Assignment - 4

Question 1

Input neurons - N

Hidden neurons - H

Output neurons - C

(a) Total weight = $N \times H + H \times C$

(b) Total weight = $N \times H + H \times C + N \times C$

Question 2

(a) Network A being shallower, offering reduced computational complexity with fewer layers, it requires less processing power and memory, leading to faster training times and more straightforward implementation. This efficiency is beneficial when working with limited computational resources or when rapid prototyping is required

(b) Network B with its additional layers, possesses an enhanced capacity to model complex functions. The increased depth allows the network to represent more intricate patterns and relationships within the data.

Question 3

$$\varphi(x) = \frac{1}{1+e^{-ax}} = \frac{1}{1+e^{-2x}}$$

$$V_1 = x_1 W_1 + x_2 W_2 + b_2 \cdot 1 = -0.1 + 0.5 + b_1 = b_1 + 0.4$$

$$\varphi(V_1) = 0.73 \Rightarrow 1 + e^{-2V_1} = \frac{1}{0.73}$$

$$\Rightarrow e^{-2V_1} = 0.3698$$

$$\Rightarrow V_1 = 0.497311$$

$$\Rightarrow \boxed{b_1 = 0.097311}$$

Question 4

(a) Input $x \in \{0, 1\}$ Output $= \begin{cases} 1, & x=0 \\ 0, & x=1 \end{cases}$

Choosing $y = \text{step}(z) = \begin{cases} 0, & z<0 \\ 1, & z \geq 0 \end{cases}$ as our activation function

$y = \text{step}(wx + b) \Rightarrow z = wx + b$

Choosing weight $w = -1$, bias $b = 0.5$

$\therefore y = \text{step}(-x + 0.5)$

$x = 0, \quad z = 0 + 0.5 = 0.5 \qquad y = 1$

$x = 1, \quad z = -1 + 0.5 = -0.5 \qquad y = 0$

$\therefore$ Satisfied

(b)

Input x_1	Input x_2	Output (NAND)
0	0	1
0	1	1
1	0	1
1	1	0

Perceptron eq^n :- $y = \text{step}(w_1 x_1 + w_2 x_2 + b)$

Choosing $w_1 = -1, w_2 = -1, b = 1.5$

$y = \text{step}(-x_1 - x_2 + 1.5)$

x_1	x_2	$z = -x_1 - x_2 + 1.5$	y
0	0	1.5	1
0	1	0.5	1
1	0	0.5	1
1	1	-0.5	0

$\therefore$ Satisfied

Question 5

Parity function -

x_1	x_2	x_3	Output
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

Input layer

x_1, x_2, x_3 (each 0 or 1)

Perceptron is needed for $y = x_1 \oplus x_2 \oplus x_3$, which is not linearly separable.

Using hidden layers is necessity. Now we can model it in such a way that we have a hidden layer neuron for each case, which fires (output 1) only when specific pattern is present

Hidden Layers

We use 4 neurons in hidden layers, each with a weighted sum and bias followed by a step function as activation function

