

# 딥러닝의 기본 작동 방식

*How AI make a computational prediction about structured information?*

- I. 인공신경망(Artificial Neural Network; ANN)과 활용 목적
- II. (비)지도 학습((Un-)supervised Learning)
- III. 순전파와 역전파 (Forward and Back Propagation)
- IV. 로지스틱 회귀 (Logistic Regression)
- V. 경사하강법 (Gradient Descent)
- VI. 비선형 활성화 함수 (Non-linear Activation Function)
- VII.(초)매개변수 ((Hyper-)parameter)

## BRIEF RESEARCH REPORT article

Front. Psychol., 25 January 2023

Sec. Language Sciences

Volume 14 - 2023 |

<https://doi.org/10.3389/fpsyg.2023.1049266>

This article is part of the Research Topic

Using Artificial Intelligence Technology for Language Learning

[View all 4 Articles >](#)

# Fusion weighted features and BiLSTM- attention model for argument mining of EFL writing



Jincai Yang<sup>1</sup>,



Meng Zheng<sup>1</sup> and



Yingliang Liu<sup>2\*</sup>

<sup>1</sup> School of Computer Science, Central China Normal University, Wuhan, China

<sup>2</sup> School of Foreign Languages, Wuhan University of Technology, Wuhan, China

Argument mining (AM), an emerging field in natural language processing (NLP), aims to automatically extract arguments and the relationships between them in texts. In this study, we propose a new method for argument mining of argumentative essays.

Keyword:

논증 마이닝(Argument Mining)

툴민 논증 모형(Toulmin's argument model)

외국어로서의 영어(English as a Foreign Language;EFL)

자동 주석과 평가(Automatic Annotation and Evaluation)

(3) TV가 지금 우리 생활 주변에 있어요 어디든지 언제든지 다 볼 수 있어요. 지금 세상은 TV가 없으면 안되는 세상입니다. TV를 보는 사람들이 점점 올라서 TV가 우리 생활 필수품이 됐어요. 그래서 무엇보다 TV프로그램을 잘 선택해서 보는 것이 중요해요. 지금 TV의 장단점을 알아보겠어요.

TV의 장점이 진짜 많아요. 첫째, 세상 여러 나라에서 일어나는 일들을 빨리 알 수 있어요. 우리가 가 본 적도 없고 갈 수도 없는 곳을 TV로 다 알 수 있어요. 동물 다큐멘터리하고 여러 나라 시사를 다 볼 수 있어요. 둘째, TV를 보면 스트레스를 풀 수 있어요. 사람들은 드라마나 연예, 오락 프로그램을 보면서 함께 웃고 함께 울고 감동도 받을 수 있어요. 생활에서 받은 스트레스가 풀릴 거예요.

하지만 TV의 장점만큼 단점도 많아요. 첫째, TV를 오랫동안 보면 공부하고 일할 시간이 빼앗길 것 같아요. 아무것도 안 하고 그냥 TV만 봐요. 그리고 전기 요금도 많이 올라요. 둘째, TV는 나쁜 프로그램도 많아요. 나오는 나쁜 말과 나쁜 행동을 따라할 때도 있어요. 특히 청소년들이 그렇게 많이 할 것 같아요. 우리는 TV를 봐서 얻을 수도 있지만 잃을 수도 있어요. 무엇보다 좋은 프로그램을 선택해서 보는 것이 중요해요. 우리는 좋은 TV 프로그램을 보면 TV에서 얻을 것이 많아질 것 같아요.

자료 삽입항 → 자료(논거) ————— 결론(테제)

추론규칙      예외조건

지원(근거)

<그림 1> 논증형 주제 전개 모형

〈표 3〉 예문 (3)의 논증형 주제 전개 모형에 따른 분석

자료 삽입항	TV는 우리 생활에서 어디에서든지 볼 수 있을 만큼 대중화되었다.
결론(테제)	좋은 TV프로그램을 선택해서 보는 것이 중요하다.
논거	1. TV를 통해서 여러 나라에서 일어나는 일들을 빨리 알 수 있고 많은 간접 경험을 할 수 있다. 또 TV를 보면서 스트레스를 풀 수 있다.
	2. 그러나 TV를 오래 볼 경우시간을 많이 빼앗기고 나쁜 프로그램을 통해서 나쁜 영향을 받을 수도 있다.
추론규칙	TV를 보면서 많은 것을 얻을 수도 있지만 잃을 수도 있다.

안정호. (2012). 학문 목적 한국어 작문 교육을 위한 학습자 텍스트 구조 분석 연구-논  
증형 텍스트를 중심으로. *작문연구*, 15, 101-126.

✓ 딥러닝 또는 심층학습은 어떤 거대한 인공 신경망을 가리킨다

(Deep learning refers to large artificial neural networks).

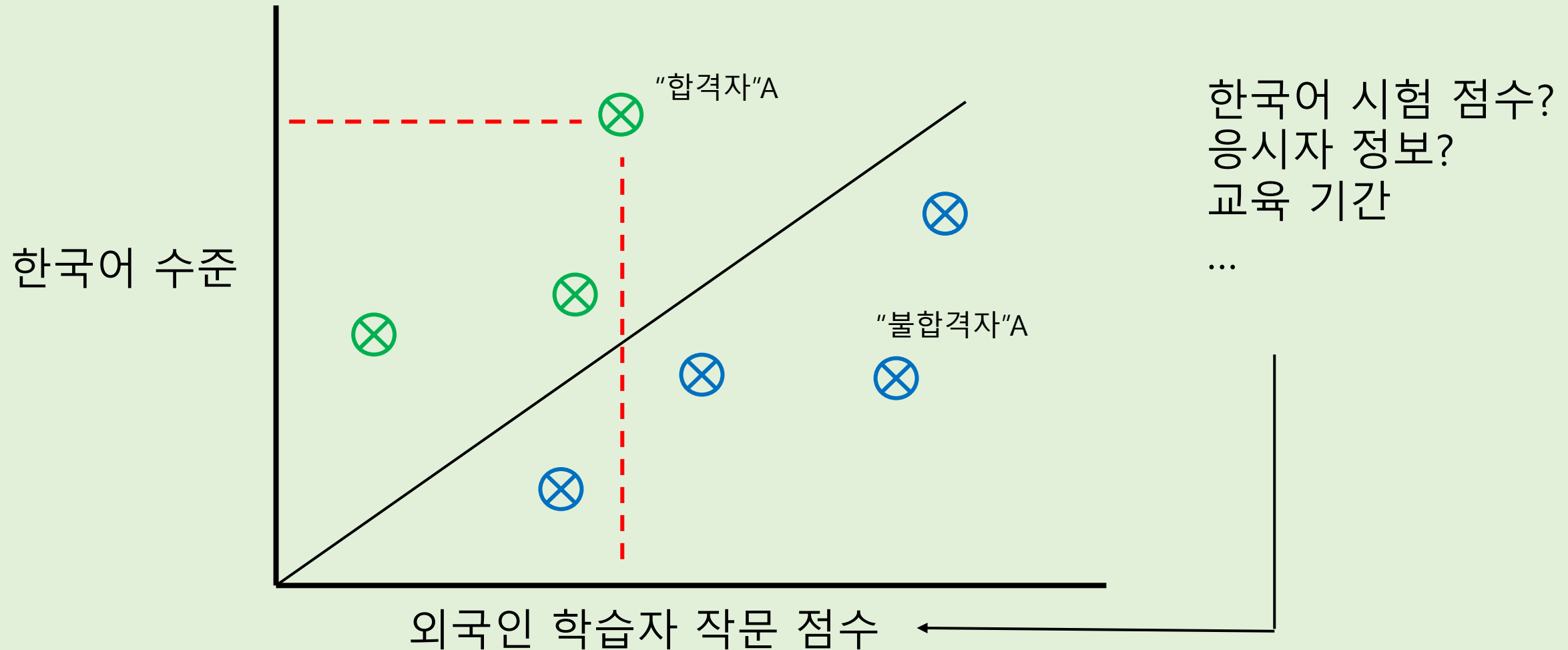
✓ 거대한 신경망은 매우 작은 인공 뉴런을 쌓아 올려서 만든다

(large neural networks are built by stacking tiny artificial neurons together).



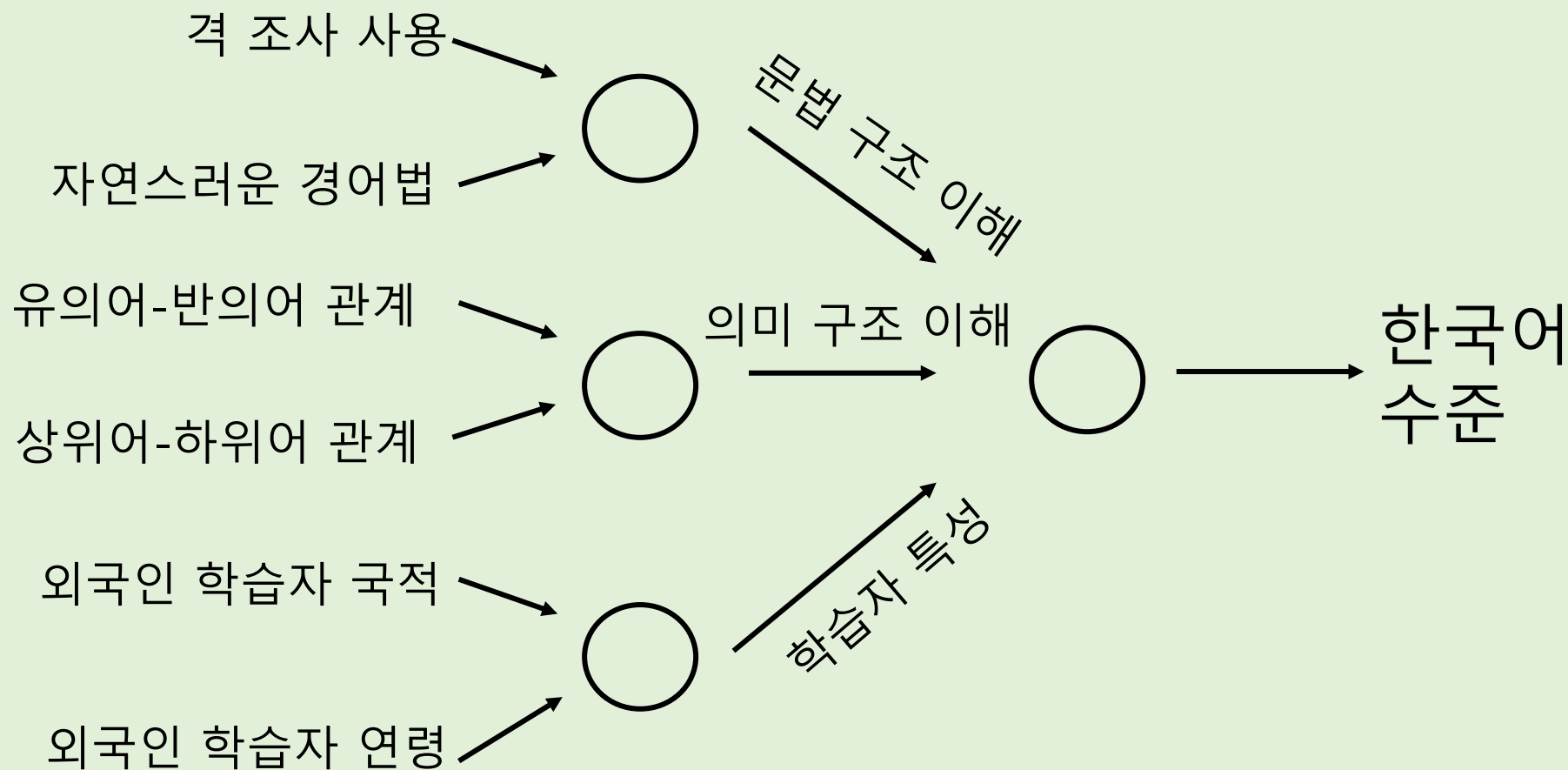
# 외국인 학습자 한국어 수준 예측

인공 뉴런이 하는 일을 평면으로 나타내면..



# 외국인 학습자 한국어 수준 예측

인공 뉴런을 쌓아 올려서 만들면..



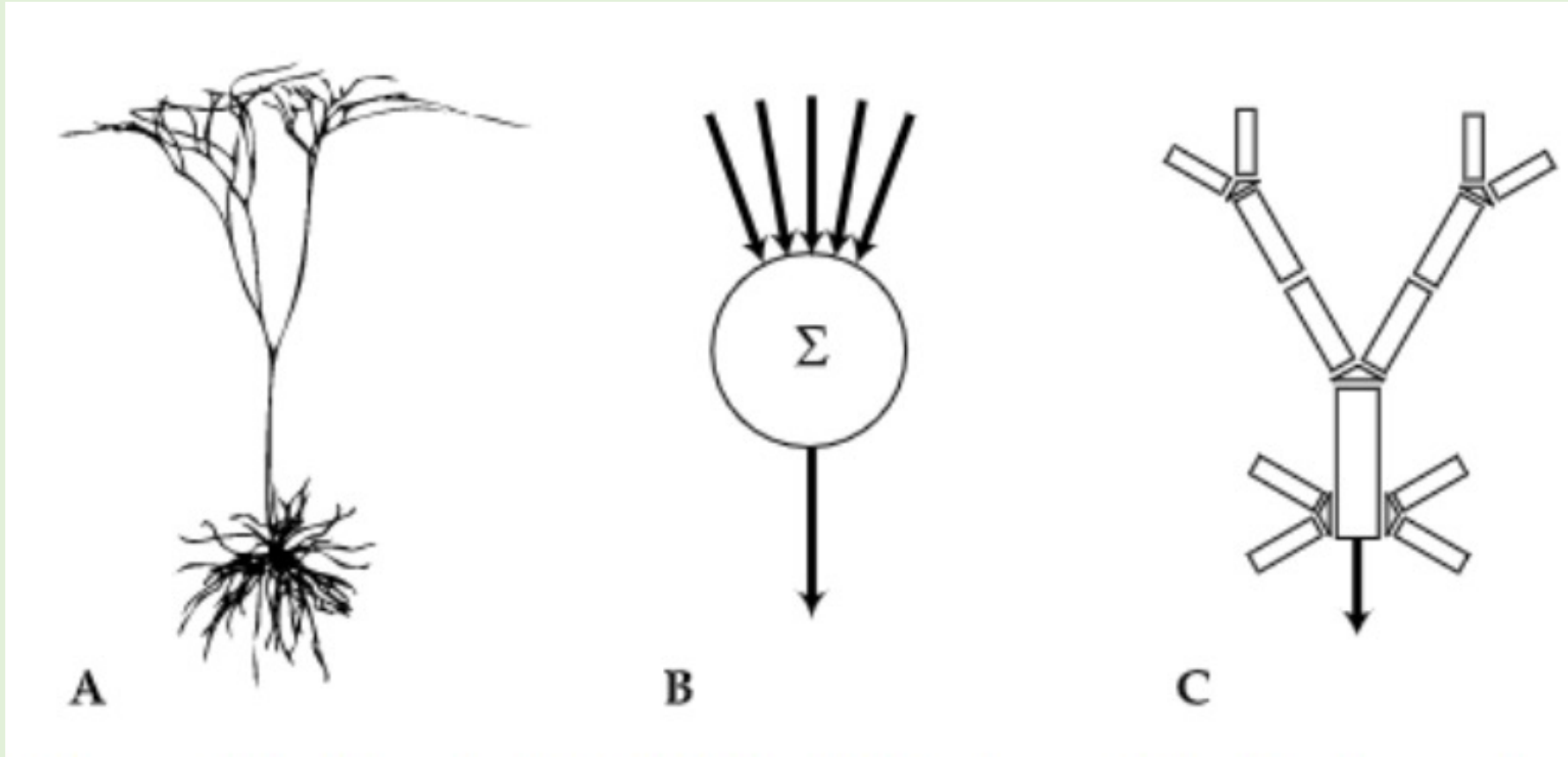




- ✓ 인공 뉴런과 신경망의 관계는 레고 블록과 완성품의 관계와 유사함!
- ✓ 차이점이 있다면, 신경망은 사람이 만든 설명서를 참조하는 것이 아닌 스스로 추론 한다는 점.



