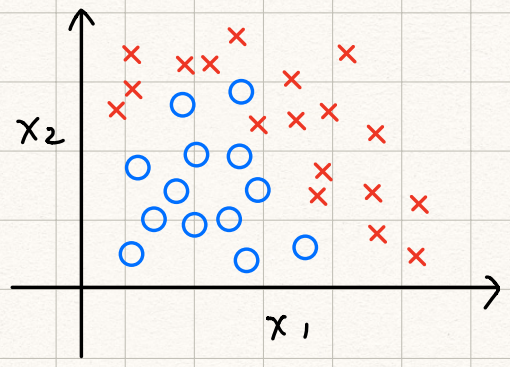
Motivations

Why need Neural Network

非線性分類器



如何定義 features 和 x_1, x_2 的關係呢？

$$g(\theta_0 + \theta_1 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 x_2 + \cdots)$$

More features ?

x_1 = size

x_2 = age

⋮

x_{100}

→ $x_1^2 + x_2^2 + x_3^2 + \cdots + x_{100}^2$

擬合不足

→ $x_1 x_2 x_3 + x_1^2 x_2 + \cdots$

太多項，計算時間爆表

Example of computer vision

50 × 50 pixel images → 2500 pixels

n = 2500 (7500 if RGB)

then, features $\simeq$ 3 million

Sensor representations in the brain

BrainPort —— Seeing with tongue

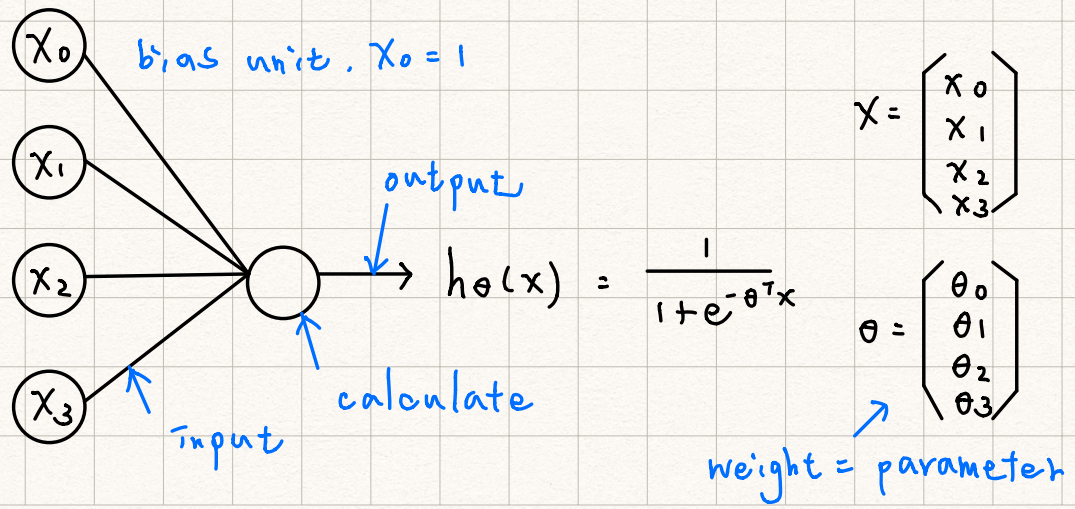
Human echolocation (sonor)

