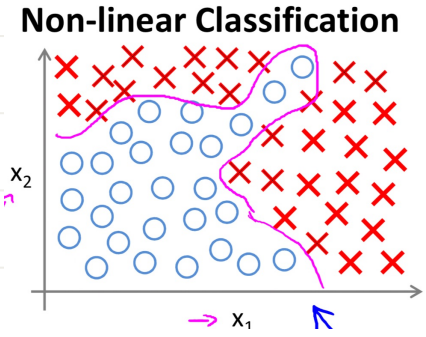
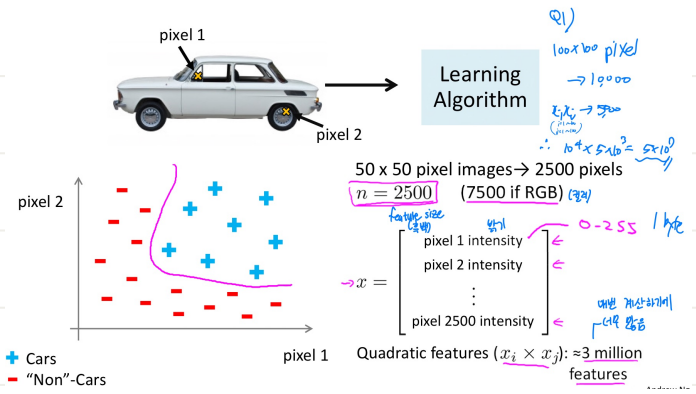
8. Neural Networks

8.1 Non-linear Classification



← 복잡한 linear classification 할 때,
feature의 수가 매우 많아지는 단점이 있음
∴ 연산량↑, overfitting 가능성↑



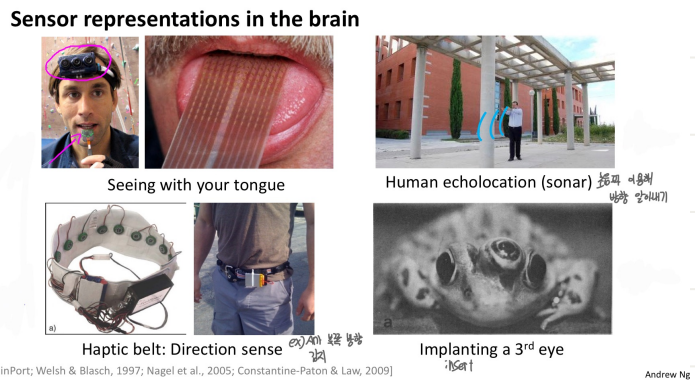
ex) computer vision에서
자동차인지 아닌지 맞히는 알고리즘
labled된 data가 아닌 data 준비 → 보통은 학습 → 학습 알고리즘 적용
50x50 pixel images x 3(RGB) → 2500 pixels
$x = \begin{bmatrix} \text{pixel 1} \\ \text{pixel 2} \\ \vdots \\ \text{pixel 2500} \end{bmatrix}$ → pixel 하나당 (0-255) 값 (feature)
≈ 3 million features 가짐 (매번 계산하기에 너무 많음)

8.2 Neurons and the brain

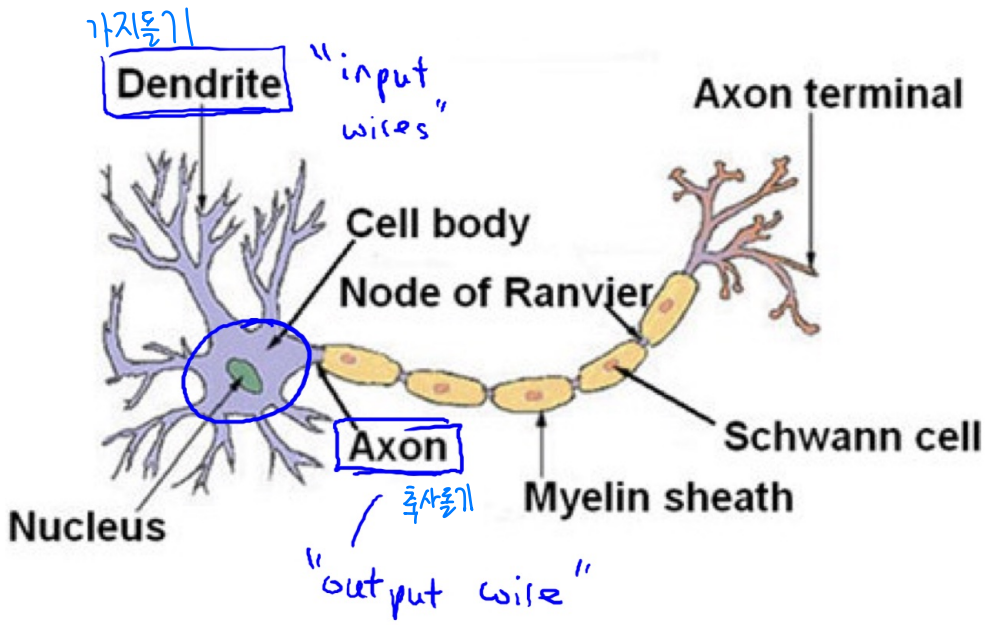
: 뇌를 모방한 알고리즘
: 최근 다시 많이 쓰이게 시작(← 컴퓨터 성능 빨라져 신경망 비용이 많이 들어도 잘 다룰 수 있게 됨)