## Consider a Spherical Neuron...



Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of life reviews*, 11(3), 329-364.

The old joke: Consider a Spherical Cow...



이론 물리학자



이론 물리학자



# 학습 데이터 유형

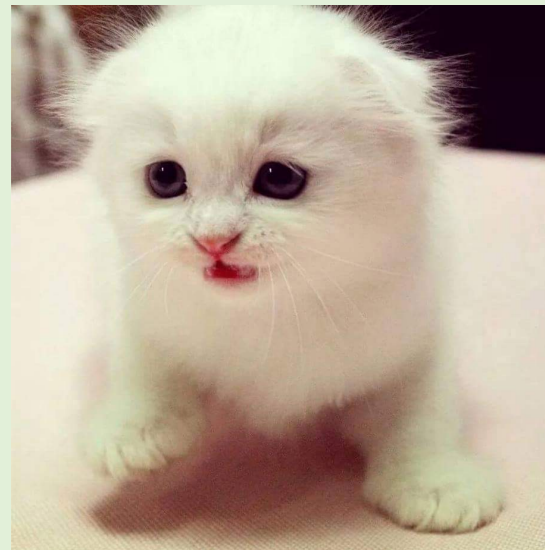
## 정형 데이터(Structured Data)

Input <sub>1</sub>	Input <sub>2</sub>	...	Output
격조사	담화 표지	...	점수
33건	32건		814점
64건	15건		560점
21건	46건		910점
...	...		...

Input <sub>1</sub>	Input <sub>2</sub>	...	Output
한국어 점수	국적	...	합격 여부
42	중국		0
28	베트남		1
72	미국		0
...	...		

## 비정형 데이터(Unstructured Data)

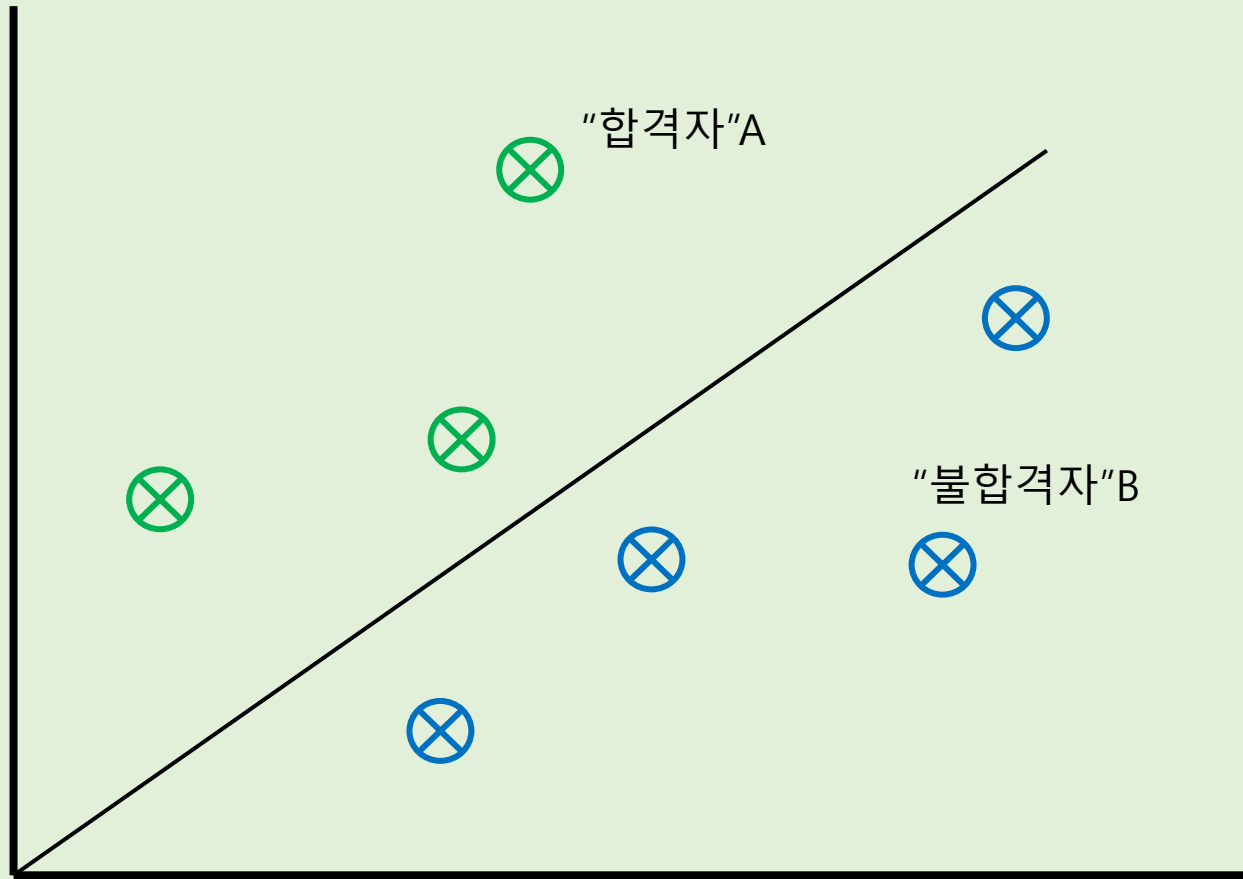
Image



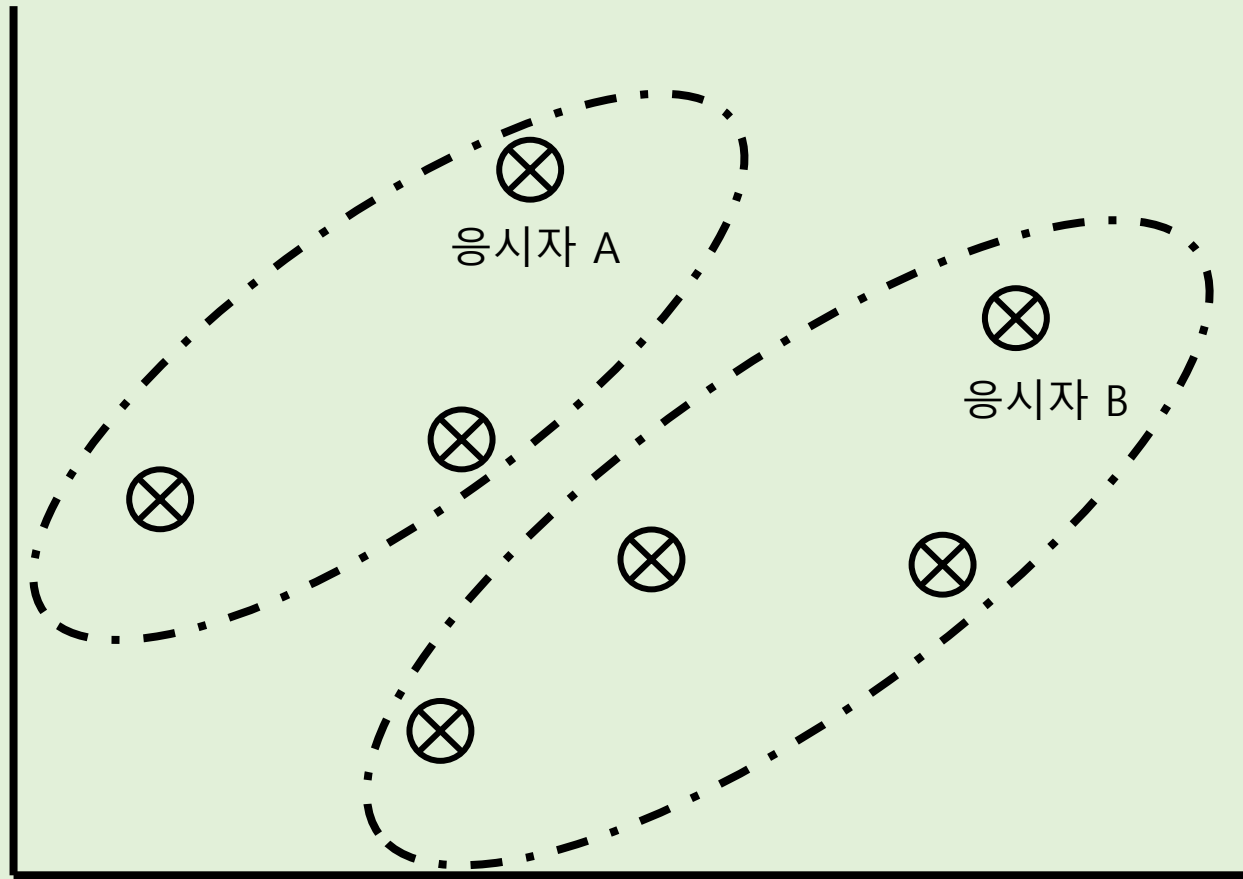
Text

The key to the cabinets are ...

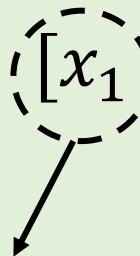
## 지도 학습(Supervised Learning)



## 비지도 학습(Unsupervised Learning)



BookCorpus  $[x_1, x_2, x_3, x_4, x_5, \dots, x_T]$

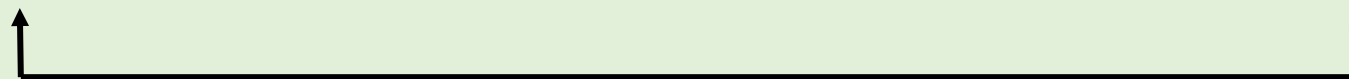


The key to the cabinets ...

$[s_1, s_2, s_3, s_4, s_5, \dots, s_{n-k-1}]$

... **is/are** on the table

$[s_{n-k}, s_{n-k+1}, \dots, s_n]$



$p(s_{n-k}, s_{n-k+1}, \dots, s_n \mid s_1, s_2, \dots, s_{n-k-1})$

*output*

*input*

베이지 정리 참조

$s_T$ 는 "(Linguistic) Symbol"의 약어로 단어, 형태소, 접사가 될 수 있다.

# 로지스틱 회귀(logistic Regression)

응시자  $m$  명의 정보

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \dots \\ x_{m-1} \\ x_m \end{bmatrix}$$

응시자 정보  $x$ 에 대한 가중치

$$\begin{bmatrix} b \\ w_1 \\ w_2 \\ w_3 \\ \dots \\ w_{m-2} \\ w_{m-1} \end{bmatrix}$$

고정값

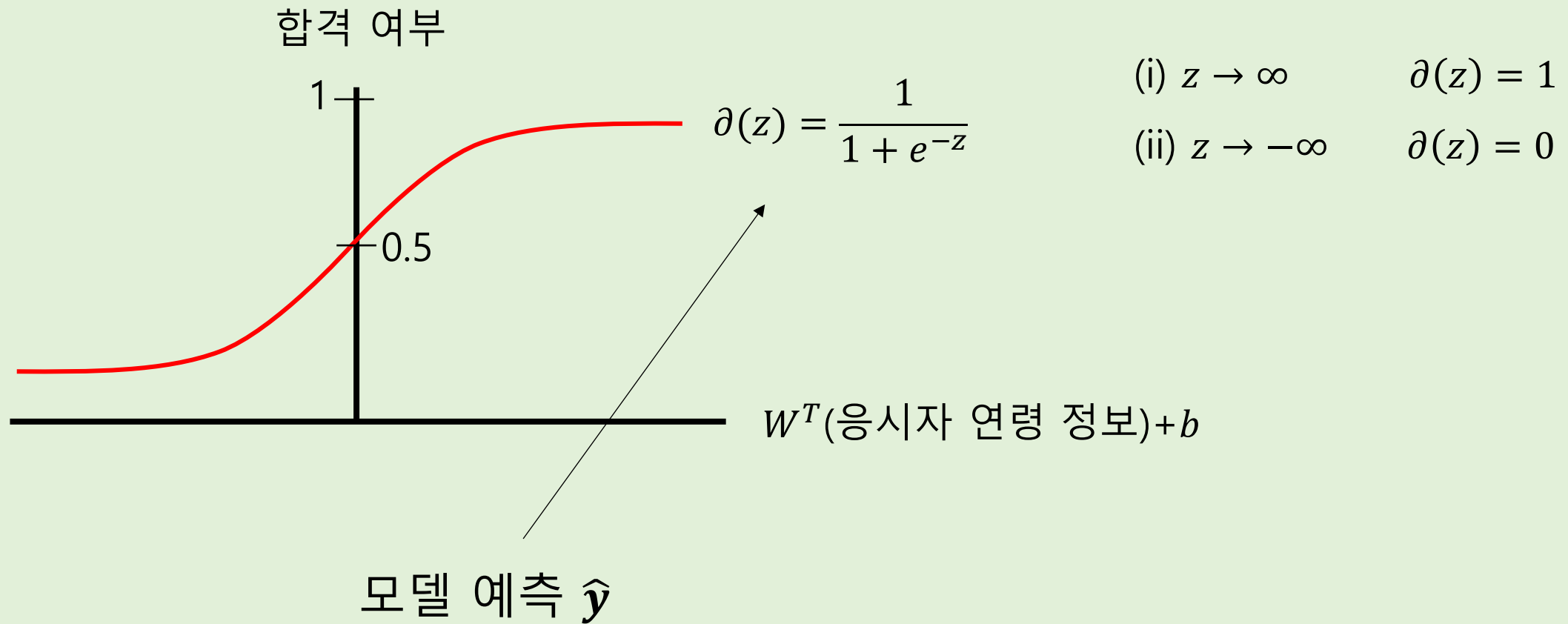
변동값

"Transpose"

$$\longrightarrow Z = w_1 x_1 + w_2 x_2 \dots w_{m-1} x_{m-1} + b$$
$$= \mathbf{W}^T + \mathbf{b}$$



