

신경망

Table of
Contents

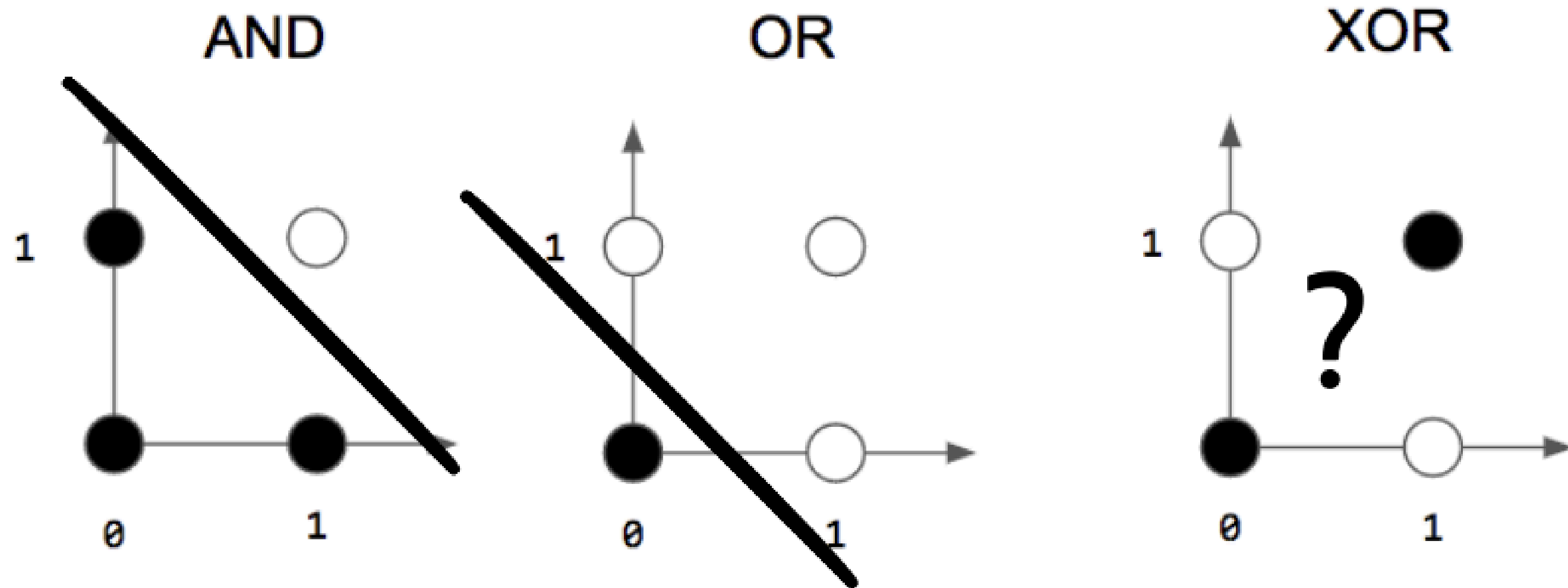
	Page
I 활성화 함수	3
II 다차원 배열의 계산	6
III 다층 신경망과 출력층	9

