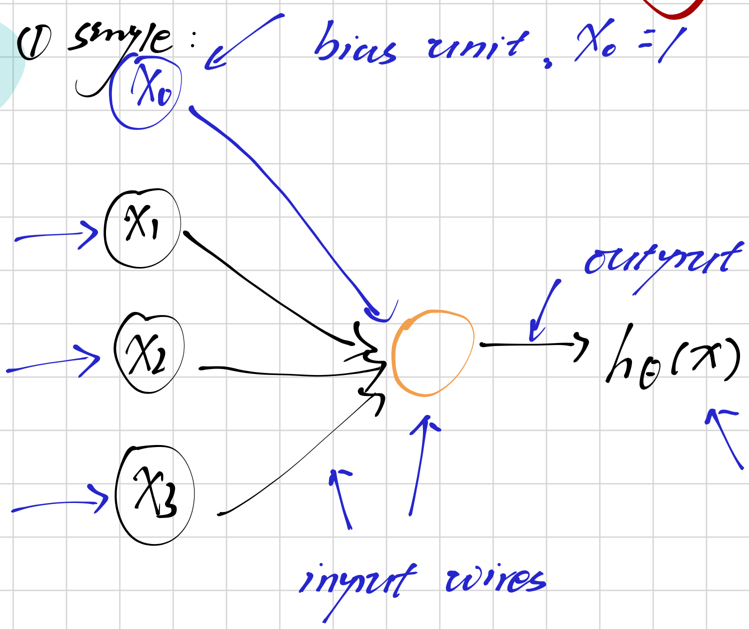
§ Machine Learning – Week 4

1.

① single:



$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

Sigmoid (logistic) activation function

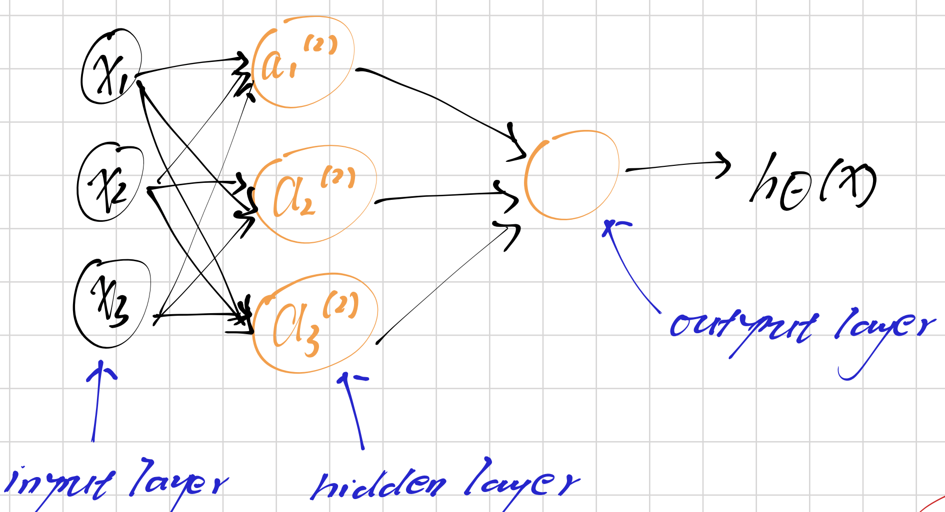
$$g(z) = \frac{1}{1+e^{-z}}$$

"weights"

‖

parameters

② neural network:



$a_i^{(j)}$ = activation of unit i in layer j

$\Theta^{(j)}$ = matrix of weights controlling function mapping from layer j to layer $j+1$

$$\begin{cases} a_1^{(2)} = g(\underbrace{\Theta_{10}^{(1)} x_0 + \Theta_{11}^{(1)} x_1 + \Theta_{12}^{(1)} x_2 + \Theta_{13}^{(1)} x_3}_{z_1^{(2)}}) \\ a_2^{(2)} = g(\underbrace{\Theta_{20}^{(1)} x_0 + \Theta_{21}^{(1)} x_1 + \Theta_{22}^{(1)} x_2 + \Theta_{23}^{(1)} x_3}_{z_2^{(2)}}) \\ a_3^{(2)} = g(\underbrace{\Theta_{30}^{(1)} x_0 + \Theta_{31}^{(1)} x_1 + \Theta_{32}^{(1)} x_2 + \Theta_{33}^{(1)} x_3}_{z_3^{(2)}}) \end{cases}$$

$$\Rightarrow \underbrace{\Theta^{(1)}}_{\in \mathbb{R}^{3\times 4}} = \begin{bmatrix} \Theta_{10} & \Theta_{11} & \Theta_{12} & \Theta_{13} \\ \Theta_{20} & \Theta_{21} & \Theta_{22} & \Theta_{23} \\ \Theta_{30} & \Theta_{31} & \Theta_{32} & \Theta_{33} \end{bmatrix}$$

$$h_\Theta(x) = a_1^{(3)} = g(\underbrace{\Theta_{10}^{(2)} a_0^{(2)} + \Theta_{11}^{(2)} a_1^{(2)} + \Theta_{12}^{(2)} a_2^{(2)} + \Theta_{13}^{(2)} a_3^{(2)}}_{z^{(3)}})$$

