

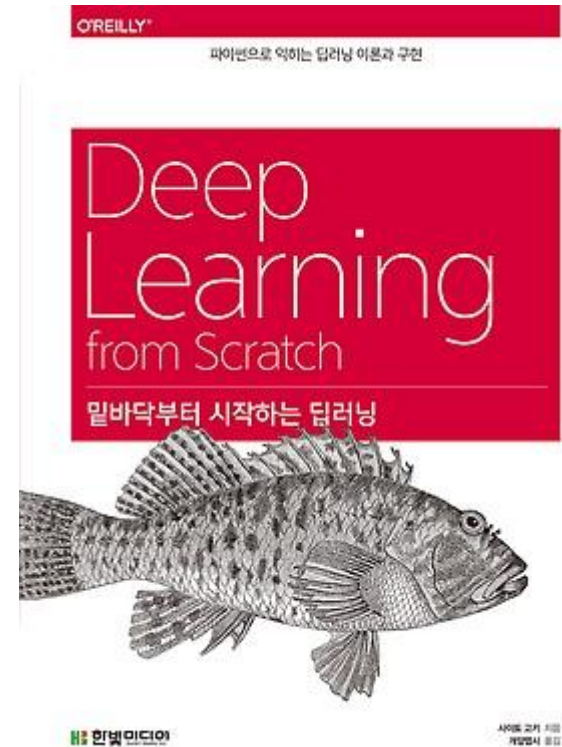
딥러닝

2021.07.28 AAI Lab. 세미나

김민준

참고 자료

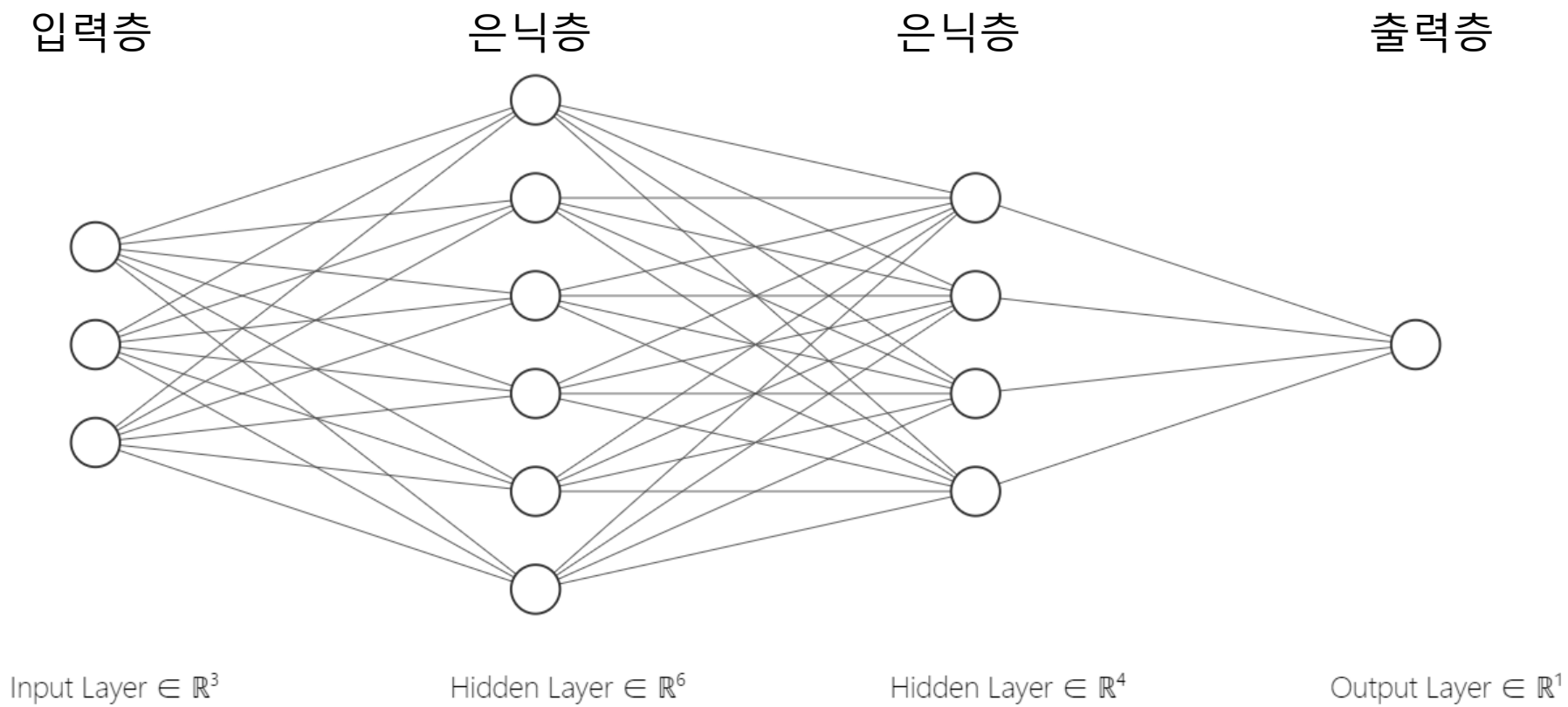
- 유튜브 NeoWizard 머신러닝 강의
- 밑바닥부터 시작하는 딥러닝 - 사이토 고키