$$h_j = \text{step}\,(W_{j1} x_1 + W_{j2} x_2 + W_{j3} x_3 + b_j)$$

$$\text{step}\,(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

$h_1 = \text{step}\,(-x_1 - x_2 + x_3 - 0.5)$

Neuron ($h1$) $\longrightarrow$ Detects $\bar{x}_1 \wedge \bar{x}_2 \wedge x_3$

$\therefore$ $h1 = 1$ iff $x_1 = 0, x_2 = 0, x_3 = 1$

$W_{11} = -1, \; W_{12} = -1, \; W_{13} = 1, \; b_1 = -0.5$

Neuron ($h2$) $\rightarrow$ Detects $\bar{x}_1 \wedge x_2 \wedge \bar{x}_3$

$h_2 = \text{step}(-x_1 + x_2 - x_3 - 0.5)$

$\therefore h2 = 1$ iff $x_1 = 0,\ x_2 = 1,\ x_3 = 0$

$W_{21} = -1,\ W_{22} = 1,\ W_{23} = -1,\ b_2 = -0.5$

Neuron ($h3$) $\rightarrow$ Detects $x_1 \wedge \bar{x}_2 \wedge \bar{x}_3$

$h_3 = \text{step}(x_1 - x_2 - x_3 - 0.5)$

$\therefore h3 = 1$ iff $x_1 = 1,\ x_2 = 0,\ x_3 = 0$

$W_{31} = 1,\ W_{32} = -1,\ W_{33} = -1,\ b_3 = -0.5$

Neuron ($h4$) $\rightarrow$ Detects $x_1 \wedge x_2 \wedge x_3$

$h_4 = \text{step}(x_1 + x_2 + x_3 - 2.5)$

$\therefore h4 = 1$ iff $x_1 = 1,\ x_2 = 1,\ x_3 = 1$

$W_{41} = 1,\ W_{42} = 1,\ W_{43} = 1,\ b_4 = -2.5$

Output Neuron

Output Neuron combines the output of hidden neurons like an OR function. Final output is 1 if any hidden neuron is 1.

$y = \text{step}(h_1 + h_2 + h_3 + h_4 - 0.5)$ as $\sum h_i - 0.5 = -0.5$

Question 6

For the MLP network,

Input : $x \in \mathbb{R}^D$

η = learning rate

Hidden layer : H units

Weights : $W_1 \in \mathbb{R}^{H \times D}$

Biases : $b_1 \in \mathbb{R}^H$

Pre activation : $z_i = \sum_{j=1}^{D} W_{1,ij}\, x_j + b_{1i}$ for $i = 1, \ldots, H$

Activation : $h_i = \text{ReLU}(z_i) = \max(0, z_i)$

Output layer :

Weights : $W_2 \in \mathbb{R}^H$

Bias : $b_2 \in \mathbb{R}$

Output : $\hat{y} = \sum_{i=1}^{H} W_{2i}\, h_i + b_2$

Loss function : Squared error $L = \frac{1}{2}(\hat{y} - y)^2$

<u>Gradients</u> :-

(A) Output layer : $\frac{\partial L}{\partial w_{2i}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_{2i}} = (\hat{y} - y)\, h_i$

$$\frac{\partial L}{\partial \hat{y}} = \hat{y} - y$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b_2} = \hat{y} - y$$

∴ Updated equations :- (with learning rate η)

$$w_{2i}^{new} = w_{2i} - \eta(\hat{y} - y)\, h_i$$

$$b_2^{new} = b_2 - \eta(\hat{y} - y)$$

(B) For a hidden unit i, which has pre-activation z_i and activation h_i

$$h_i = \text{ReLU}(z_i), \quad \text{ReLU}'(z_i) = \begin{cases} 1, & z_i > 0 \\ 0, & z_i \le 0 \end{cases}$$

$$\frac{\partial L}{\partial w_{1,ij}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_i} \cdot \frac{\partial h_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_{1,ij}}$$

$$\frac{\partial \hat{y}}{\partial h_i} = w_{2i}, \quad \frac{\partial h_i}{\partial z_i} = \text{ReLU}'(z_i) \cdot I(z_i > 0)$$

$I(\cdot)$ is an indicator function

$$\frac{\partial z_i}{\partial w_{1,ij}} = x_j$$

$$\frac{\partial L}{\partial w_{1,ij}} = (\hat{y} - y)\, w_{2i}\, I(z_i > 0) \cdot x_j$$

$$\frac{\partial L}{\partial b_{1i}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_i} \cdot \frac{\partial h_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial b_{1i}} = (\hat{y} - y) \cdot w_{2i} \cdot I(z_i > 0)$$

Updated equations :-

$$W_{1,ij}^{new} = W_{1,ij} - \eta (\hat{y} - y) W_{2i} \cdot I(z_i > 0) \cdot x_j$$

$$b_{1i}^{new} = b_{1i} - \eta (\hat{y} - y) W_{2i} \cdot I(z_i > 0)$$

Here $I(z_i > 0)$ is 1 if $z_i > 0$

0 otherwise