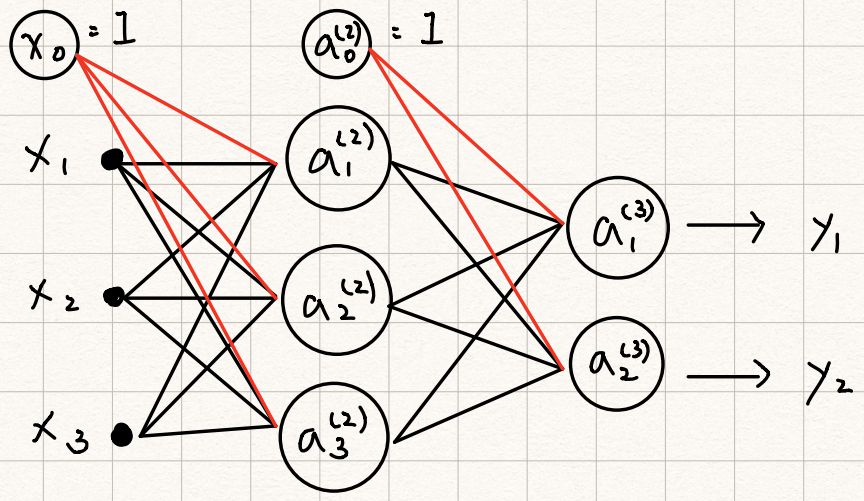
Haptic belt: Direction sense

Implanting a 3rd eye

Neural Networks

Model representation





3 hidden unit

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

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

$$a_1^{(2)} = g(\theta_{10}^{(1)} x_0 + \theta_{11}^{(1)} x_1 + \theta_{12}^{(1)} x_2 + \theta_{13}^{(1)} x_3)$$

$$a_2^{(2)} = g(\theta_{20}^{(1)} x_0 + \theta_{21}^{(1)} x_1 + \theta_{22}^{(1)} x_2 + \theta_{23}^{(1)} x_3)$$

$$a_3^{(2)} = g(\theta_{30}^{(1)} x_0 + \theta_{31}^{(1)} x_1 + \theta_{32}^{(1)} x_2 + \theta_{33}^{(1)} x_3)$$

$$h_\theta(x) = a_1^{(3)}$$
$$= g(\theta_{10}^{(2)} a_0^{(2)} + \theta_{11}^{(2)} a_1^{(2)} + \theta_{12}^{(2)} a_2^{(2)} + \theta_{13}^{(2)} a_3^{(2)})$$

if network has s_j units in layer j, s_{j+1} units in layer $j+1$, then $\theta^{(j)}$ will be dimension $s_{j+1} \times (s_j + 1)$.

$$a_1^{(2)} = g(z_1^{(2)})$$

$$a_2^{(2)} = g(z_2^{(2)})$$

$$\vdots$$

$$x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad z^{(2)} = \begin{pmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \end{pmatrix}$$