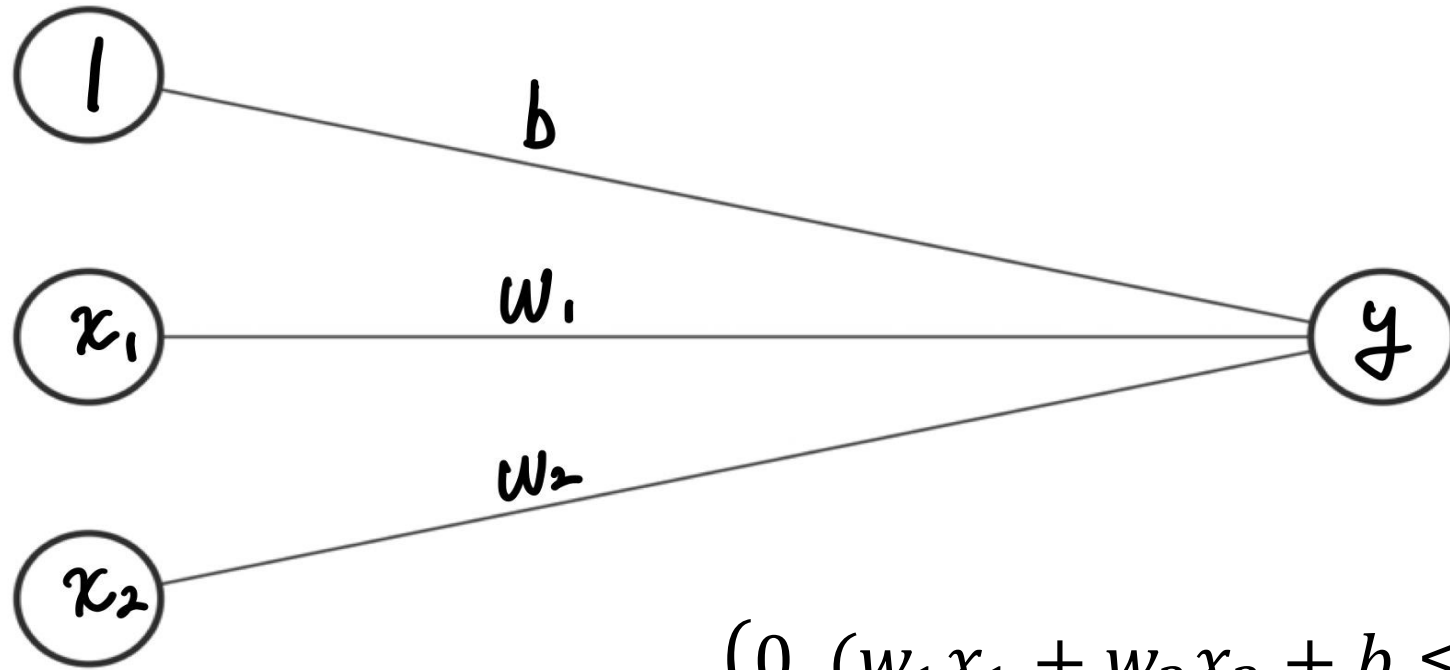
신경망

- 가중치 매개변수의 적절한 값을 데이터로부터 자동으로 학습

신경망



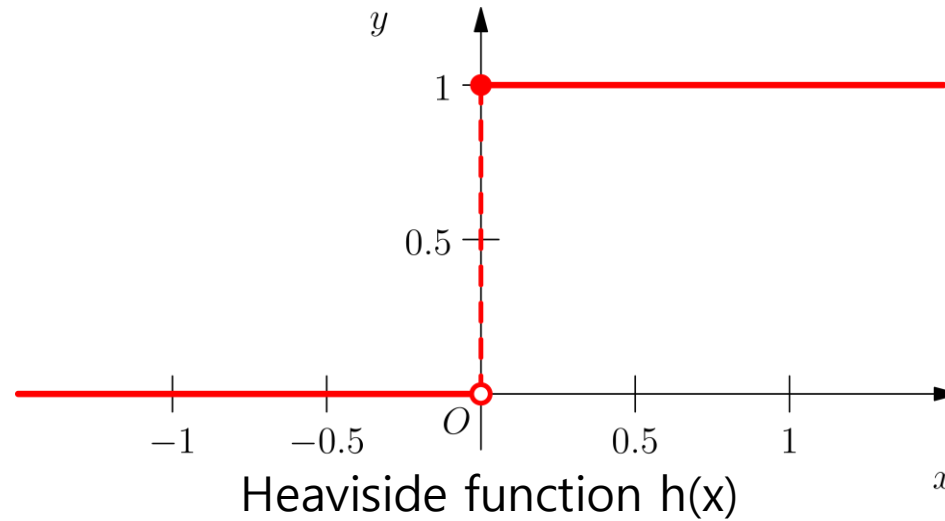
퍼셉트론



$$y = \begin{cases} 0 & (w_1x_1 + w_2x_2 + b \leq 0) \\ 1 & (w_1x_1 + w_2x_2 + b > 0) \end{cases}$$

퍼셉트론

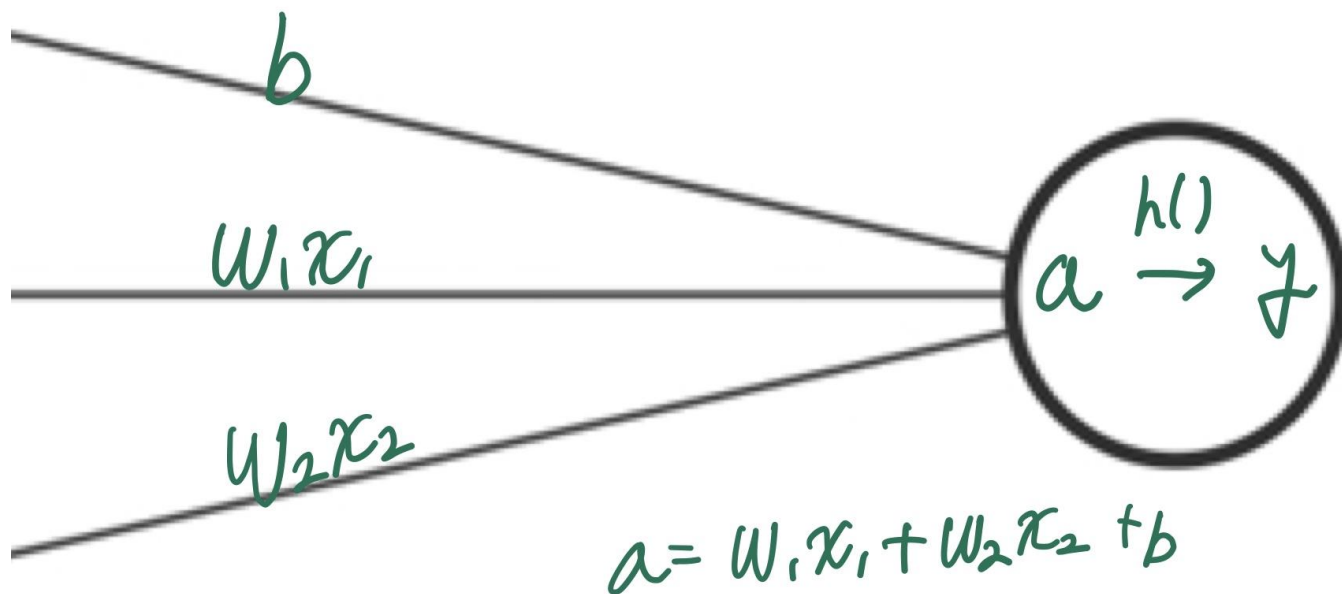
- $$h(x) = \begin{cases} 0 & (x \leq 0) \\ 1 & (x > 0) \end{cases}$$