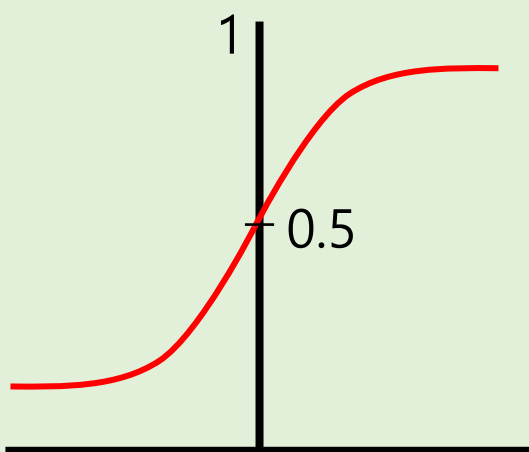
## 로지스틱 회귀(Logistic Regression)



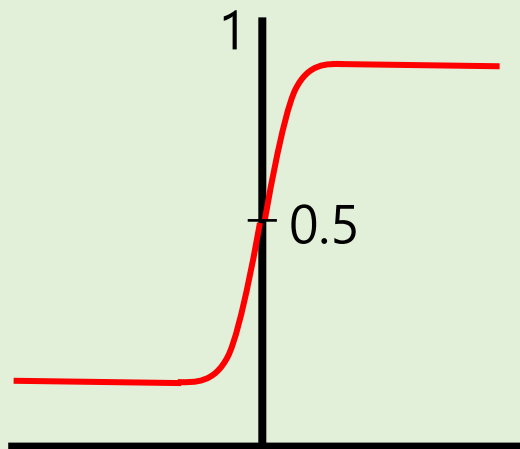
# 로지스틱 회귀(Logistic Regression)

시그모이드 함수  
(Sigmoid function)

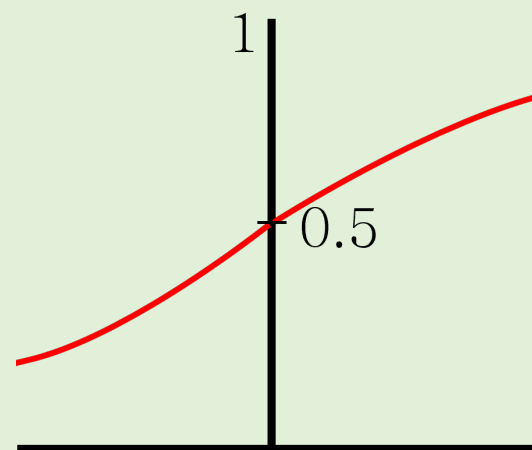
합격 여부



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

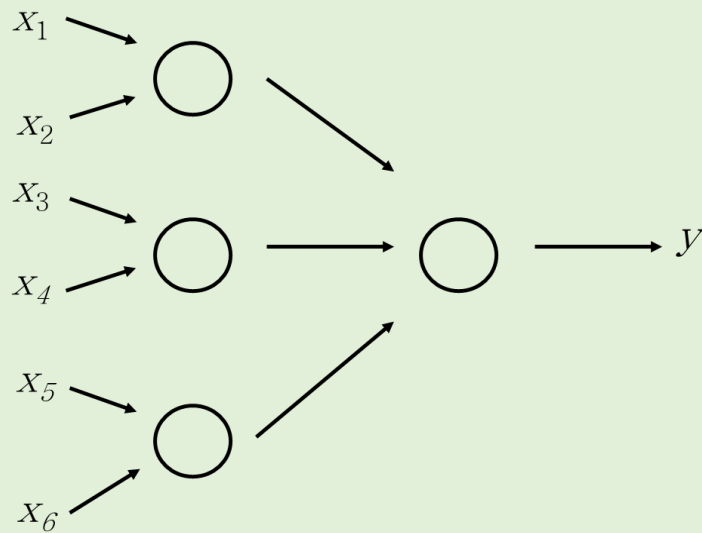


$$\sigma(z) = \frac{1}{1 + e^{-2z}}$$

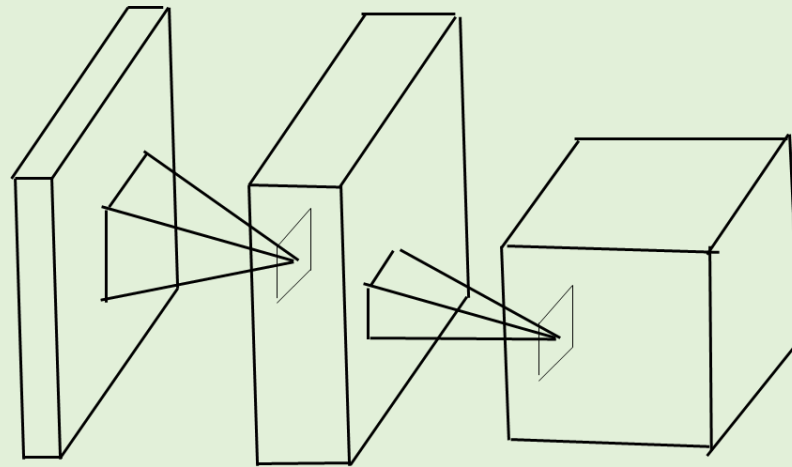


$$\sigma(z) = \frac{1}{1 + e^{-1/2z}}$$

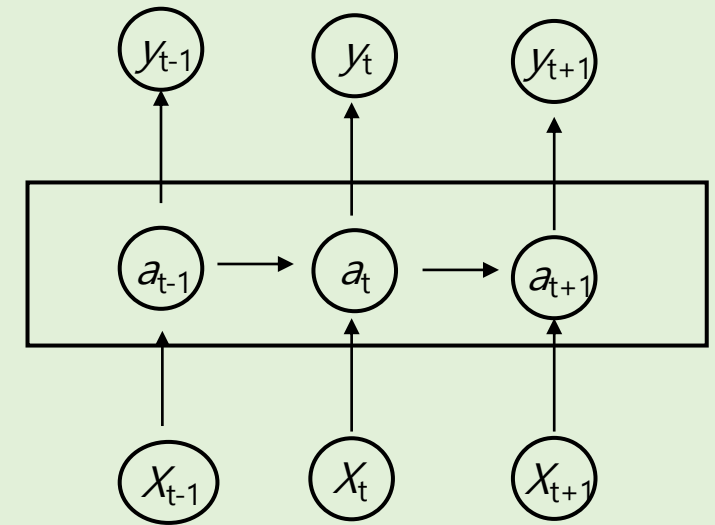
# 인공 신경망(Artificial Neural Networks; ANNs)



표준 신경망  
(Standard Neural Network)

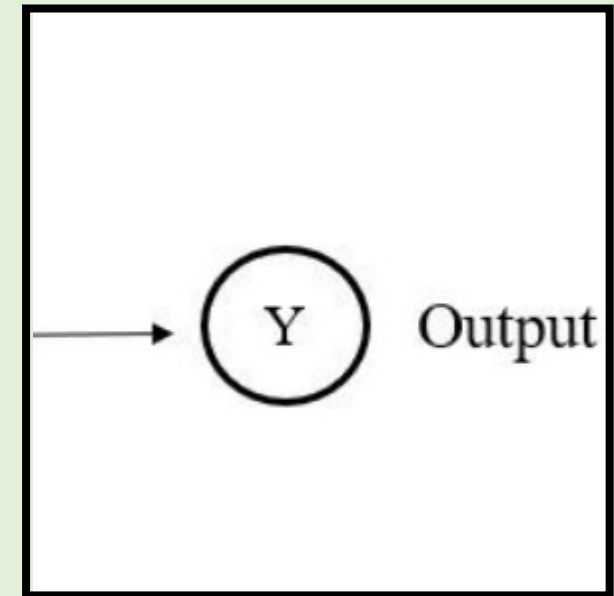
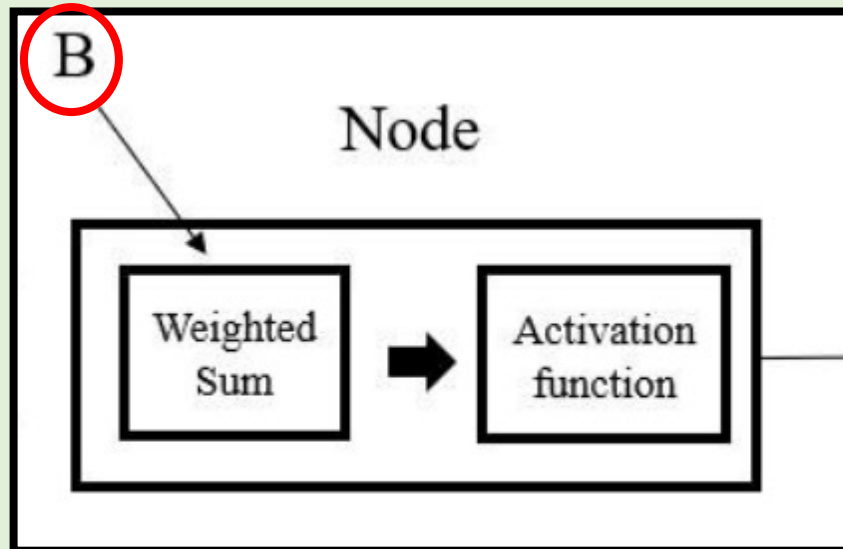
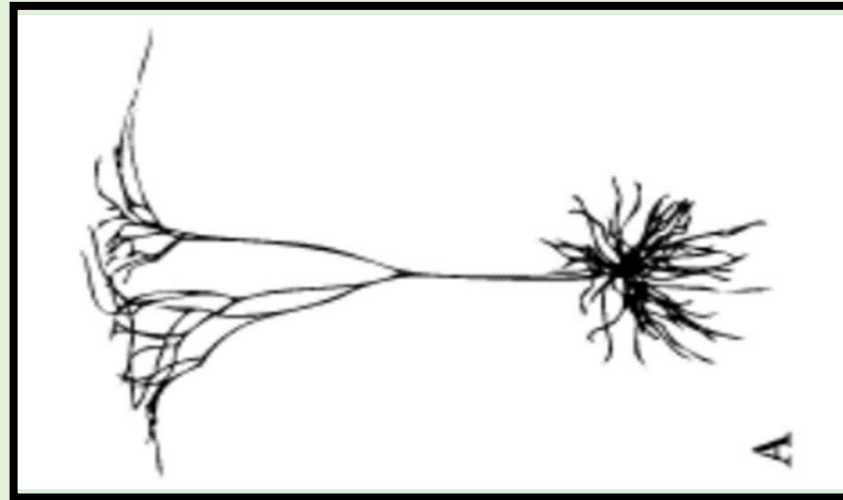
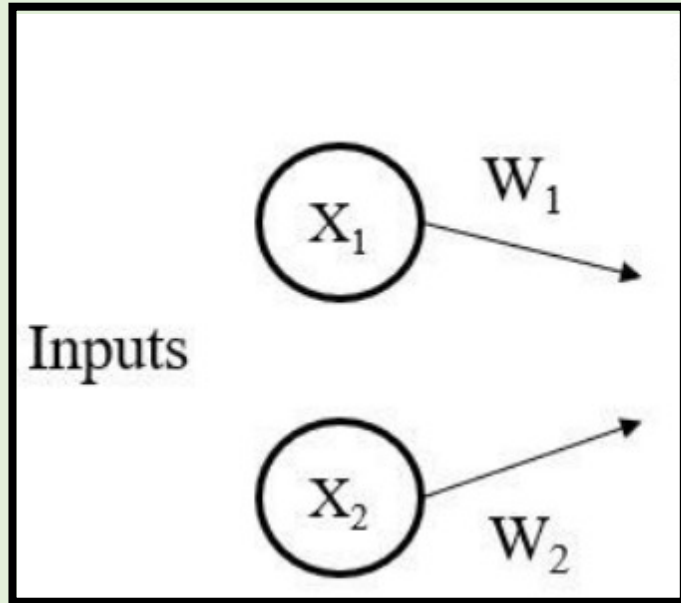


합성곱 신경망  
(Convolutional Neural Network)



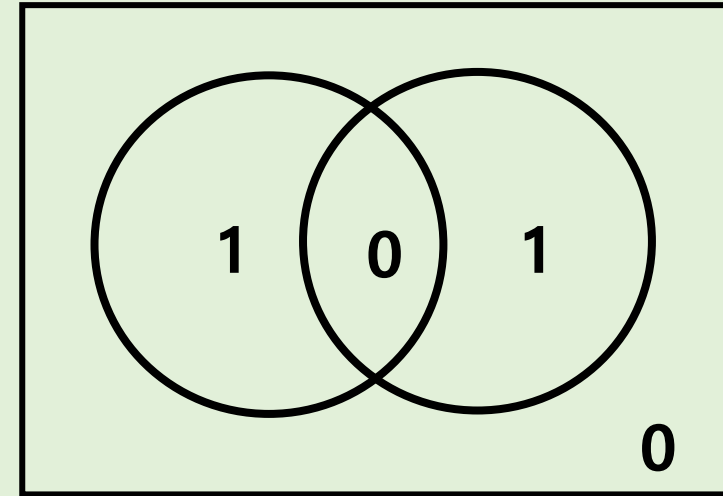
순환 신경망  
(Recurrent Neural Network)

# 퍼셉트론(Perceptron)



## 배타적 논리합(XOR) 문제

$x_1$	$x_2$	$y$
0	0	0
0	1	1
1	0	1
1	1	0



사실 : A는 신체검사를 통과하지 못했다.