1. 활성화 함수

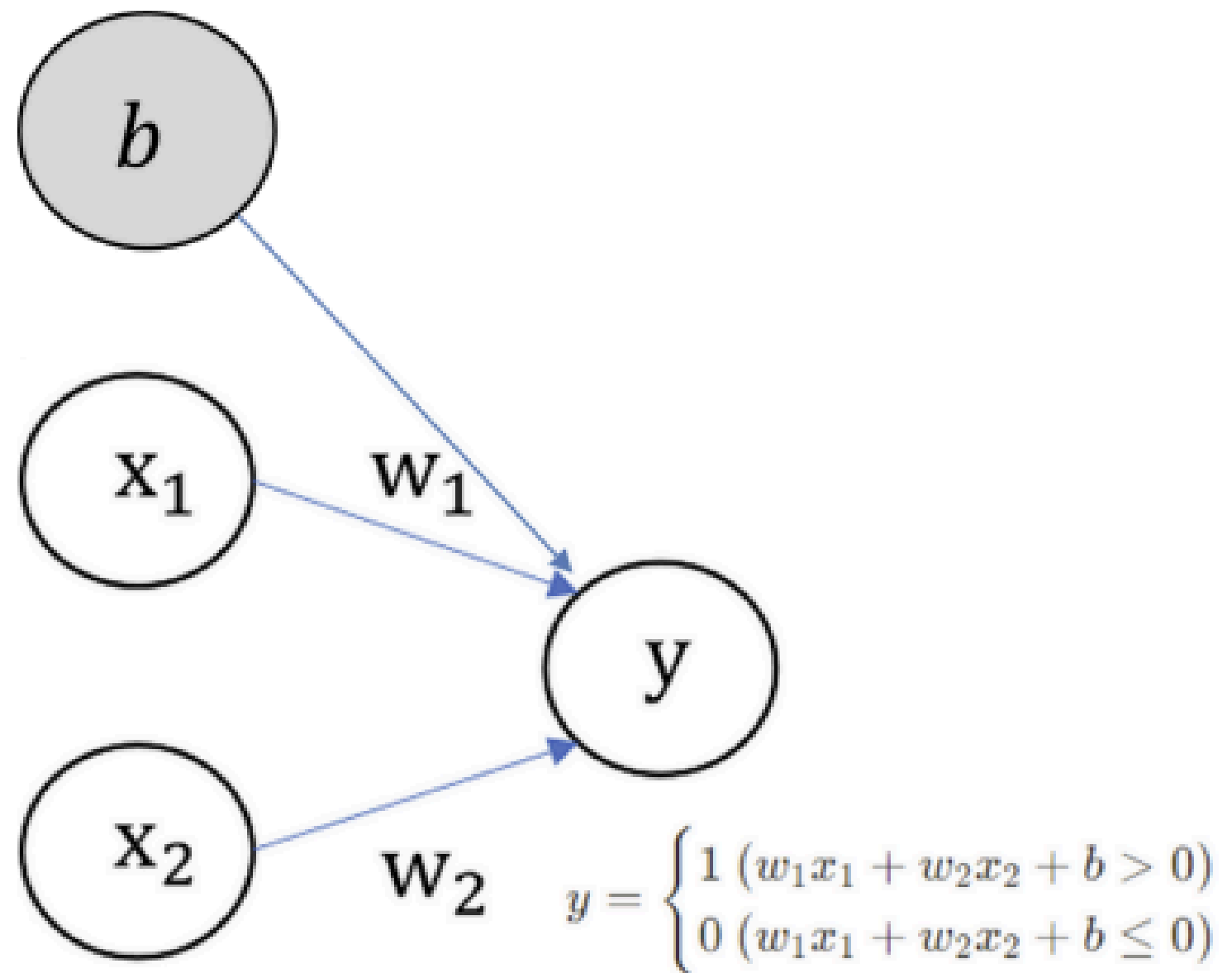
단층 퍼셉트론의 한계

XOR 게이트는 선형 영역이 아닌 비선형 영역으로 나뉘야 한다

-> 단층 퍼셉트론에 층을 더 쌓아서 다층 퍼셉트론으로 표현



가중치와 편향

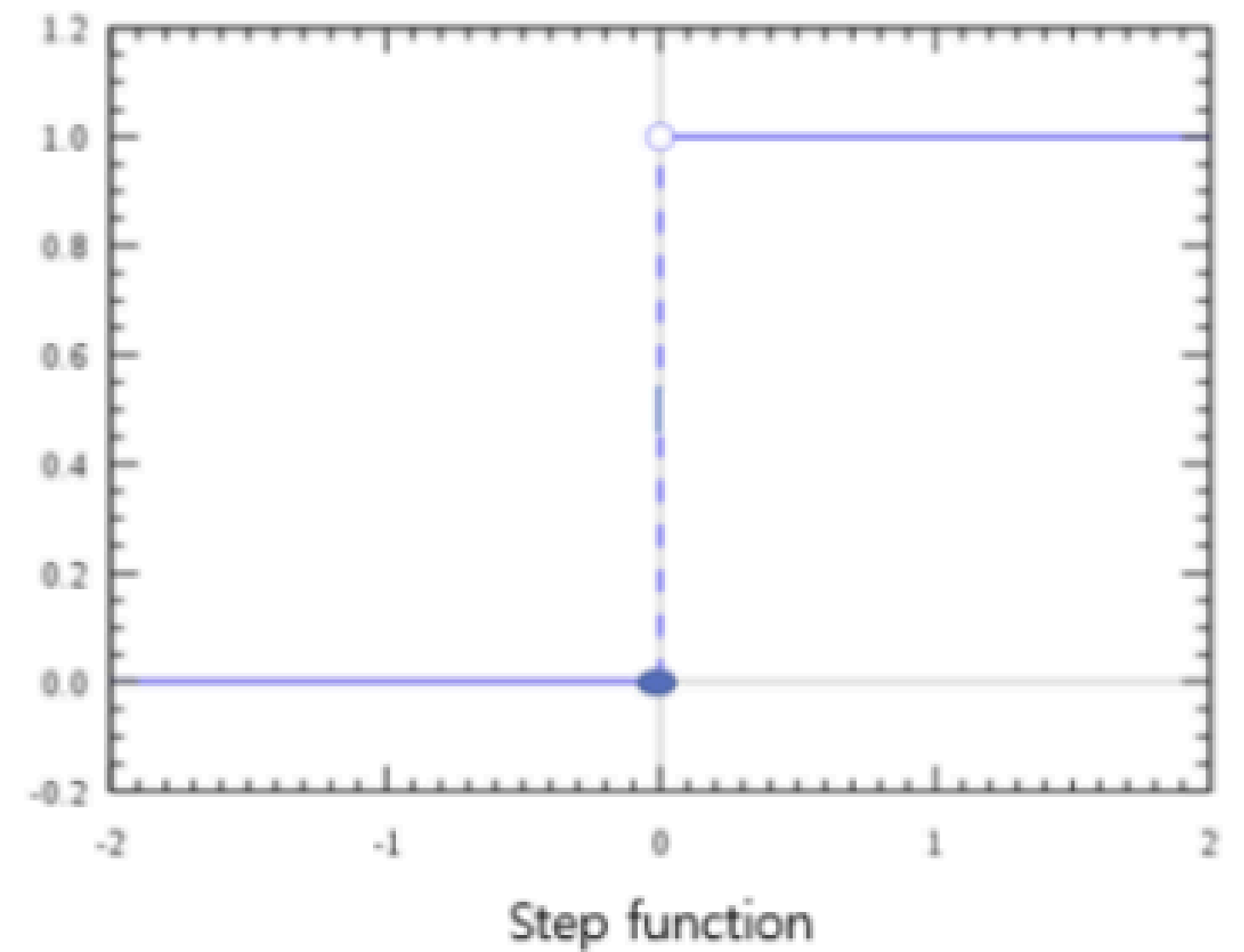
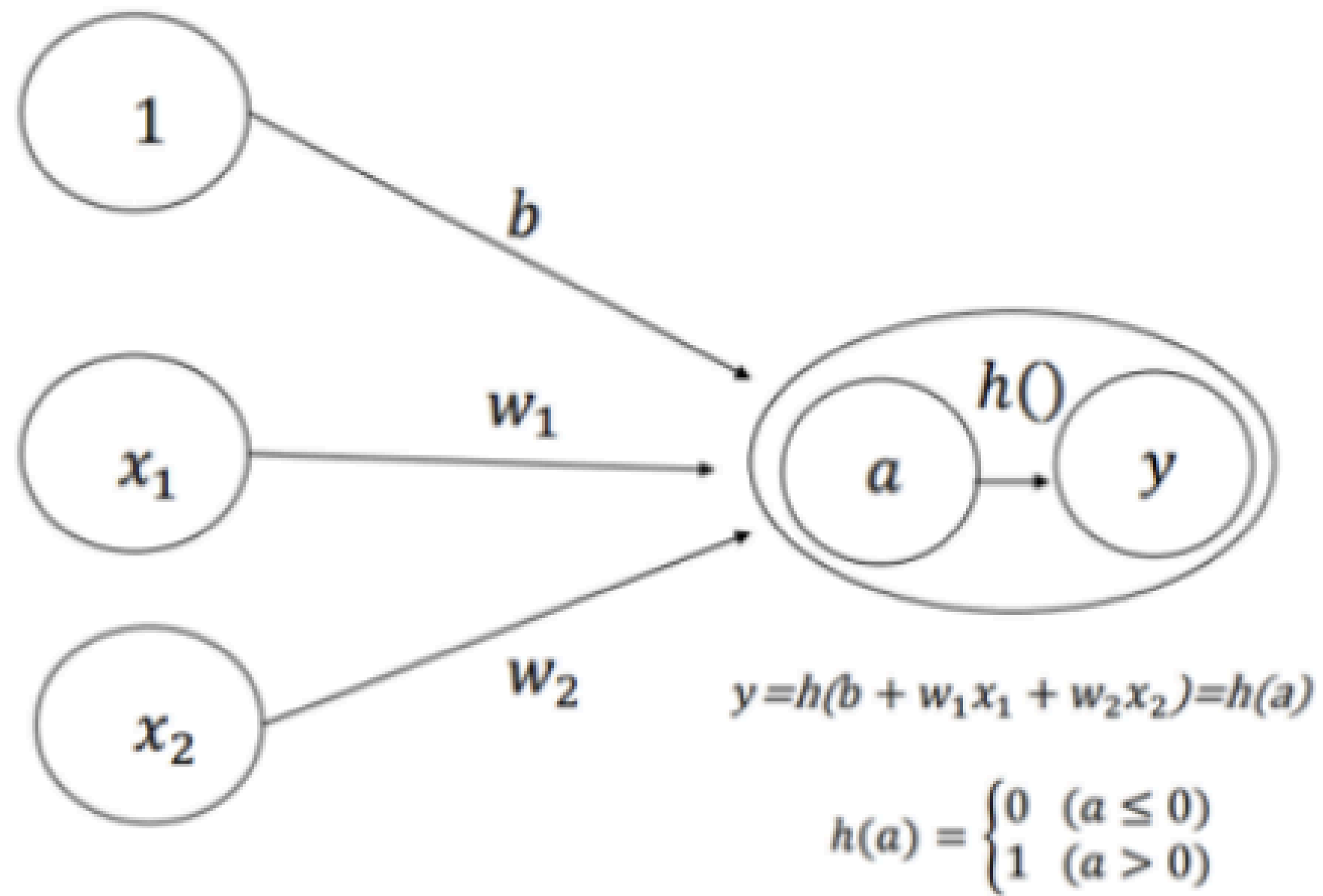