$$y = \begin{cases} 0 & (w_1x_1 + w_2x_2 + b \leq 0) \\ 1 & (w_1x_1 + w_2x_2 + b > 0) \end{cases}$$

$$\rightarrow y = h(w_1x_1 + w_2x_2 + b)$$

퍼셉트론

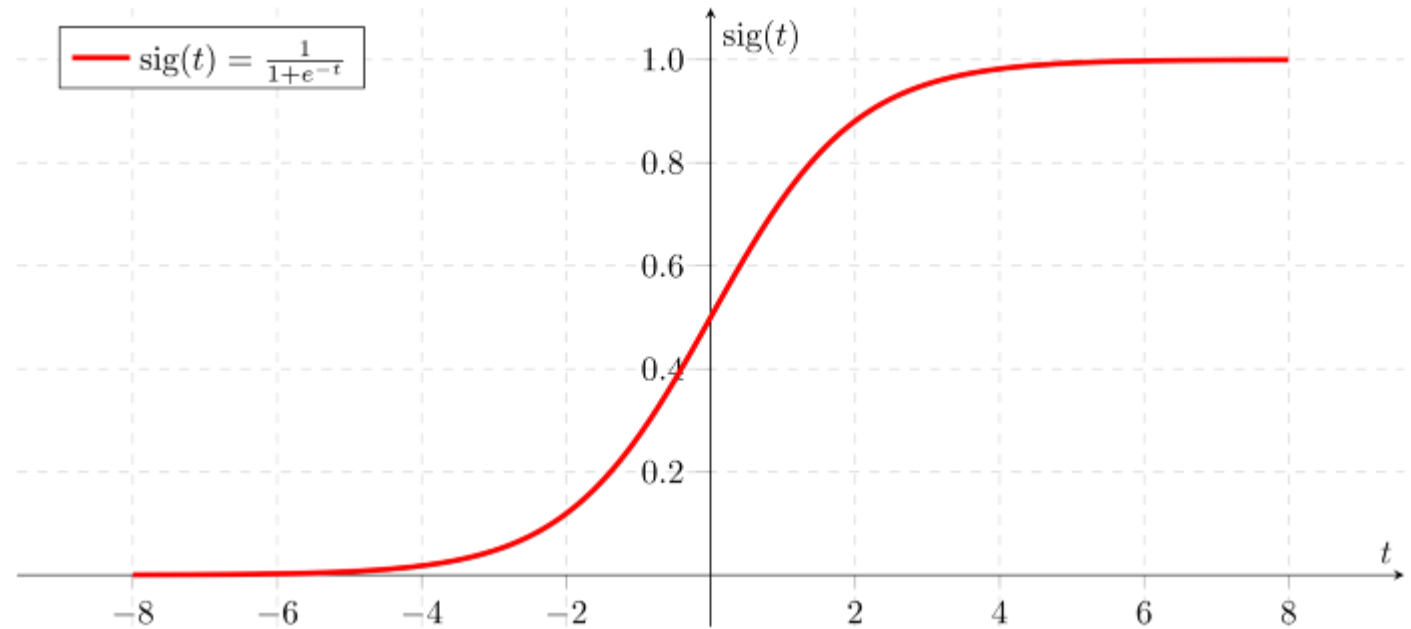


활성화 함수

- Sigmoid
- $\text{sigmoid}(x) = \frac{1}{1+e^{-x}}$
- $0 \leq \text{sigmoid}(x) \leq 1$

계단함수와 유사한 모양

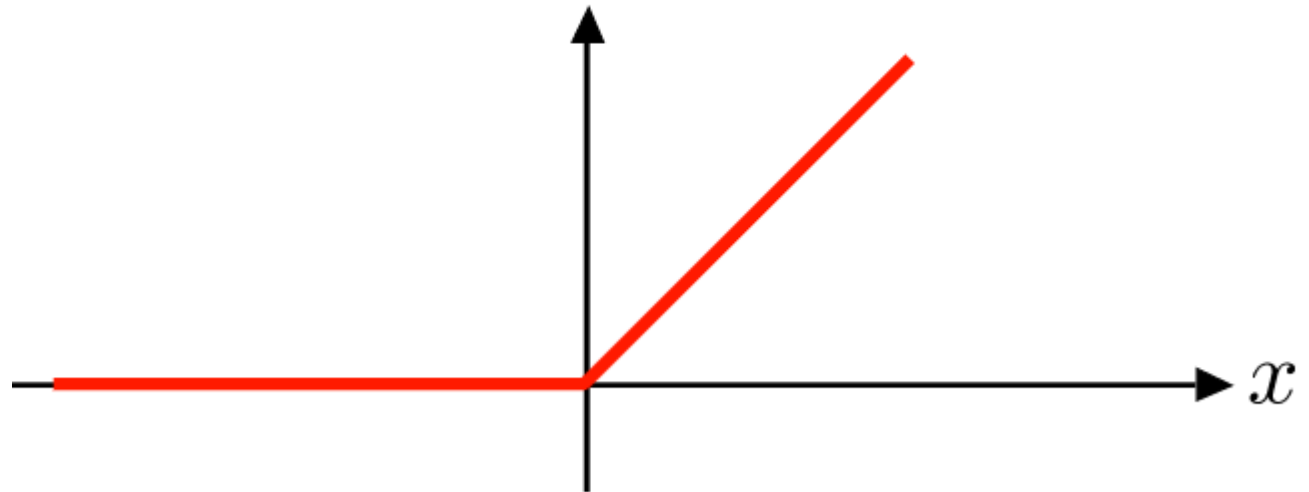
$$\text{sigmoid}'(x) = \text{sigmoid}(x)(1 - \text{sigmoid}(x))$$



