CS3244 Tutorial 4

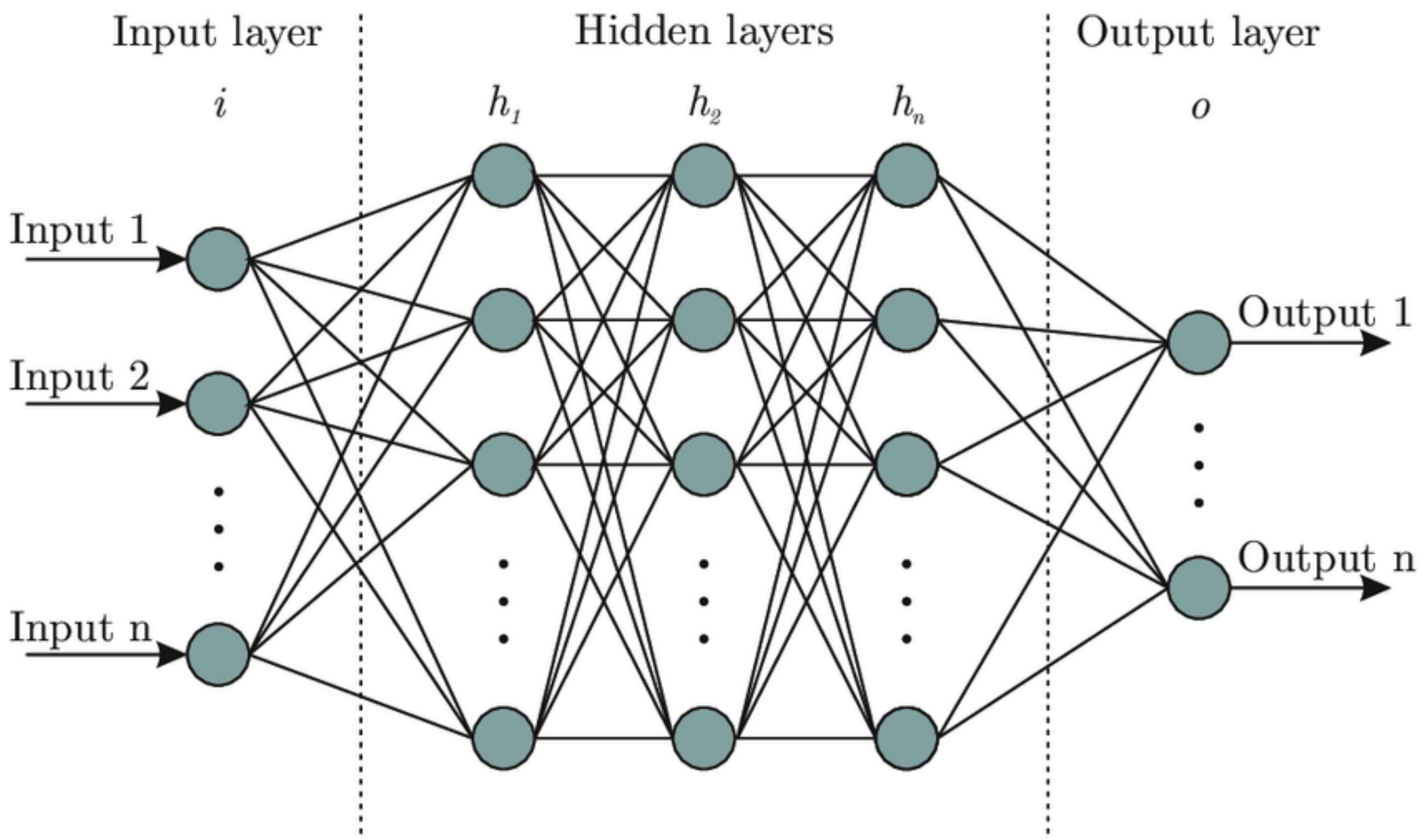
Wu Zhaoxuan

wu.zhaoxuan@u.nus.edu

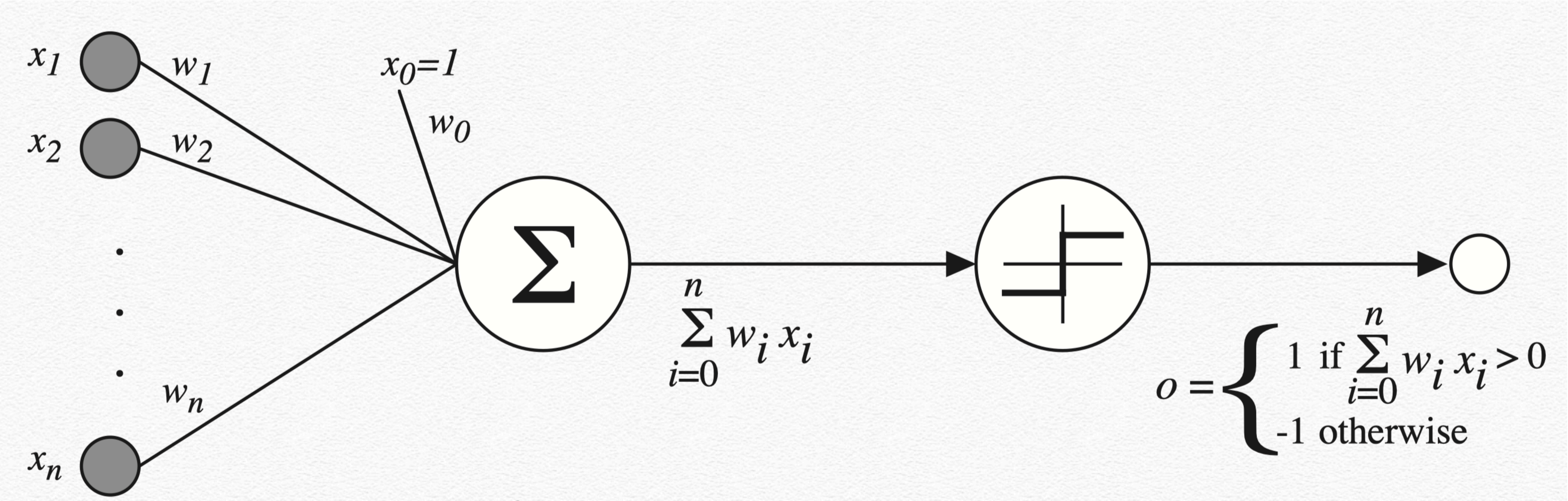
2022.3.4

Neural Networks

- Properties of neural networks
  - Real vector, high dimensional inputs and outputs
  - Restricted hypothesis space: by the number of hidden units or parameters
  - Long training time
  - Black-box model with almost no interpretability



Perceptron Unit



- In vector form, $o(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} > 0, \\ -1 & \text{otherwise} \end{cases}$
- Where both $\mathbf{w}$ and $\mathbf{x}$ are column vectors,
  - $\mathbf{w} = (w_0, w_1, \ldots, w_n)^\top$ and $\mathbf{x} = (1, x_1, x_2, \ldots, x_n)^\top$

BL 6

*Supposing the weights w_1 and w_2 of a perceptron are both set to the value of 0.5, derive the **<u>largest</u>** possible range of the values of w_0 that can be set for the perceptron to represent the AND gate (i.e., $AND(x_1, x_2)$). Assume that the inputs x_1 and x_2 and output $o(x_1, x_2)$ of the perceptron are Boolean with the values of 1 or -1. Show the steps of your derivation. No marks will be awarded for not doing so.*

BL 6

- Case 1: $x_1 = x_2 = 1$, so we need $o(x_1, x_2) = 1$

$$x_1 + x_2 = 1 + 1$$

$$\Rightarrow w_0 + 0.5x_1 + 0.5x_2 = w_0 + 1$$

  We need $o(\cdot) = 1 \Rightarrow w_0 + 1 > 0 \Rightarrow w_0 > -1$.

- Case 2: At least one of x_1 or x_2 is -1, so we need $o(x_1, x_2) = -1$

$$x_1 + x_2 \leq -1 + 1$$

$$\Rightarrow w_0 + 0.5x_1 + 0.5x_2 \leq w_0$$

  We need $o(\cdot) = -1 \Rightarrow w_0 \leq 0$.

- Overall, we have $-1 < w_0 \leq 0$.

BL 7 (Final AY17/18)

Supposing the weights w_1 and w_2 of a perceptron are both set to the value of -1, derive the ***largest*** *possible range of the values of w_0 that can be set for the perceptron to represent the NAND gate (i.e., $NAND(x_1, x_2)$). Assume that the inputs x_1 and x_2 and output $o(x_1, x_2)$ of the perceptron are Boolean with the values of 1 or -1. Show the steps of your derivation. No marks will be awarded for not doing so.*