If network has s_j units in layer j, s_{j+1} units in layer $j+1$, then $\Theta^{(j)}$ will be of dimension $s_{j+1} \times (s_j + \underbrace{1}_{\text{bias unit}})$

⇓ vectorized implementation

$$z^{(2)} = \begin{bmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \end{bmatrix} = \Theta^{(1)} x \leftarrow a^{(1)}$$

$$\underset{\mathbb{R}^3}{a^{(2)}} = \begin{bmatrix} a_1^{(2)} \\ a_2^{(2)} \\ a_3^{(2)} \end{bmatrix} = \underset{\mathbb{R}^3}{g(z^{(2)})}$$

Add $a_0^{(2)} = 1$

$$z^{(3)} = \Theta^{(2)} \underbrace{a^{(2)}}_{\mathbb{R}^4}$$

$$h_\Theta(x) = a^{(3)} = g(z^{(3)})$$

$\Rightarrow$ setting $x = a^{(1)}$, then:

$$z^{(j)} = \Theta^{(j-1)} a^{(j-1)}$$

$$a^{(j)} = g(z^{(j)})$$

add bias unit $a_0^{(j)} = 1$

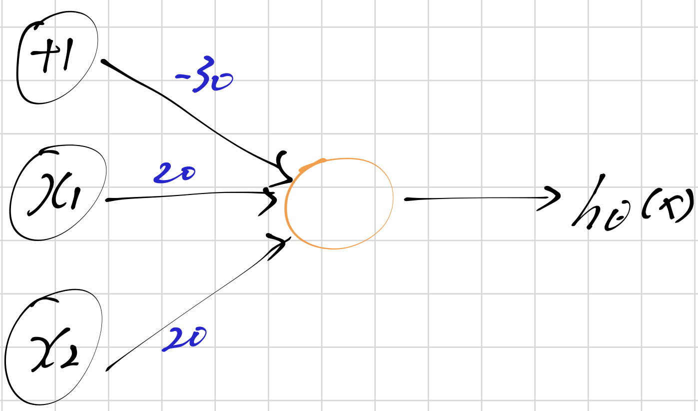
$$z^{(j+1)} = \Theta^{(j)} a^{(j)}$$

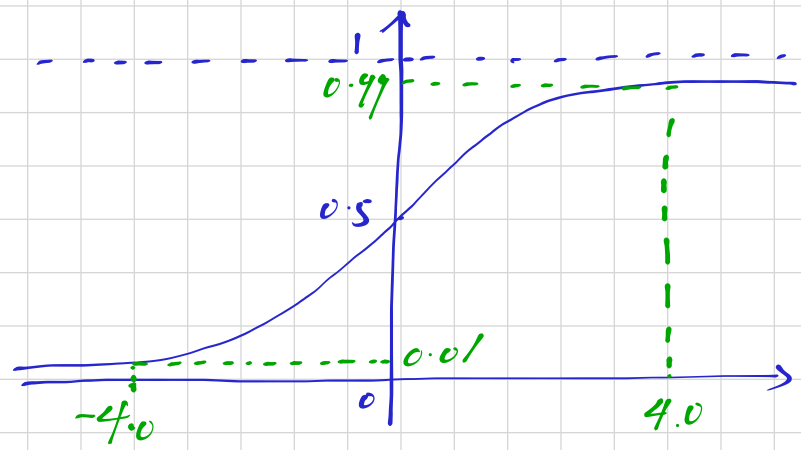
$$h_\Theta(x) = a^{(j+1)} = g(z^{(j+1)})$$

2. examples

(1) AND

eg.





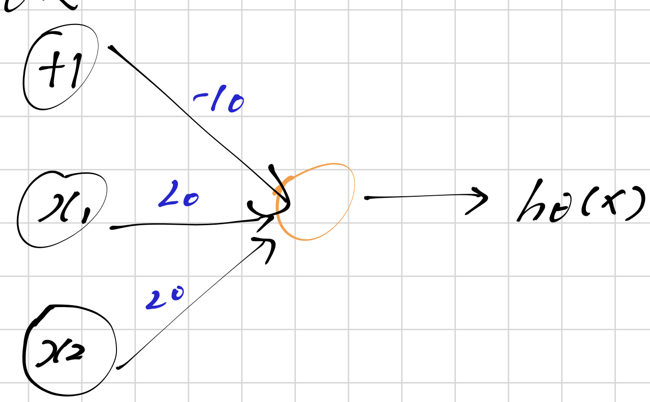
$$h_\Theta(x) = g(-30 + 20x_1 + 20x_2)$$

x_1	x_2	$h_\Theta(x)$
0	0	$g(-30) \approx 0$
0	1	$g(-10) \approx 0$
1	0	$g(-10) \approx 0$
1	1	$g(10) \approx 1$