진술 : A는 병역의무가 있다.

진술 : A는 남학생이다.

논리합: A는 병역의무가 없고, A는 남학생이 아니다.

<False>

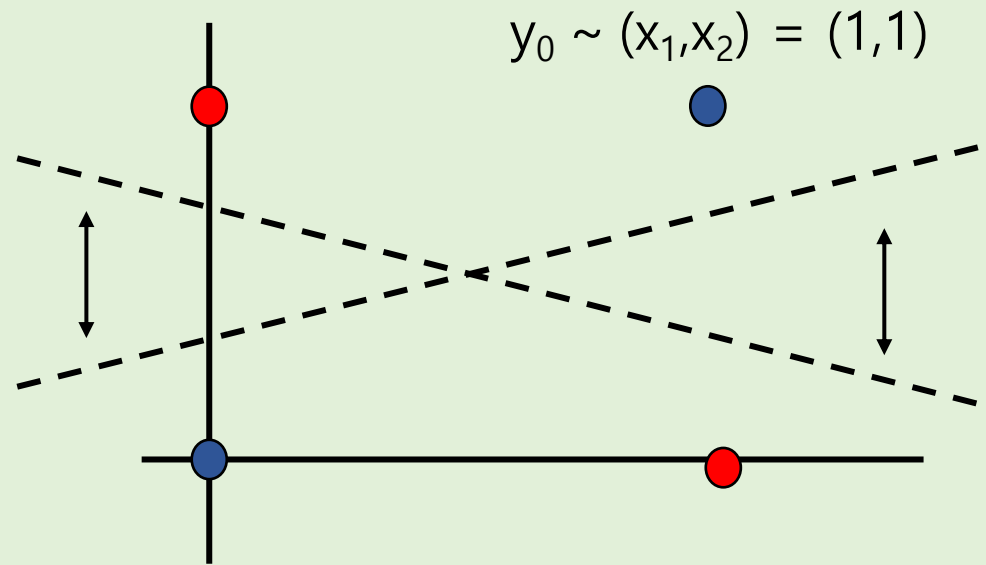
배타적 논리합: A는 병역의무가 있고, A는 남학생이다.

<True>

배타적 논리합: A는 병역의무가 없고, A는 남학생이 아니다.

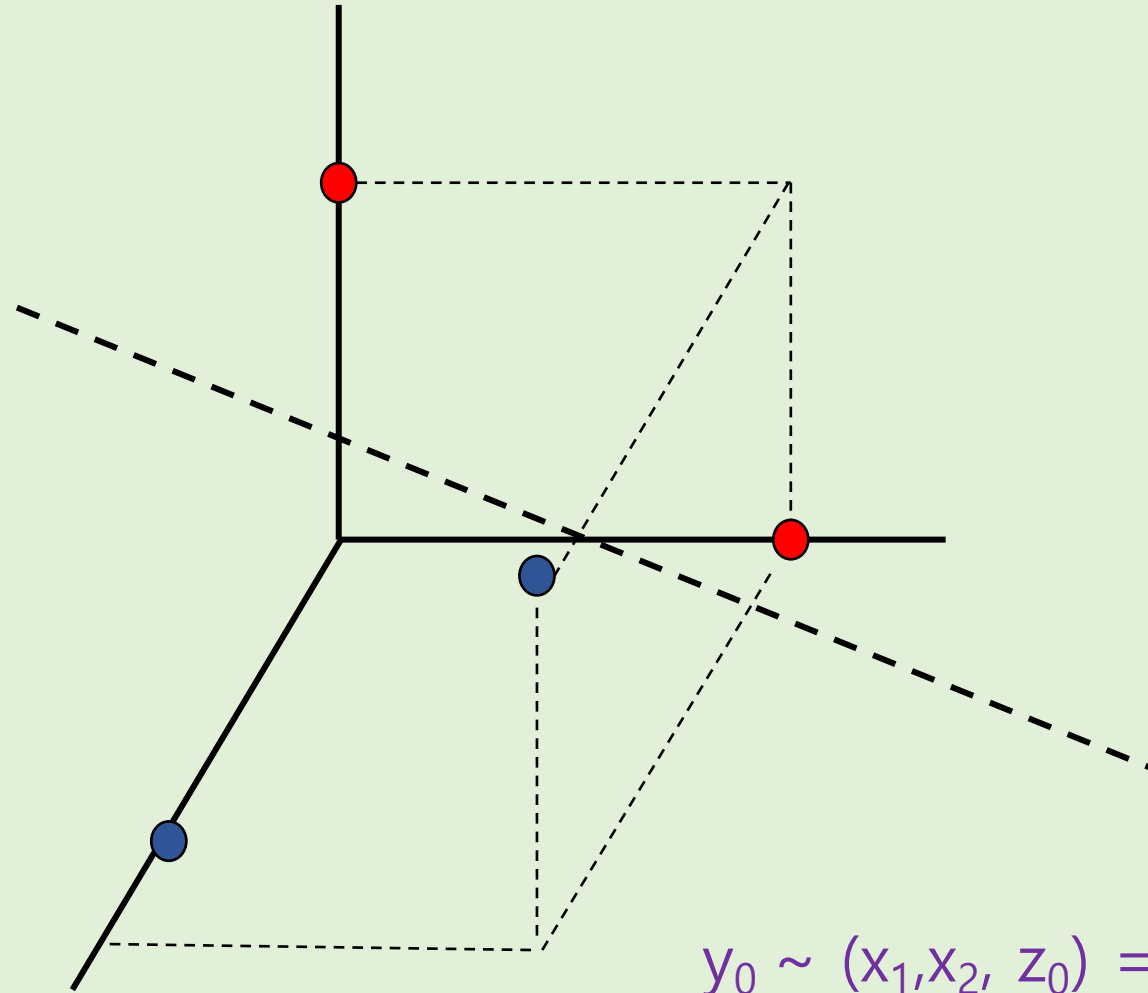
<True>

## $N$ 차원에서의 배타적 논리합(XOR) 문제



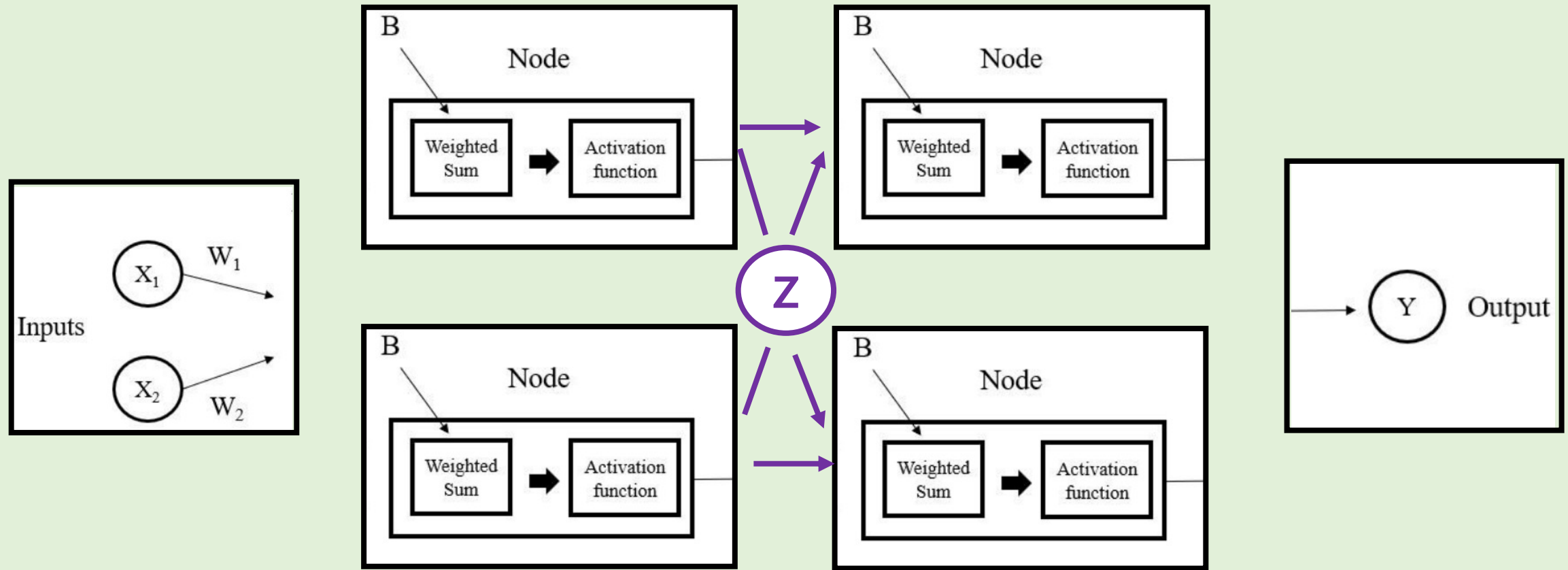
하나의 분류선을 그려서 빨강(1)과 파랑(0)을 구별할 수 없음!

$N$ 차원에서  $N+1$ 차원으로 확장





# 다층 퍼셉트론(Multi-layer Perceptron)



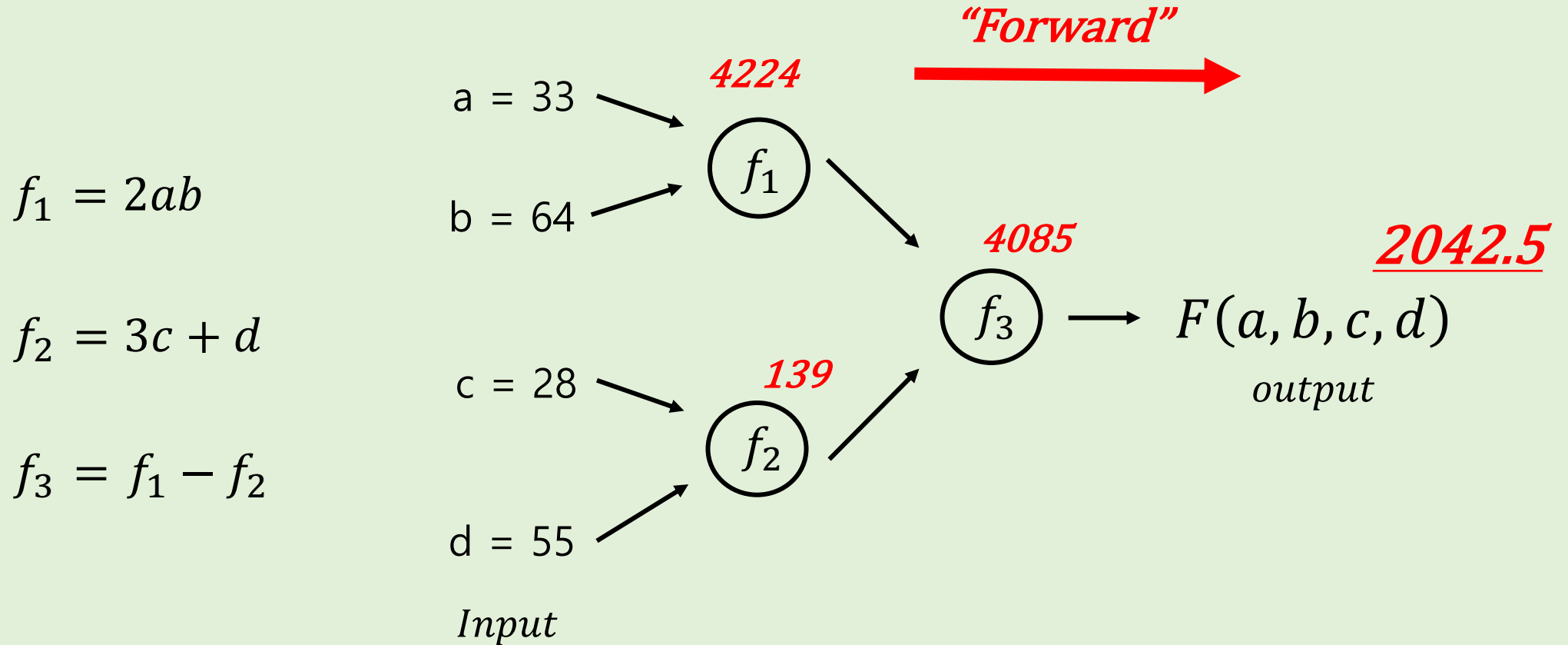
정답을 예측하는, 또는 논리 문제를 풀 수 있는  
최적의 회귀선은 어떻게 찾나요?

(i) 가중합 (*Weighted Sum*)를 계산하고 (ii) 활성화 함수  
(*activation function*)으로 판별한다.

