

뉴럴 네트워크 기초

Basic of Neural Network

김영진

you359@cbnu.ac.kr

충북대학교 산업인공지능연구센터



CONTENTS



Introduction of Neural Network

Neural Network Concepts



Basic Neural Network Architecture

Perceptron, Multi-layered Perceptron, Activation Function, Logits



Neural Network Training

Gradient Decent and Back Propagation, Loss, optimization algorithms



Neural Network Evaluation

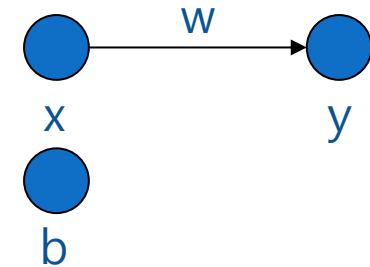
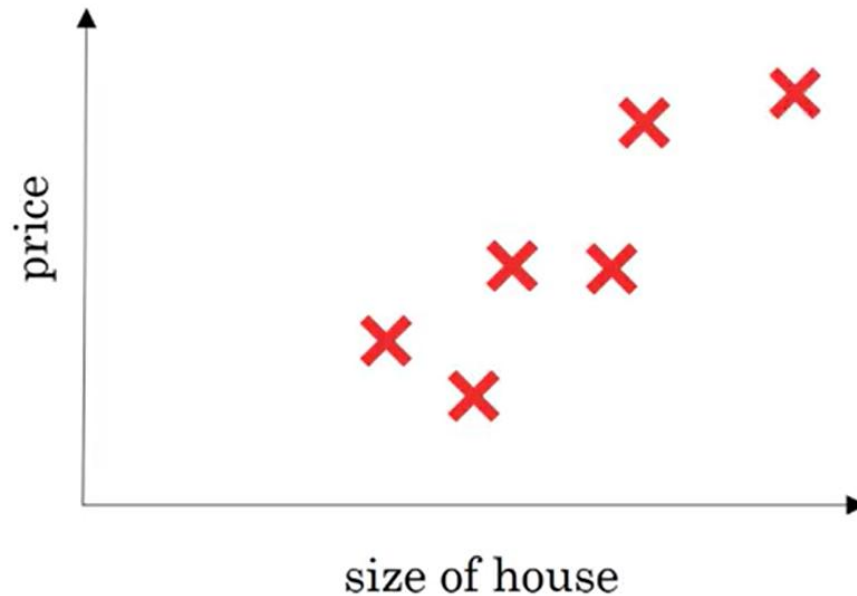
Generalization, Loss curve

■ 앞으로 2주간...

주차	수업내용	수업방식
1	Lecture Introduction	대면수업
2	Introduction to Pattern Recognition	대면수업
3	Basic of Neural Network → Neural Network의 기본 구조와 동작 방식	대면수업
4	Learning Strategies: Supervised Learning → 지도학습에서의 하이퍼파라미터 튜닝, 규제화, 최적화 방법 학습	대면수업
5	Learning Strategies: Unsupervised Learning → 비지도학습의 동작방식과 학습 방법, 응용 방안 등	대면수업
6	Model Architecture: Convolutional Neural Networks	대면수업
7	Model Architecture: Sequence Models	대면수업
8	Midterm exam	대면수업
9	Pattern Recognition Application: Visual Anomaly Detection	대면수업
10	Pattern Recognition Application: Temporal Anomaly Detection	대면수업
11	Pattern Recognition Application: Predictive Maintenance with Time Series	대면수업
12	Advanced Pattern Recognition	대면수업
13	Final Project: Proposal	대면수업
14	Final Project: Discussion	대면수업
15	Final Project: Presentation	대면수업