(The "one learning algorithm" hypothesis)

ex) 청각피질 or 체성감각 피질 등이 시신경 연결 → 해당 피질이 보는것 학습
∴ 어떤 감각 기관에서 받은 신호든 단 하나의 알고리즘에 따라 처리한다는 가설



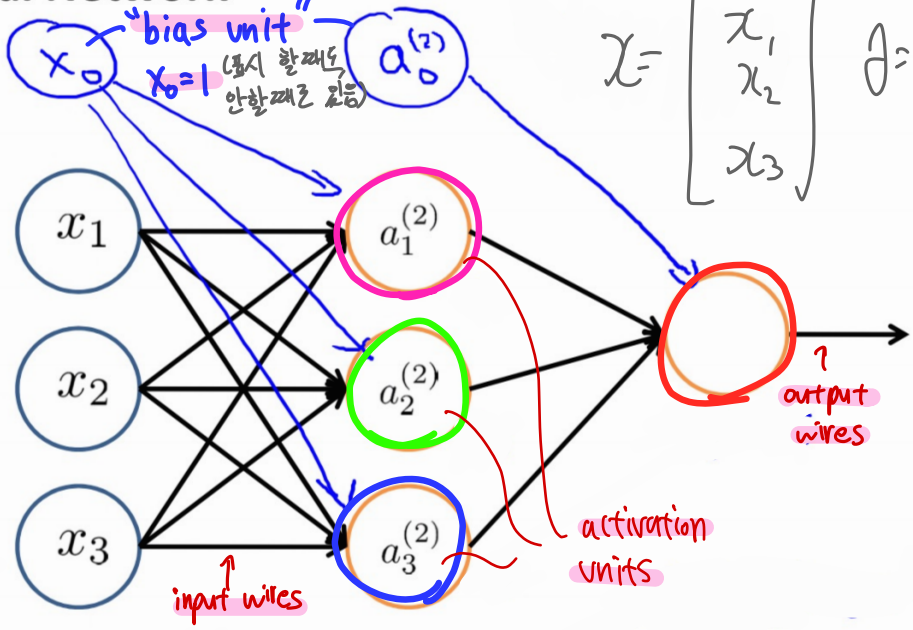
ex) · 혀를 통해 봄 (촉각 → 시각)
· 초음파 이용해 방향 감지 (청각→시각)
· 방향 센서 이용해 방향 감지 (촉각→시각)



8. Neural Networks

8.3 Model representation

Neural Network



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

"weights" (NN에서 θ 이렇게 부름)

$h_\Theta(x) = \frac{1}{1 + e^{-\theta^T x}}$ (sigmoid 로 activation f-tn)

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

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

NN이 layer j에 s_j개 units
layer $j+1$에 s_{j+1}개 units 까지면
⇒ $\Theta^{(j)}$는 $s_{j+1} \times (s_j + 1)$ 의 dimension 갖는다.

bias unit 때문에 생김 (input 층에 포함O output 층 때 "X)

ex) $\theta^{(1)} \in \mathbb{R}^{3\times 4}$ $(s_1 = 3,\ s_2 = 3)$

Layer 1 — input layer
Layer 2 — Hidden layer (input, output 제외한 모든 layer)
Layer 3 — output layer

$$\rightarrow a_1^{(2)} = g(\Theta_{10}^{(1)} x_0 + \Theta_{11}^{(1)} x_1 + \Theta_{12}^{(1)} x_2 + \Theta_{13}^{(1)} x_3) = g(z_1^{(2)})$$

$$\rightarrow a_2^{(2)} = g(\Theta_{20}^{(1)} x_0 + \Theta_{21}^{(1)} x_1 + \Theta_{22}^{(1)} x_2 + \Theta_{23}^{(1)} x_3) = g(z_2^{(2)})$$

$$\rightarrow a_3^{(2)} = g(\Theta_{30}^{(1)} x_0 + \Theta_{31}^{(1)} x_1 + \Theta_{32}^{(1)} x_2 + \Theta_{33}^{(1)} x_3) = g(z_3^{(2)})$$

$$\rightarrow h_\Theta(x) = a_1^{(3)} = g(\Theta_{10}^{(2)} a_0^{(2)} + \Theta_{11}^{(2)} a_1^{(2)} + \Theta_{12}^{(2)} a_2^{(2)} + \Theta_{13}^{(2)} a_3^{(2)}) = g(z_1^{(3)})$$

