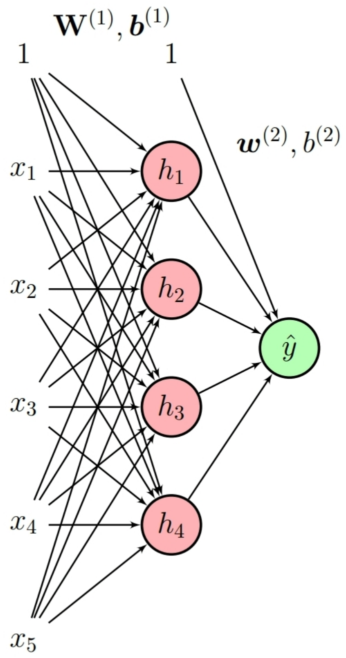
1) We need $x_1 > x_2 > x_3 > x_4 > x_5$, x_i are real numbers

One way of ensuring $x_i > x_{i+1}$ is to ensure that their difference

$$d_i = x_i - x_{i+1} > 0 \quad \text{for } i = 1, 2, 3, 4$$



We have 4 nodes in the hidden layer and each of them could be fed with one difference d_i. We could then expect the hidden node h_i to activate only if $d_i > 0$, however, given our output activation fires even when $d_i = 0$, this architecture will not work when $x_i = x_{i+1}$ (i.e. $d_i = 0$)

Therefore, consider $d_i = x_{i+1} - x_i$. Now, when $d_i < 0$, the output would be zero, else the output would be 1

$$d_i \geq 0 \rightarrow h_i \rightarrow 1$$

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

If all $d_i < 0$, then the output of all hidden activations would be zero, i.e. $h_i = 0$, and hence the output would be set 1

$z = 0 \Rightarrow \sigma(z) = 1$

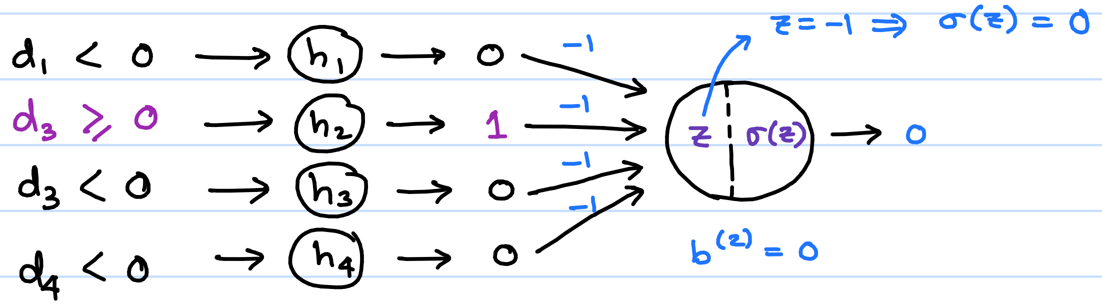
$$d_1 < 0 \rightarrow h_1 \rightarrow 0$$
$$d_2 < 0 \rightarrow h_2 \rightarrow 0$$
$$d_3 < 0 \rightarrow h_3 \rightarrow 0$$
$$d_4 < 0 \rightarrow h_4 \rightarrow 0$$

$z \mid \sigma(z) \rightarrow 1$

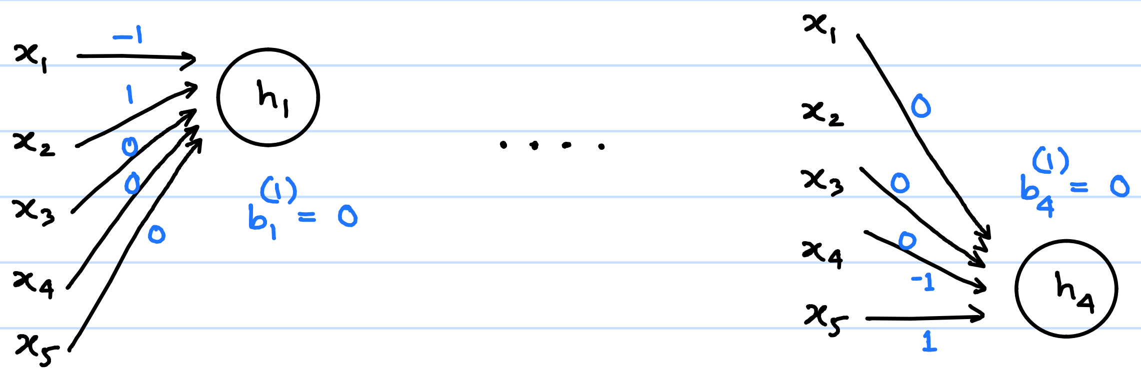
$b^{(2)} = 0$

However, if any $d_i \geq 0$, then the output of the corresponding hidden node would be 1, i.e. $h_i = 1$. To ensure that the output activation of $\hat{y}$ does not fire when any of the h_i's $= 1$, one could set the weights of the 2nd layer to some negative value