활성화 함수

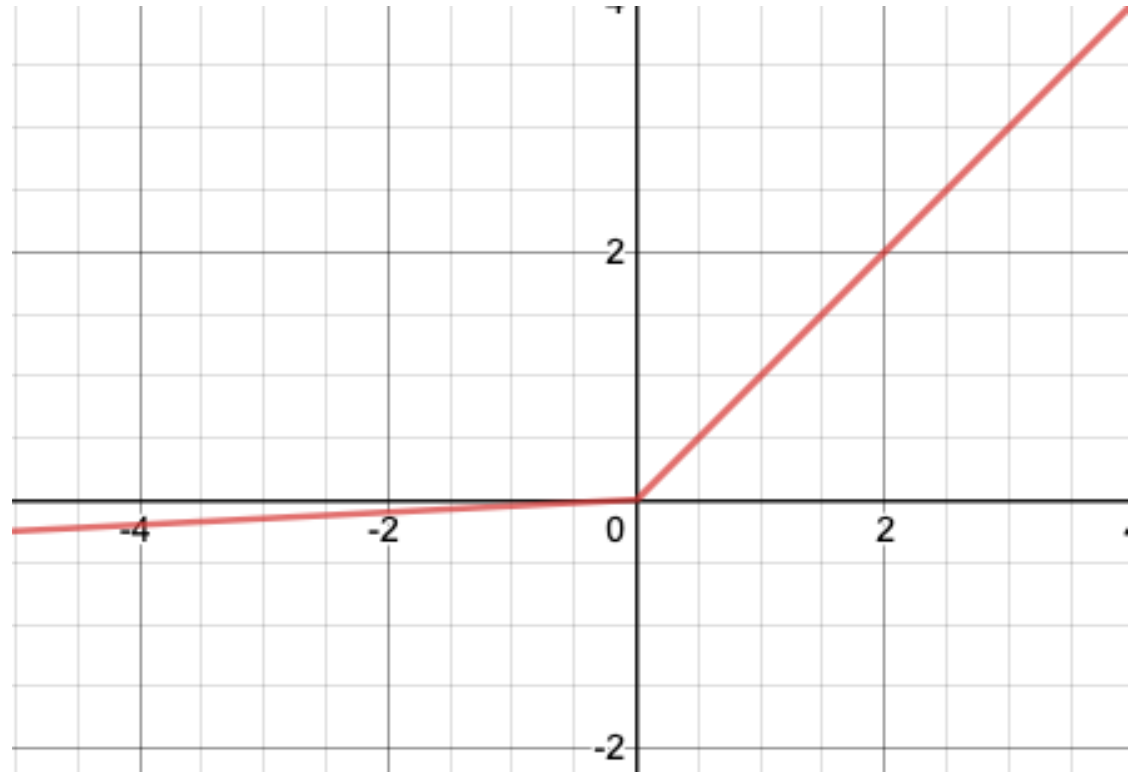
- ReLU
- Solve vanishing gradient

$$\text{ReLU}(x) \triangleq \max(0, x)$$



활성화 함수

- Leaky ReLU
- Solve vanishing gradient

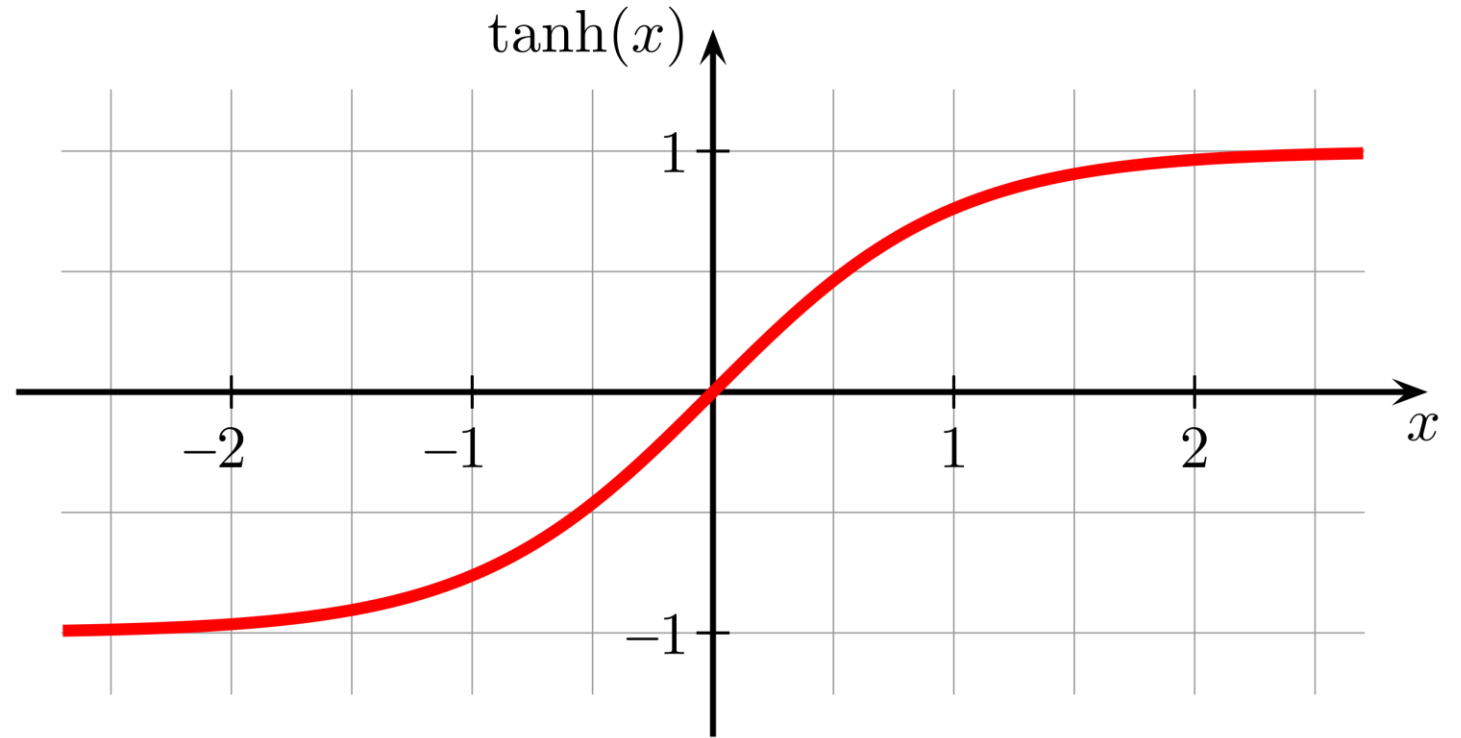


활성화 함수

- Tanh