〈forward propagation〉 순전파

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

$z^{(2)} = \underset{(3\times 4)}{\Theta^{(1)}} \underset{(4\times 1)}{a^{(1)}}$

$\underset{\mathbb{R}^3}{a^{(2)}} = g(\underset{\mathbb{R}^3}{z^{(2)}})$

$\rightarrow a_0^{(2)} = 1$ 더해주기 (input에만 포함)

$\underset{(3\times 1)}{z^{(3)}} = \underset{(1\times 4)}{\Theta^{(2)}} \underset{(4\times 1)}{a^{(2)}}$

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

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

NN은 자기 자신이 feature들을 학습한다.

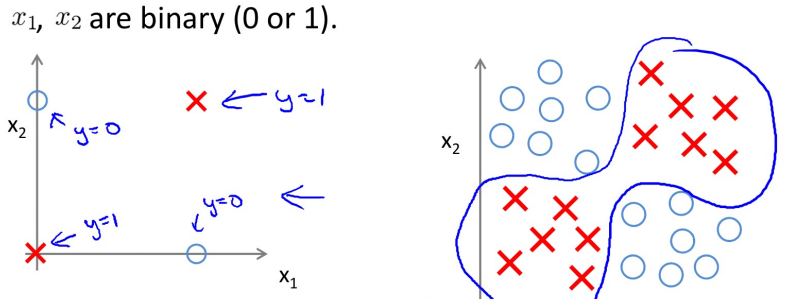
↑ logistic regression에서 봤던 $h(x)$의 비슷한 형태 (θ 대신 Θ, $x_1, x_2, x_3 \rightarrow a_1^{(2)}, a_2^{(2)}, a_3^{(2)}$ 사용)

8. Neural Networks

8.4 Examples and intuitions

Non-linear classification example: XOR/XNOR

→ x_1, x_2 are binary (0 or 1).



$y = x_1 \text{ XOR } x_2$ (XOR : 입력된 2개중 1개만 참일때 참)

$\Leftrightarrow x_1 \text{ XNOR } x_2$

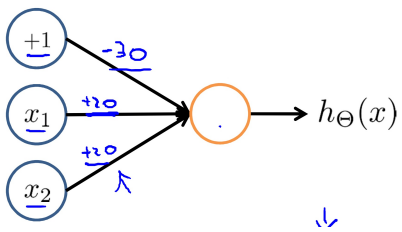
$\Leftrightarrow \text{NOT } (x_1 \text{ XOR } x_2)$ (XNOR : 서로 같으면 참)

ex1) $x_1, x_2 \in \{0, 1\}$

$y = x_1$ AND x_2

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

$\Theta^{(1)}_{10}$ $\Theta^{(1)}_{11}$ $\Theta^{(1)}_{12}$

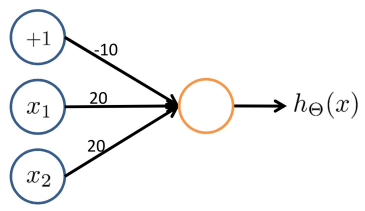


x_1	x_2	$h_\Theta(x)$
0	0	$g(-30) \approx 0$
→ 0	1	$g(-10) \approx 0$
1	0	$g(-10) \approx 0$
→ 1	1	$g(10) \approx 1$

$h_\Theta(x) \approx (x_1 \text{ AND } x_2)$

ex2) $y = x_1$ OR x_2

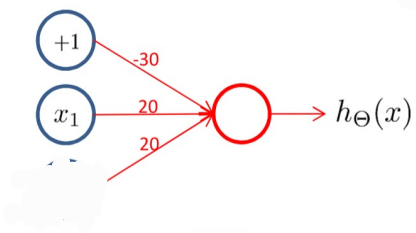
$h_\Theta(x) = g(-10 + 20x_1 + 20x_2)$



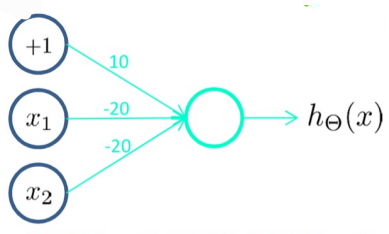
x_1	x_2	$h_\Theta(x)$
0	0	$g(-10) \approx 0$
0	1	$g(10) \approx 1$
1	0	$g(10) \approx 1$
1	1	$g(30) \approx 1$

