Homework -3

1. Method 1 :- This method has a very small dataset, so, the model's performance might not generalize well, leading to instability in error estimation. The error estimate e_1 may not be very reliable due to the small dataset size.

Method 2 :- This also uses a small training set. The large testing set makes e_2 a more stable: but the model will likely perform poorly due to insufficient training data.

Method 3 :- With the large dataset, the model is expected to have more representative training and testing, making e_3 more accurate. This method is likely to provide the most reliable error estimate.

So; the answers are $e_1 > e_2, e_1 > e_3, e_2 > e_3$.

3. Step 1 :- Derivative of $g(h)$

case 1 :- $h \geq 0$.

In this case $|h| = h$, & $g(h) = 1/1+h$

derivative $\rightarrow g'(h) = \frac{-1}{(1+h)^2}$

Substituting $g(h) = 1/1+h$

$\Rightarrow g'(h) = -o^2$ (since $o = g(h)$).

case 2 :- $h < 0$

In this case $|h| = -h$ & $g(h) = 1/1-h$

$g'(h) = (1/(1-h)^2)$

Substituting $g(h) = 1/(1-h)$

$\Rightarrow g'(h) = o^2$ (since $o = g(h)$)

Thus, we have

$$\Rightarrow g'(h) = \begin{cases} -o^2 & \text{if } h \geq 0 \\ o^2 & \text{if } h < 0. \end{cases}$$

Step 2 :- Deriving the Error Term δ

In perceptron training it is defined as the difference between the desired output y and the actual output O.

$$\Rightarrow \text{Error} = y - O$$

Update rule - $\Delta w_i = \epsilon \frac{\partial O}{\partial w_i} = \epsilon \delta x_i$

Apply chain rule $\Rightarrow \frac{\partial O}{\partial w_i} = \frac{\partial O}{\partial h} \cdot \frac{\partial h}{\partial w_i}$

Since $h = \sum_{i=1}^{n} w_i x_i$, we have $\partial h / \partial w_i = x_i$

$$\therefore \frac{\partial O}{\partial w_i} = g'(h) x_i$$

$$\Rightarrow \Delta w_i = \epsilon g'(h)(y-O) x_i$$

substituting $g'(h) \Rightarrow \delta = \begin{cases} -O^2(y-O) & \text{if } h \geq 0 \\ O^2(y-O) & \text{if } h < 0 \end{cases}$

For $i = 1, ---, n$, the weights are updated as

$$w_i \leftarrow w_i + \epsilon \delta x_i$$

where δ is

$$\Rightarrow \begin{cases} -O^2(y-O) & \text{if } h \geq 0 \\ O^2(y-O) & \text{if } h < 0 \end{cases}$$

4. Initial weights :- $w_1 = 0, w_2 = 1$

$\epsilon = 0.1$

Input vector :- $x_1 = 2, x_2 = 0$

Desired output; $y = 0$.

Activation function :- $g(h) = \frac{1}{1+e^{-2h}}$ for $\beta = 1$

wkt $y = h = w_1 x_1 + w_2 x_2$

$$\Rightarrow h = 0 * 2 + 1 * 0 = 0.$$

$$g(h) = \frac{1}{1+e^{-2*0}} = \frac{1}{2}.$$

$$\text{Error} = y - O \Rightarrow 0 - \frac{1}{2} = -\frac{1}{2}.$$

calculating $g'(h) = -\frac{1}{(1+e^{-2\beta h})^2} \cdot (-2\beta e^{-2\beta h})$

$$\Rightarrow \frac{2\beta e^{-2\beta h}}{(1+e^{-2\beta h})^2}$$

calculating $1-g'(h)$

$$\Rightarrow 1 - \frac{1}{(1+e^{-2\beta h})} = \frac{e^{-2\beta h}}{1+e^{-2\beta h}}$$

If we multiple $g(h)$ & $g'(h)$

$$\Rightarrow \frac{e^{-2\beta h}}{(1+e^{-2\beta h})^2}$$

Substituting the value in our $g'(h)$

$$\Rightarrow g'(h) = 2\beta g(h)(1-g(h))$$

$$\Rightarrow g'(h) = 1/2$$

wkt $w_i \leftarrow w_i + \epsilon \cdot Error \cdot g'(h) \cdot x_i$

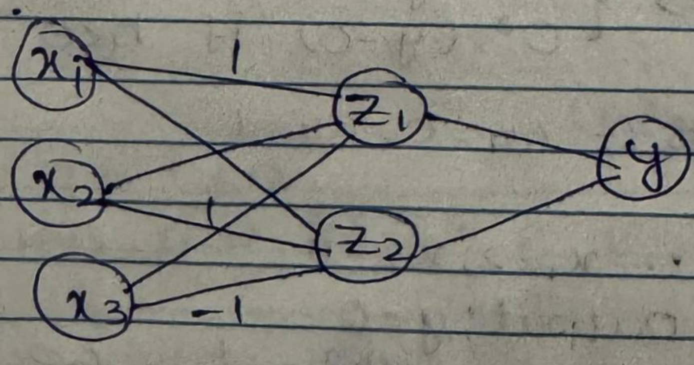
For $w_1 = 0 + 0.1 * (-1/2) 1/2 * 2 = 0 - 0.05 = -0.05$

$w_2 = 1 + 0.1 * (-1/2) 1/2 * 0 = 1 - 0 = 1$

Hence $w_1 = -0.05$

$w_2 = 1$

2.



$w_{11} = 1$

$w_{12} = 1$ bias $= 0.5$ $\begin{cases} 1 & \text{if } z_1 \geq 0.5 \\ 0, & \text{Otherwise} \end{cases}$

$w_{13} = -1$

$w_{22} = -1$ bias $= -0.5$ $\begin{cases} 1, & \text{if } z_2 \geq 0.5 \\ 0 & \text{Otherwise} \end{cases}$

$w_{01} = 1$ bias $= 1.5$ $\begin{cases} 1 & \text{if } z_{output} \geq 1.5 \\ 0, & \text{otherwise} \end{cases}$

$w_{02} = 1$

$x = 1, \quad x_2 = 0, \; x_3 = 0;$

$z_1 = 1.1 + 1.0 + (-1).0 = 1$ (Output 1)

$z_2 = 0.$ (Output 1)

$z_{output} = 1 + 1 = 2.$

Truth Table:-

x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	0	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	0