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

- $-1 \leq \tanh(x) \leq 1$

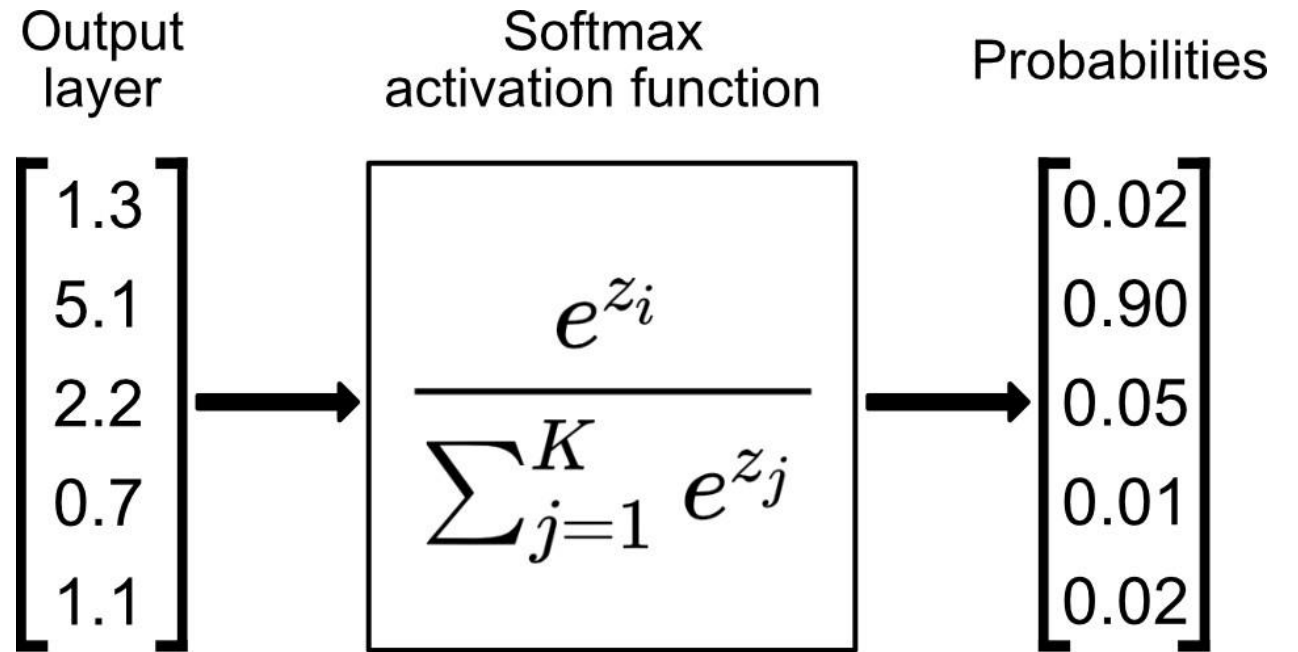


활성화 함수

- Softmax

- $softmax(x) = \frac{e^{x_i}}{\sum e^{x_i}}$

- 출력층에서 확률 값을 표현하기 위해 사용



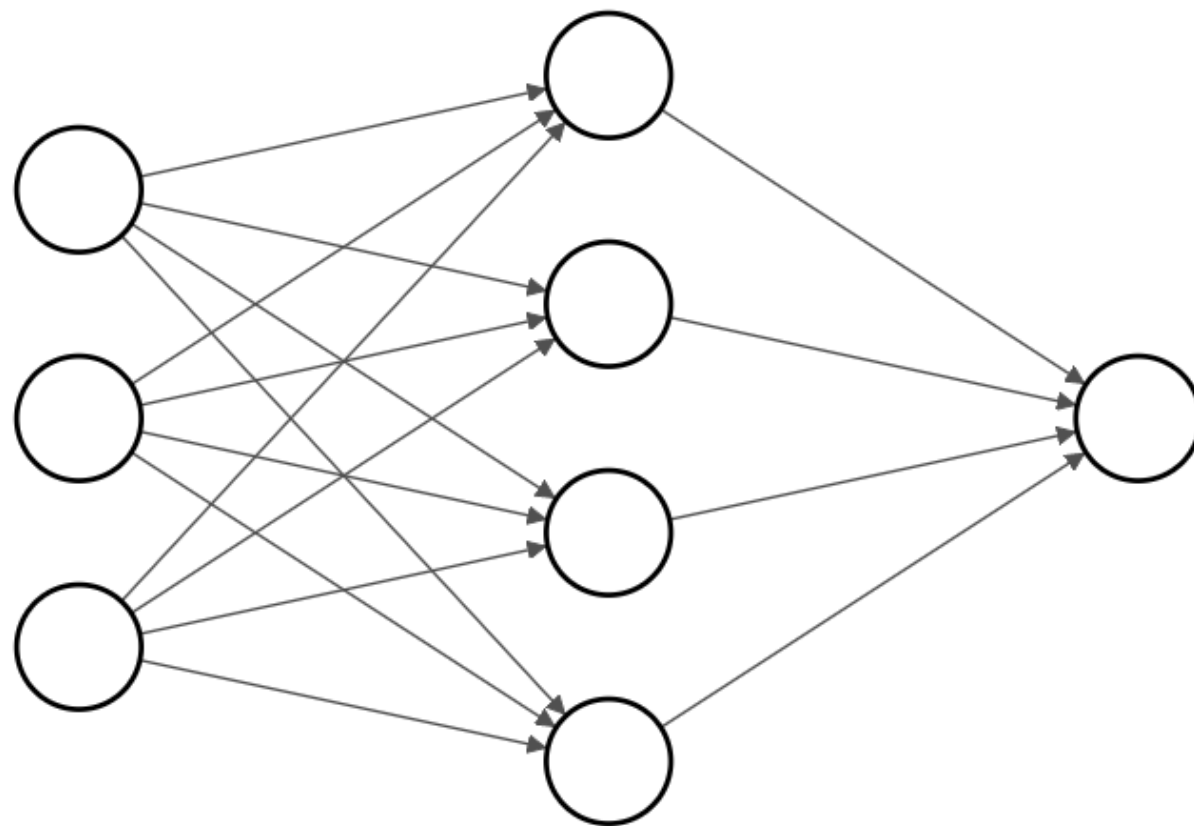
활성화 함수, why?