## 순전파와 역전파(Forward and Backward Propagation)

응시자 A,B,C,D의 정보로 어떤 임의의 모형  $F(x)$  또는 “가중합”를 평가한다면?

$$F(a, b, c, d) = 0.5\{2ab - (3c + d)\}$$



## 순전파와 역전파(Forward and Backward Propagation)

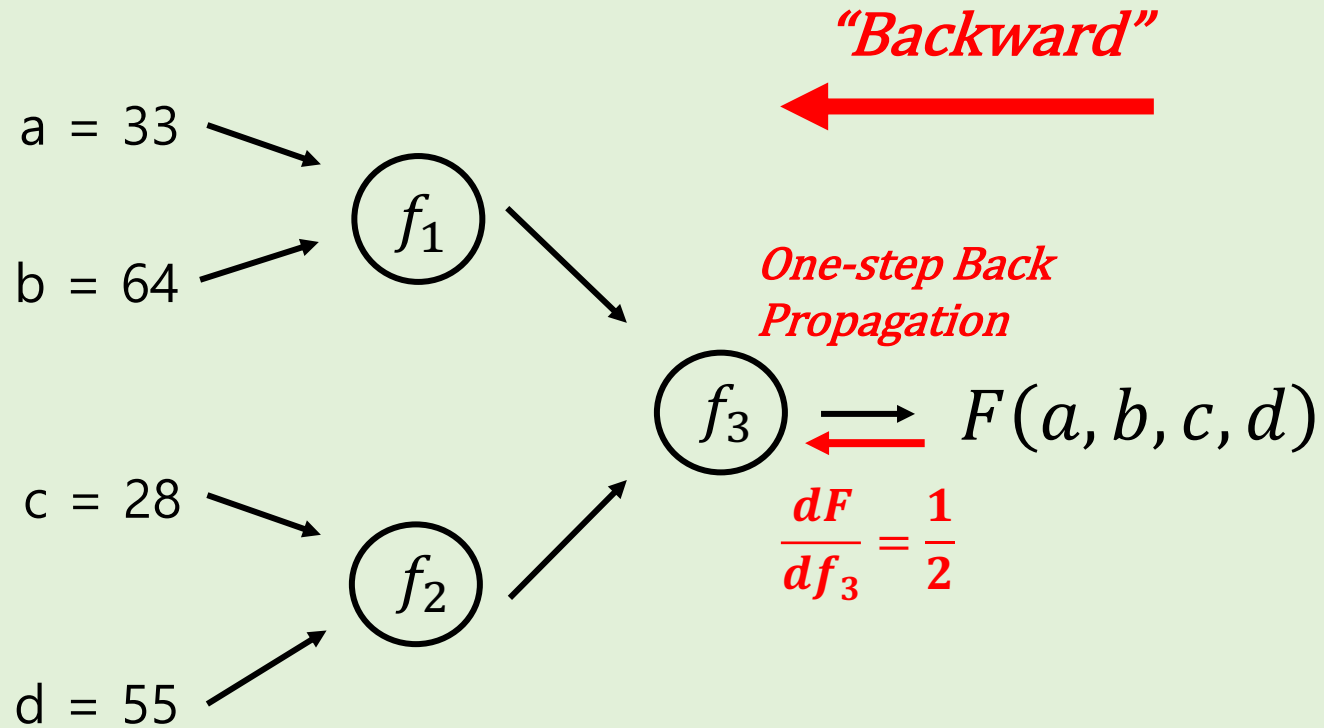
$$F(a, b, c, d) = 0.5\{2ab - (3c + d)\}$$

$$f_1 = 2ab$$

$$f_2 = 3c + d$$

$$f_3 = f_1 - f_2$$

$$F = 0.5 \times f_3$$



## 순전파와 역전파(Forward and Backward Propagation)

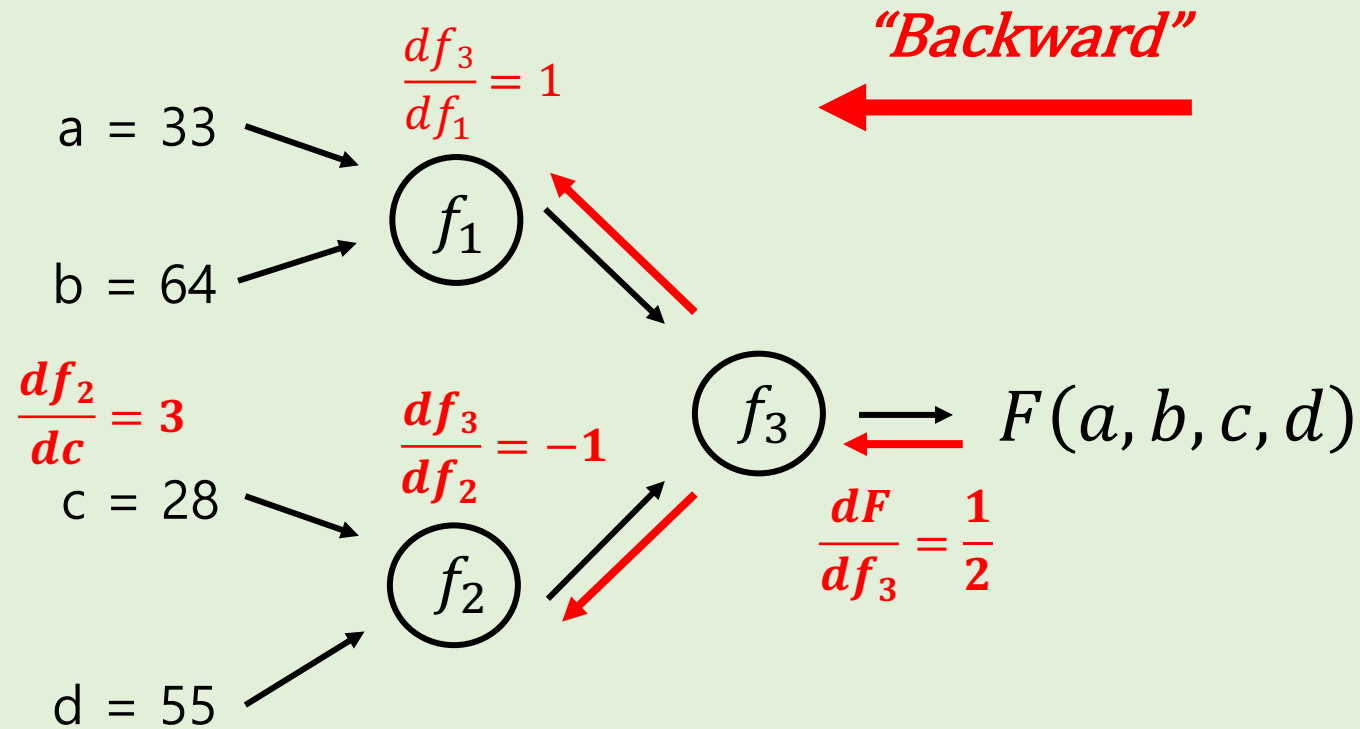
$$F(a, b, c, d) = 0.5\{2ab - (3c + d)\}$$

$$f_1 = 2ab$$

$$f_2 = 3c + d$$

$$f_3 = f_1 - f_2$$

$$F = 0.5 \times f_3$$



## 순전파와 역전파(Forward and Backward Propagation)

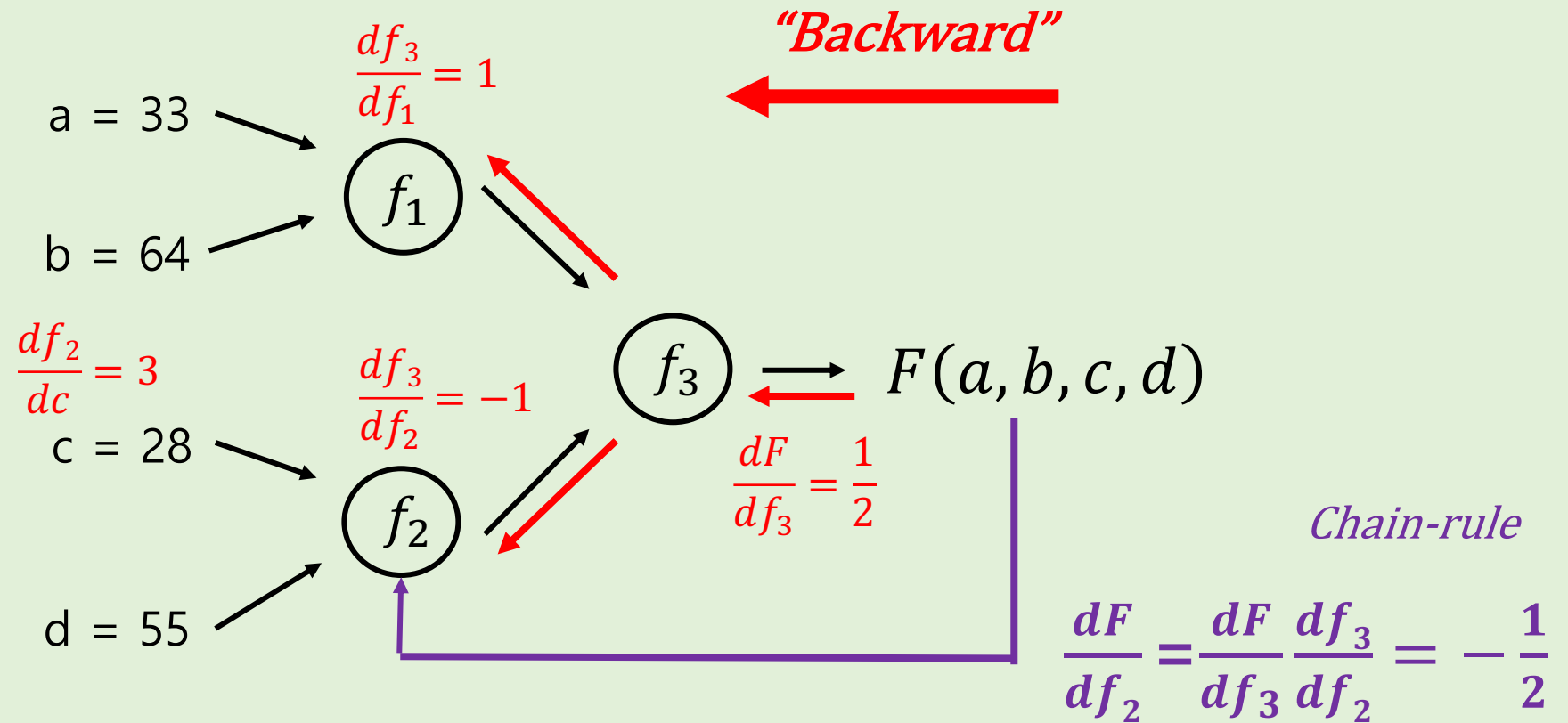
$$F(a, b, c, d) = 0.5\{2ab - (3c + d)\}$$

$$f_1 = 2ab$$

$$f_2 = 3c + d$$

$$f_3 = f_1 - f_2$$

$$F = 0.5 \times f_3$$



## 비용 함수(Cost Function)

:비용 함수는 모든 입력 정보에 대한 손실 함수(Loss Function)의 평균값

M개의 입력 정보  $(x^1, y^1) (x^2, y^2) \dots (x^m, y^m)$  에 대하여

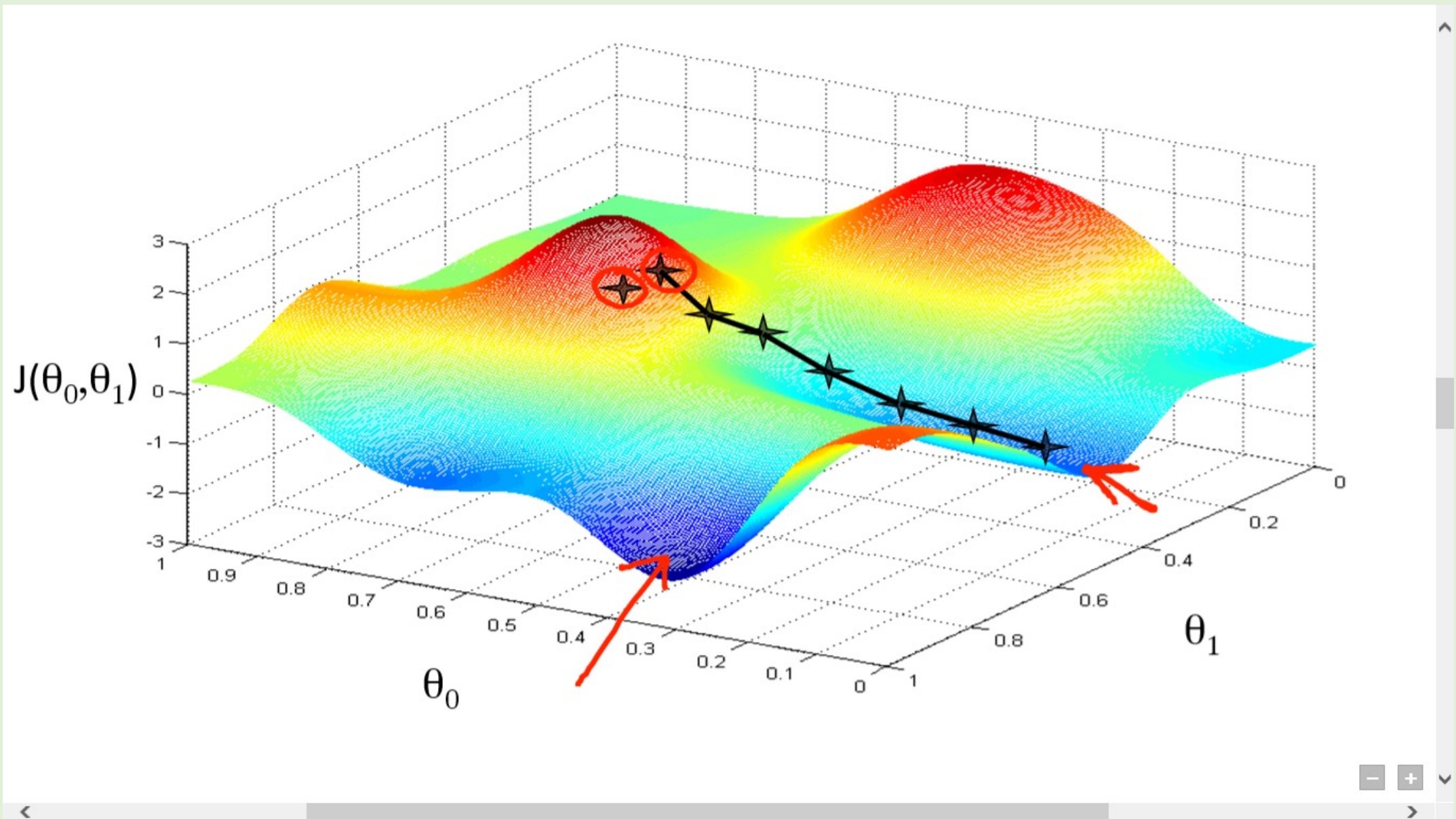
손실 함수  $L$  은 합격 여부  $y$  와 모델 예측  $\hat{y}$  의 차이에 대한 함수이다.