■ Housing Price Prediction

- 데이터 x : size of house (1개 값)
- 예측값 y : price



뉴런

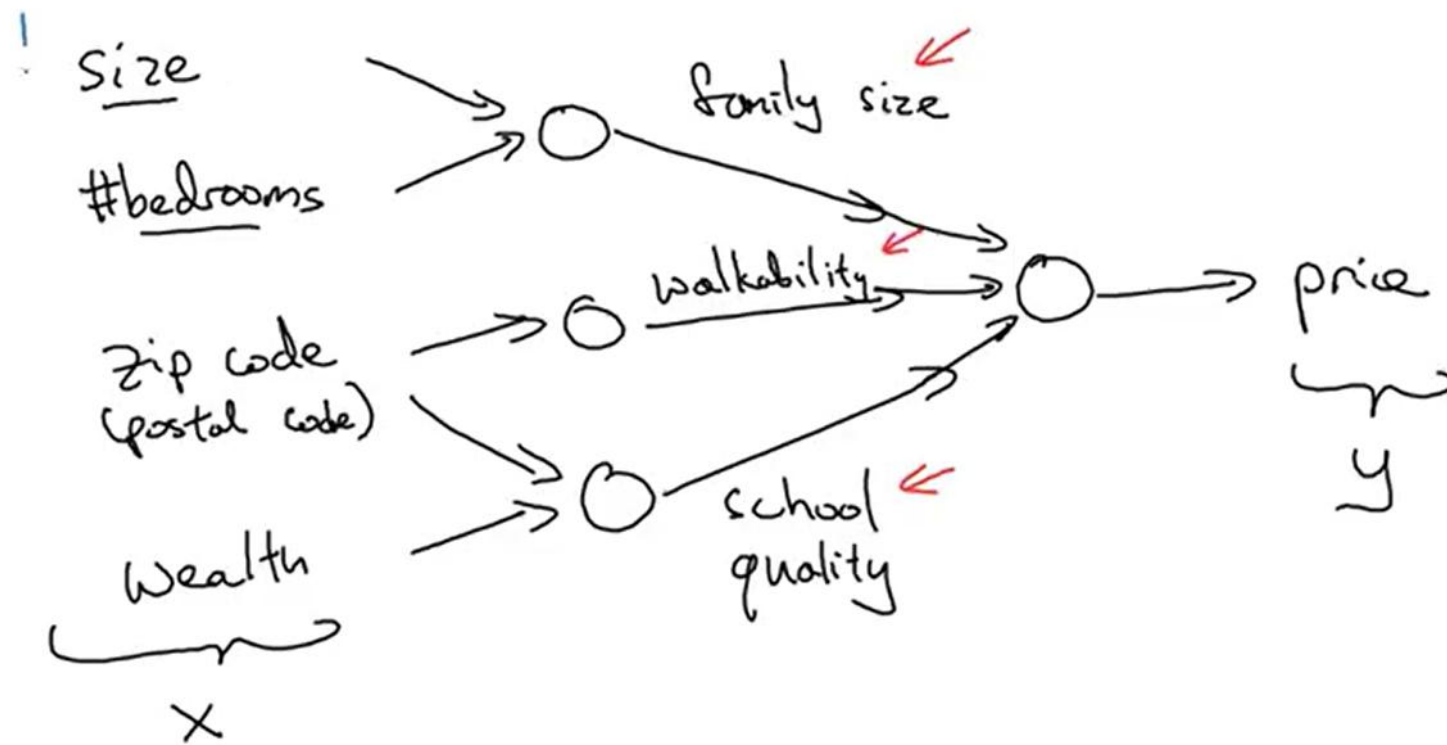
$$Y = wx + b$$

선형 방정식

$$Y = ax + b$$

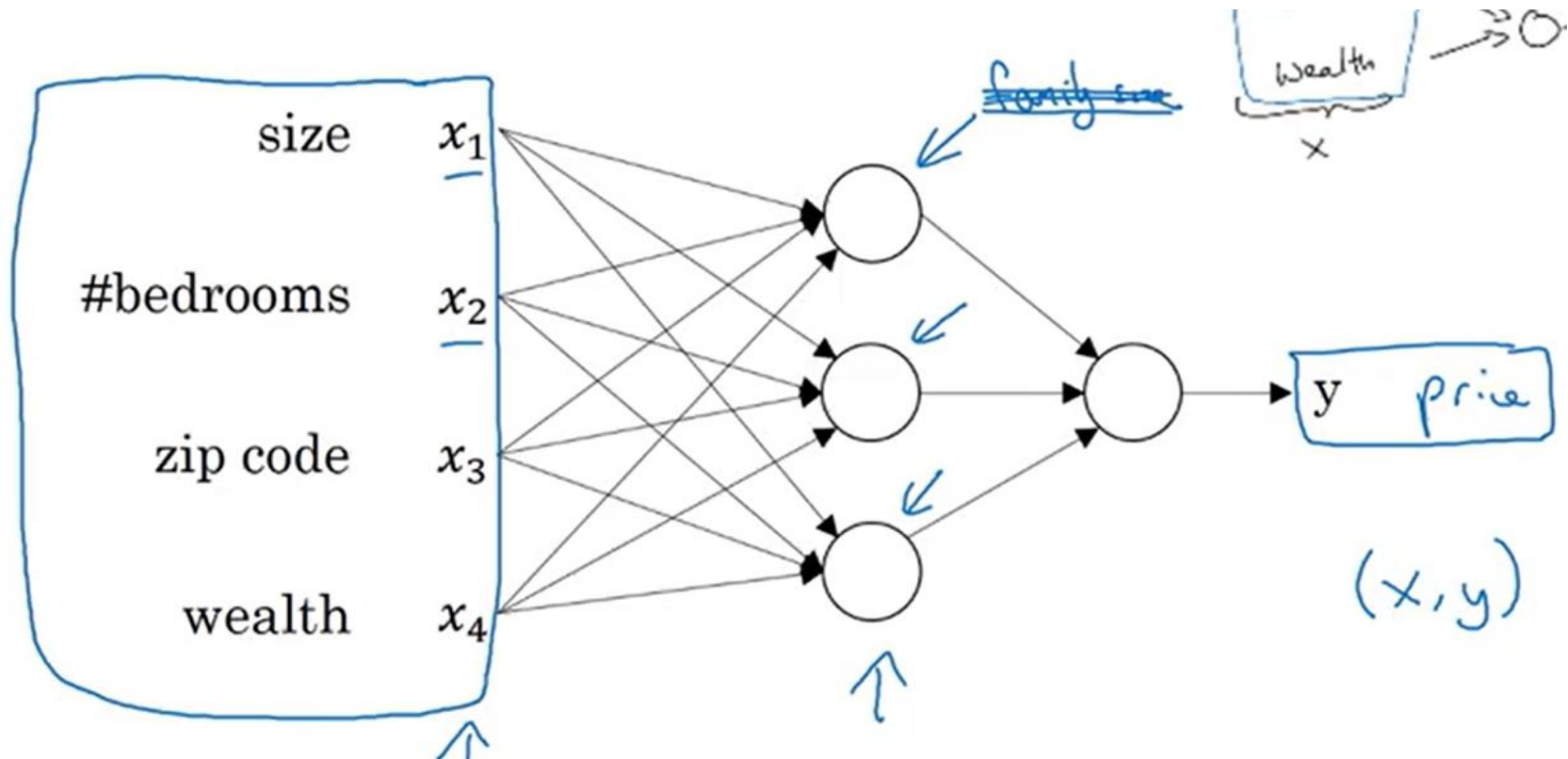
■ Housing Price Prediction

- 데이터 x : size, # bedrooms, zip code, wealth (4개)
- 예측값 y : price



■ House Price Prediction

- Neural Network는 입력 (x)로부터 예측 값 (y)를 만들어내는 함수 $f_{\theta}()$ 이자 f 를 구성하는 θ



■ Supervised Learning

Supervised Learning

Input(x) ↙	Output (y) ↙	Application
Home features	Price	Real Estate
Ad, user info ↙	Click on ad? (0/1)	Online Advertising
Image	Object (1,...,1000)	Photo tagging
<u>Audio</u>	Text transcript	Speech recognition
English	Chinese	Machine translation
Image, Radar info	Position of other cars	Autonomous driving

