CS 7643 - HW 2

Name: Niranjan Thakurdesai

Solutions discussed with Srijan Sood, Karan Shah and Nupur Chatterji

1-1 The general strategy is to write inequalities corresponding to each row of the truth table, and choose w and b such that all the inequalities are satisfied.

AND:

~~$0 \cdot w_1 + 0 \cdot w_2 +$~~

Let $w = [w_1, w_2]$

$0 \cdot w_1 + 0 \cdot w_2 + b < 0$

i.e. $b < 0$ —— (1)

$0 \cdot w_1 + 1 \cdot w_2 + b < 0$

i.e. $w_2 + b < 0$ —— (2)

$1 \cdot w_1 + 0 \cdot w_2 + b < 0$

i.e. $w_1 + b < 0$ —— (3)

$1 \cdot w_1 + 1 \cdot w_2 + b \geq 0$

i.e. $w_1 + w_2 + b \geq 0$ —— (4)

The inequalities above suggest that b should be negative, and w_1 and w_2 should be positive but lesser in magnitude than b. Let's choose w and b as follows (there are many possible solutions):

$w = [1, 1]$, $b = -2$

$b < 0$

$w_2 + b = 1 - 2 = -1 < 0$

$w_1 + b = 1 - 2 = -1 < 0$

$w_1 + w_2 + b = 1 + 1 - 2 = 0 \geq 0$

Thus, the inequalities above are satisfied

$$\Rightarrow \boxed{\begin{array}{l} w_{AND} = [1, 1] \\ b_{AND} = -2 \end{array}}$$

~~1~~ 2 For OR, Let's write down the inequalities like we did for AND ~~in part~~ 1.

$$0 \cdot w_1 + 0 \cdot w_2 + b < 0$$

i.e. $b < 0$ ——— ①

$$0 \cdot w_1 + 1 \cdot w_2 + b \geq 0$$

i.e. $w_2 + b \geq 0$ ——— ②

$$1 \cdot w_1 + 0 \cdot w_2 + b \geq 0$$

i.e. $w_1 + b \geq 0$ ——— ③

$$1 \cdot w_1 + 1 \cdot w_2 + b \geq 0$$

i.e. $w_1 + w_2 + b \geq 0$ ——— ④

The inequalities above suggest that b should be negative, and w_1 and w_2 should be positive with a greater absolute value than or equal to that of b. Let's choose w and b as follows (there are many possible solutions):

$w = [1, 1]$, $b = -1$

$b < 0$

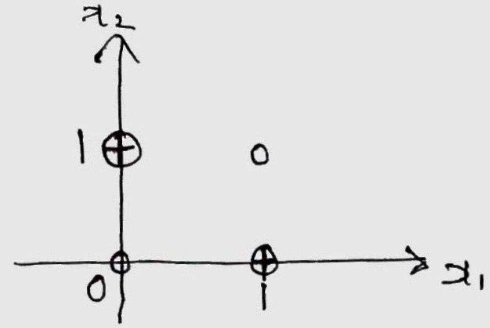
$w_2 + b = 1 - 1 = 0 \geq 0$

$w_1 + b = 1 - 1 = 0 \geq 0$

$w_1 + w_2 + b = 1 + 1 - 1 = 1 \geq 0$

Thus, the inequalities above are satisfied.

$$\Rightarrow \boxed{\begin{array}{l} w_{OR} = [1, 1] \\ b_{OR} = -1 \end{array}}$$



As we can see above, there's no linear boundary that can separate the examples. Let's prove this analytically.
We will start by writing down the inequalities as in part 1 and prove that they are inconsistent.

$0 \cdot w_1 + 0 \cdot w_2 + b < 0$

i.e. $b < 0$ ——— (1)

$0 \cdot w_1 + 1 \cdot w_2 + b \geq 0$

i.e. $w_2 + b \geq 0$ ——— (2)

$1 \cdot w_1 + 0 \cdot w_2 + b \geq 0$

i.e. $w_1 + b \geq 0$ ——— (3)

$1 \cdot w_1 + 1 \cdot w_2 + b \geq 0$

i.e. $w_1 + w_2 + b < 0$ ——— (4)

Adding (2) and (3),

$w_1 + w_2 + 2b \geq 0$ ——— (5)

From (1),

$-b > 0$ ——— (6)

Adding (5) and (6),

$w_1 + w_2 + b > 0$ ——— (7)

From (4) and (7), we can see that there is a contradiction.

$\Rightarrow$ XOR cannot be represented using a linear model with the same form as (1) in the question.

2-1 The general strategy is to do a forward pass, computing intermediate outputs ~~in~~ as a function of the input x.

$$x = 1$$

$$h_1(x) = W^{(1)}x + b^{(1)} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5x \\ 0.5x+1 \end{bmatrix}$$

$$a_1(x) = \max\{0, W^{(1)}x + b^{(1)}\} = \begin{bmatrix} 0.5x \\ 0.5x+1 \end{bmatrix} \quad \text{as } x > 0$$

$$h_2(x) = W^{(2)}a_1(x) + b^{(2)} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5x \\ 0.5x+1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} x+1 \\ x+1 \end{bmatrix}$$

$$a_2(x) = \max\{0, h_2(x)\} = \begin{bmatrix} x+1 \\ x+1 \end{bmatrix} \quad \text{as } x > 0$$

$$h(x) = W^{(3)}a_2(x) + b^{(3)} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x+1 \\ x+1 \end{bmatrix} + 1 = 2x+3$$