So, we could have $w_i^{(2)} = -1$



As for the weights of the first layer, we can define individual differences d_i as follows:



Therefore, one possible solution would be:

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

(b) The bias vector for first layer can be all zeros

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

(c) Weight vector for 2nd layer, $\underline{w}^{(2)} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$

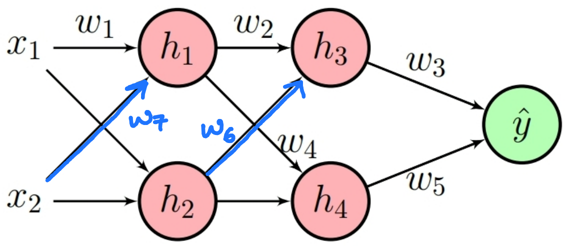
(d) Bias for 2nd layer, $b^{(2)} = 0$

Note: There could be many possible solutions of this problem!

Marking: If model works for $x_1 \geq x_2 \geq x_3 \geq x_4 \geq x_5 \rightarrow 1.5/3$

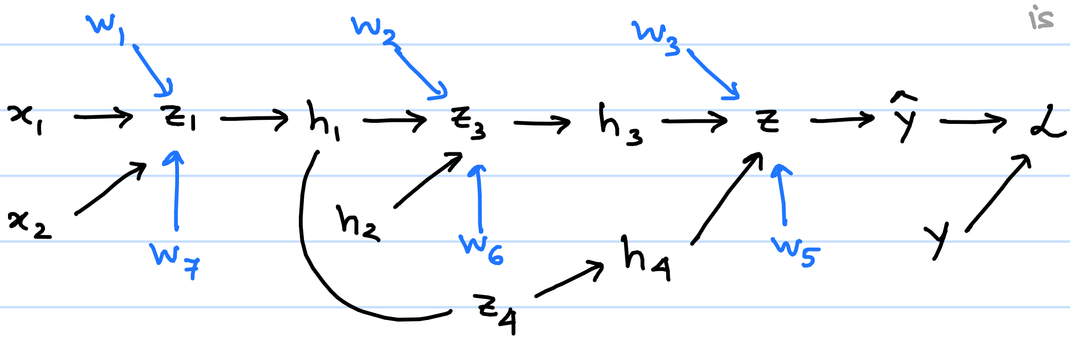
↳ for $x_1 > x_2 > x_3 > x_4 > x_5 \rightarrow 3/$

2) You may choose to define additional weights for convenience

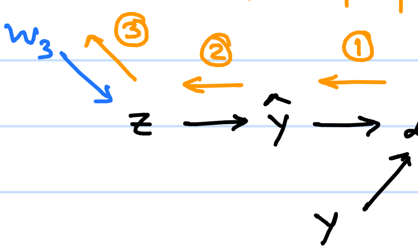


Note : Each node is associated with a ReLU activation

Let's draw a partial computation graph (drawing the graph is not necessary)



Backprop



(a) $$\frac{\partial \mathcal{L}}{\partial w_3} = \underbrace{\frac{\partial \mathcal{L}}{\partial y}}_{1} \times \underbrace{\frac{\partial y}{\partial z}}_{2} \times \underbrace{\frac{\partial z}{\partial w_3}}_{3}$$

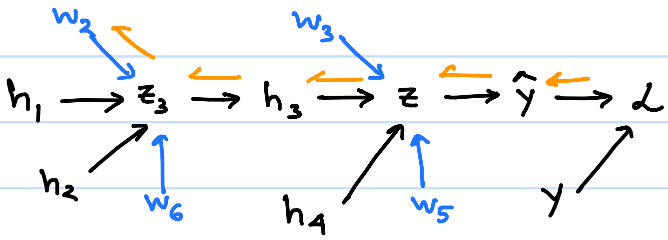
During forward computation $z_1 = -1$ and $h_1 = 0$. All other variables need not be zero

