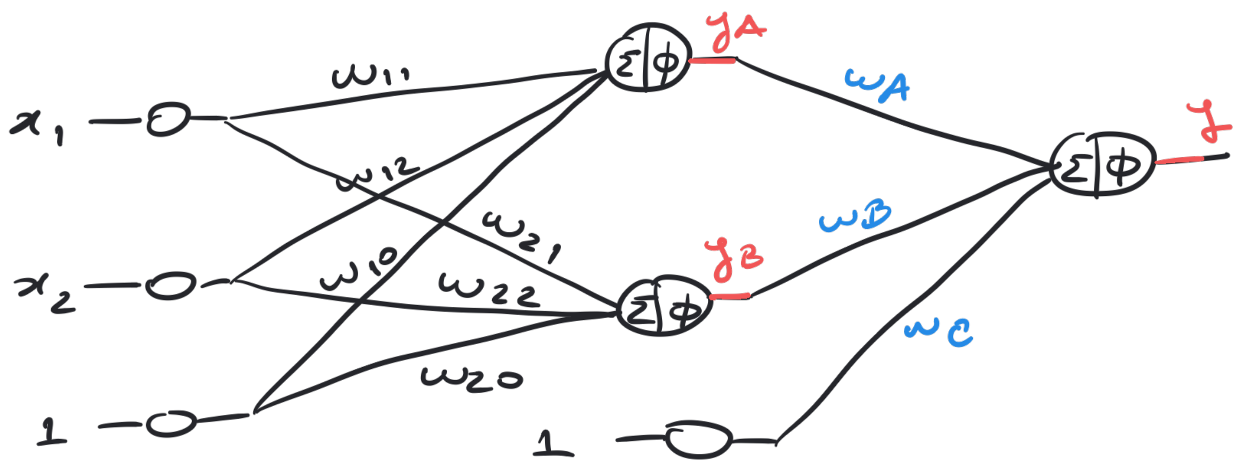


input layer

Hidden layer

output layer

$\phi(x) \equiv$ ACTIVACTION function

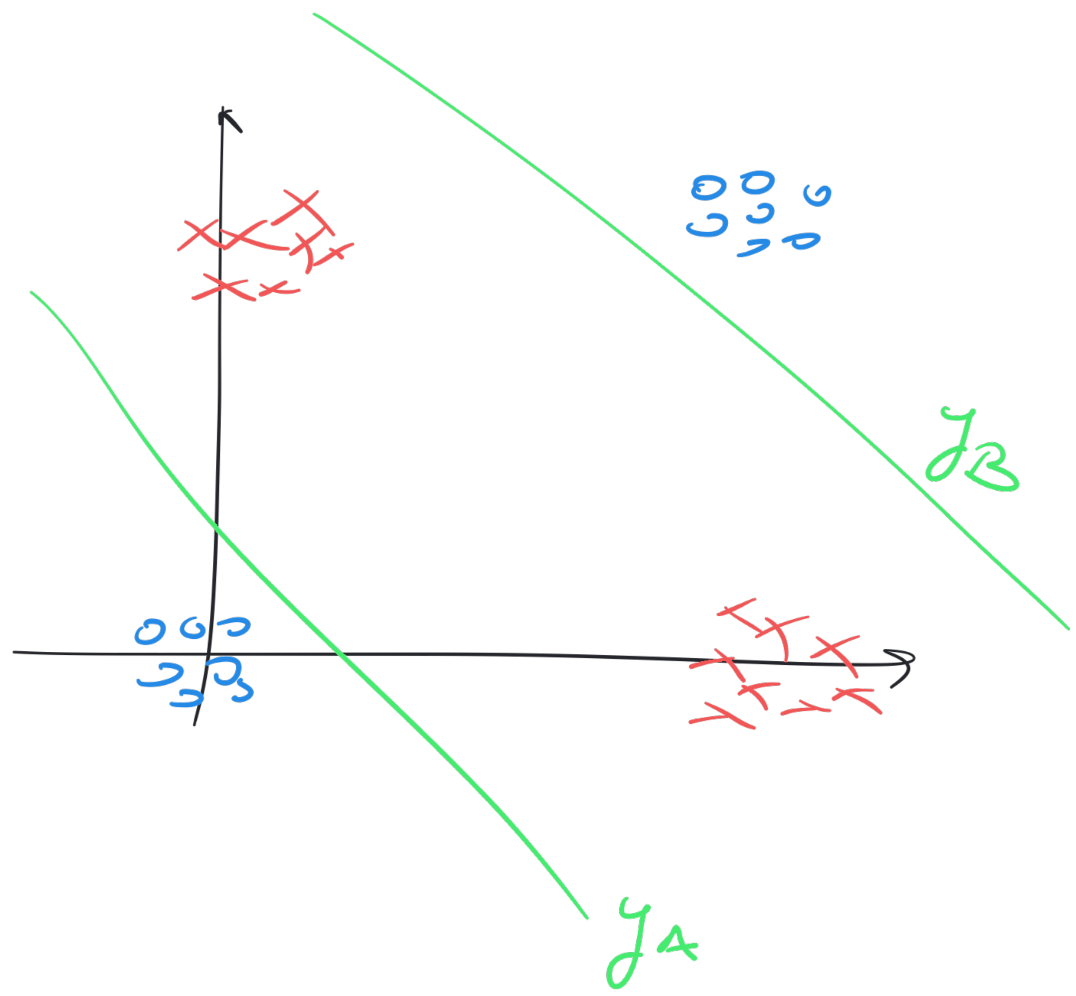
e.g.