$$a = \sum (\text{weights} \cdot \text{inputs}) + \text{bias}$$

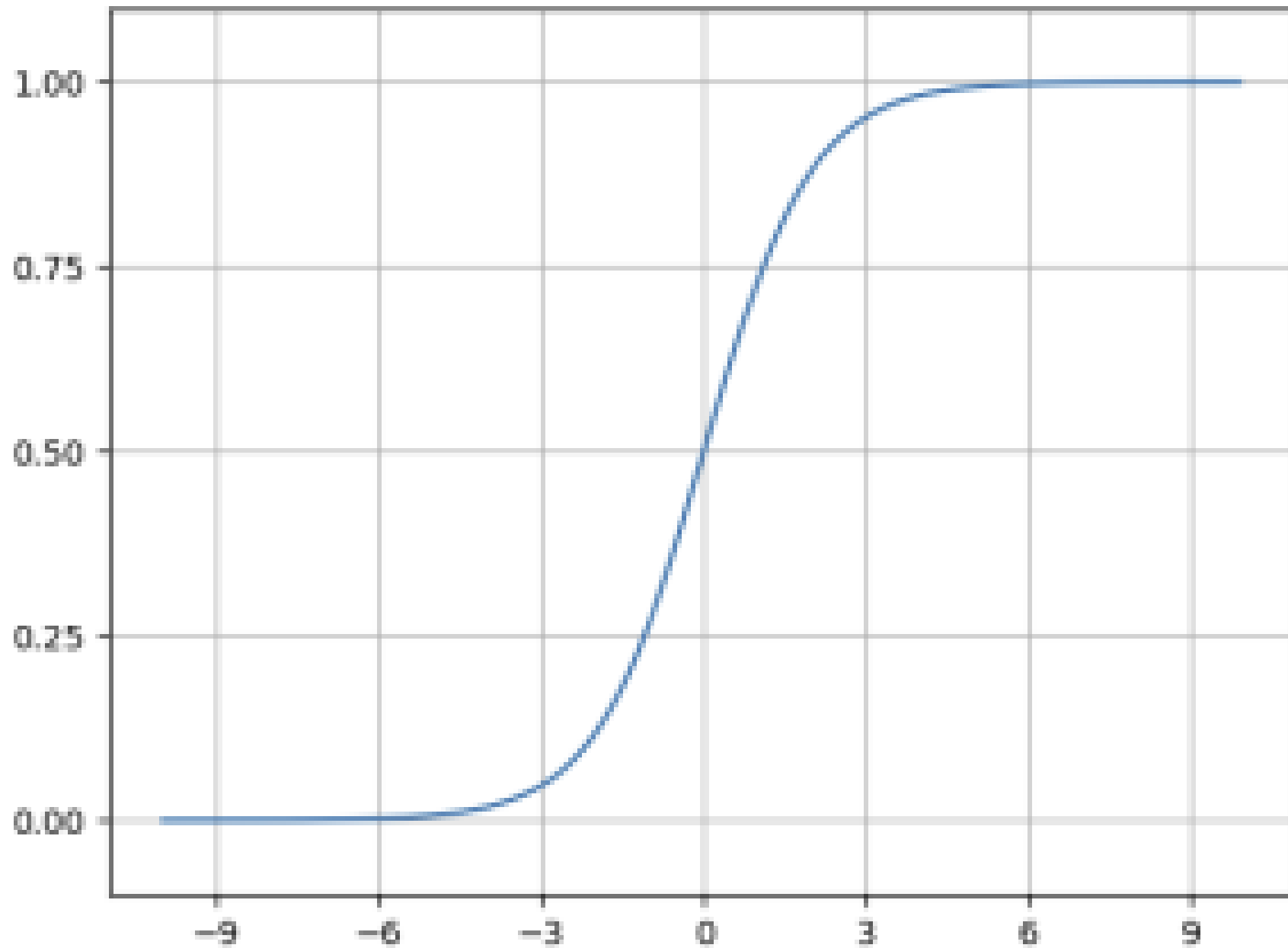
$$\text{output} = h(a)$$

$h = \text{activation function}$

I Activation Function & Step Function

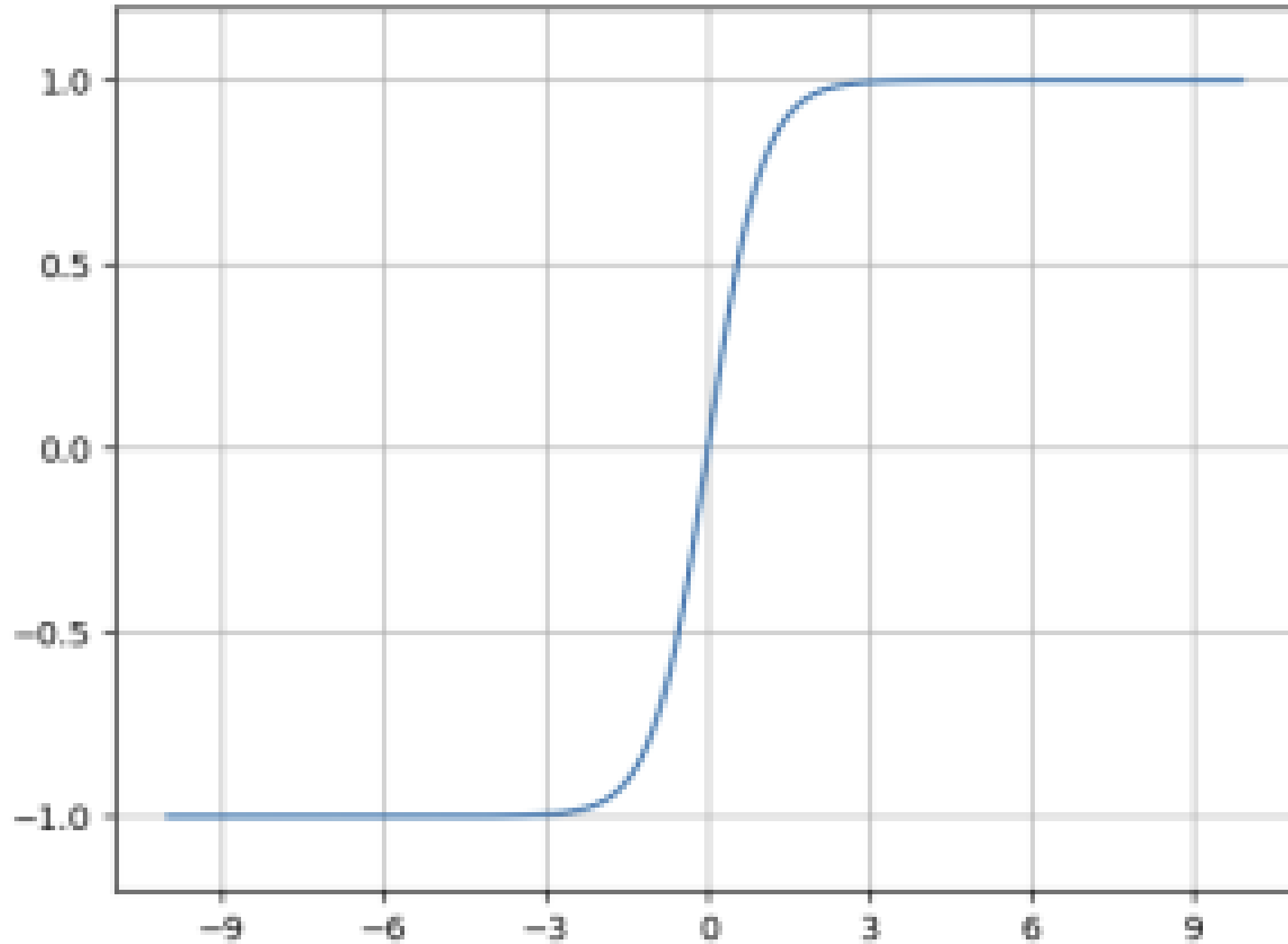


I sigmoid function



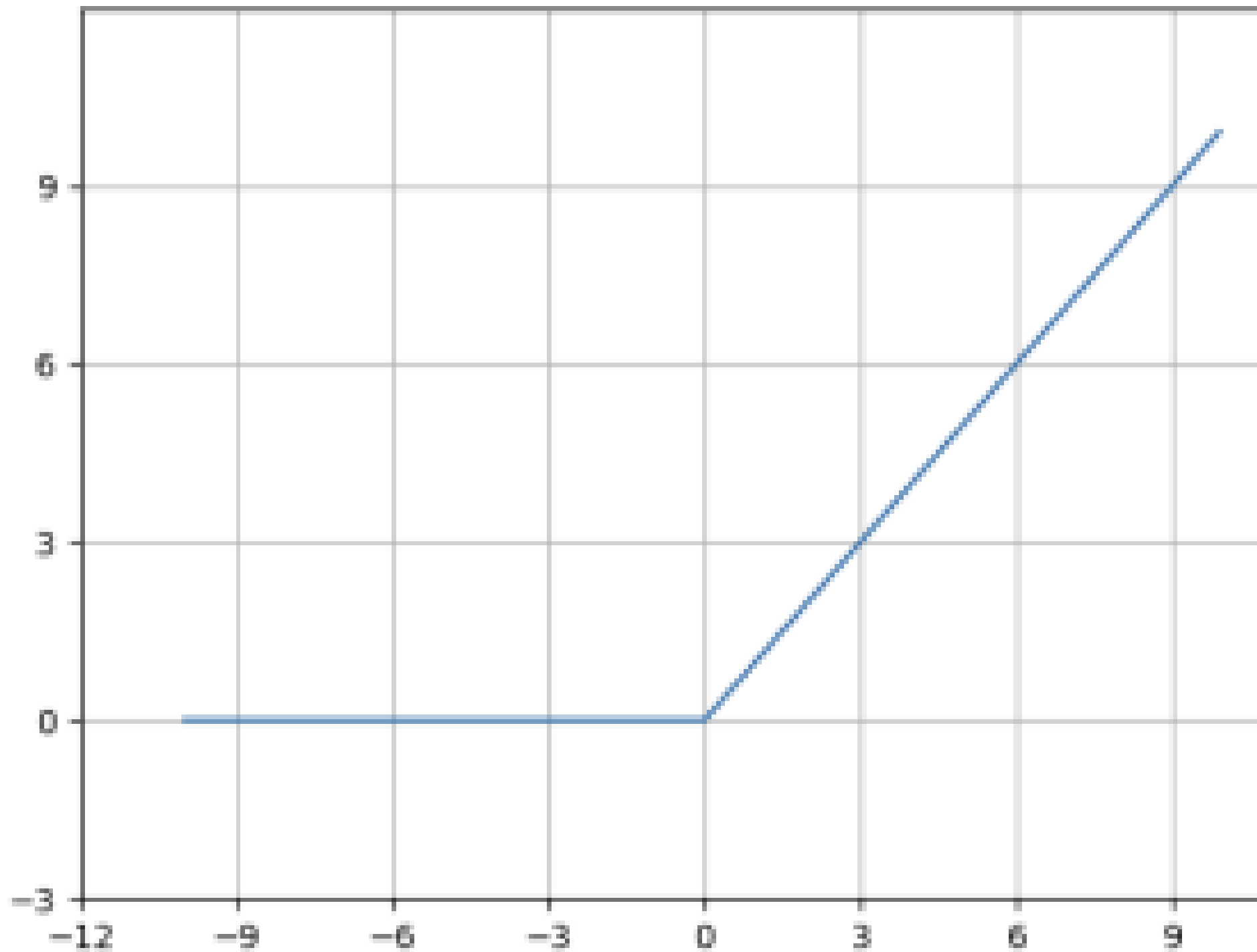
- $g(x) = \frac{1}{1+e^{-x}}$
- The output value ranges $[0, 1]$.
- Problem 1. Saturation
 - The derivative value goes to zero when saturated.
 - Backpropagation will send zero signal to the lower layers without updates on parameters.
- Problem 2. Off-zero-centered
 - Recall the backpropagation algorithm.
 - $\frac{\partial L}{\partial w_{jk}} = \frac{\partial L}{\partial o_k} \frac{\partial o_k}{\partial net_k} \frac{\partial net_k}{\partial w_{jk}} = \frac{\partial L}{\partial o_k} \frac{\partial o_k}{\partial net_k} o_j$
 - With o_j being always positive.
