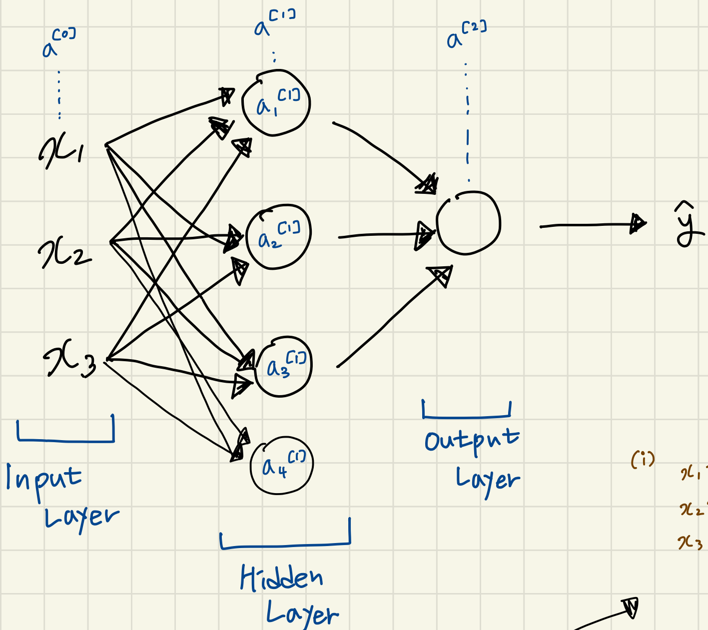
Neural Network Representation



"2 Layer N.N"

☺ Input layer을 count X

* Layer는 총 3개가 맞음

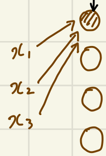
(Input, hidden, output)

단지, count 하지 않는 것일뿐!!

$z_1^{[1]} = w_1^{[1]T} x\ b_1^{[1]}$

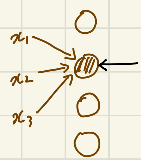
$a_1^{[1]} = \sigma(z_1^{[1]})$

(i)



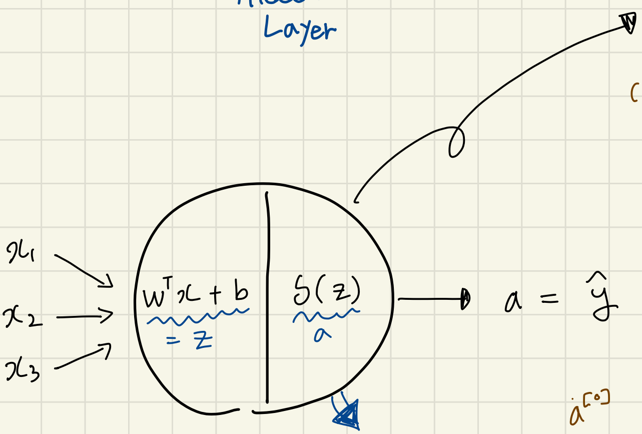
☺ $a_i^{[l]}$ ← Layer, ← Node in Layer

(ii)



$z_2 = w_2^{[1]T} x\ b_2^{[1]}$

$a_2^{[1]} = \sigma(z_2^{[1]})$



$z = W^T X + b$

$a = \sigma(z)$

☺ $z^{[1]} = W^{[1]} \underset{\sim}{X} + b^{[1]}$, $X = a^{[0]}$

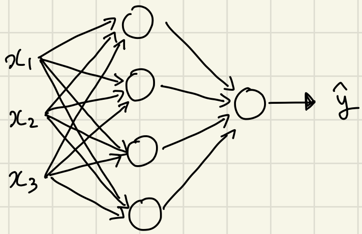
$a^{[1]} = \sigma(z^{[1]})$

$z^{[2]} = W^{[2]} \underset{\sim}{a^{[1]}} + b^{[2]}$

$a^{[2]} = \sigma(z^{[2]})$

⋮

Vectorizing



$a^{[0]} = X \longrightarrow \hat{y} = a^{[2]}$

$x^{(1)} \dashrightarrow \hat{y}^{(1)} = a^{[2](1)}$

$x^{(2)} \dashrightarrow \hat{y}^{(2)} = a^{[2](2)}$

$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$

$x^{(m)} \dashrightarrow \hat{y}^{(m)} = a^{[2](m)}$

$a^{[i](j)}$ ← j th training samples.
↑ i th Layer

<code>

for $i = 1$ to m :