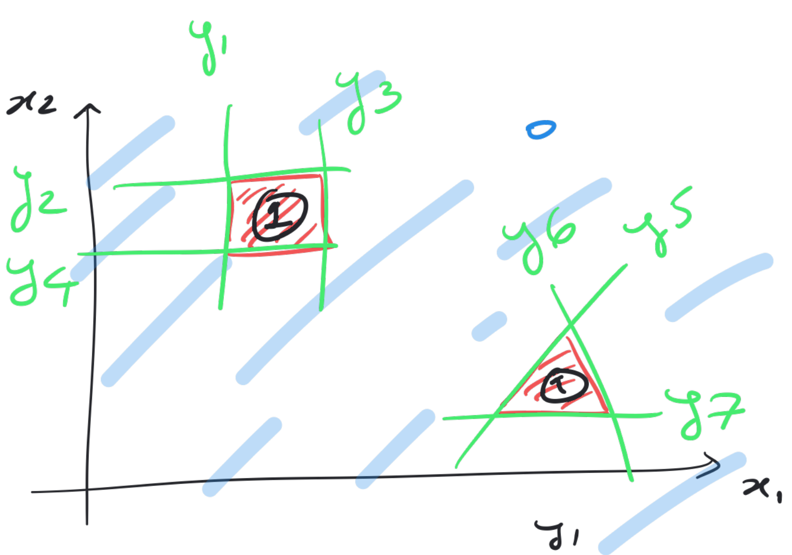
$$\phi(x) = \begin{cases} 1, & x > 0 \\ -1, & x \leq 0 \end{cases}$$