BL 6

- Case 1: $x_1 = x_2 = 1$, so we need $o(x_1, x_2) = -1$

$$x_1 + x_2 = 1 + 1$$
$$\Rightarrow w_0 - x_1 - x_2 = w_0 - 2$$

We need $o(\cdot) = -1 \Rightarrow w_0 - 2 \leq 0 \ \ \Rightarrow w_0 \leq 2.$

- Case 2: At least one of x_1 or x_2 is -1, so we need $o(x_1, x_2) = 1$

$$x_1 + x_2 \leq -1 + 1$$
$$\Rightarrow w_0 - x_1 - x_2 \geq w_0$$

We need $o(\cdot) = 1 \Rightarrow w_0 > 0.$

- Overall, we have $0 < w_0 \leq 2.$

BL 8

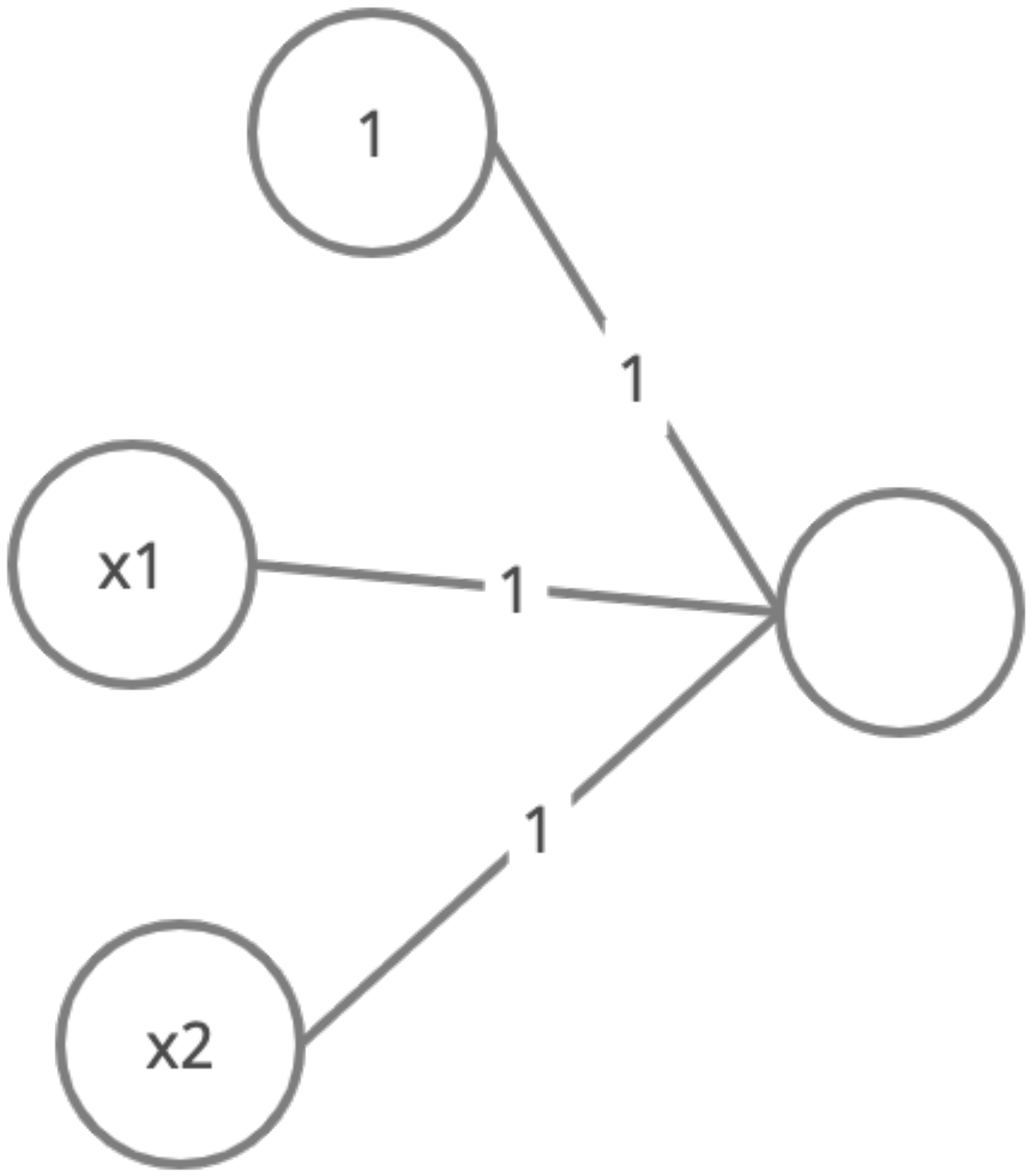
- *In every multi-layer network of perceptron units described below,*
  - *There should be only one (Boolean) output unit and an input unit for every (Boolean) input.*
  - *A Boolean is –1 if false, and 1 if true.*
  - *The activation function of every (non-input) unit is a –1 to 1 step function, including that of the output unit (see page 6 of the "Neural Networks" lecture slides).*
  - *Your weights must be integers and kept small, but possibly negative.*
  - *Keep your networks as symmetric as possible – doing this in questions a and b may help you in question c.*
  - *You don't have to draw edges with weight 0.*

BL 8

(a) Construct and draw a perceptron network with no hidden layers that implements $(x_1\ OR\ x_2)$.

Solution

Either $x_1 = 1$ or $x_2 = 1$ for the activation to fire:



BL 8

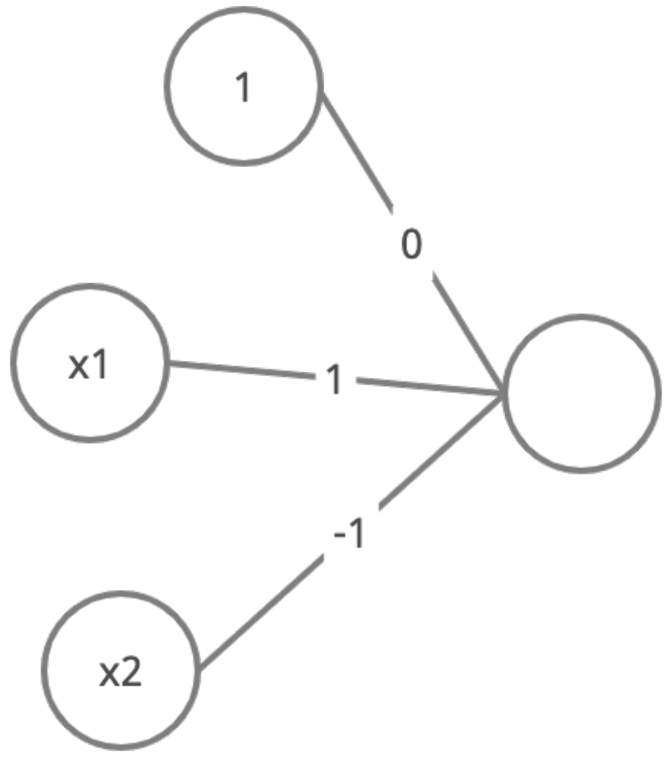
(b) Construct and draw a perceptron network with ONE hidden layers (with two units) that implements $(x_1\, XOR\, x_2)$.