$$z^{[1](i)} = W^{[1]} x^{(i)} + b^{[1]}$$

$$a^{[1](i)} = \sigma(z^{[1](i)})$$

$$z^{[2](i)} = W^{[2]} a^{[1](i)} + b^{[2]}$$

$$a^{[2](i)} = \sigma(z^{[2](i)})$$

$$X = \begin{bmatrix} \vdots & & & \\ x^{(1)} & x^{(2)} & \cdots & x^{(m)} \\ \vdots & & & \end{bmatrix}$$

$$Z^{[1]} = \begin{bmatrix} z^{[1](1)} & z^{[1](2)} & \cdots & z^{[1](m)} \end{bmatrix}$$

다른 입력 특성

다른 훈련 샘플

$$\boxed{Z^{[1]} = W^{[1]} X^{[1]} + b^{[1]}}$$

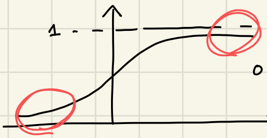
$$z^{[1](1)} = W^{[1]} x^{[1]} + \cancel{b}^{[1]} \to 0, \quad z^{[1](2)} = W^{[1](2)} + \cancel{b}^{[1]} \to 0, \quad z^{[1](3)} = W^{[1](3)} + \cancel{b}^{[1]} \to 0$$

$$W^{[1]} x^{(1)} = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \qquad W^{[1]} x^{(2)} = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \qquad W^{[1]} x^{(3)} = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

$$W^{[1]} \underbrace{\begin{bmatrix} | & | & | \\ x^{(1)} & x^{(2)} & x^{(3)} \\ | & | & | \end{bmatrix}}_{X} = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} = \begin{bmatrix} | & | & | \\ z^{[1](1)} & z^{[1](2)} & z^{[1](3)} \\ | & | & | \end{bmatrix} = Z^{[1]}$$

Activation function

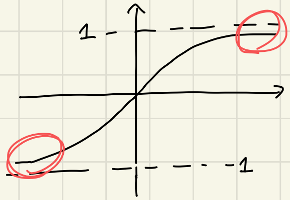
① Sigmoid



$$f = \frac{1}{1+e^{-z}}$$

$0 < f < 1$

② tanh(z)



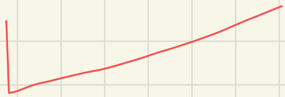
$$f = \tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

$-1 < f < 1$

* 이진분류 출력층 이외에 사용 권장 X

tanh(z) 함수가 더 나음

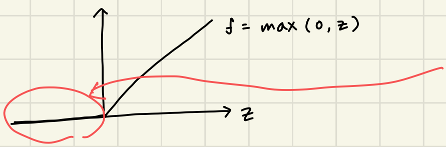
* tanh(z)가 시그모이드보다 거의 항상 좋음



시그모이드 & tanh(z)는 z값이 너무 크거나 작아지면

도함수가 0에 수렴..

③ ReLU



$f = \max(0, z)$

z<0 인 부분에서는 기울기가 0 이지만,

실제로는 대부분의 z의 값이 0보다 크기 때문에

잘 작동.

* 보통, 은닉층 활성화 함수로 ReLU 함수가 쓰인다.

Activation function Derivative

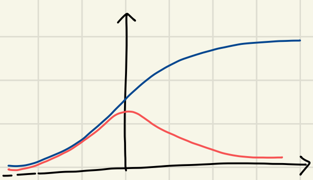
— : Derivative

— : original

① Sigmoid

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

$$g'(z) = \frac{d}{dz} g(z) = g(z)\,(1-g(z))$$



② tanh(z)

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

$$g'(z) = 1 - (g(z))^2$$

③ ReLU

$$g(z) = \max(0, z)$$

$$g'(z) = \begin{cases} 0 & (z < 0) \\ 1 & (z \geq 0) \end{cases}$$