 - The sign of the derivative is decided by the sign of $\frac{\partial L}{\partial o_k} \frac{\partial o_k}{\partial net_k}$ for all the weights.
 - The same sign for all the weights
→ Zig-zag updates!

I Hyperbolic Tangent Function



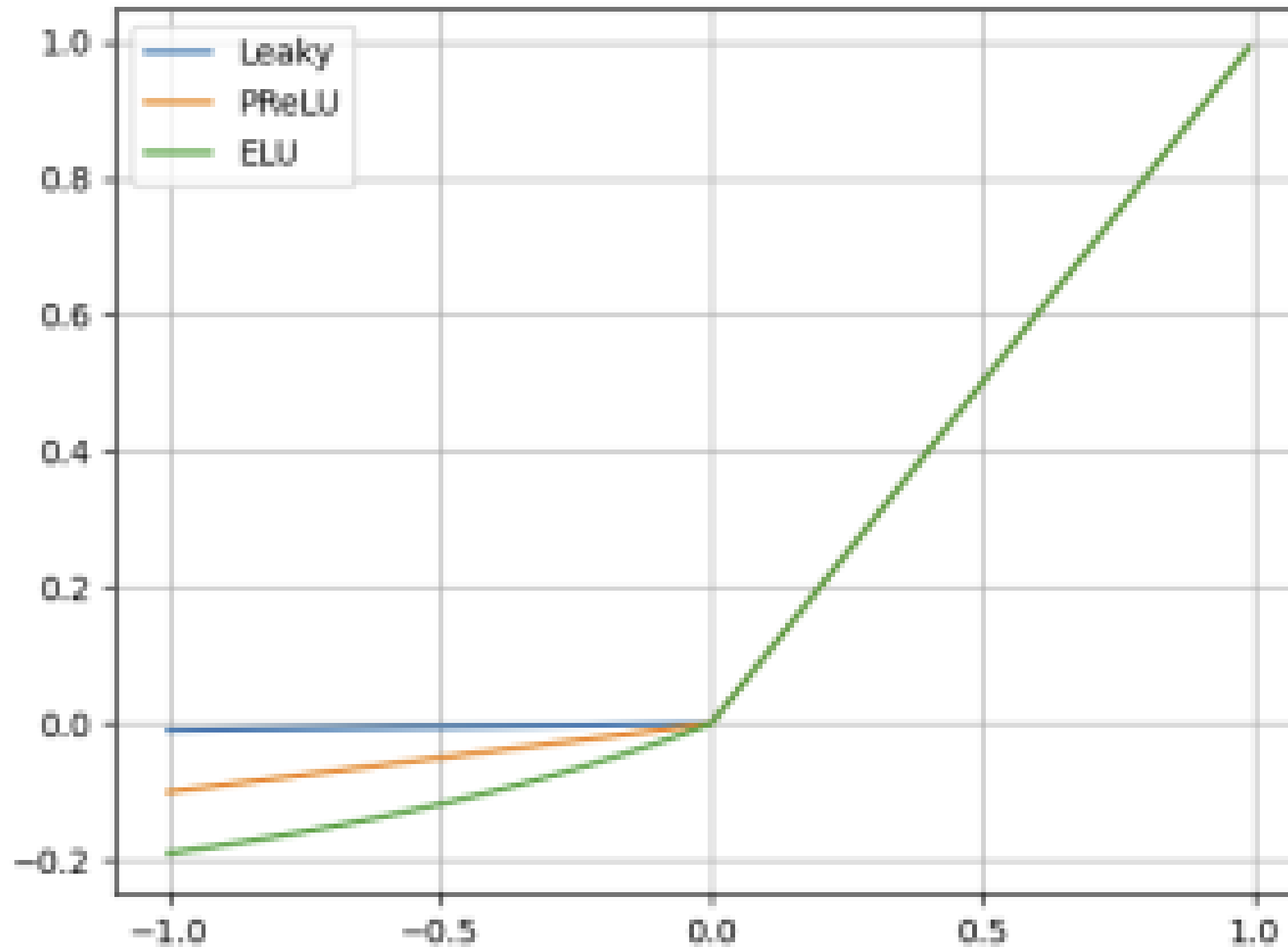
- $g(x) = \frac{e^{2x}-1}{e^{2x}+1}$
- $g(x) = \tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
- The output value ranges $[-1, 1]$.
- It is zero-centered.
- However, it still kills gradients when saturated.

I ReLU Fuction



- $g(x) = \max(0, x)$
- Biologically plausible function
 - Actual neurons in a brain rarely reach the maximum saturation.
- It does not saturate in $x \geq 0$ region.
 - However, still saturating in $x < 0$ region.
- ReLU tends to activate only a subset of neuron in a layer, leading to sparse activation.
 - It prevents overfitting.