Solution

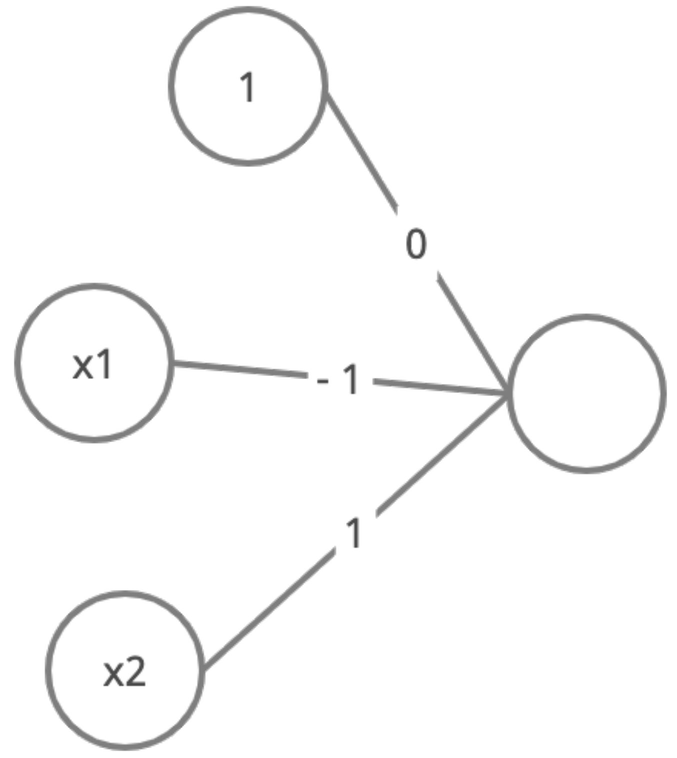
Now there are two conditions: $(x_1 = 1 \wedge x_2 = -1) \vee (x_1 = -1 \wedge x_2 = 1)$

Notice that we already know how to construct OR gate, so what is left is to construct the each conjunction as a perception network.

For $(x_1 = 1 \wedge x_2 = -1)$,



For $(x_1 = -1 \wedge x_2 = 1)$,

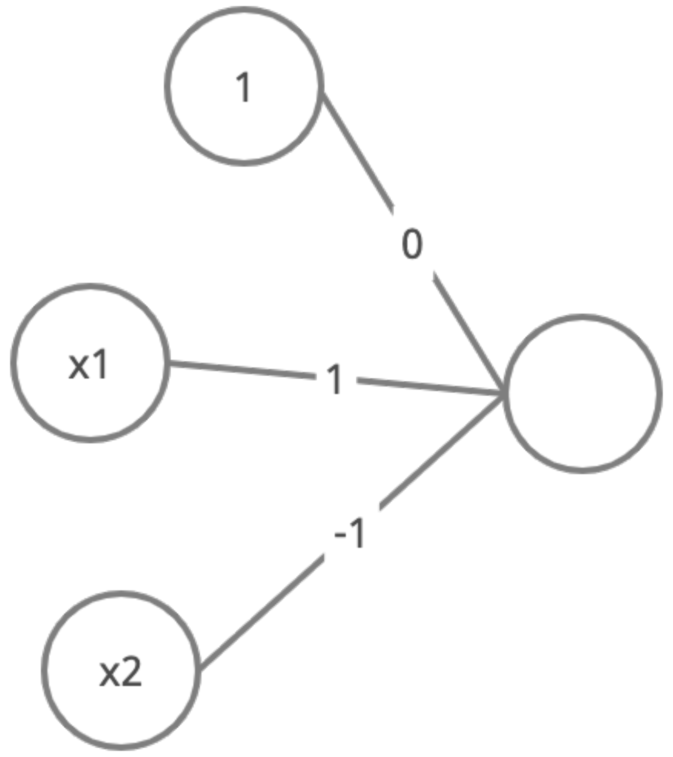


BL 8

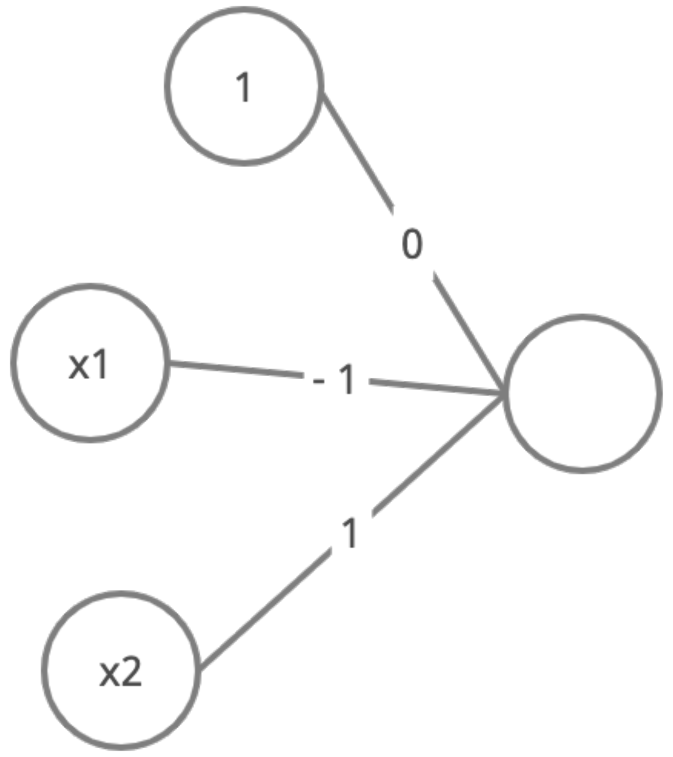
(b) Construct and draw a perceptron network with ONE hidden layers that implements $(x_1\, XOR\, x_2)$.