- $\frac{\partial \mathcal{L}}{\partial y}$ need not be zero
- $\frac{\partial y}{\partial z} = \frac{\partial}{\partial z} \text{ReLU}(z) = \frac{\partial}{\partial z} \max(z, 0)$ need not be zero
- $\frac{\partial z}{\partial w_3} = \frac{\partial}{\partial w_3}(w_3 h_3 + w_5 h_4) = h_3 \leftarrow$ need not be zero

$$\therefore \frac{\partial \mathcal{L}}{\partial w_3} = 0 \quad \text{(NO)}$$

0.5

(b) $$\frac{\partial \mathcal{L}}{\partial w_2} = \frac{\partial \mathcal{L}}{\partial z} \cdot \frac{\partial z}{\partial h_3} \cdot \frac{\partial h_3}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_2}$$



already proven to be not necessarily zero

$$\frac{\partial z}{\partial h_3} = \frac{\partial (w_3 h_3 + w_5 h_4)}{\partial h_3} = w_3 \leftarrow \text{need not be zero}$$

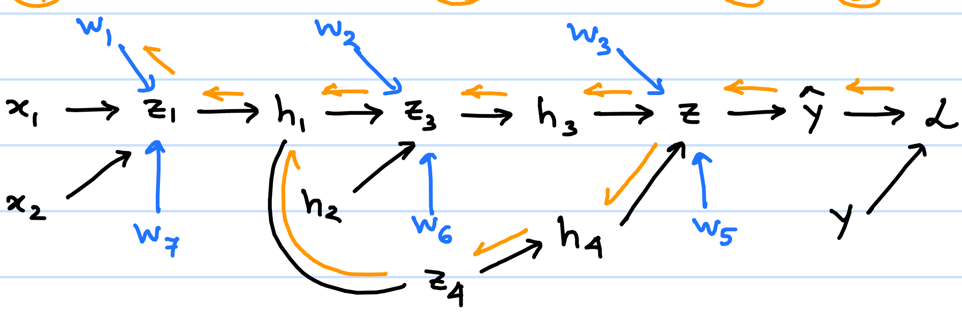
$$\frac{\partial h_3}{\partial z_3} = \frac{\partial}{\partial z_3} \mathrm{ReLU}(z_3) = \frac{\partial}{\partial z_3} \max(0, z_3) \leftarrow \text{need not be zero}$$

$$\frac{\partial z_3}{\partial w_2} = \frac{\partial}{\partial w_2}(w_2 h_1 + w_6 h_2) = h_1 = 0 \quad \text{(given)}$$

$$\therefore \quad \frac{\partial \mathcal{L}}{\partial w_2} = 0 \quad \text{(YES)}$$

0.75

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



(A) $\frac{\partial L}{\partial z} \leftarrow$ need not be zero (proved previously)

(B₁)
- $\frac{\partial z}{\partial h_3} = \frac{\partial}{\partial h_3}(w_3 h_3 + w_5 h_4) = w_3 \leftarrow$ need not be zero
- $\frac{\partial h_3}{\partial z_3} = \frac{\partial}{\partial z_3} \mathrm{ReLU}(z_3) = \frac{\partial}{\partial z_3} \max(0, z_3) \leftarrow$ need not be zero
- $\frac{\partial z_3}{\partial h_1} = \frac{\partial}{\partial h_1}(w_2 h_1 + w_3 h_2) = w_2 \leftarrow$ need not be zero

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

(C) $$\frac{\partial h_1}{\partial z_1} = \frac{\partial}{\partial z_1} \text{ReLU}(0, z_1) = \frac{\partial}{\partial z_1} \max(0, \overset{-1}{z_1}) = \frac{\partial}{\partial z_1}(0) = 0$$