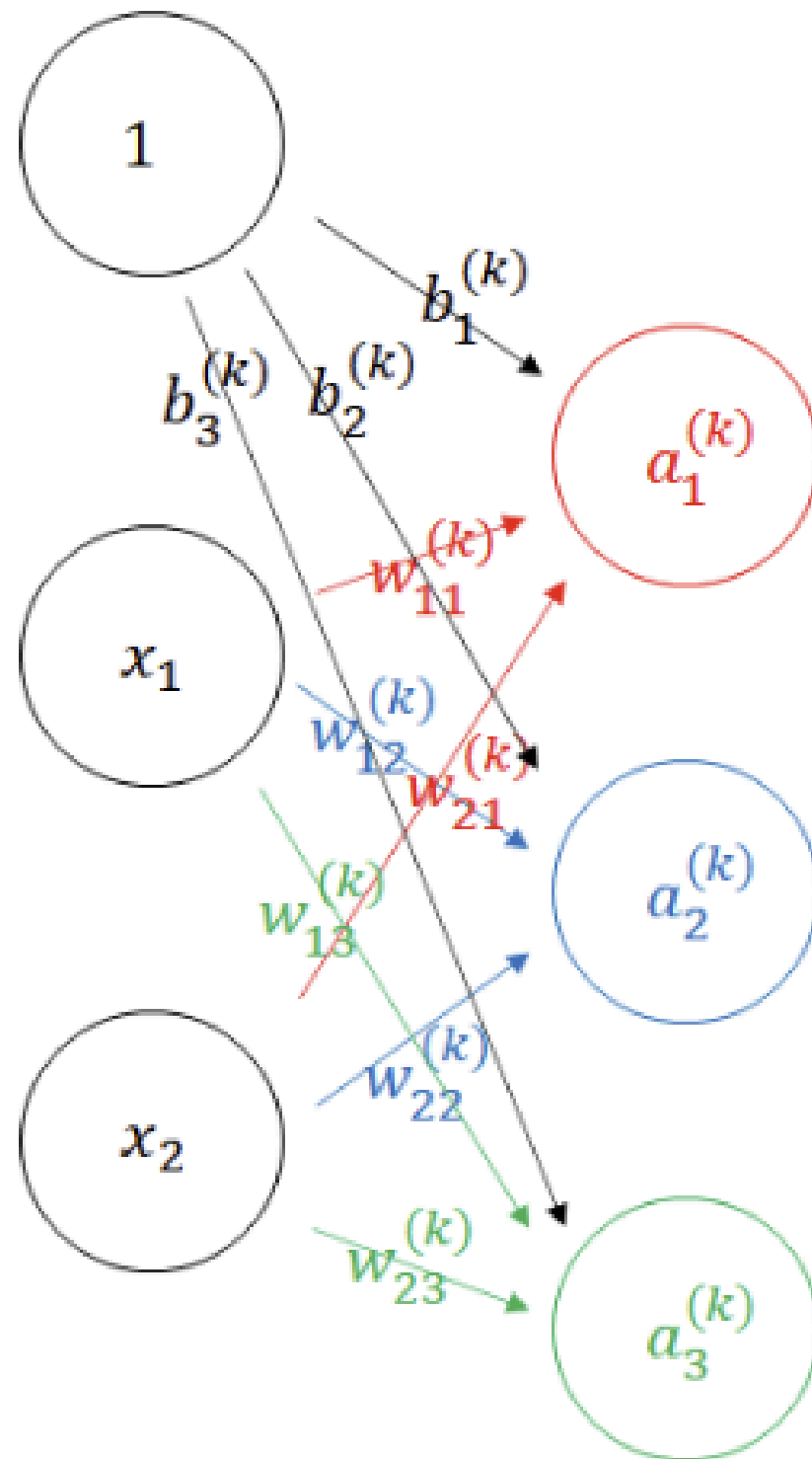
I Leaky ReLU / PReLU (Parametric ReLU) / ELU (Exponential LU) Function



- $g_{Leaky}(x) = \max(0.01x, x)$
 - Non-zero gradients in the negative regime.
- $g_{PReLU}(x) = \max(\alpha x, x)$
 - α is trained by backpropagation.
- $g_{ELU}(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(\exp(x) - 1) & \text{if } x \leq 0 \end{cases}$
 - ELU is smooth everywhere and has a non-zero gradient for all inputs.

2. 다차원 배열의 계산

II 다차원배열의 계산



$W^{(\text{층})}$
입력, 출력

$$a_1^{(k)} = w_{11}^{(k)} x_1 + w_{21}^{(k)} x_2 + b_1^{(k)}$$

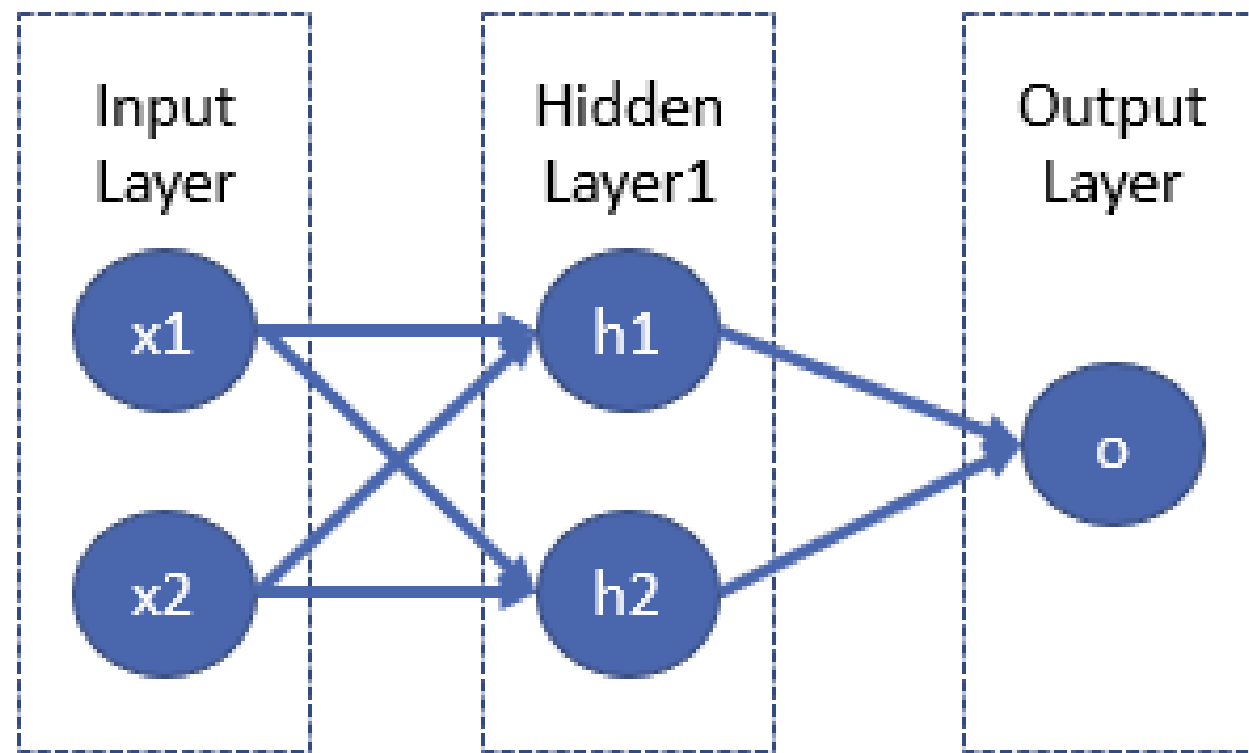
$$a_2^{(k)} = w_{12}^{(k)} x_1 + w_{22}^{(k)} x_2 + b_2^{(k)}$$

$$a_3^{(k)} = w_{13}^{(k)} x_1 + w_{23}^{(k)} x_2 + b_3^{(k)}$$

\Leftrightarrow

$$\begin{pmatrix} a_1^{(k)} & a_2^{(k)} & a_3^{(k)} \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} w_{11}^{(k)} & w_{12}^{(k)} & w_{13}^{(k)} \\ w_{21}^{(k)} & w_{22}^{(k)} & w_{23}^{(k)} \end{pmatrix} + \begin{pmatrix} b_1^{(k)} & b_2^{(k)} & b_3^{(k)} \end{pmatrix}$$

II 다차원배열의 계산



- $h_1 = g(0.5x_1 + 0.5x_2 - 50)$
- $h_2 = g(-0.5x_1 - 0.5x_2 - 50)$
- $o = g(100h_1 + 100h_2 - 100)$

$g = \frac{1}{1+\exp(-x)}$, as activation function.