Solution

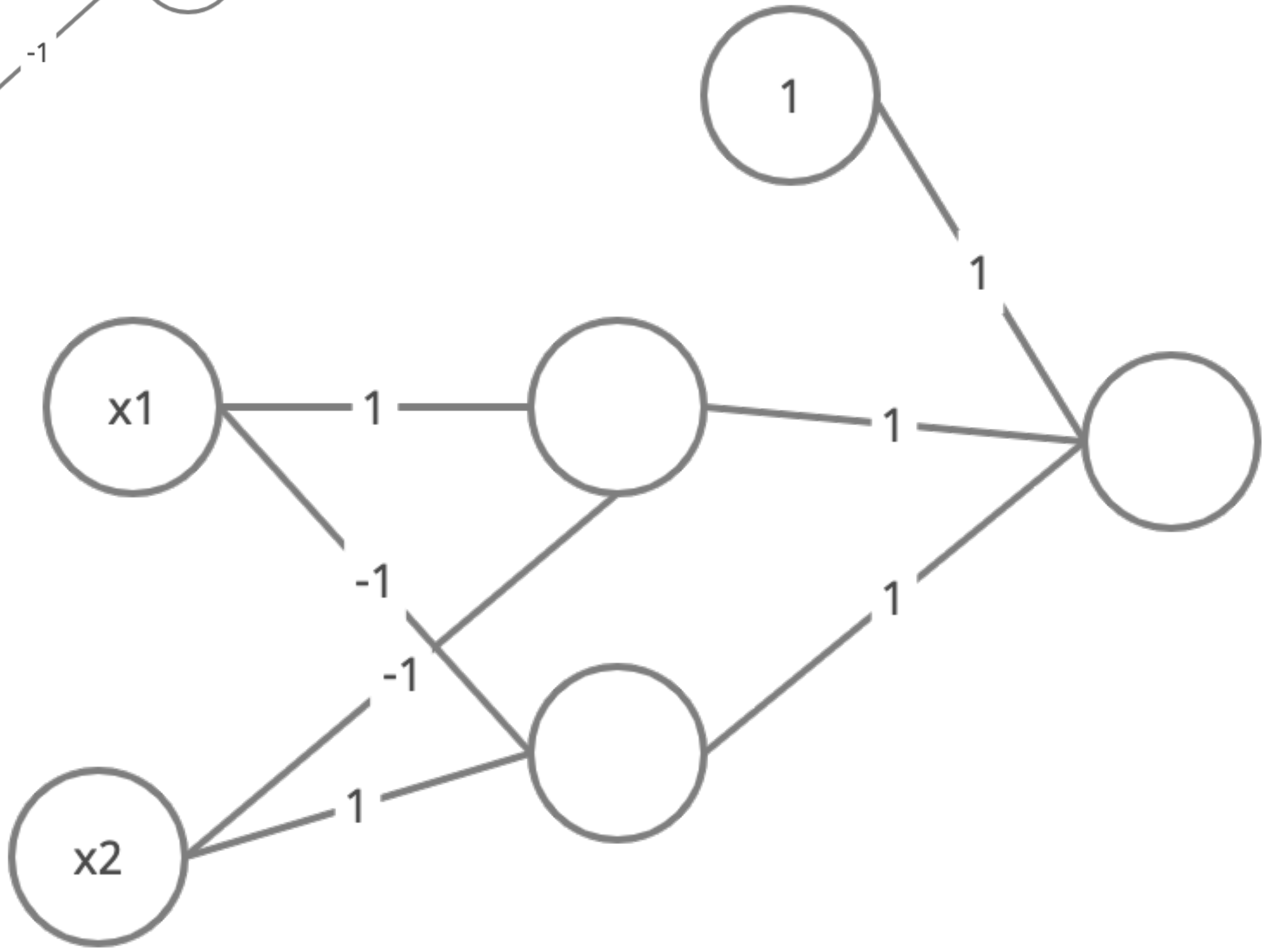
For $(x_1 = 1 \wedge x_2 = -1)$,



For $(x_1 = -1 \wedge x_2 = 1)$,



Append to the OR gate:

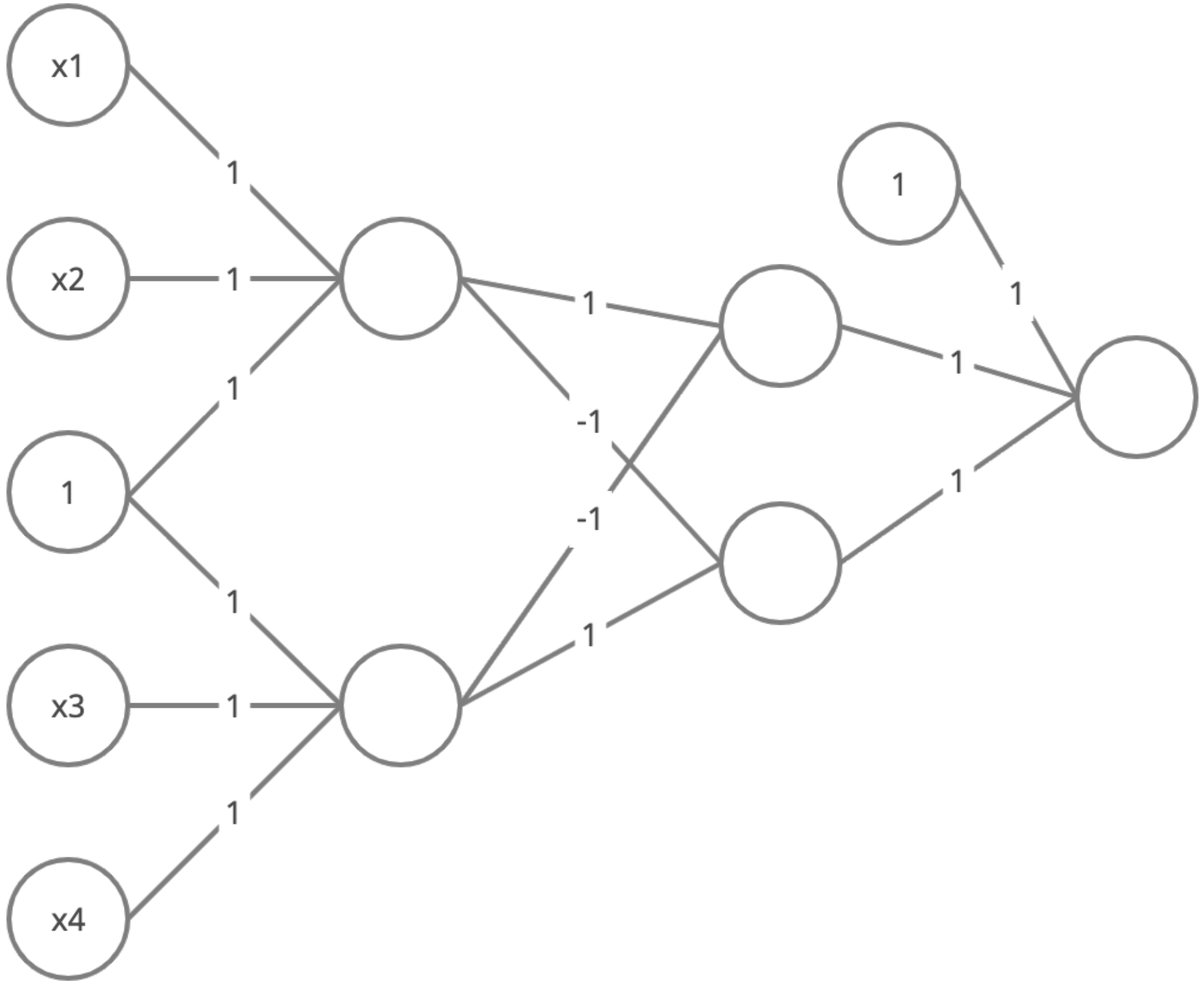


BL 8

(c) Construct and draw a perceptron network with TWO hidden layers that implements $(x_1\ OR\ x_2)\ XOR\ (x_3\ OR\ x_4)$.

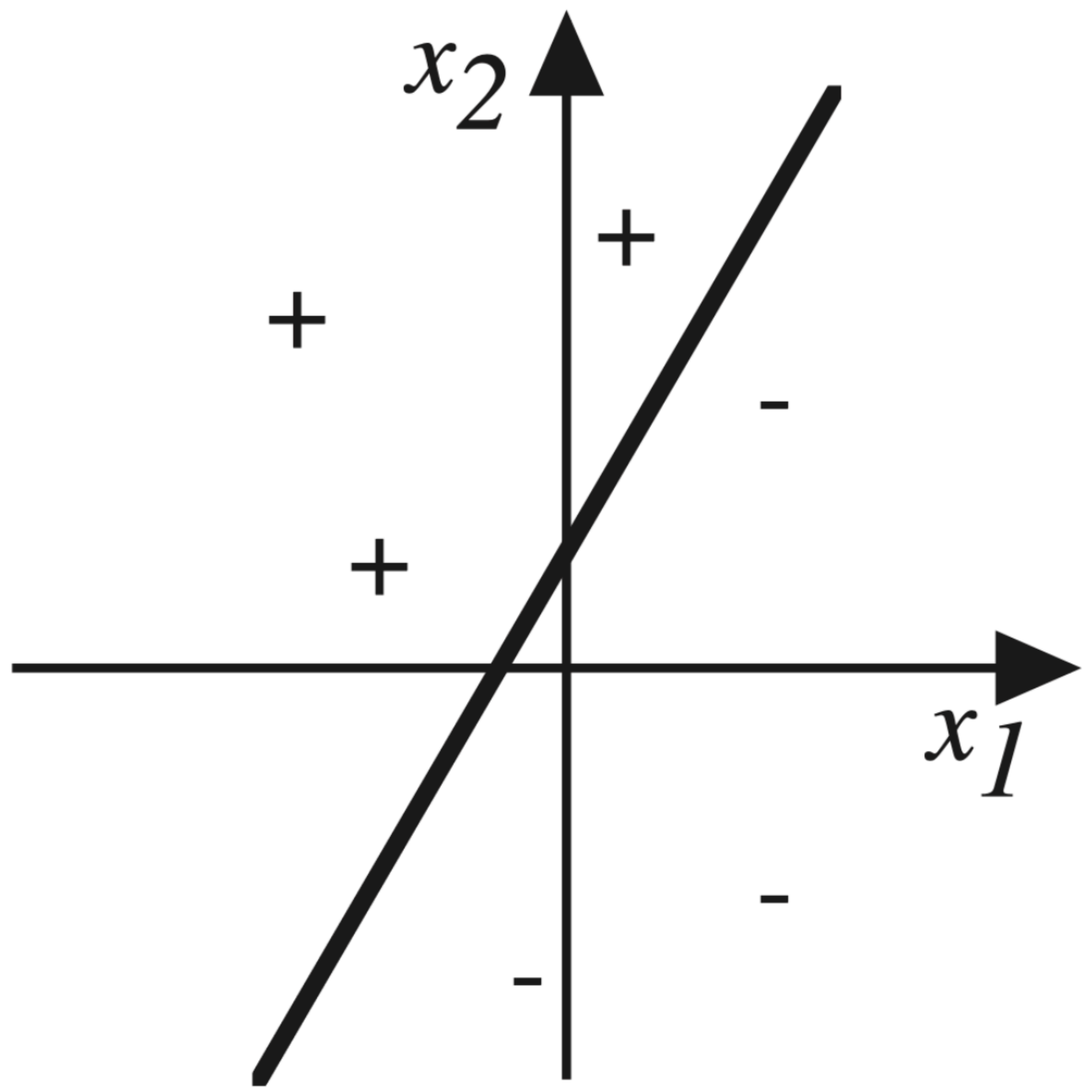
Solution

Building two OR gates on top of XOR gate.