$$L(\hat{y}, y) = - (y \log \hat{y} + (1 - y) \log(1 - \hat{y}))$$

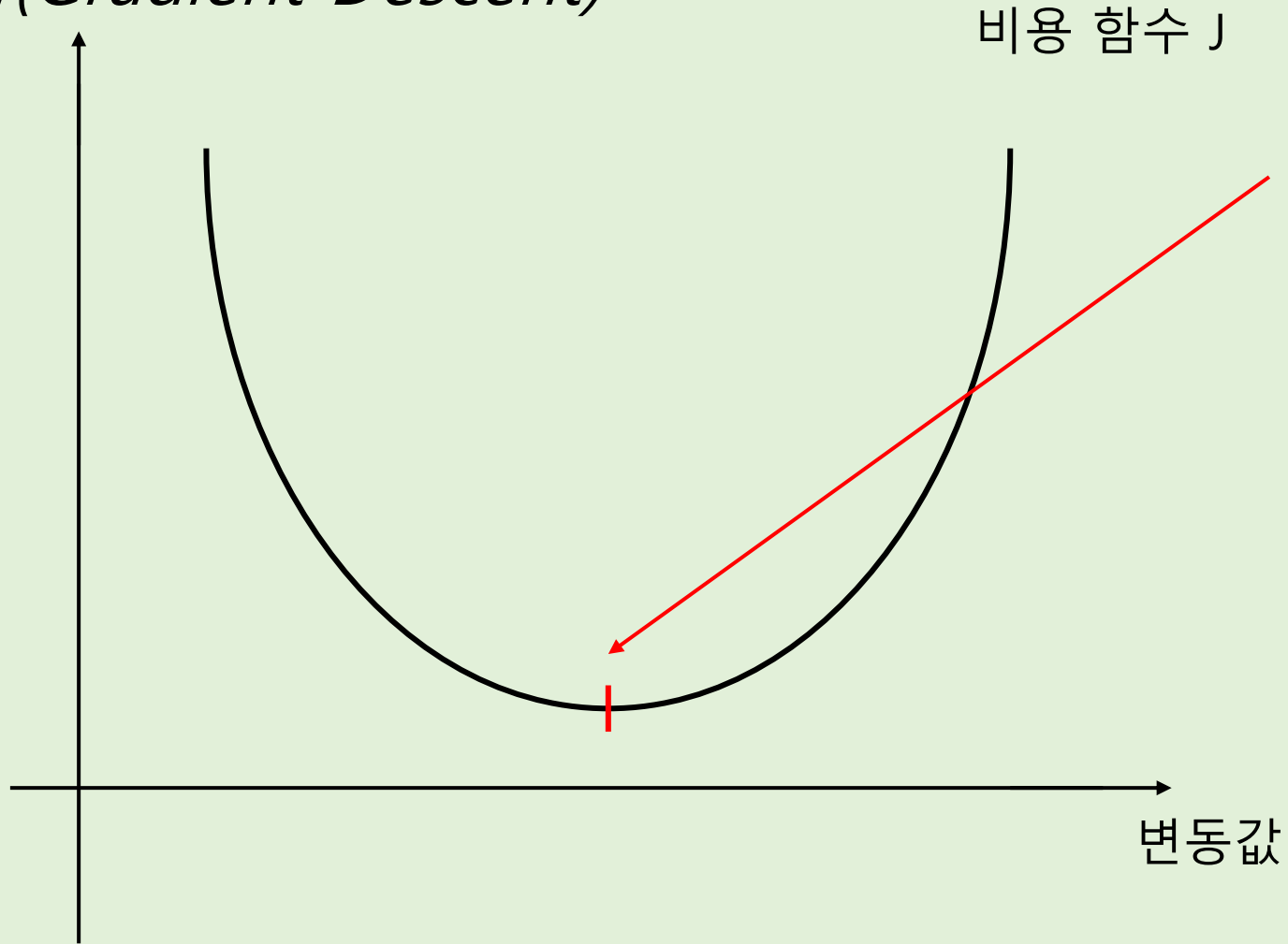
손실 함수 또는 "오류의 크기"를 최소화하는 모델을 만드는 것이 목적!

합격 (1) 이라면 모델 예측  $\hat{y}$  는 커져야 하고, 불합격 (0) 이라면 모델 예측  $\hat{y}$  는 작아져야 한다.



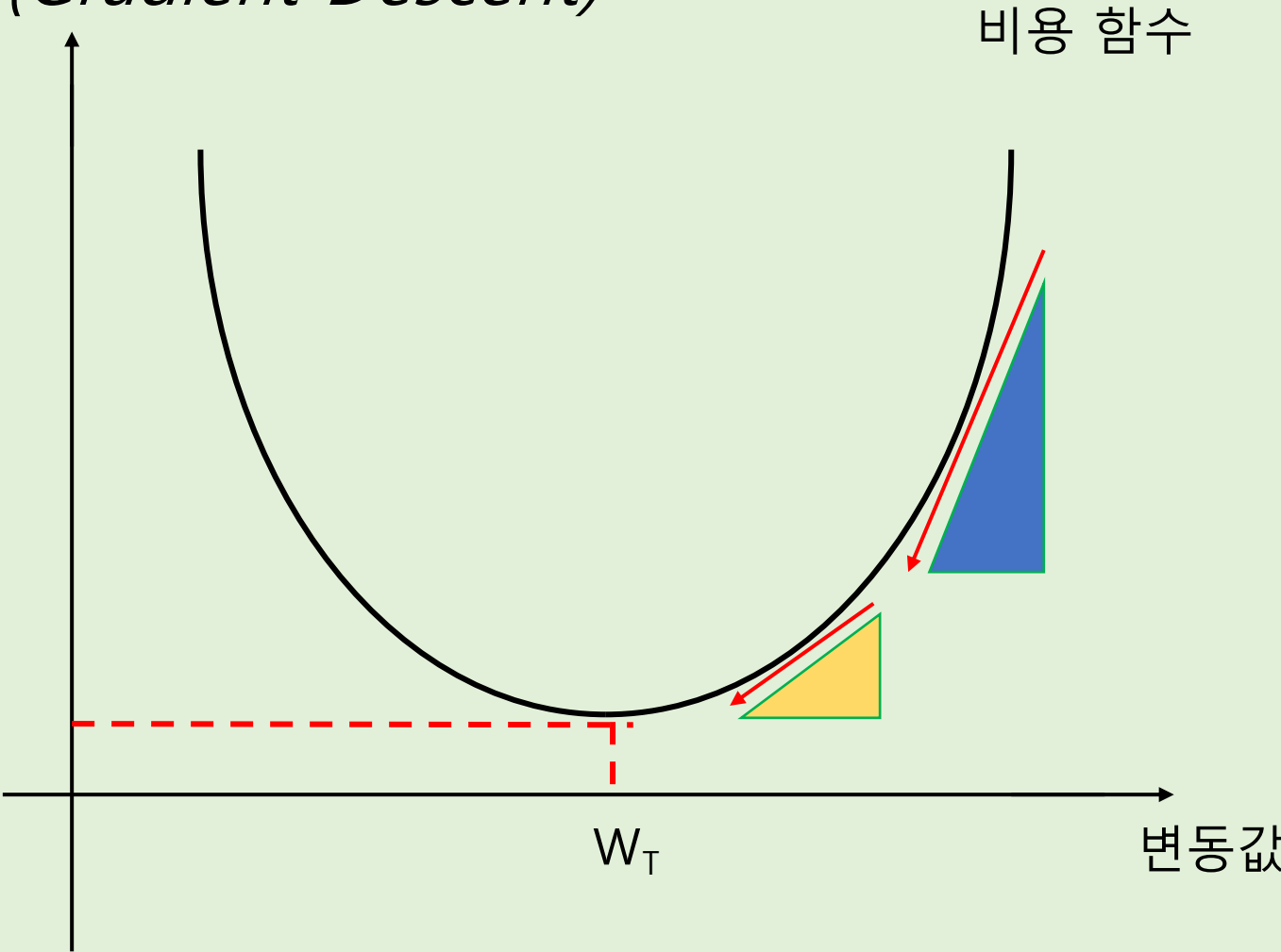


# 경사하강법(Gradient Descent)



비용 함수  $J$ 의 값이 가장 작아지는 변동값  $W$ 를 찾는 것이 목적

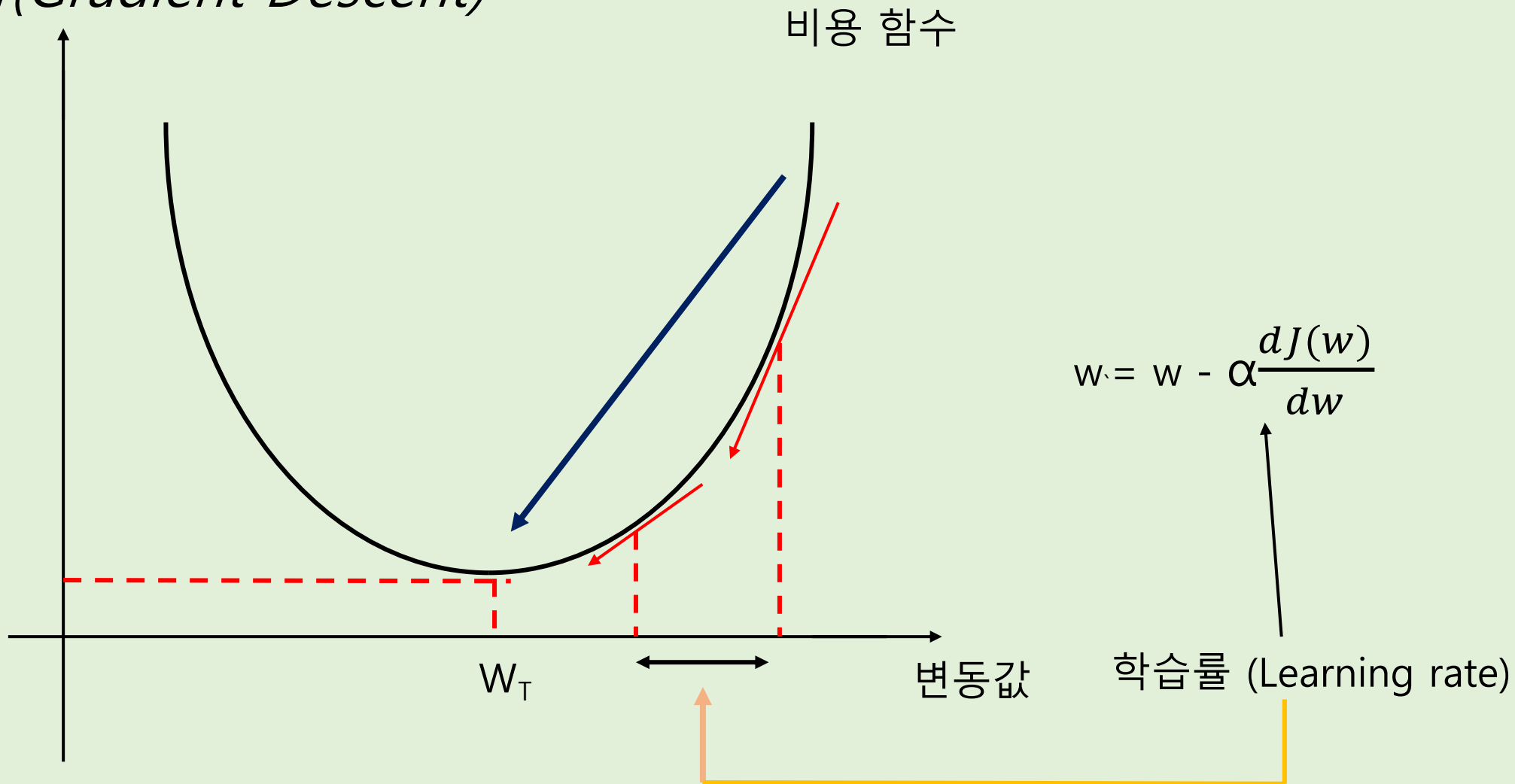
# 경사하강법(Gradient Descent)