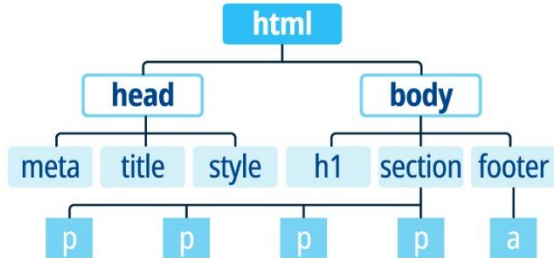
} *Standard NN*

} *CNN*

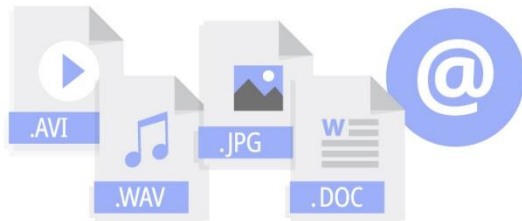
■ Structured Data, Unstructured Data

ID	Name	AGE	SEX
01	KIM	32	M
02	LEE	26	F
03	PARK	72	F
04	CHOI	15	M

structured
data



semi-
structured
data



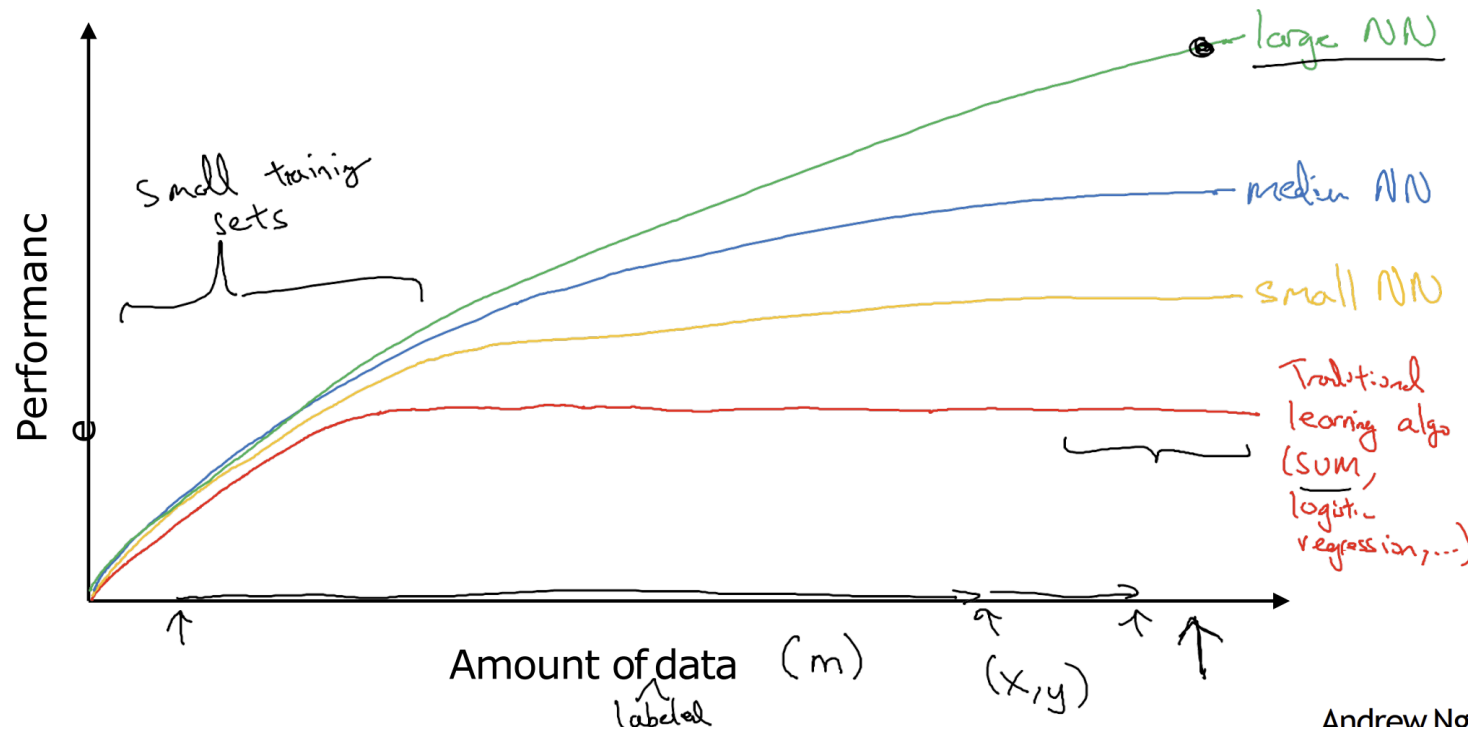
unstructured
data

딥 러닝

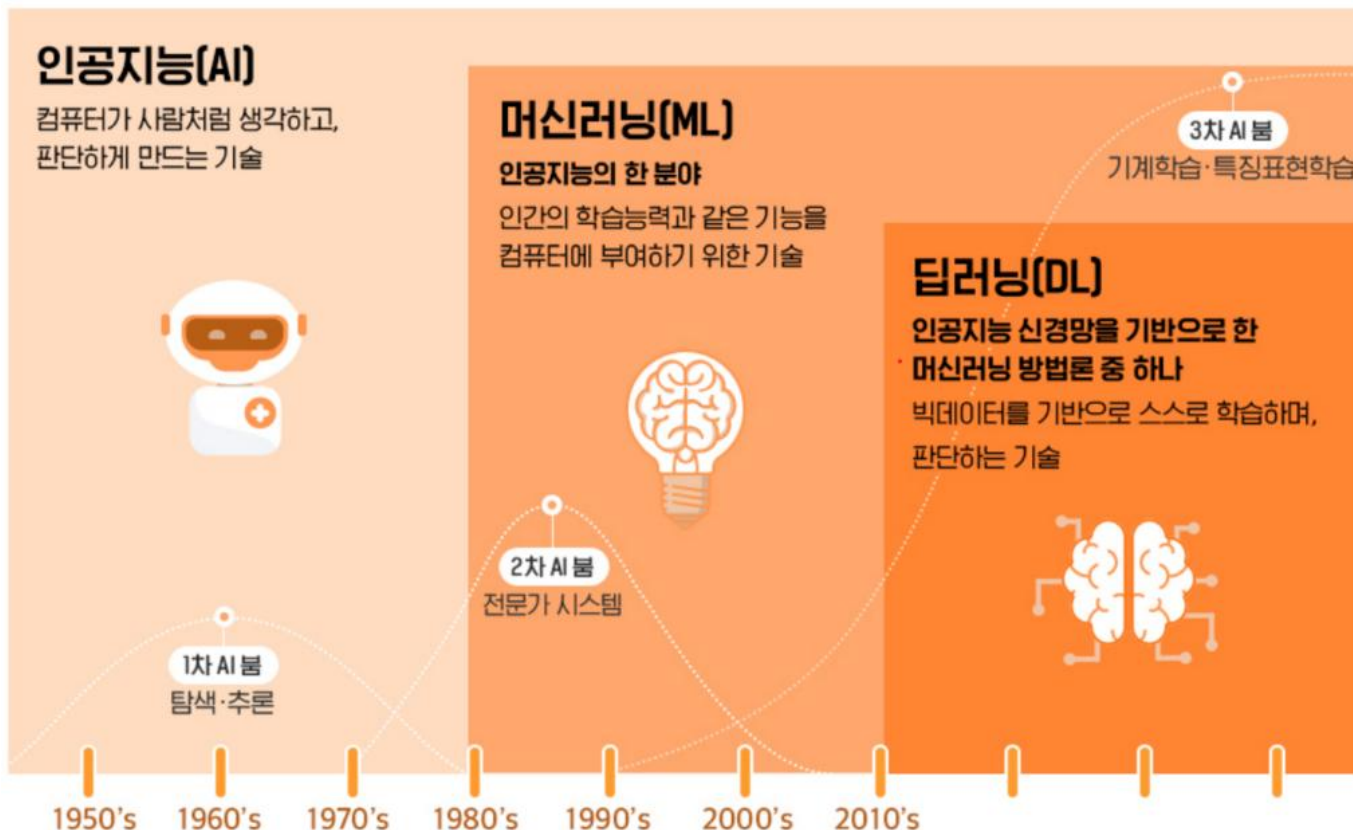
<https://aliencoder.tistory.com/99>

■ 왜 딥러닝이 떠오르나?

Scale drives deep learning progress



■ 왜 딥러닝이 떠오르나?



[출처 : https://live.lge.co.kr/live_with_ai_01/]