Visualizing Perceptron (and TM 4.1)

- We can visualize learned perceptrons when the number of dimensions is small
- For $\mathbf{x} = (1, x_1, x_2)^\top$ and $\mathbf{w} = (w_0, w_1, w_2)^\top$, the decision surface (or line in 2D) is just $\mathbf{w}^\top \mathbf{x} = \mathbf{w} \cdot \mathbf{x} = 0$.
- Expanding: $x_2 = -(\frac{w_1}{w_2})x_1 - \frac{w_0}{w_2}$
- The weight vector $(w_1, w_2)^\top$ is perpendicular to the decision surface. Why?
    - Take two points $\mathbf{x}^A$ and $\mathbf{x}^B$on the line, check $(w_1, w_2)^\top(\mathbf{x}^A - \mathbf{x}^B) = 0$
    - Because $(w_1, w_2)^\top \mathbf{x}^A + w_0 = 0$ and $(w_1, w_2)^\top \mathbf{x}^B + w_0 = 0$
- Let us define the weight vector to point towards the +ve points
    - If it points to -ve x_1 axis and +ve x_2 axis, then
        - Smaller x_1 implies larger sum => $w_1 < 0$
        - Larger x_2 implies larger sum => $w_2 > 0$
- Note that when $w_0 = 0$, the line should pass the origin
    - when $x_1 = x_2 = 0$ and sum is negative => $w_0 < 0$



TM 4.1

What are the values of weights w_0, w_1 and w_2 for the perceptron whose decision surface is illustrated in Figure (a) on page 7 of the "Neural Networks" lecture slides? Assume the surface crosses the x_1 axis at -1, and the x_2 axis at 2.

Solution

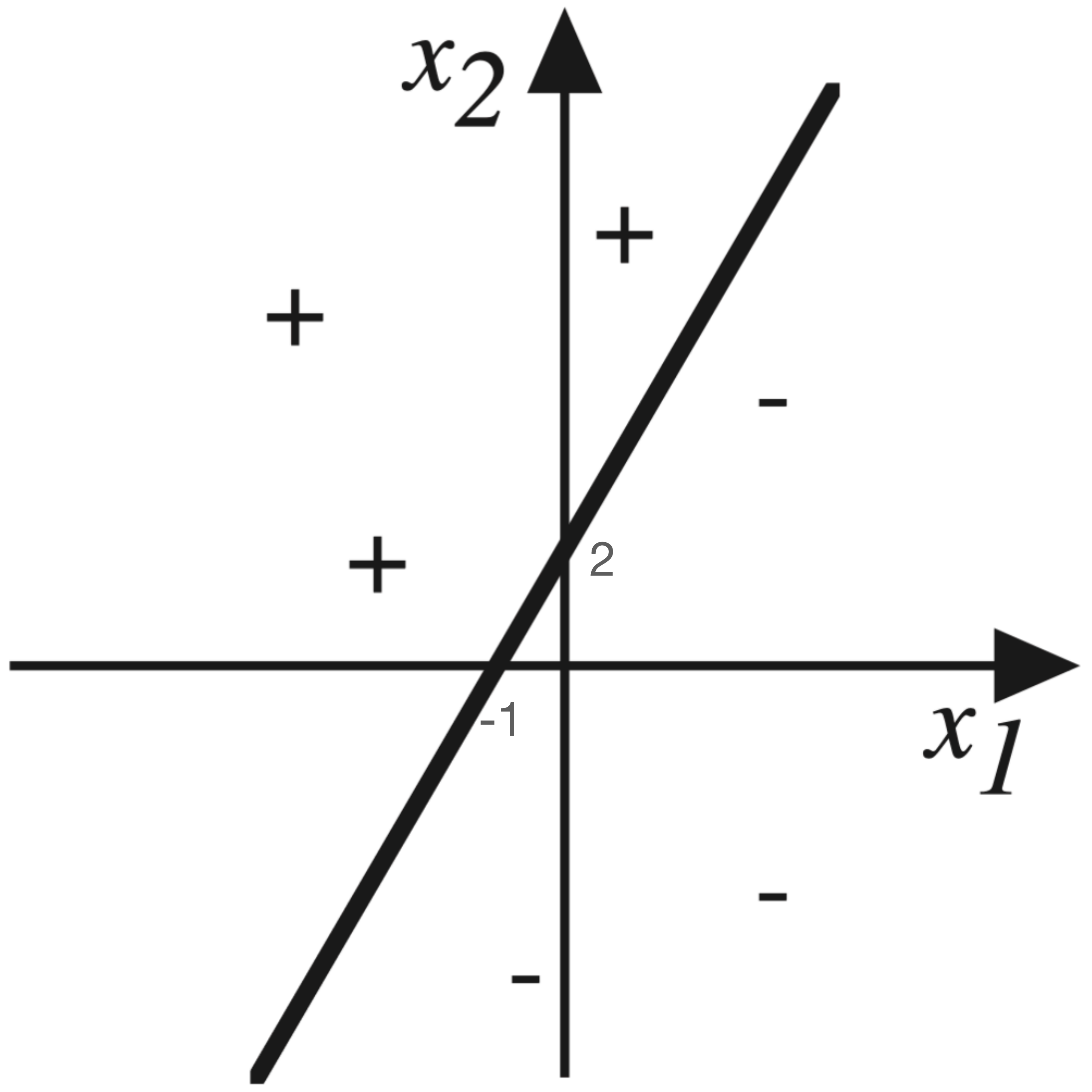
1. We know that the decision surface is represented by the line $w_0 + w_1x_1 + w_2x_2 = 0$.

2. It crosses $(-1{,}0)$, then $w_0 - w_1 = 0$. That is $w_0 = w_1$.

3. It crosses $(0{,}2)$, then $w_0 + 2w_2 = 0$. That is $w_0 = -2w_2$.

4. Therefore, $w_0 = w_1 = -2w_2$.

5. We know from the previous page that $w_0 < 0$, $w_1 < 0$ and $w_2 > 0$, one possible solution would be $w_0 = w_1 = -2$ and $w_2 = 1$.