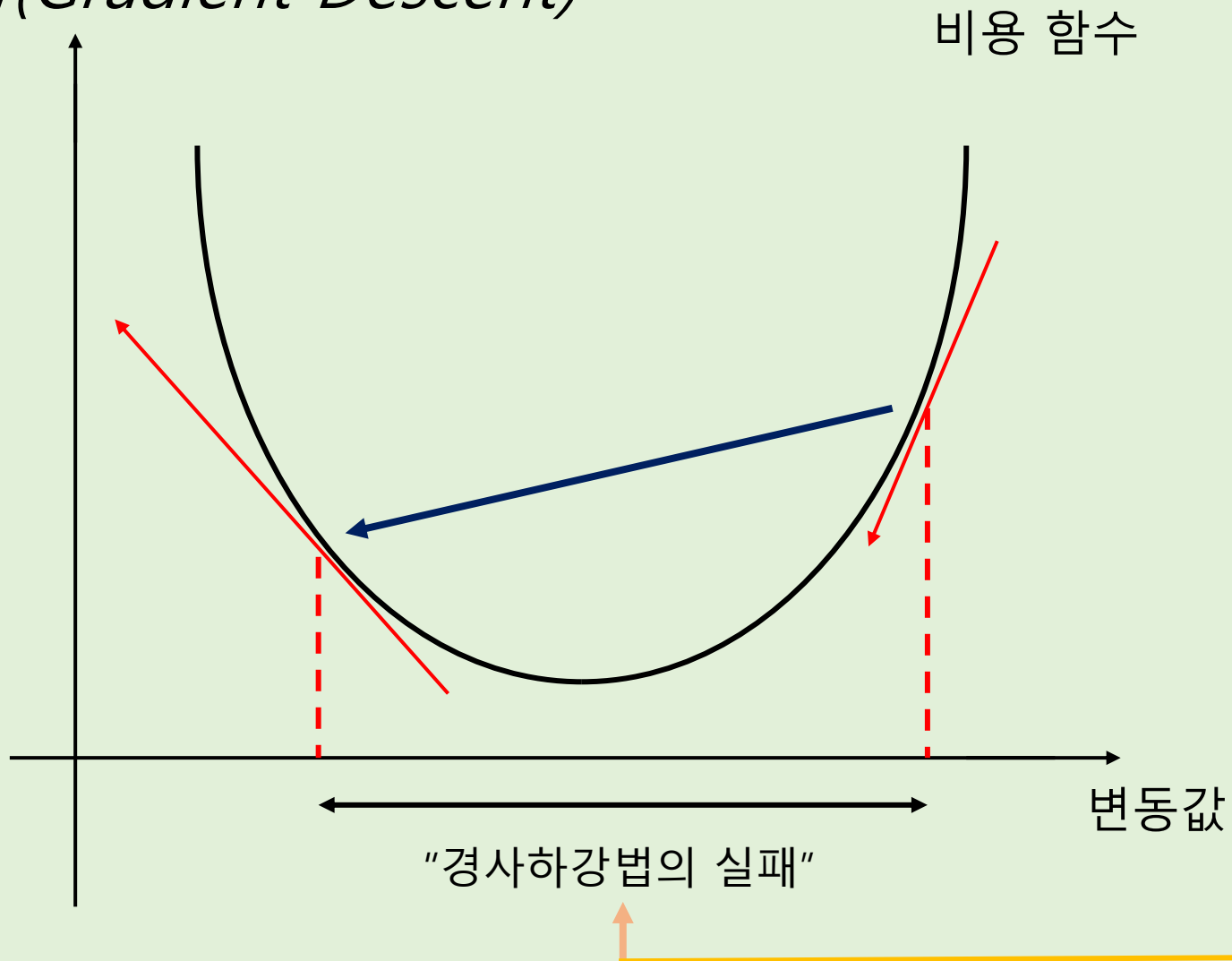
비용 함수

$$w' = w - \alpha \frac{dJ(w)}{dw}$$

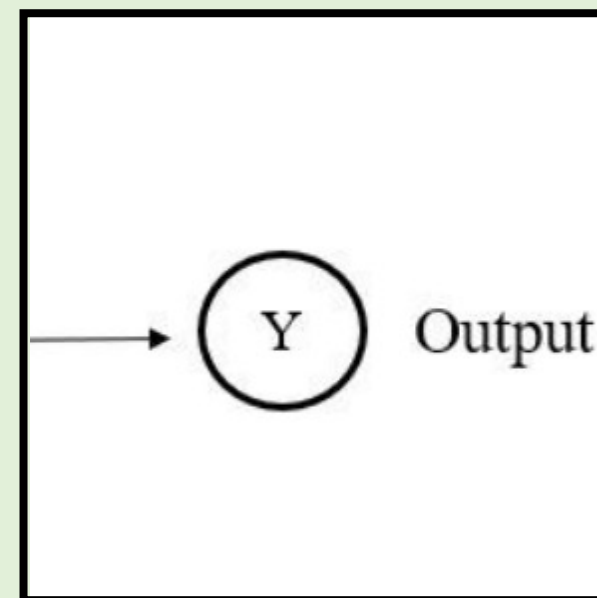
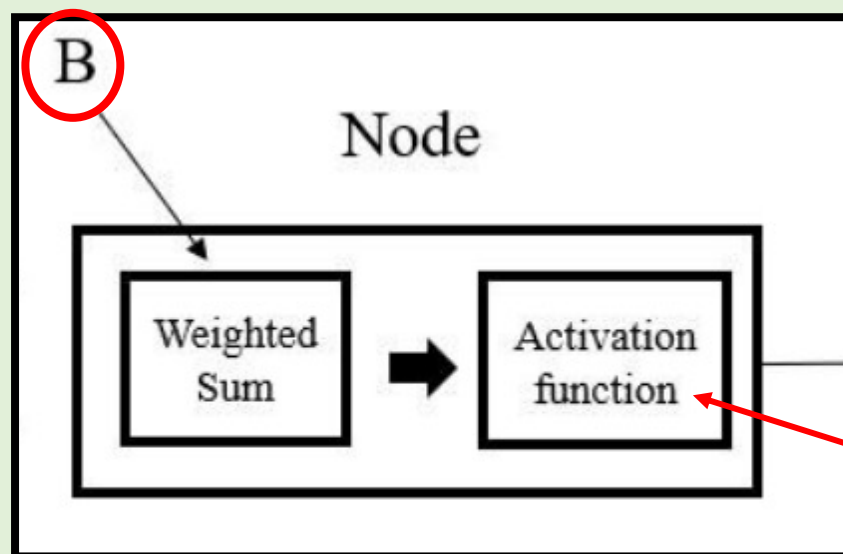
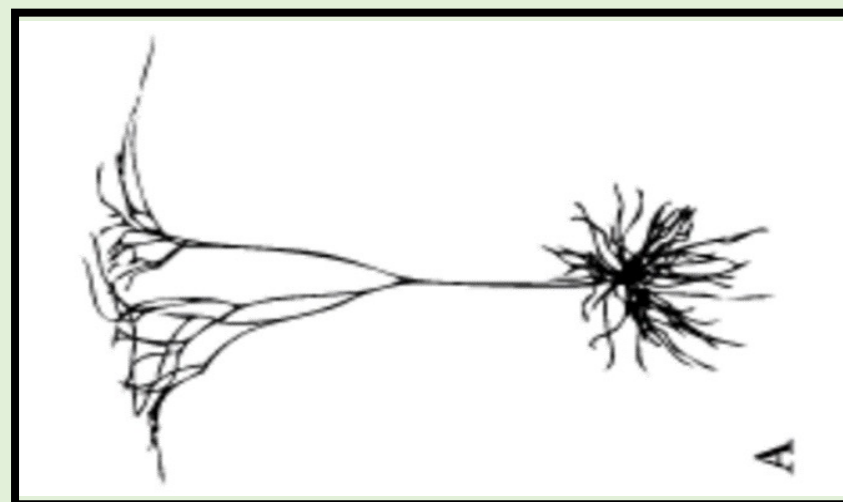
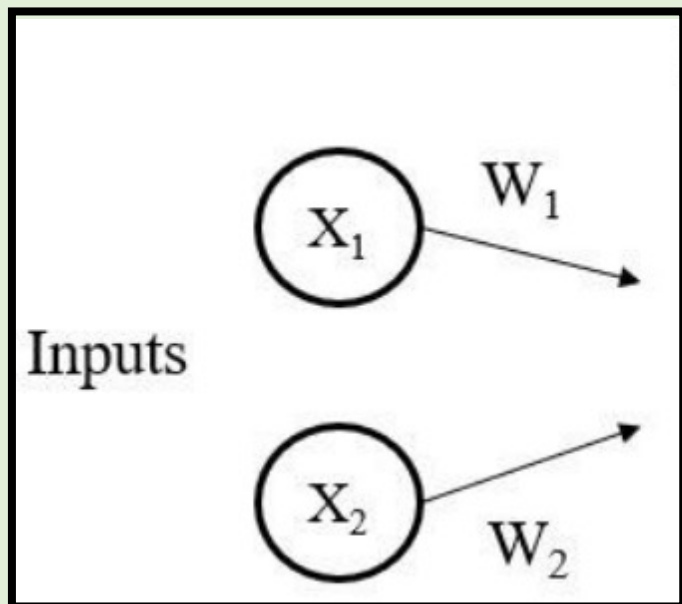
# 경사하강법(Gradient Descent)



# 경사하강법(Gradient Descent)

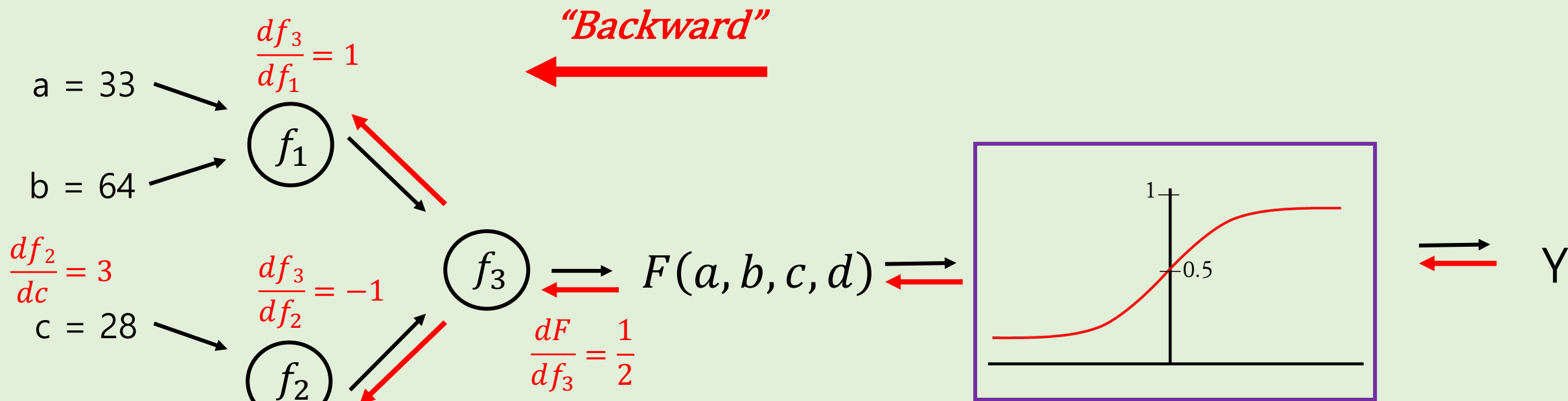


$$w = w - \alpha \frac{dJ(w)}{dw}$$



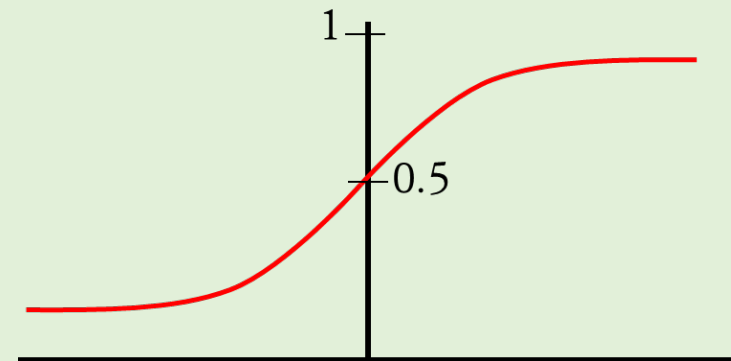
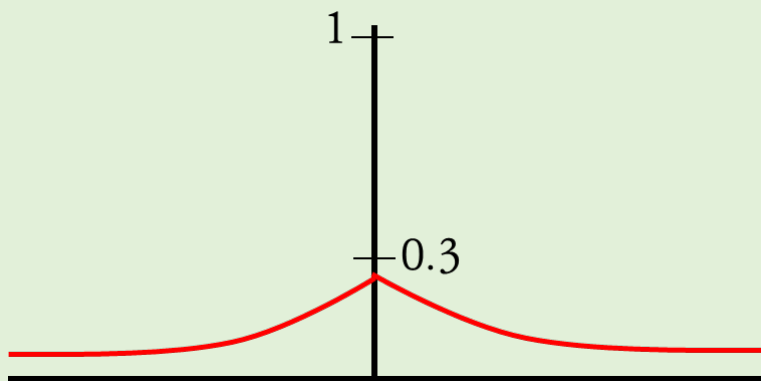
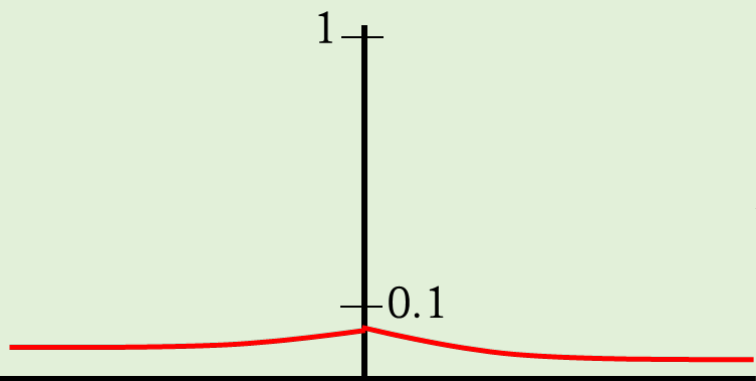
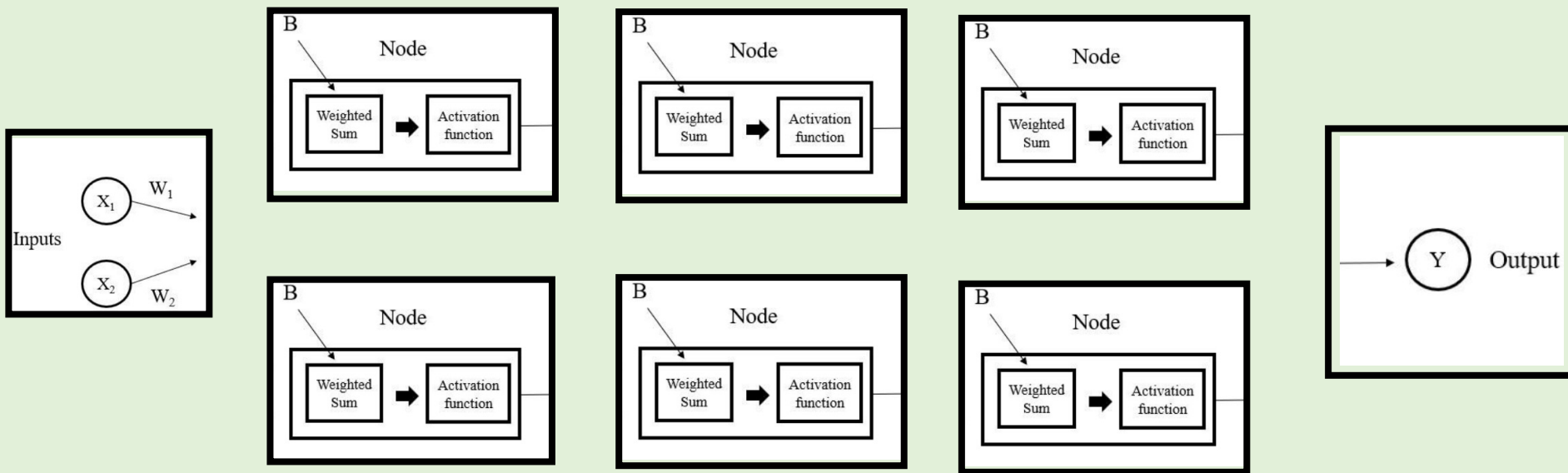
활성 함수

## 비선형 활성화 함수(Non-linear Activation Function)



오차역전파 과정에서 활성화함수도 미분됨!