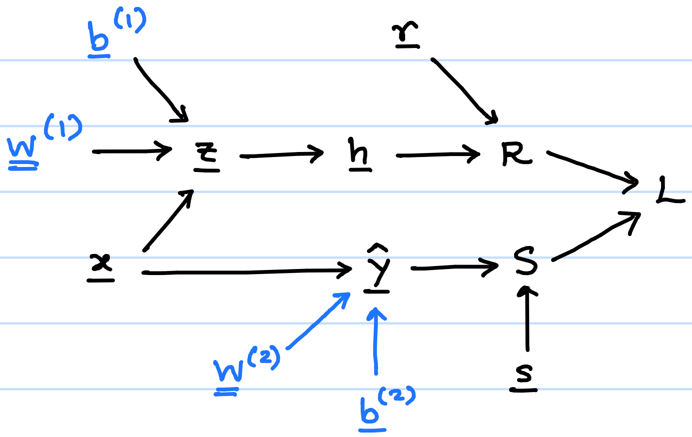
So the entire product turns out to be zero because of (C)

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

0.75

3)

a) Computation graph



0.9

$$z = \mathbf{W}^{(1)}x + b^{(1)}$$
$$h = \sigma(z)$$
$$\hat{y} = x + \mathbf{W}^{(2)}h + b^{(2)}$$
$$L = S + R$$
$$S = \frac{1}{2}\|\hat{y} - s\|_2^2$$
$$R = r^T h$$

b) $\bar{L} = \frac{\partial L}{\partial L} = 1$, $\bar{R} = \frac{\partial L}{\partial R} = 1$, $\bar{S} = \frac{\partial L}{\partial S} = 1$

0.45

$$\bar{\hat{y}} = \frac{\partial L}{\partial \hat{y}} = \bar{S}\frac{\partial S}{\partial \hat{y}} = \bar{S}\begin{bmatrix} \partial S/\partial y_1 \\ \vdots \\ \partial S/\partial y_n \end{bmatrix} = \bar{S}\begin{bmatrix} y_1 - s_1 \\ \vdots \\ y_n - s_n \end{bmatrix} = \bar{S}\cdot(y - s)$$

0.3

$$\bar{\underline{h}} = \frac{\partial \mathcal{L}}{\partial \underline{h}} = \frac{\partial R}{\partial \underline{h}} \bar{R} + \left(\frac{\partial \underline{y}}{\partial \underline{h}}\right)^T \bar{\underline{y}}$$

$$= \begin{bmatrix} \partial R/\partial h_1 \\ \vdots \\ \partial R/\partial h_K \end{bmatrix} \bar{R} + \begin{bmatrix} w_{11}^{(2)} & \cdots & w_{1K}^{(2)} \\ \vdots & \ddots & \vdots \\ w_{n1}^{(2)} & \cdots & w_{nK}^{(2)} \end{bmatrix}^T \bar{\underline{y}}$$

$$= \begin{bmatrix} r_1 \\ \vdots \\ r_K \end{bmatrix} \bar{R} + \begin{bmatrix} w_{11}^{(2)} & \cdots & w_{1K}^{(2)} \\ \vdots & \ddots & \vdots \\ w_{n1}^{(2)} & \cdots & w_{nK}^{(2)} \end{bmatrix}^T \bar{\underline{y}}$$

$$= \underline{r}\, \bar{R} + \underline{\underline{W}}^{(2)^T} \bar{\underline{y}}$$

0.6

$$\bar{\underline{z}} = \frac{\partial \mathcal{L}}{\partial \underline{z}} = \left(\frac{\partial \underline{h}}{\partial \underline{z}}\right)^T \bar{\underline{h}} \quad \begin{bmatrix} \frac{\partial h_1}{\partial z_1} & \cdots & \frac{\partial h_K}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_K}{\partial z_1} & \cdots & \frac{\partial h_K}{\partial z_K} \end{bmatrix}^T \bar{\underline{h}} = \begin{bmatrix} \sigma'(z_1) & & 0 \\ & \ddots & \\ 0 & & \sigma'(z_K) \end{bmatrix} \bar{\underline{h}}$$

$$= \bar{\underline{h}} \circ \sigma'(\underline{z})$$

0.3

elementwise product

$$\bar{\underline{x}} = \frac{\partial \mathcal{L}}{\partial \underline{x}} = \left(\frac{\partial \underline{z}}{\partial \underline{x}}\right)^T \bar{\underline{z}} + \left(\frac{\partial \underline{y}}{\partial \underline{x}}\right)^T \bar{\underline{y}}$$

identity matrix

$$= \underline{\underline{W}}^{(1)^T} \bar{\underline{z}} + \underline{\underline{I}}\, \bar{\underline{y}}$$

$$= \underline{\underline{W}}^{(1)^T} \bar{\underline{z}} + \bar{\underline{y}}$$

0.45