3차 AI 붐을 견인한 요소

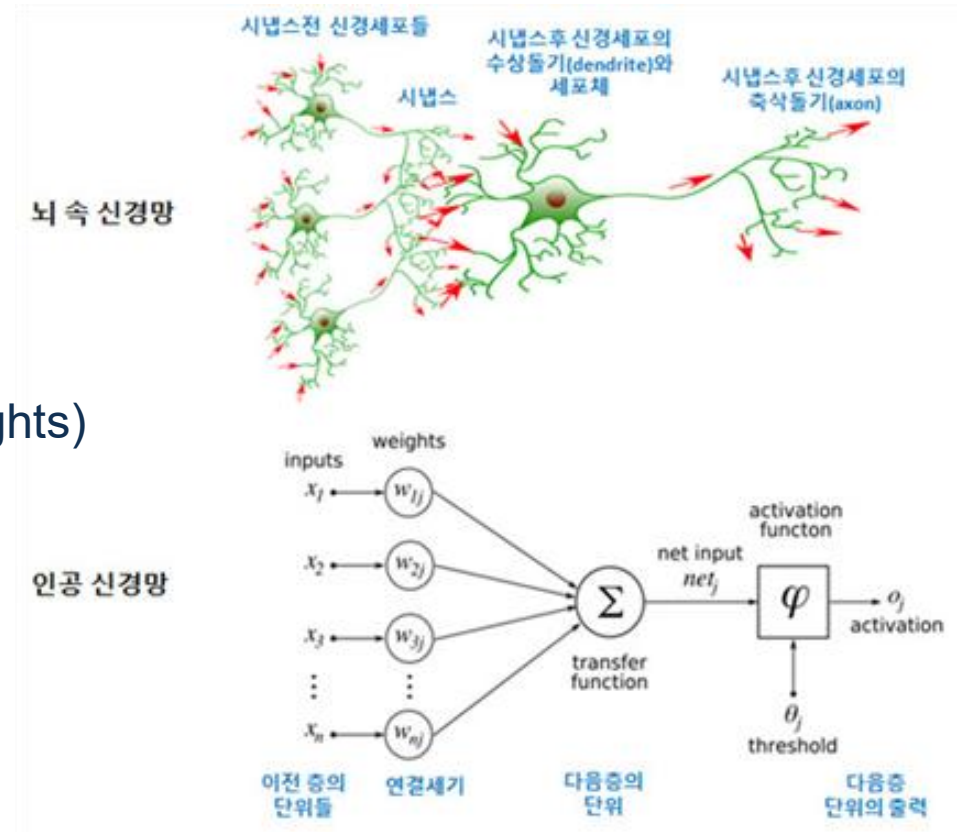


Algorithm

■ 인공신경망 (Artificial Neural Network)

➤ 인공신경망 (Artificial Neural Network)는 인간 두뇌의 신경망을 수학적으로 모델링한 것

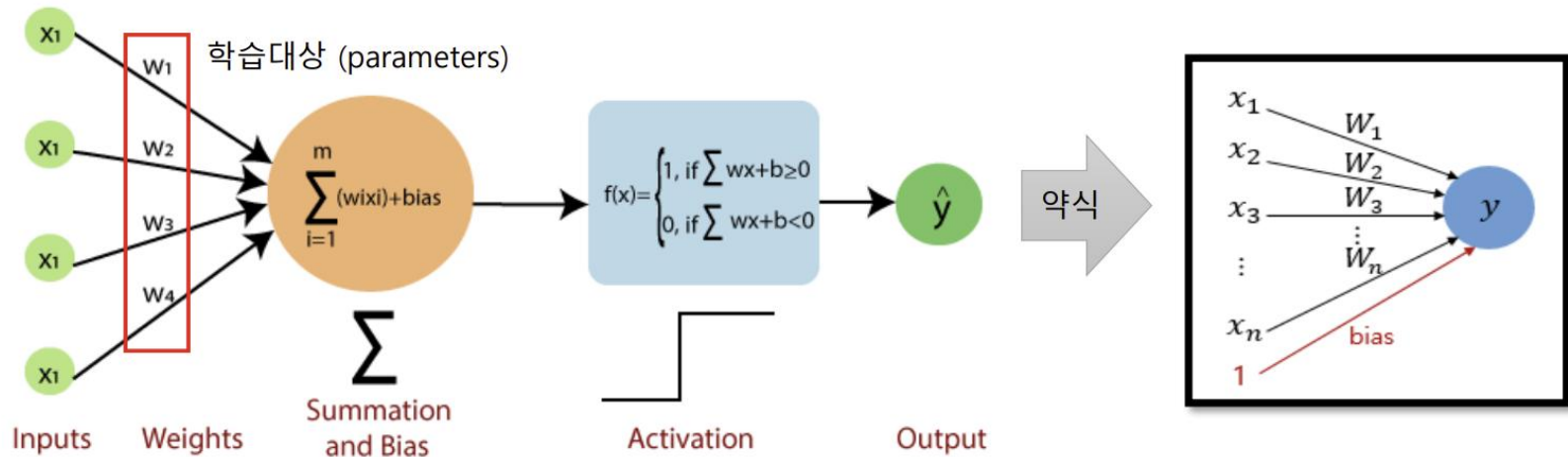
- 뇌/신경세포 (Neuron)
↔ 인공신경망/노드(Node)
- 뇌/시냅스 (Synapse)
↔ 인공신경망/연결세기(weights)
- 뇌/신호전달 (임계값)
↔ 인공신경망/활성함수(activation function)