$$\Rightarrow \boxed{\begin{matrix} W = 2 \\ b = 3 \end{matrix}}$$

$$\boxed{\frac{dh}{dx} = W = 2}$$

2-2 $\quad x = -1$

$$h_1(x) = W^{(1)}x + b^{(1)} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5x \\ 0.5x+1 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}$$

$$a_1(x) = \max\{0, h_1(x)\} = \begin{bmatrix} 0 \\ 0.5x+1 \end{bmatrix}$$

$$h_2(x) = W^{(2)}a_1(x) + b^{(2)} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0.5x+1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5x+1 \\ 0.5x+1 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$$a_2(x) = \max\{0, h_2(x)\} = \begin{bmatrix} 0.5x+1 \\ 0.5x+1 \end{bmatrix}$$

$$h(x) = W^{(3)}a_2(x) + b^{(3)} = \begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} 0.5x+1 \\ 0.5x+1 \end{bmatrix} + 1$$

$$= x+3$$

$$\Rightarrow \boxed{\begin{matrix} W = 1 \\ b = 3 \end{matrix}}$$

$$\boxed{\frac{dh}{dx} = W = 1}$$

2-3 $x = -0.5$

$$h_1(x) = W^{(1)}x + b^{(1)} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5x \\ 0.5x+1 \end{bmatrix} = \begin{bmatrix} -0.25 \\ 0.75 \end{bmatrix}$$

$$a_1(x) = \max\{0, h_1(x)\} = \begin{bmatrix} 0 \\ 0.5x+1 \end{bmatrix}$$

$$h_2(x) = W^{(2)}a_1(x) + b^{(2)} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0.5x+1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5x+1 \\ 0.5x+1 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 0.75 \end{bmatrix}$$

$$a_2(x) = \max\{0, h_2(x)\} = \begin{bmatrix} 0.5x+1 \\ 0.5x+1 \end{bmatrix}$$

$$h(x) = W^{(3)}a_2(x) + b^{(3)} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5x+1 \\ 0.5x+1 \end{bmatrix} + 1 = x+3$$

$$\Rightarrow \boxed{\begin{matrix} W = 1 \\ b = 3 \end{matrix}}$$

$$\boxed{\frac{dh}{dx} = W = 1}$$

3-1 As W is a diagonal matrix,

$W^{(1)}x + b = \cancel{2x+1}\ 2x-1$

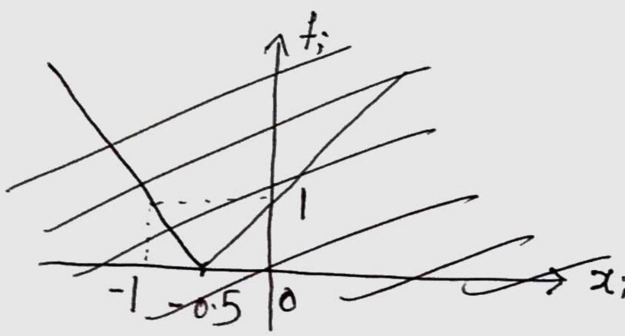
$\Rightarrow f_1(x) = \cancel{|2x+1|}\ |2x-1|$

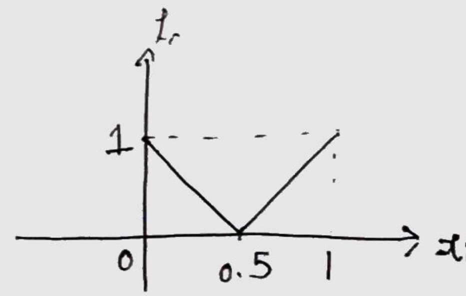
As $x \in \mathbb{R}^d$, $f_1(x) \in \mathbb{R}^d$. Let f_i be the i^{th} element of $f_1(x)$ $(i \in \{1, 2, \dots d\})$

$f_i = |2x_i - 1|$ where x_i is the i^{th} element of input x.

$$= \begin{cases} 2x_i - 1 & \text{if } x_i \geq 0.5 \\ -2x_i + 1 & \text{otherwise} \end{cases}$$

Thus, if we consider a single dimension of the output, it identifies ~~the input in~~ 2 regions of the input. This is shown in the figure below:





The 2 regions are $R_1 = \{(0, 0.5), (0.5, 1)\}$ for x_i

This applies to each dimension of the output. For every dimension, we can pick one of the two regions identified based on the value of the input in that dimension. This will result in a total of $\boxed{2^d}$ regions of the input that are identified onto $O = (0,1)^d$ by $f_1(\cdot)$. We can combine the regions of the input for each dimension in this way as they are independent of each other, i.e. f_i depends only on x_i.

3-2 According to the definition, given ~~$g(x_0)$, it is~~ an output value of g, say g_0, it could have come from any of the n_g regions identified by it. In the composition, the output of $g(\cdot)$ is the input to $f(\cdot)$. Given an output value of $f\circ g(\cdot)$, the input to $f(\cdot)$ could have come from any of the n_f regions identified by $f(\cdot)$. Each of these inputs in turn could have come from any of the n_g regions identified by $g(\cdot)$. Thus, $f\circ g(\cdot)$ identifies $\boxed{n_f n_g}$ regions of its input.

3-3 As proved in part 2, for a composition of 2 functions, ~~re~~ the number of regions in the input identified is equal to the product of the number of regions identified by each of them. In the case of L layers, L functions are composed. As proved in part 1, each of them identifies 2^d regions of the input. Thus, the number of regions identified by $f(x) = h_L$ is $\prod_{i=1}^{L} 2^d = 2^{Ld}$.