

0.5 0.5 1

4) a) T 0.5

b) T 0.5

c) T 0.5

d) T 0.5

e) T 0.5

f) (D) 0.5

-0.5 for each incorrect answer

0 for no answer