Quiz 1.

```
def neuron(w1, w2, b, x1, x2):  
    return ????????????????????
```

```
def activation(x):  
    return ????????????????????
```

```
[ ] x1 = np.arange(-100, 100, 1)  
    x2 = np.arange(-100, 100, 1)  
    x1, x2 = np.meshgrid(x1, x2)
```

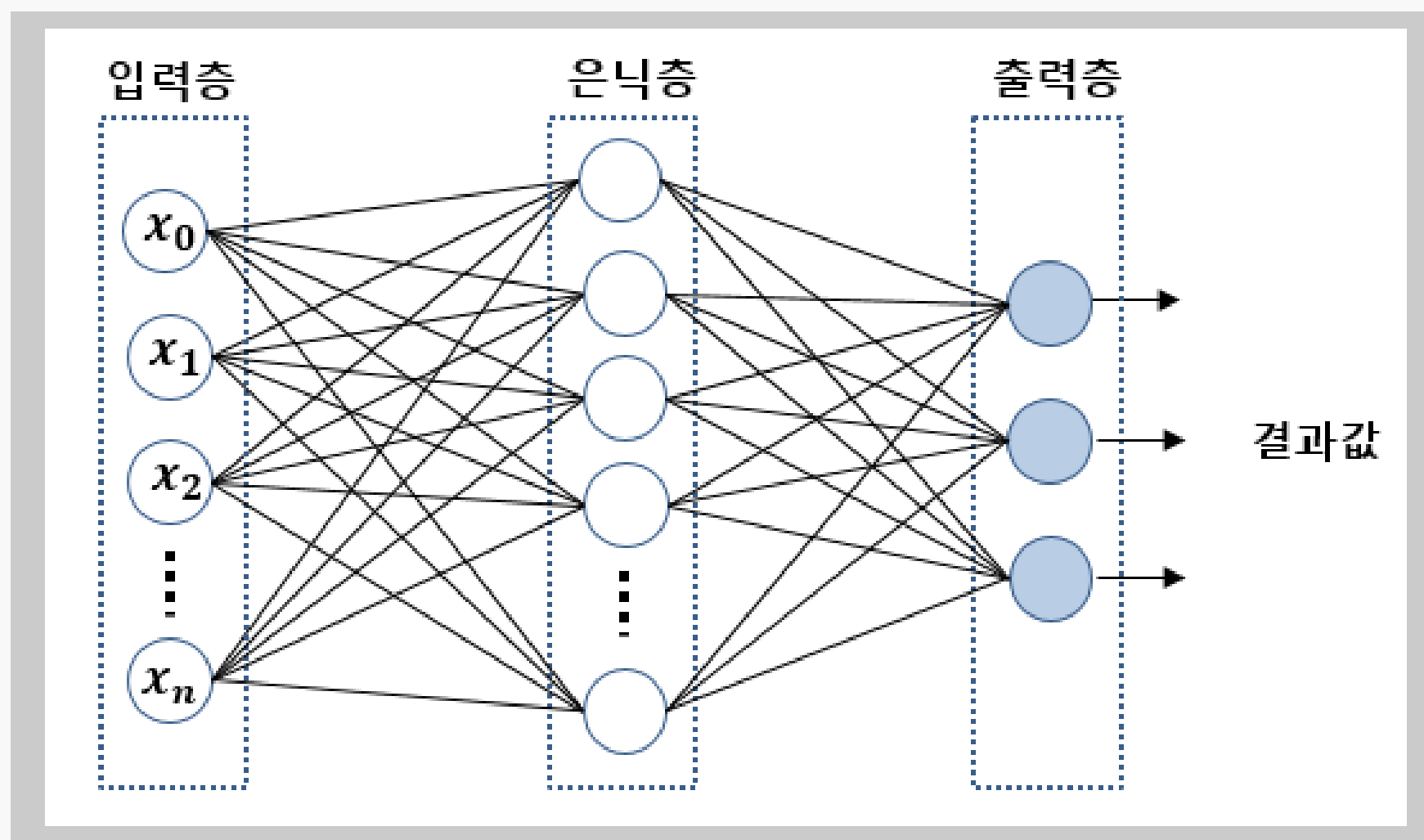
```
[ ] h1 = ????????????????????  
    h2 = ????????????????????  
    h3 = ????????????????????
```

Quiz 2.

3. 다층 신경망과 출력층

다층 신경망이란?

입력 레이어, 출력 레이어, 하나 이상의 은닉 레이어가 있는 '인공 신경망'

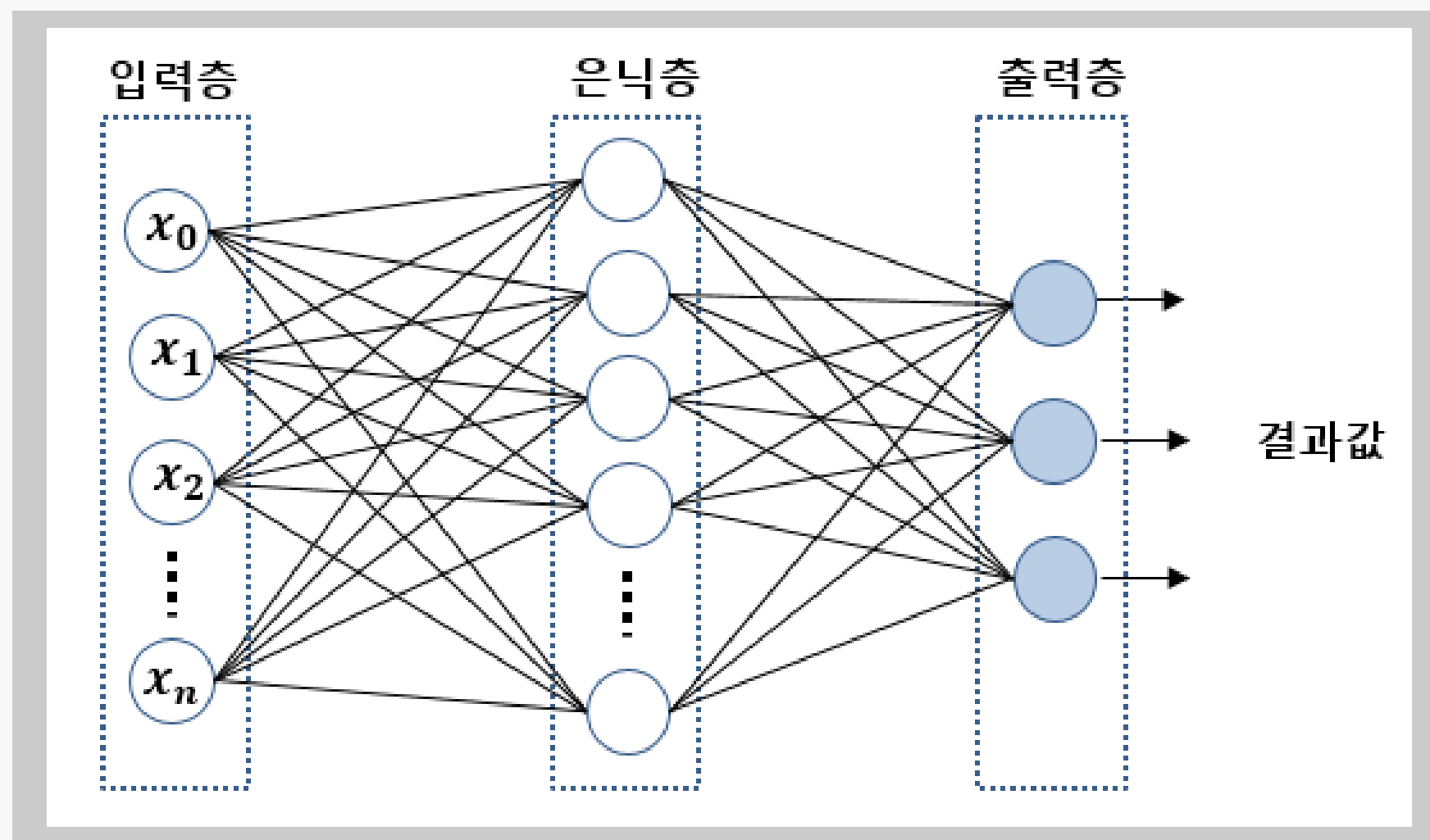


입력층: 계산을 위한 뉴런은 거의 X
신경망의 층수(depth)를 셀 때는 포함X

출력층: 은닉층에서의 출력 신호

은닉층: 입력의 특성을 파악 후,
뉴런의 가중치로 표현
계산 결과를 사용자가 볼 수 없음
※ 가중치란, '입력 패턴에 숨겨져있는 특성'을 의미

다층 신경망이란?