TM 4.3

*Consider two perceptrons defined by the threshold expression $w_0 + w_1 x_1 + w_2 x_2 > 0$. Perceptron A has weight values $w_0 = 1, w_1 = 2, w_2 = 1$ and perceptron B has the weight values $w_0 = 0, w_1 = 2, w_2 = 1$. True or false? Perceptron A is **more general than** perceptron B.*

Solution

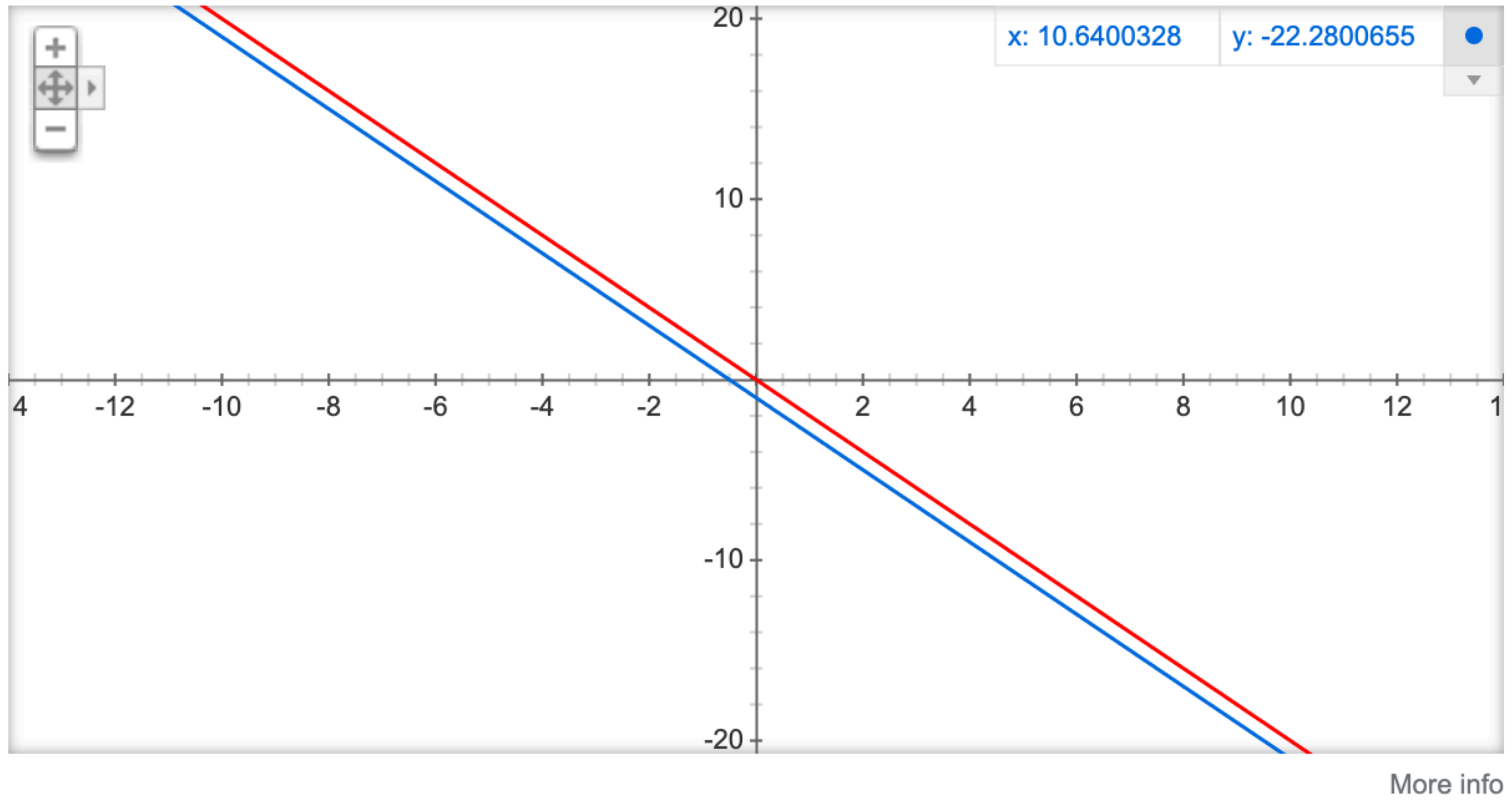
1. Choose an arbitrary $\mathbf{x} \in X$ s.t. $o(\mathbf{x}) = 1$ for B.
2. Then $2x_1 + x_2 > 0$ and $1 + 2x_2 + x_2 > 0$.
3. Therefore, $o(\mathbf{x}) = 1$ for A.
4. Therefore, A is ***more general than or equal*** to B, by definition.
5. Let $\mathbf{x} = (1,0,0)^\top$.
6. For A, $\mathbf{w}_A^\top \mathbf{x} = 1 + 2(0) + (0) = 1$.
7. For B, $\mathbf{w}_B^\top \mathbf{x} = 2(0) + (0) \ngtr 0$, so $o(\mathbf{x}) = 0$.
8. Therefore, B is not ***more general than or equal*** to A, by definition.
9. Therefore, A is ***more general than*** B.

Another visual way

A: $1 + 2x_1 + x_2 = 0 \Leftrightarrow x_2 = -2x_1 - 1$

B: $2x_1 + x_2 = 0 \Leftrightarrow x_2 = -2x_1$

Graph for (-(2*x))-1, -(2*x)



Thank you!

- Any questions?