$$a^{(2)} = g(z^{(2)}), \quad z^{(2)} = \theta^{(1)} a^{(1)}$$

$a^{(2)}$: hidden layer 的輸出值

$g(z^{(2)})$: hidden layer 的輸入值

$\theta^{(1)}$: weight

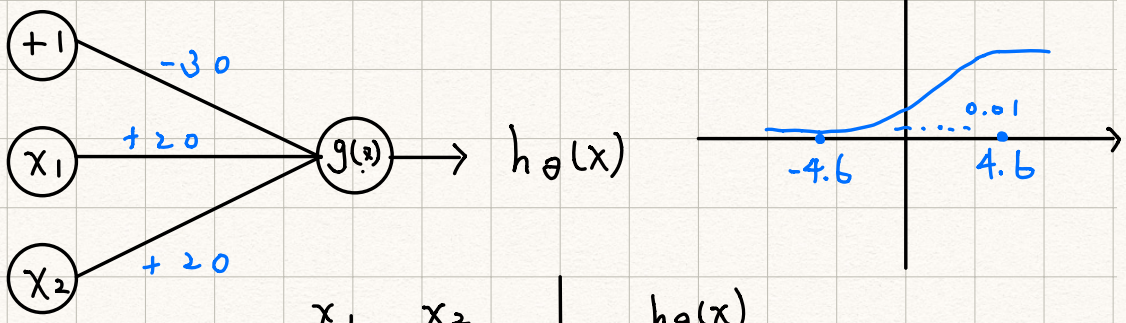
$a^{(1)}$: 輸入值

$$z^{(3)} = \theta^{(2)} a^{(2)}, \quad h_\theta(x) = a^{(3)} = g(z^{(3)})$$

Application of ML

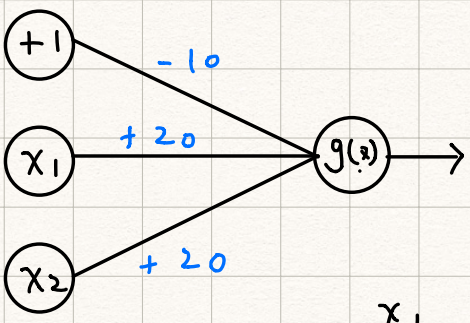
XNOR, NOR, NOT, OR

1. NOT $h_\theta(x) \approx x_1$ AND x_2



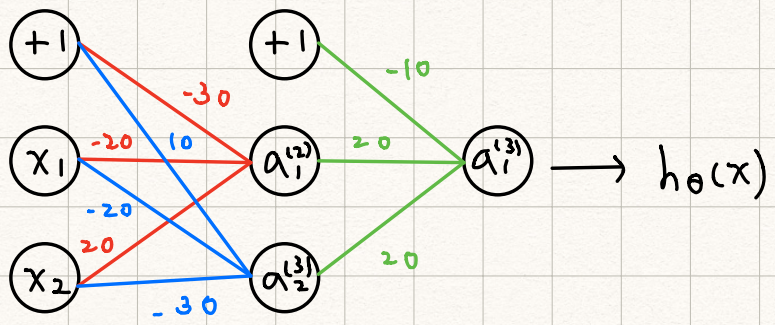
x_1	x_2	$h_\theta(x)$
0	0	$g(-30) \approx 0$
0	1	$g(-10) \approx 0$
1	0	$g(-10) \approx 0$
1	1	$g(10) \approx 1$

2. OR $h_\theta(x) \approx x_1$ OR x_2



x_1	x_2	$h_\theta(x)$
0	0	≈ 0
0	1	≈ 1
1	0	≈ 1
1	1	≈ 1

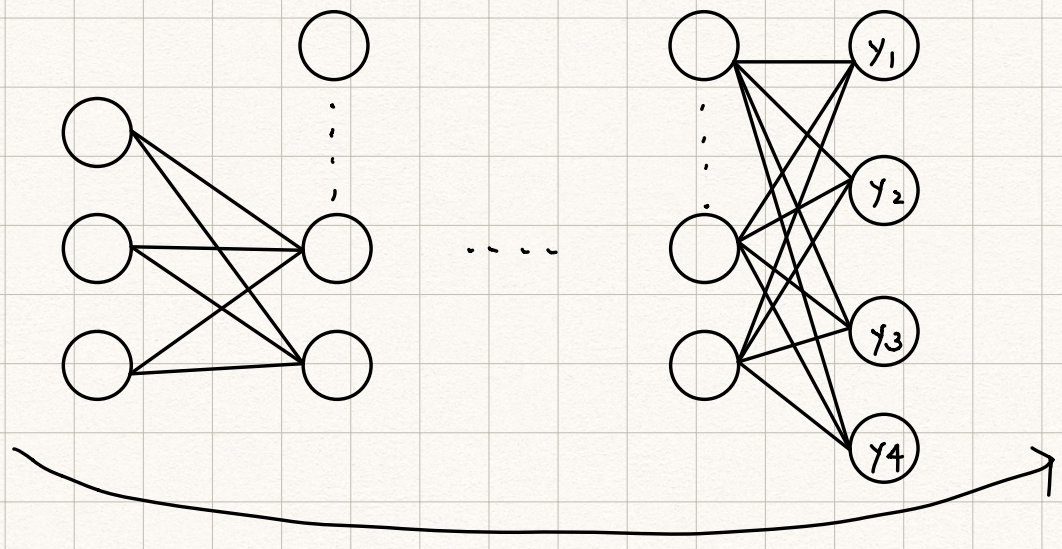
3. X_1 XNOR X_2



4 數字辨識 →

Multi-class classification

computer vision: 判別圖片是人/車/狗/貓



$$y = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$ when 人

$$y = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$ when 車

$$y = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$ when 狗 , etc