- Linear \rightarrow non-Linear
- Deep neural network에서 모든 정보를 전달하지 않고 선별적으로 전달 (ReLU의 $x < 0$ 부분)
- Vanishing gradient, gradient overflow 방지
- 확률 값 출력

딥러닝

- 신경망 구조를 바탕으로 입력층, 은닉층, 출력층으로 구성
- 은닉층이 1개 이상일 때 deep neural network
- 은닉층이 깊을수록(많을수록) 정확도가 높아짐

딥러닝

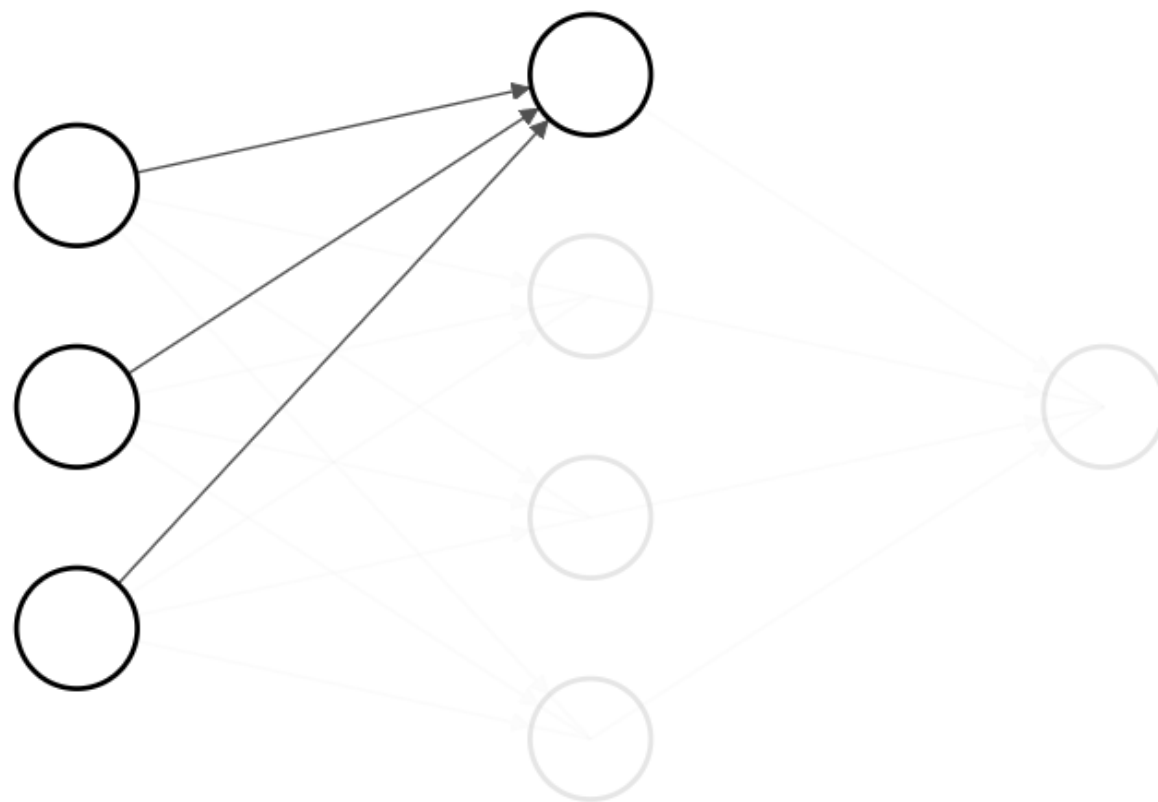


Input Layer $\in \mathbb{R}^3$

Hidden Layer $\in \mathbb{R}^4$

Output Layer $\in \mathbb{R}^1$

딥러닝



Input Layer $\in \mathbb{R}^3$

Hidden Layer $\in \mathbb{R}^4$

Output Layer $\in \mathbb{R}^1$

딥러닝