$h_\Theta(x) \approx x_1$ AND x_2

(2) OR

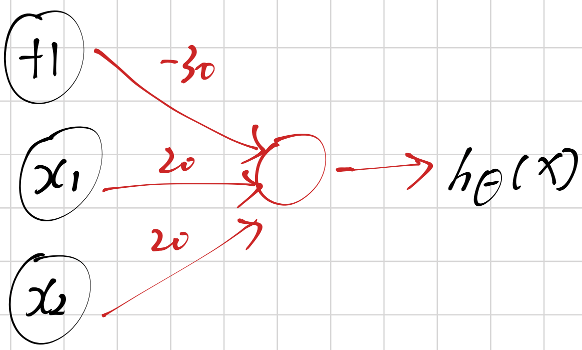
eg.



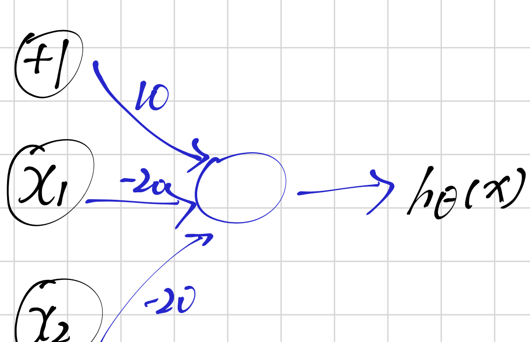
$$h_\Theta(x) = g(-10 + 20x_1 + 20x_2) \approx x_1 \text{ OR } x_2$$

x_1	x_2	$h_\Theta(x)$
0	0	0
0	1	1
1	0	1
1	1	1

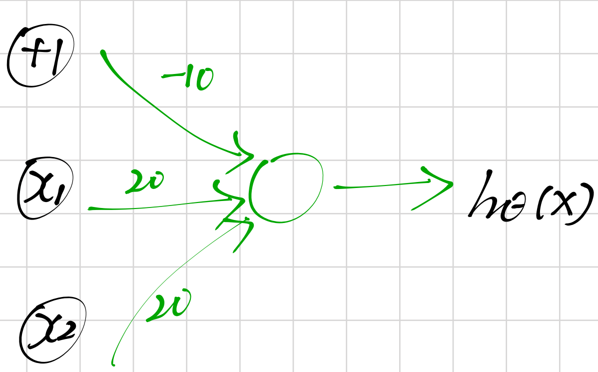
(3) More complex: x_1 XNOR x_2 ← NOT (x_1 XOR x_2)



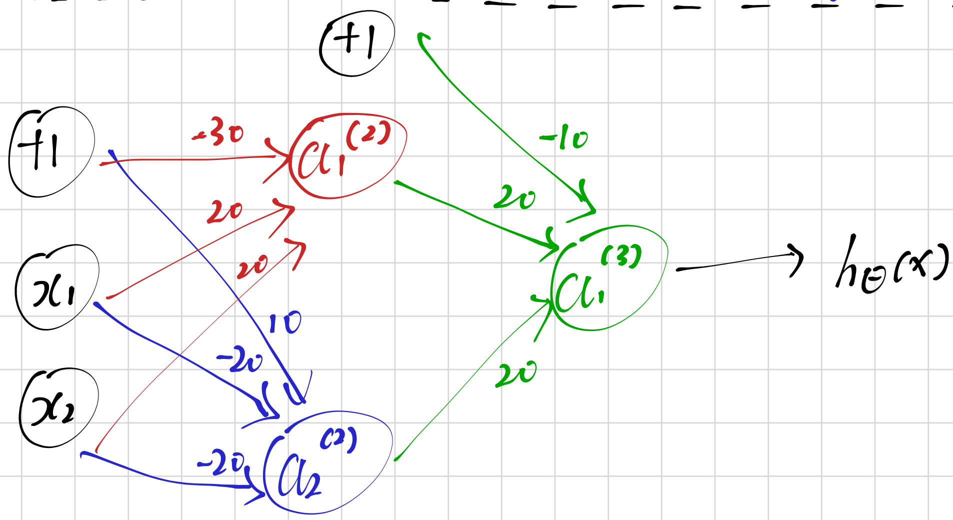
x_1 AND x_2



(NOT x_1) AND (NOT x_2)



x_1 OR x_2



x_1	x_2	$a_1^{(2)}$	$a_2^{(2)}$	$h_\Theta(x)$
0	0	0	1	1
0	1	0	0	0
1	0	0	0	0
1	1	1	0	1
		AND	NOT(AND)NOT	OR

$\underbrace{\text{AND} \quad \text{NOT(AND)NOT} \quad \text{OR}}_{\text{XNOR}}$