은닉층이 1개만 있는 다층 신경망: 얇은 신경망
(Shallow Neural Network)

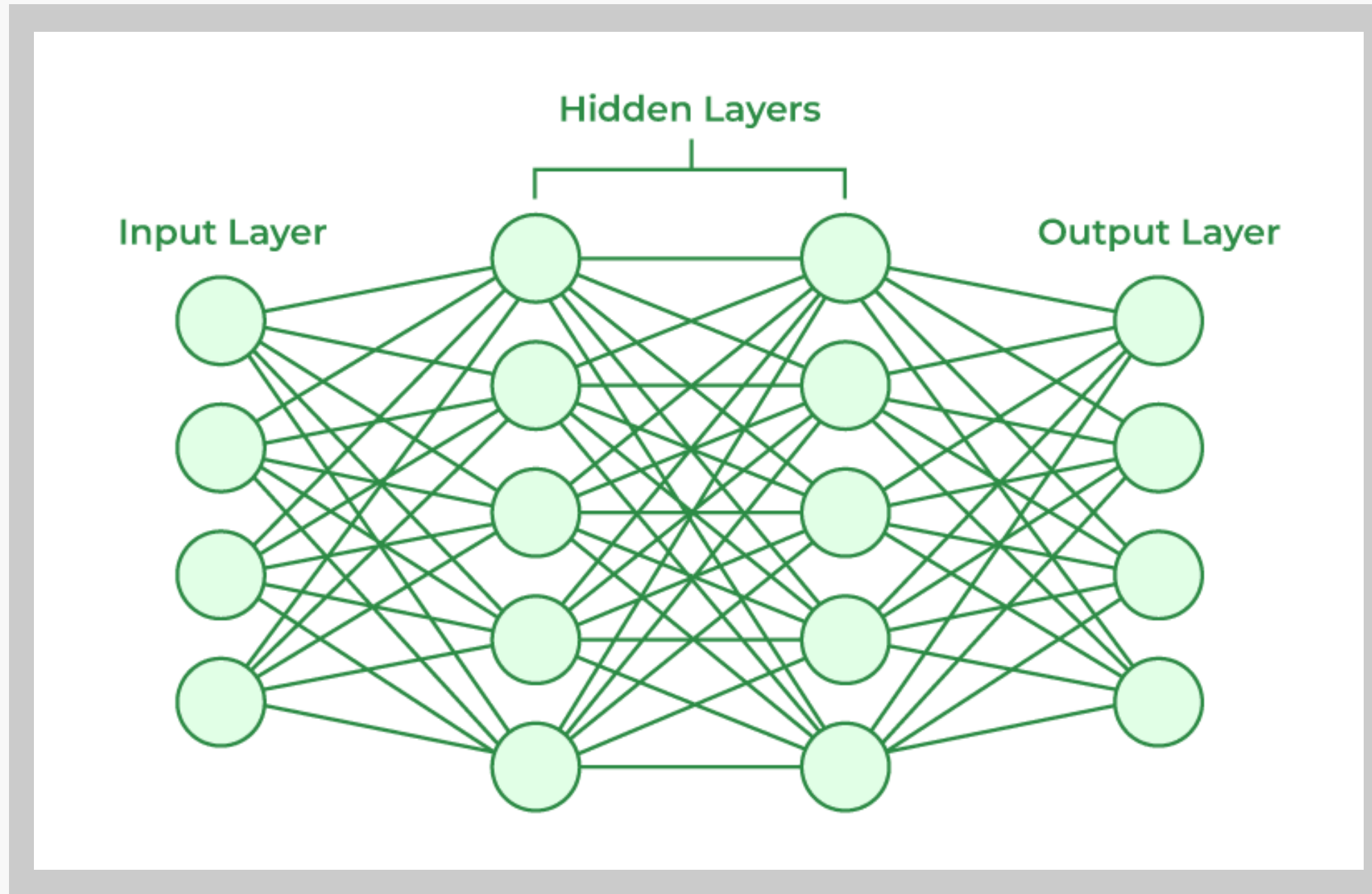
은닉층이 2개 이상인 경우: 깊은 신경망
(Deep Neural Network)

깊은 신경망을 학습 → 'Deep Learning'

단층 퍼셉트론과 달리,

비선형으로 분포하는 데이터들에 대한 학습 가능

III Multi Layer Neural Networks



Quiz 3.

- 1) 해당 신경망은 몇 층 신경망인가요?
- 2) 바이어스를 제외한 전체 노드의 개수는?

다층 신경망의 학습

퍼셉트론과 유사하게 진행

신경망은 출력 패턴을 계산

실제 목표 출력 간의 '차이'

계산에 오차가 있다면, 이 오차(Error)를 줄이도록 '가중치를 조절'

특히, 다층 신경망에서는 가중치가 여러개이며
각각의 가중치는 두 개 이상의 출력에 영향을 미침

모든 데이터 → 행렬 변환 후, 계산

✓ 텐서 형태 변환을 잘 아는 것이 중요

각 Layer 행렬 곱 수식

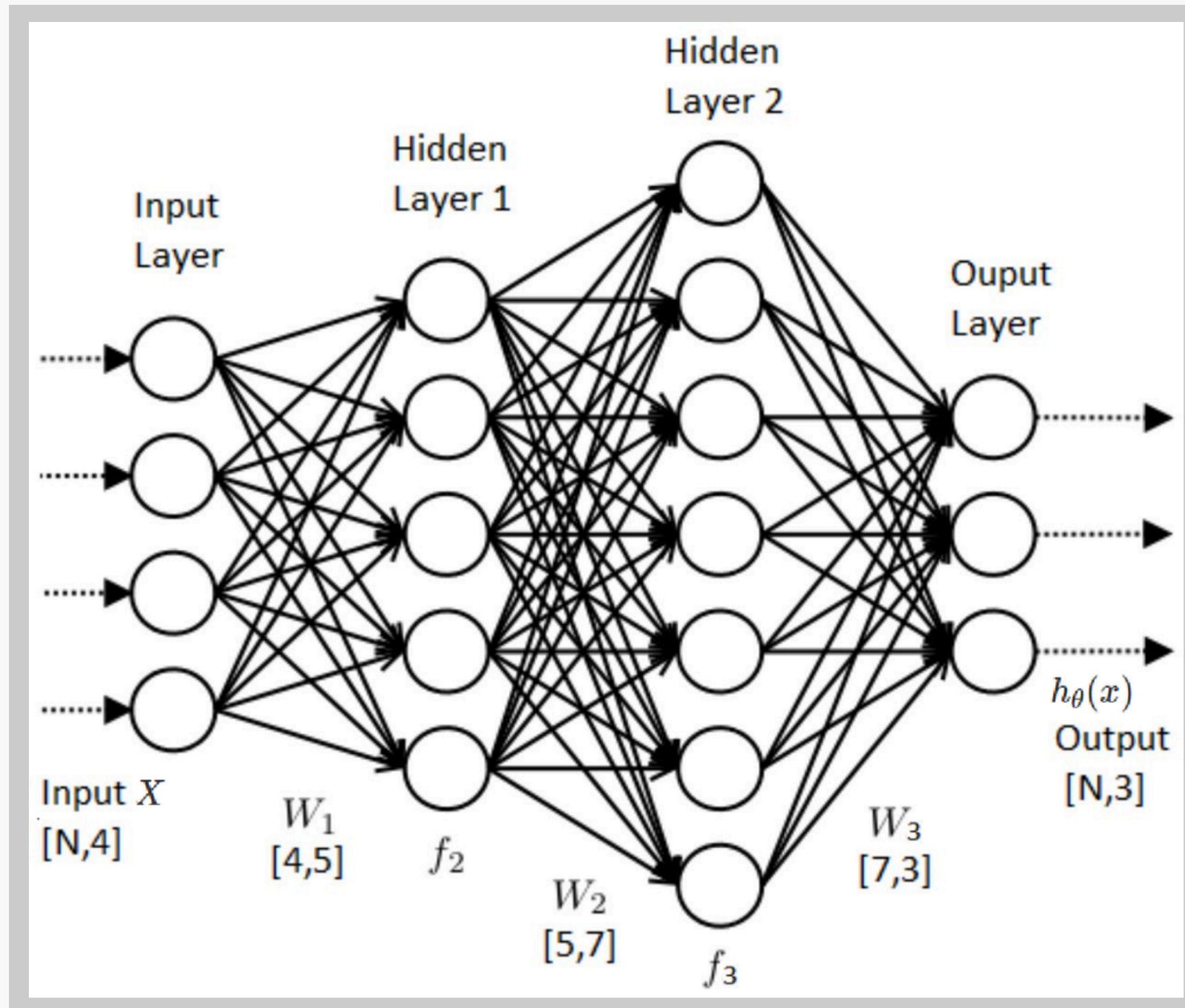
은닉 계층 입력 값 : $X_{hidden} = W_{input_hidden} * I$