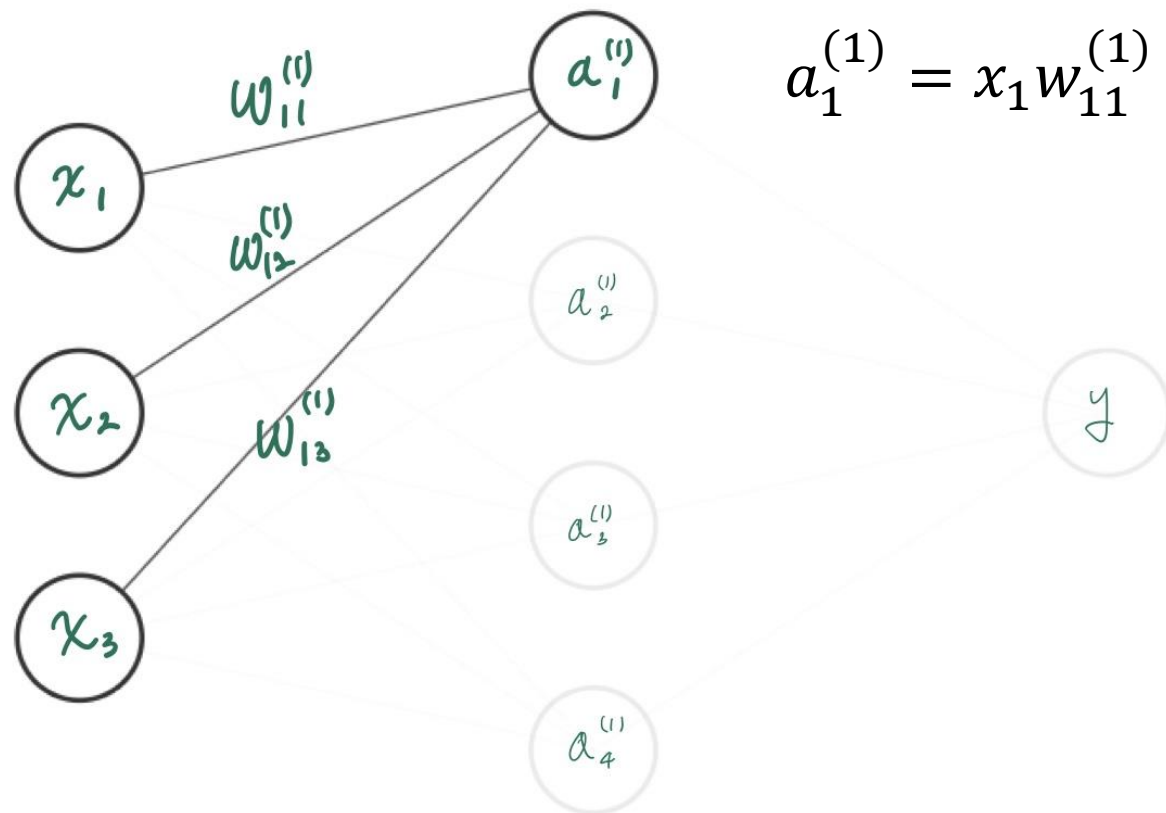
$$X = (x_1, x_2, x_3)$$

$$W^{(1)} = \begin{pmatrix} w_{11}^{(1)} & w_{21}^{(1)} & w_{31}^{(1)} & w_{41}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} & w_{32}^{(1)} & w_{42}^{(1)} \\ w_{13}^{(1)} & w_{23}^{(1)} & w_{33}^{(1)} & w_{43}^{(1)} \end{pmatrix}$$

$$A^{(1)} = (a_1^{(1)}, a_2^{(1)}, a_3^{(1)}, a_4^{(1)})$$

$$XW = A$$

$$M_{13} \times M_{34} = M_{14}$$



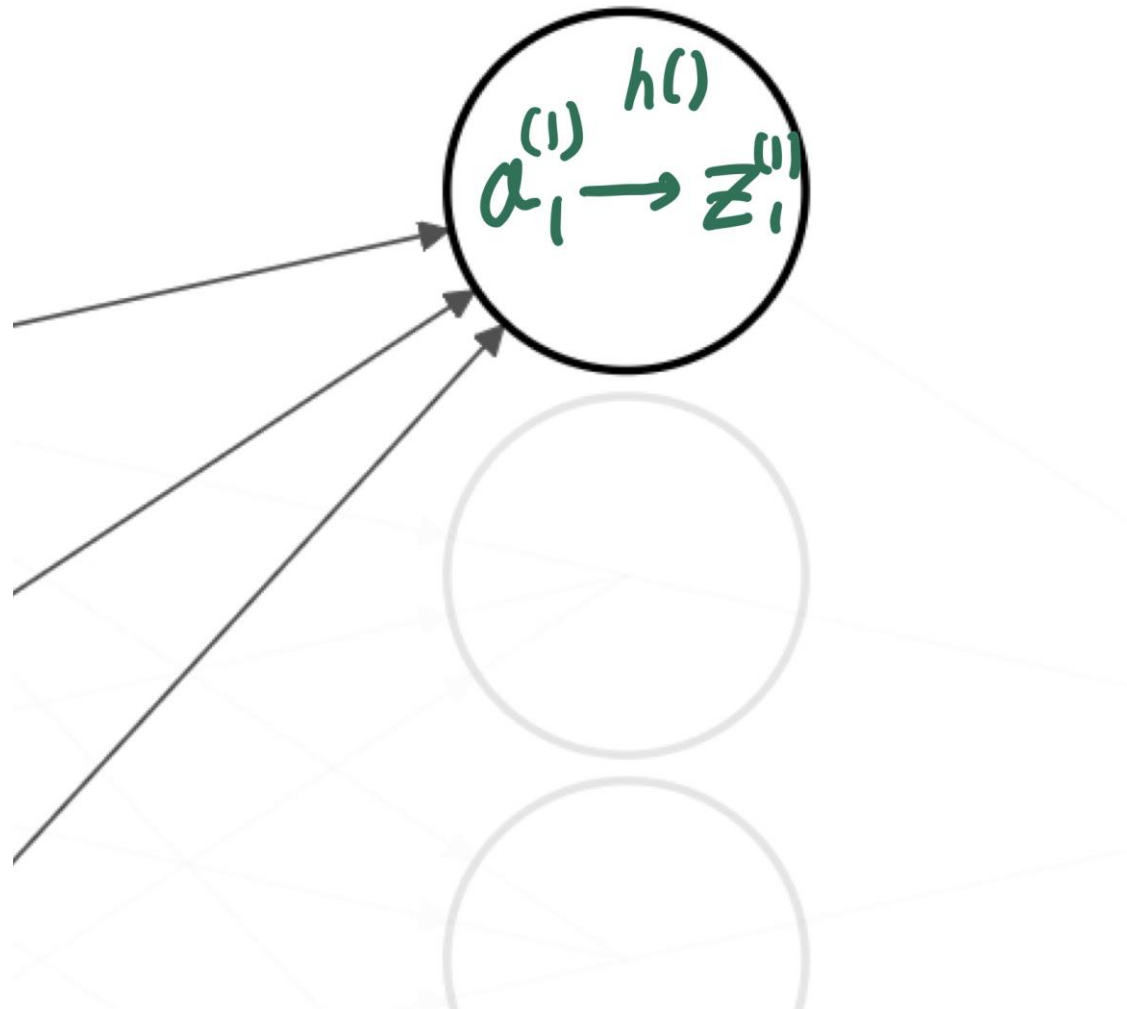
$$a_1^{(1)} = x_1 w_{11}^{(1)} + x_2 w_{12}^{(1)} + x_3 w_{13}^{(1)}$$