입력 정보에 대하여 매우 작은 값만 출력하게 되어 학습이 실패함

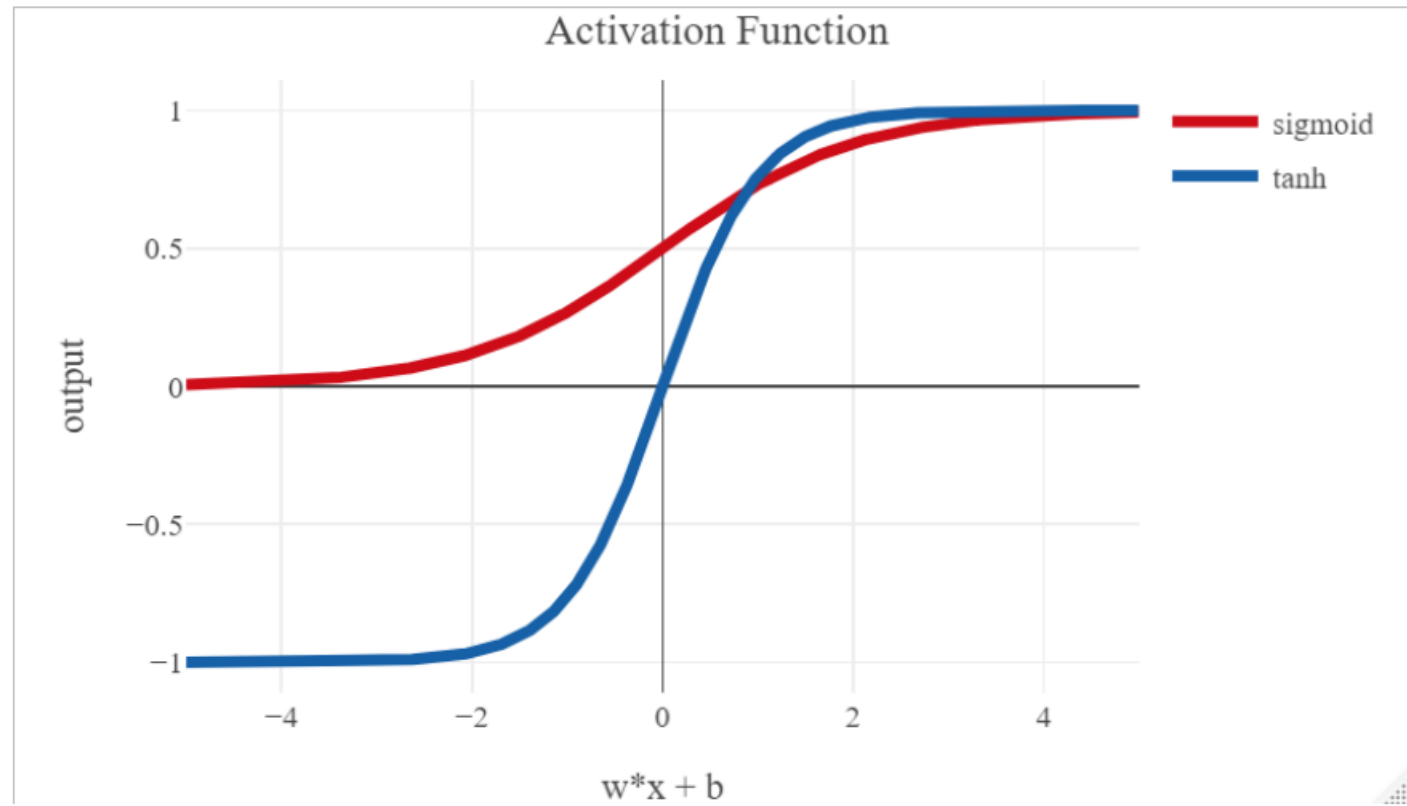


Figure 1.1. Graphs of sigmoid and hyperbolic tangent activation. The output value of the sigmoid function falls in the range  $0 < y < 1$ . The output value of the hyperbolic tangent falls in the range  $-1 < y < 1$ .

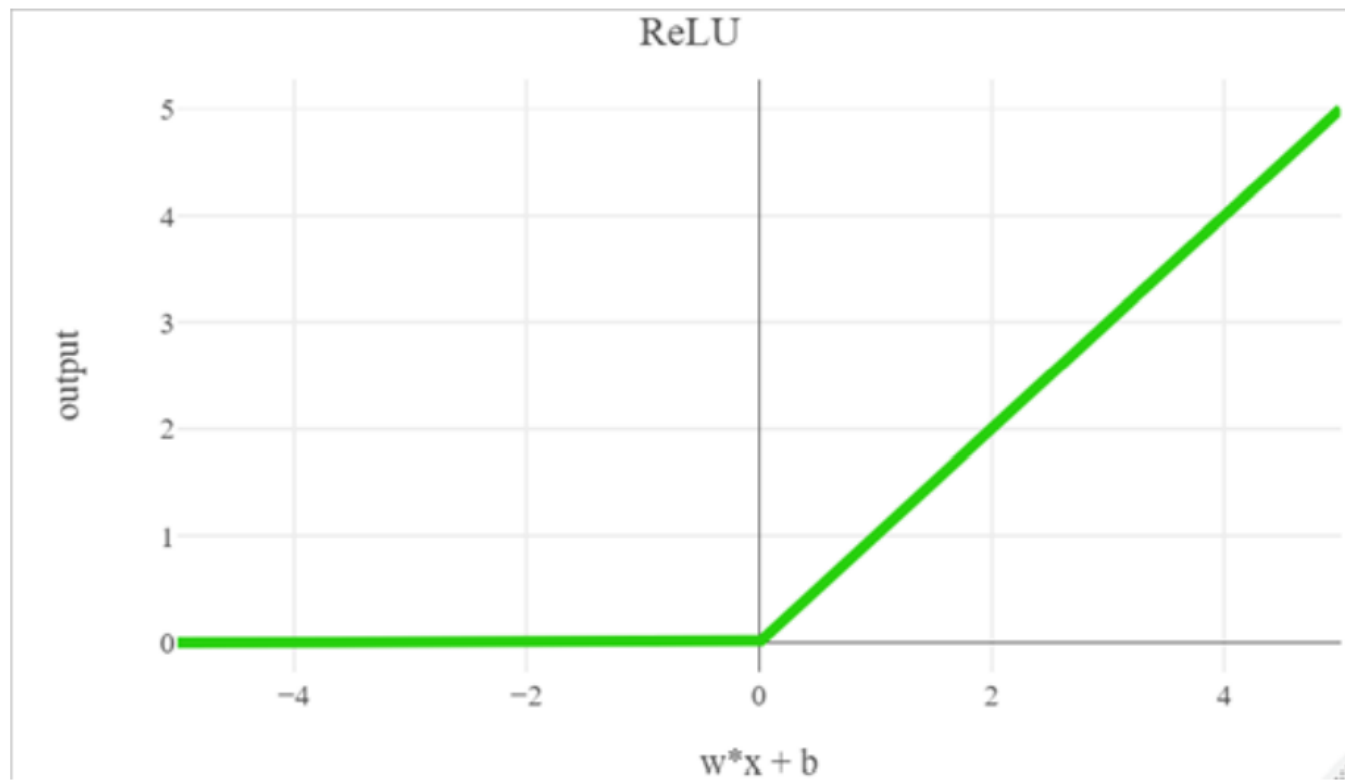
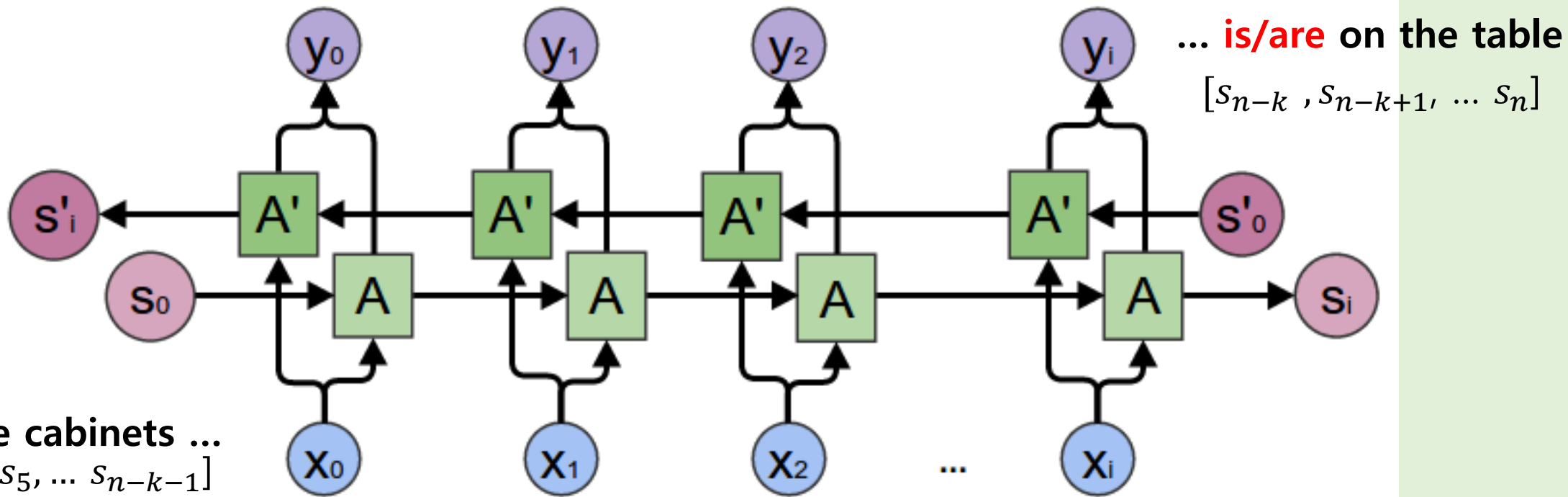


Figure 1.4. Graph of ReLU activation. ReLU is commonly used in state-of-the-art neural architecture. The output value of ReLU is sparser than that of a sigmoid or hyperbolic tangent because ReLU is approximated to zero if the inputs are negative.

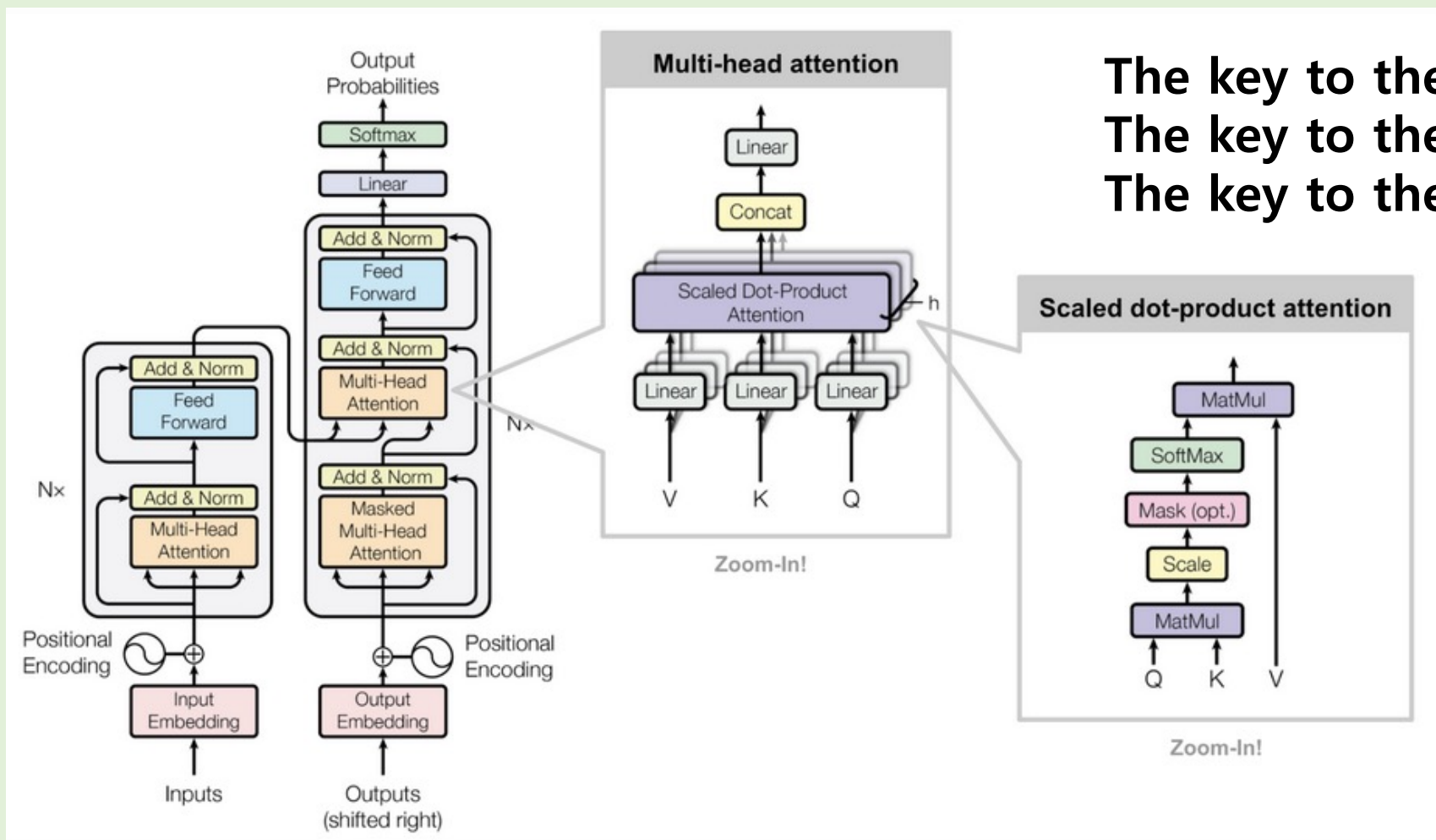
(i) Rectified Linear Unit(ReLU)는  $x$ 가 0보다 크면  $x$ 를 출력하고, 0보다 작으면 0을 출력함.

(ii) 아무리 미분하여도 도함수가 1로 고정됨!

## What to do next: $Bi_{(directional)}LSTM-RNN$



*What to do next: "deep"  $B_{(idirectional)}$ ERT*



The key to the \_\_\_ is on the table  
 The key to the cabinets \_\_\_ on the table  
 The key to the cabinets is on \_\_\_ table

...

$[s_1, s_2, s_3, [MASK], s_5, \dots s_n]$   
 $[s_1, s_2, [MASK], s_4, s_5 \dots s_n]$   
 $[s_1, [MASK], s_3, s_4, s_5, \dots s_n]$