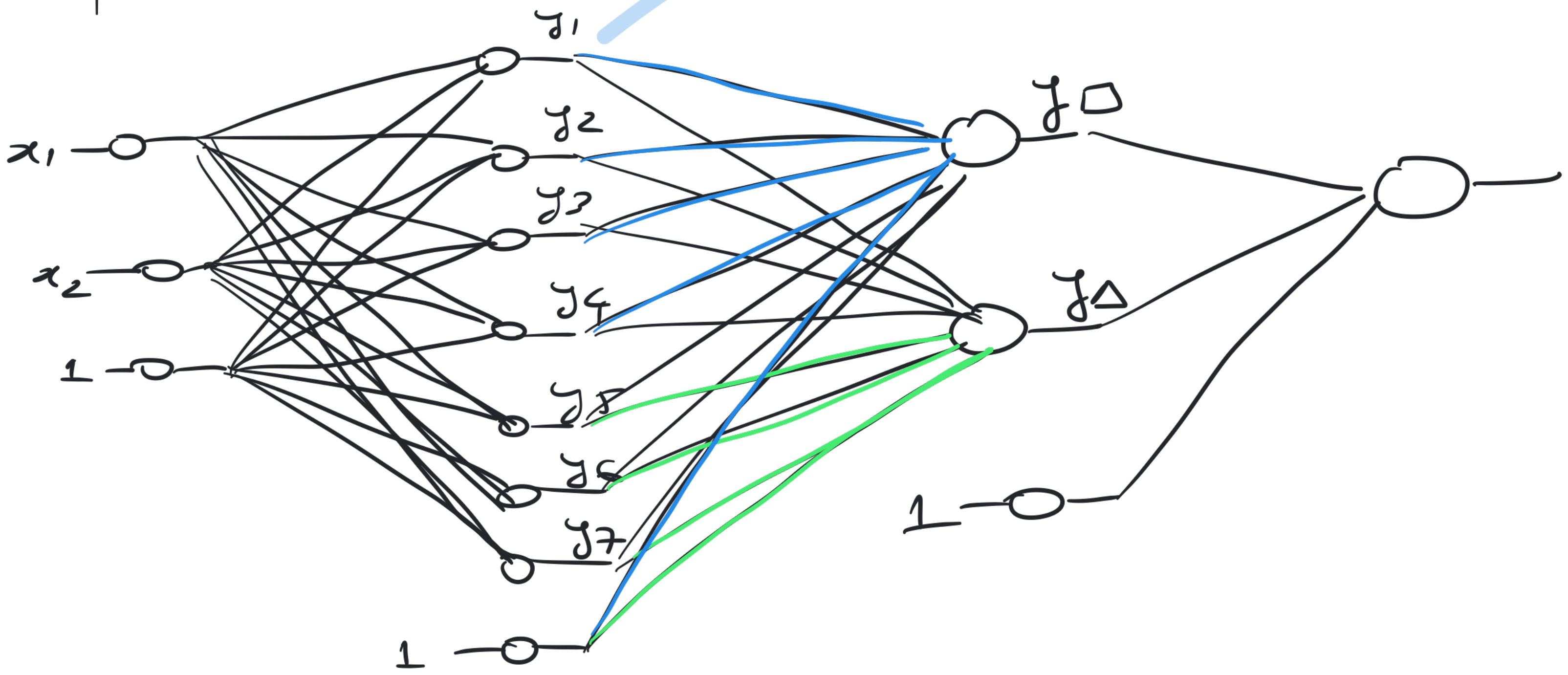
$$y_A = \phi(w_{11} \cdot x_1 + w_{12} \cdot x_2 + w_{10})$$

$$y_B = \phi(w_{21} \cdot x_1 + w_{22} \cdot x_2 + w_{20})$$

$$y = \phi(w_A \cdot y_A + w_B \cdot y_B + w_C)$$







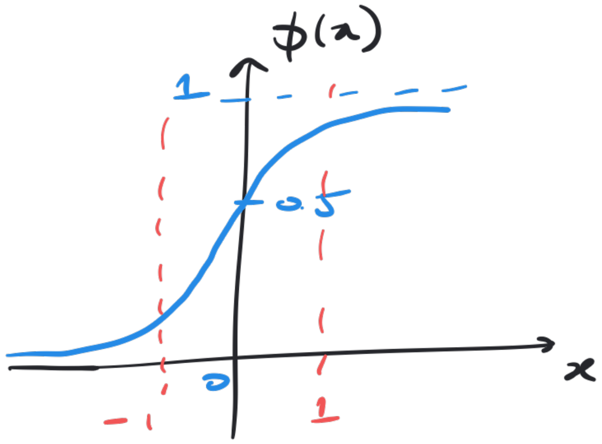
Universal Approximation theorem

A shallow network (single hidden layer) w/ a non-linear activation fct can learn arbitrarily close any function.