Input Layer $\in \mathbb{R}^3$

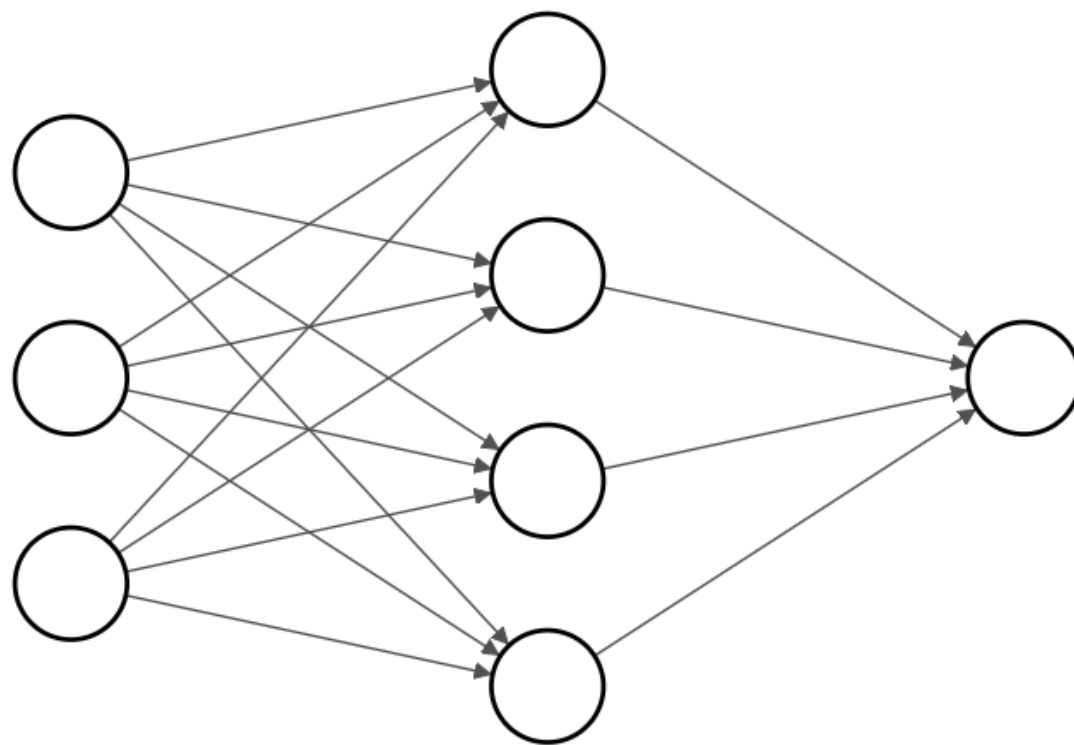
Hidden Layer $\in \mathbb{R}^4$

Output Layer $\in \mathbb{R}^1$

딥러닝



딥러닝



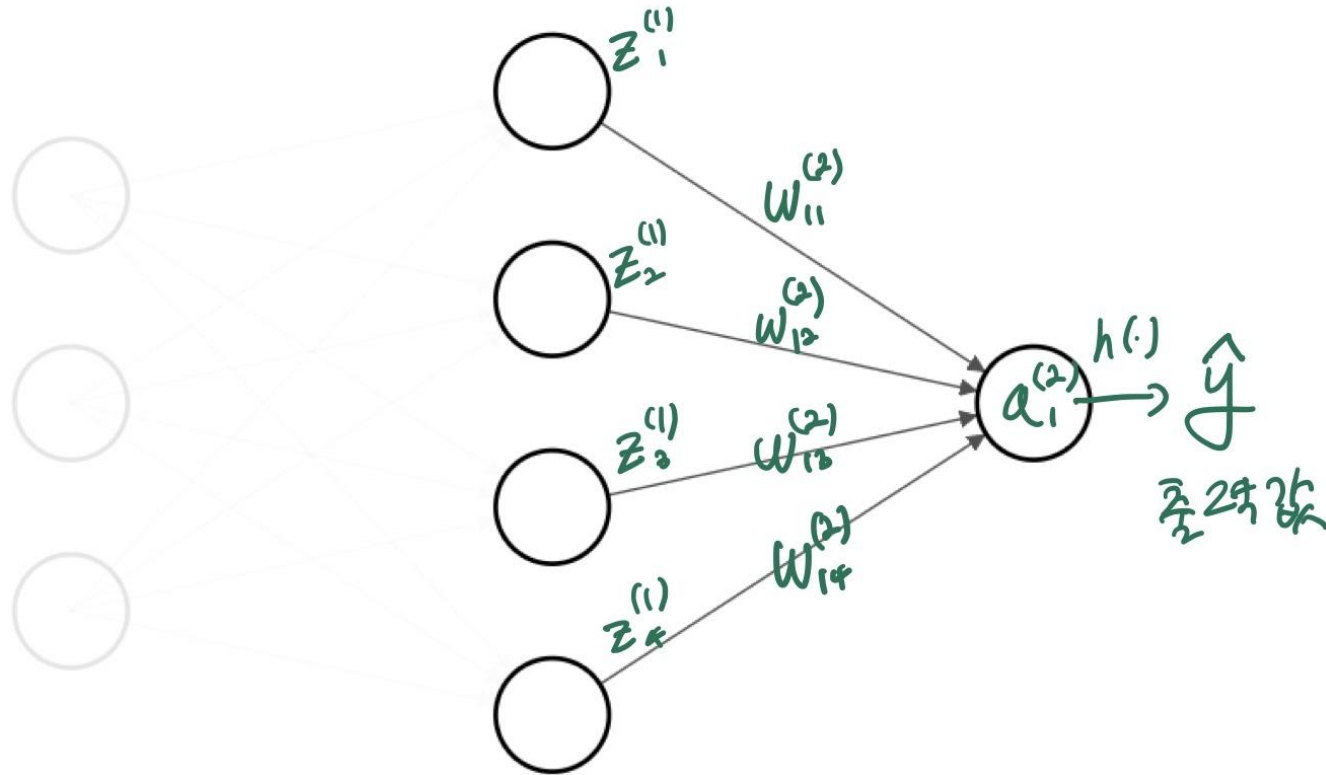
Input Layer $\in \mathbb{R}^3$

Hidden Layer $\in \mathbb{R}^4$

Output Layer $\in \mathbb{R}^1$

딥러닝

$$a_1^{(2)} = z_1^{(1)} w_{11}^{(2)} + z_2^{(1)} w_{12}^{(2)} + z_3^{(1)} w_{13}^{(2)} + z_4^{(1)} w_{14}^{(2)}$$



Input Layer $\in \mathbb{R}^3$

Hidden Layer $\in \mathbb{R}^4$

Output Layer $\in \mathbb{R}^1$

딥러닝

- 가중치 값들은 층과 층 사이의 모든 노드에 초기화 되어있음.
 - Random initialization
 - Xavier initialization
 - He initialization
- 오차가 최소값이 될 때의 가중치가 최적의 가중치

감사합니다