[그림출처: <http://matrix.skku.ac.kr/math4ai-intro/W13/>]

■ 인공신경망 (Artificial Neural Network)

- 퍼셉트론(Perceptron) : 프랑크 로젠블라트(Frank Rosenblatt)가 1957년에 제안한 초기 형태의 인공 신경망



- 퍼셉트론의 활성화함수 = Step Function → 이진 분류

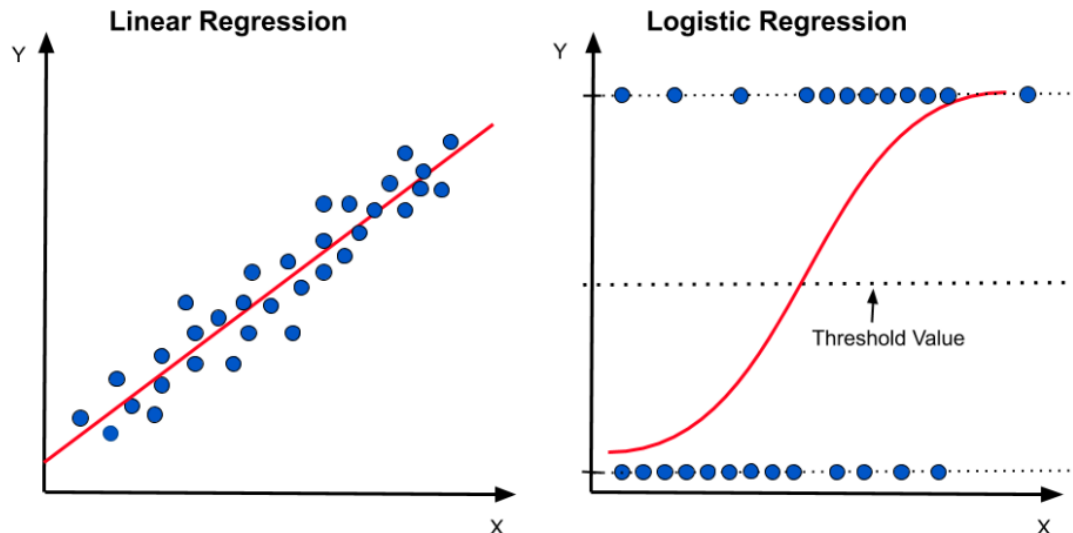
■ 인공지능망 (Artificial Neural Network)

- 선형 회귀 (Linear Regression) : 연속형 값 예측

$$\hat{y} = wx + b$$

- 로지스틱 회귀 (Logistic Regression) : 이진 분류

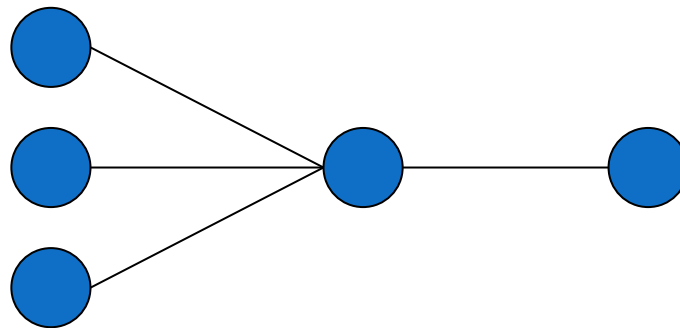
$$\hat{y} = \sigma(wx + b), \quad \sigma(z) = \frac{1}{1 + e^{-z}}$$



■ 인공신경망 (Artificial Neural Network)