Activation fct

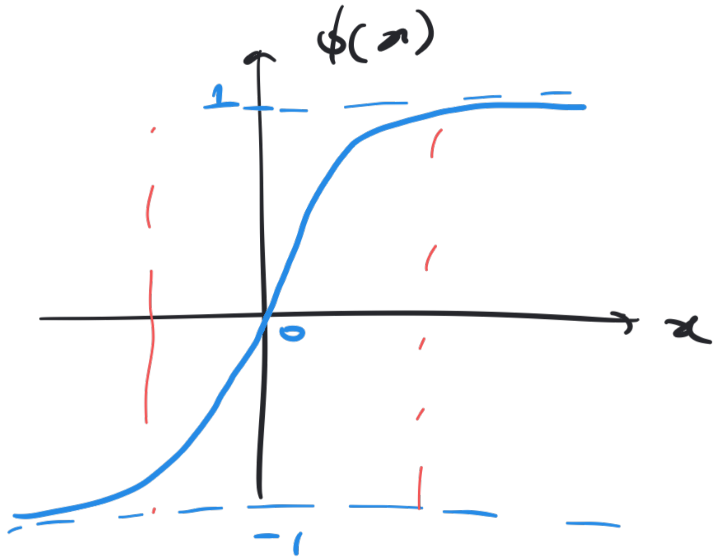
1. Sigmoid fct

$$\phi(x) = \frac{1}{1 + e^{-x}}$$



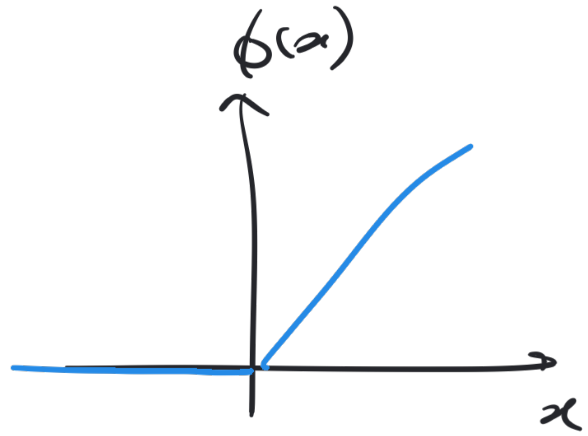
2. Hyperbolic tangent (tanh)

$$\phi(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$



3. Rectified linear Unit (ReLU)

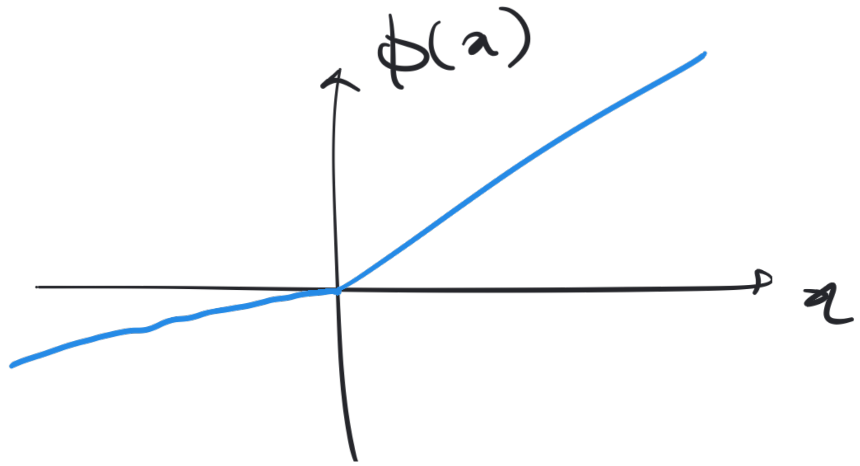
$$\phi(x) = \max(0, x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