OR fn

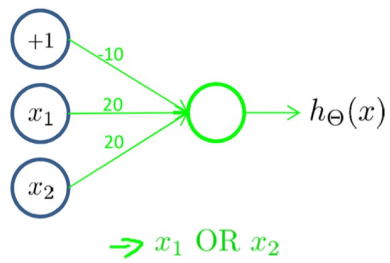
ex3) $y = x_1$ XNOR x_2



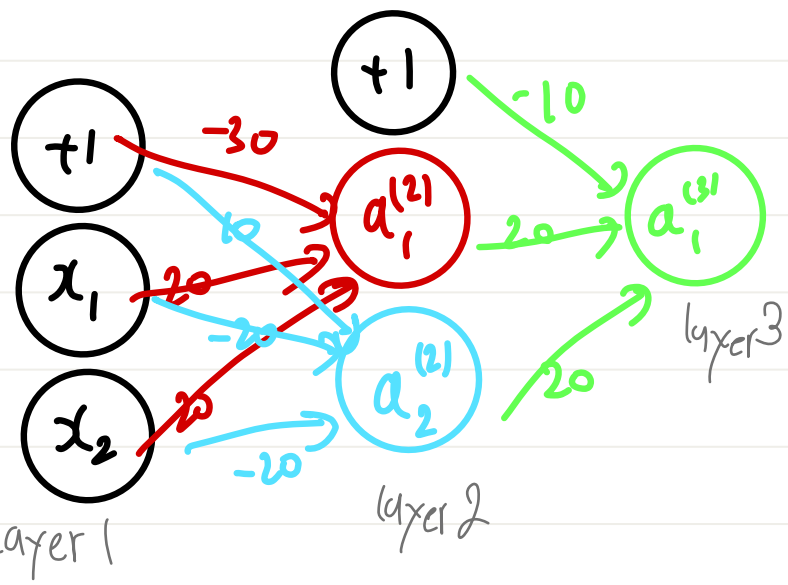
→ x_1 AND x_2



→ (NOT x_1) AND (NOT x_2)



→ x_1 OR x_2



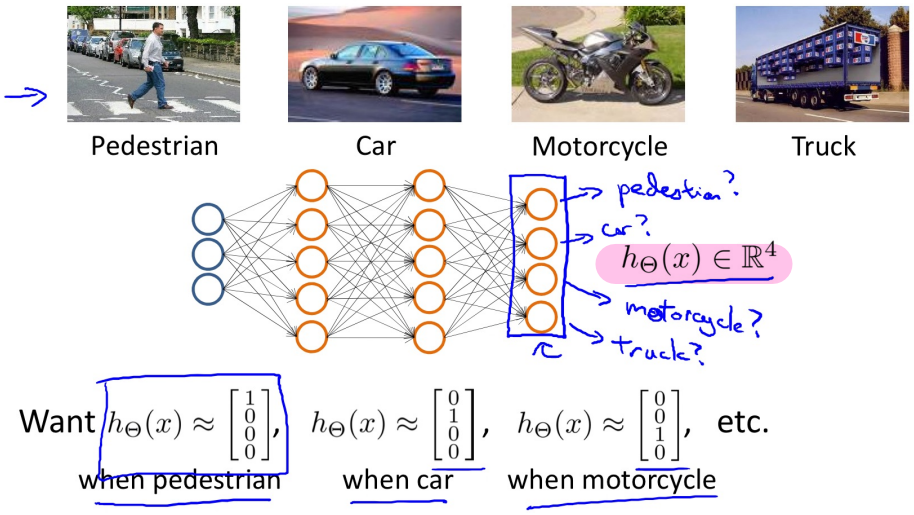
⇒

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 ←

8. Neural Networks

8.5 Multi-class classification

Multiple output units: One-vs-all.



Want $h_\Theta(x) \approx \begin{bmatrix}1\\0\\0\\0\end{bmatrix}$, $h_\Theta(x) \approx \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$, $h_\Theta(x) \approx \begin{bmatrix}0\\0\\1\\0\end{bmatrix}$, etc.

when pedestrian when car when motorcycle

training set

: $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})$

$y^{(i)}$는 $\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$ 중 하나

$h_\Theta(x^{(i)}) \approx y^{(i)}$

$\mathbb{R}^4$