은닉 계층 출력 값 : $O_{hidden} = sigmoid(X_{hidden})$

출력 계층 입력 값 : $X_{output} = W_{hidden_output} * O_{hidden}$

X: 결과값, W: 가중치, I: 입력 데이터 값

III Multi Layer Neural Networks



신경망 층의 수: 3

한 층을 구성하는 계산 단위: Quiz 4

가중치 행렬의 size:

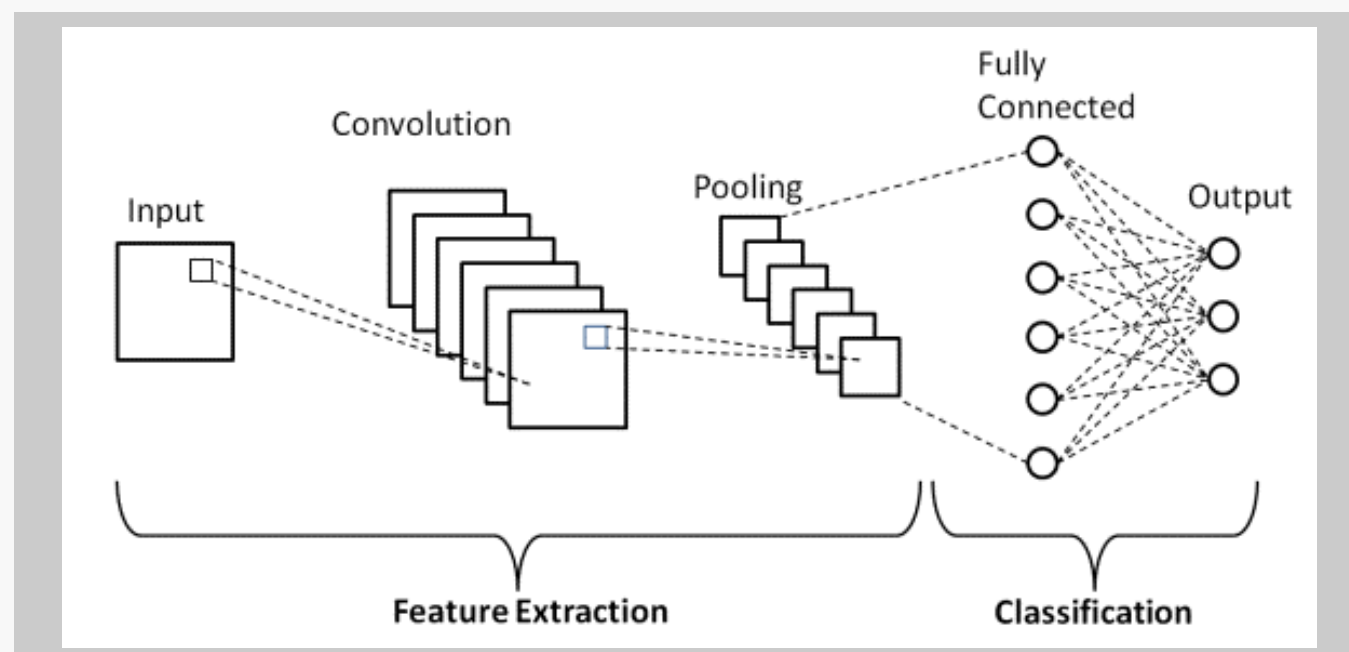
$$W_1 - 4*5$$

$$W_2 - 5*7$$

$$W_3 - 7*3$$

다층 신경망의 종류

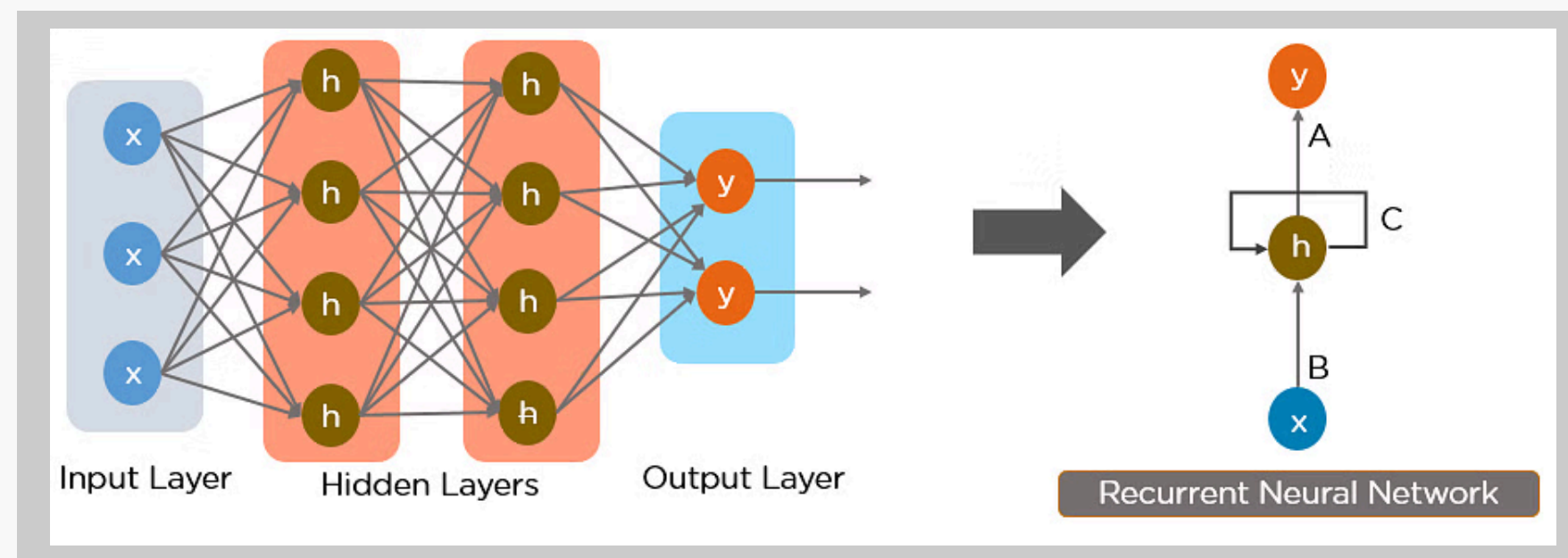
1) CNN



주로 '이미지 처리'에 사용

합성곱 층과 풀링 층을 사용해 입력 이미지의 특징 추출

2) RNN



주로 '시계열 데이터'나 '순차 데이터 처리'에 적합

이전 time의 출력이 현재 시간의 입력으로 사용되는 순환 구조

다층 신경망으로 복잡한 데이터 패턴을 학습 및 다양한 문제를 해결

IV Reference



활성화 함수, 다차원 배열

서울시립대학교 김윤영 교수님 강의교안

[ML] 딥러닝 1 - 3강 인공신경망(tistory)

<https://yonsodev.tistory.com/3>

MLP 개념

Chapter 1 - 2 다층 신경망 (MLP; Multi-Layer Perceptrons) (tistory)

<https://supermemi.tistory.com/entry/Chapter-1-%EC%8B%A0%EA%B2%BD%EB%A7%9D-%EC%9E%85%EB%AC%B8Neural-network-2-Multi-layer>

다층 신경망 (Artificial Neural Network, ANN) (tistory).

<https://0-sunny.tistory.com/73>

**Thank you
for listening!**