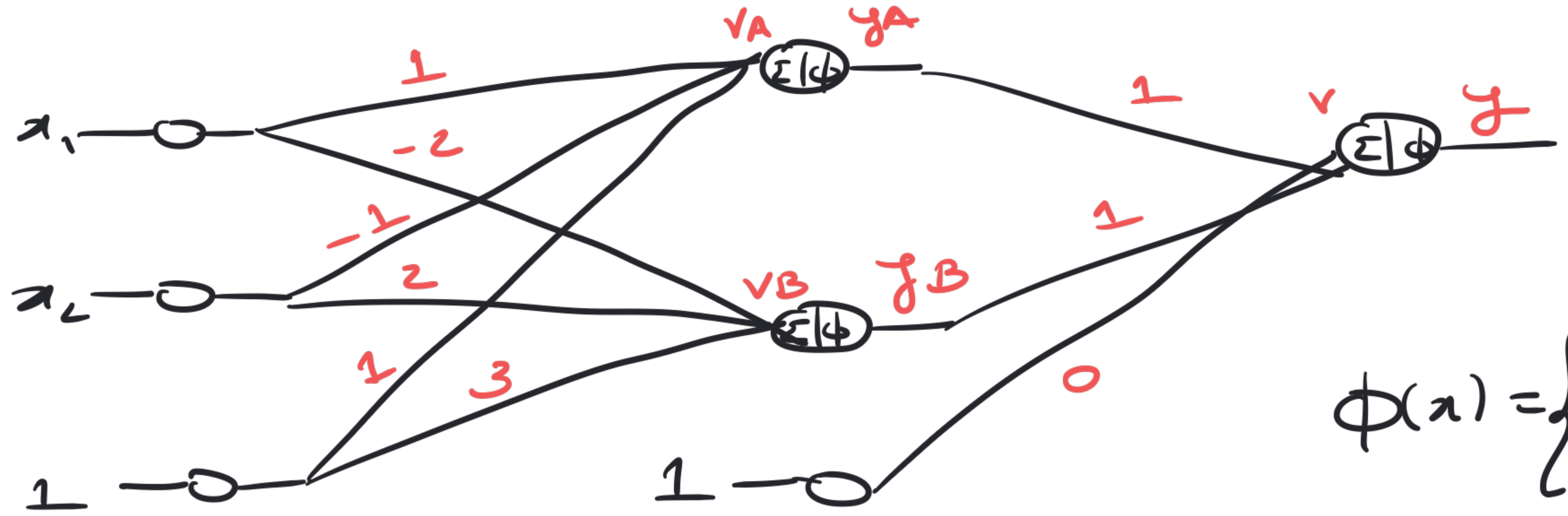


4. Leaky ReLU

$$\phi(x) = \begin{cases} x, & x \geq 0 \\ \delta . x, & x \leq 0 \end{cases}$$

$\delta = 0.01$



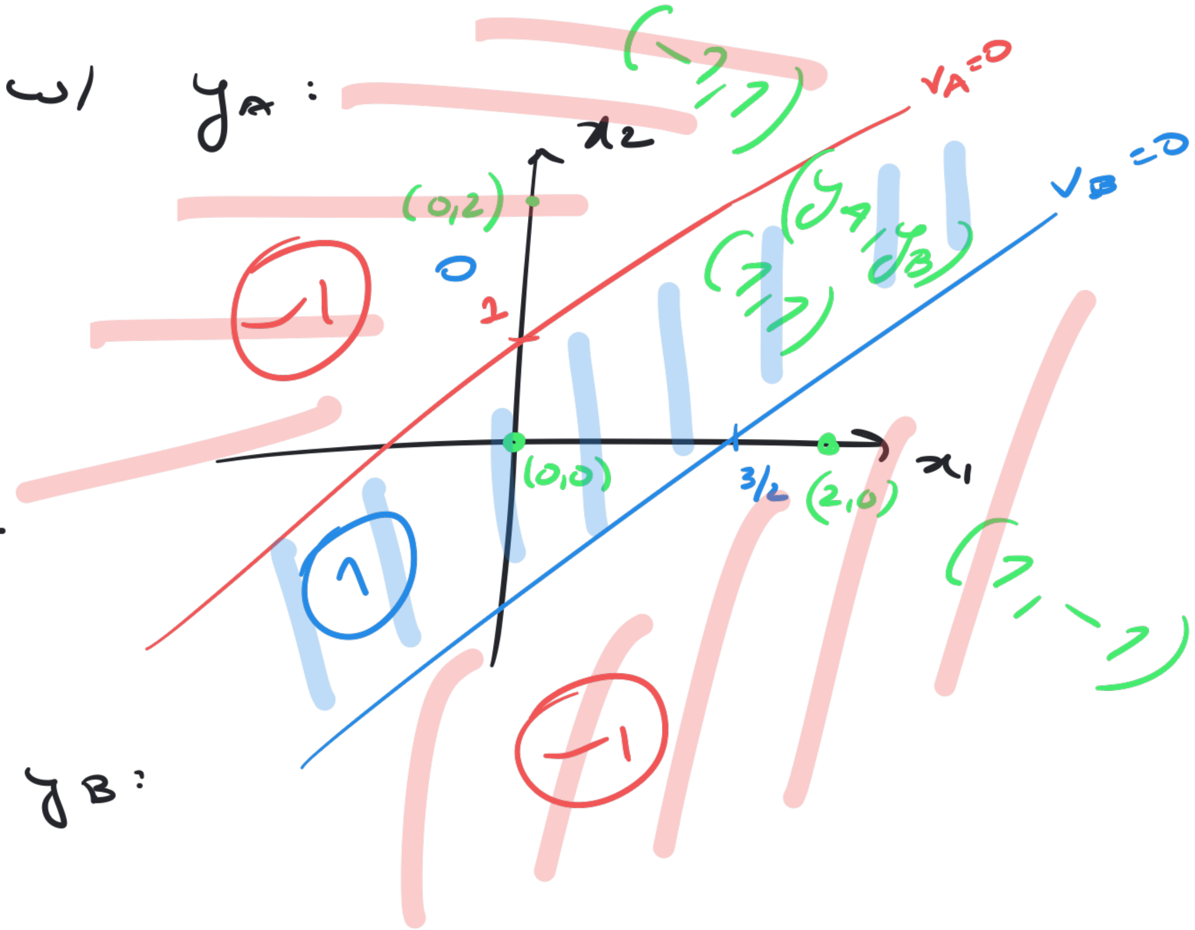


$$\phi(x) = \begin{cases} 1, & x > 0 \\ -1, & x \le 0 \end{cases}$$

$$y_A = \phi(v_A) = \phi(1 \cdot x_1 - 1 \cdot x_2 + 1)$$

$$y_B = \phi(v_B) = \phi(-2 \cdot x_1 + 2 \cdot x_2 + 3)$$

Line associated w/ y_A:



$v_A = 0$

$\iff x_1 - x_2 + 1 = 0$

$\iff x_2 = x_1 + 1$

Line associated w/ y_B:

$v_B = 0$

$\iff -2x_1 + 2x_2 + 3 = 0$

$\iff x_2 = x_1 - \frac{3}{2}$

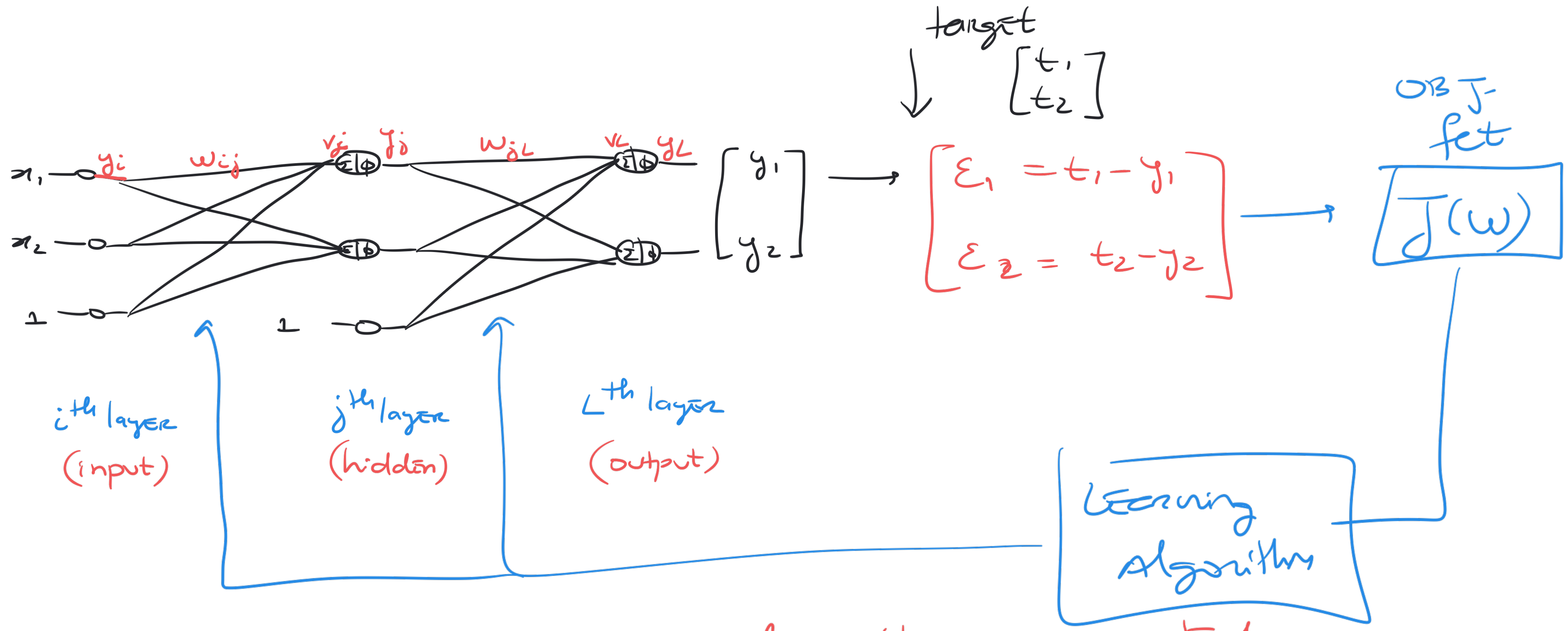
x_1	x_2	y_A	y_B
0	0	$\phi(0-0+1) = 1$	$\phi(0+0+3) = 1$
0	2	$\phi(-2+1) = -1$	$\phi(4+3) = 1$
2	0	$\phi(2+1) = 1$	$\phi(-4+3) = -1$

<u>output:</u> $y = \phi(y_A + y_B)$

y_A	y_B	y
-1	-1	$\phi(-2) = -1$
-1	1	$\phi(-1+1) = -1$
1	-1	$\phi(1-1) = -1$
1	1	$\phi(2) = 1$

AND

gate



STARTING w/ weights directly connected to output layer:

OBJ. fct: $$J(w) = \frac{1}{2}\sum_{i=1}^{N} \varepsilon_i^2$$

output layer: $$y_L = \phi(v_L)$$

weighted sum: $$v_L = \sum_j w_{jL} \cdot y_j$$

ERROR: $$\varepsilon_i = t_i - y_i$$

$$\frac{\partial J(w)}{\partial w_{jL}} = \underbrace{\frac{\partial J}{\partial e_L}}_{e_L} \cdot \underbrace{\frac{\partial e_L}{\partial y_L}}_{(-1)} \cdot \underbrace{\frac{\partial y_L}{\partial v_L}}_{\phi'(v_L)} \cdot \underbrace{\frac{\partial v_L}{\partial w_{jL}}}_{y_j}$$

gradient descent update:

$$\boxed{w_{jL}^{(t+1)} \longleftarrow w_{jL}^{(t)} + \eta \cdot e_L \cdot \phi'(v_L) \cdot y_j}$$

$(t) \equiv$ iteration t