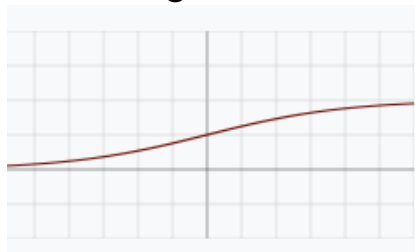
➤ 인공 신경망에서의 뉴런 (Neuron)

$$\hat{y} = f(wx + b),$$

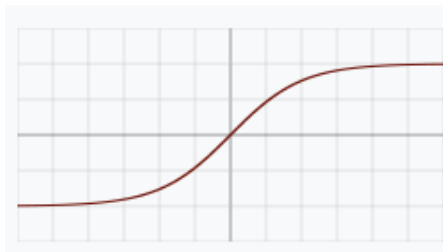


➤ 인공 신경망 뉴런(Neuron)의 다양한 활성화 함수

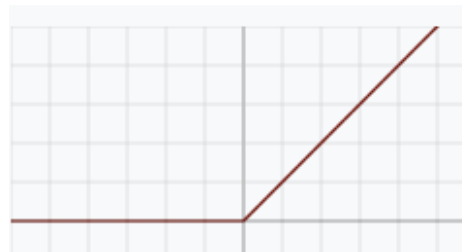
sigmoid



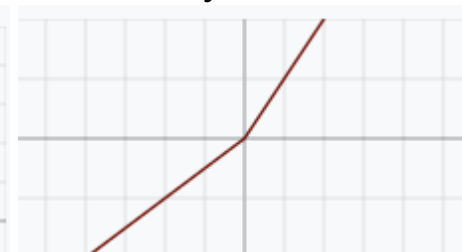
tanh



ReLU

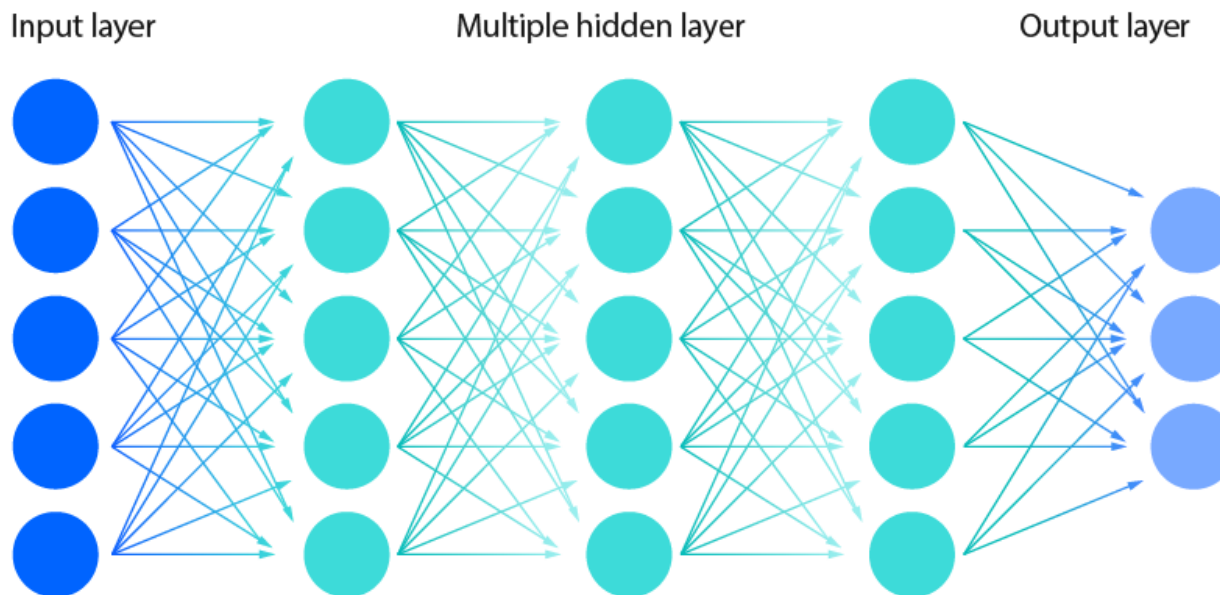


Leaky ReLU



■ 인공신경망 (Artificial Neural Network)

- 다층 퍼셉트론 (Multi-Layered Perceptron)
- 피드포워드 신경망 (Feed-forward Neural Network)

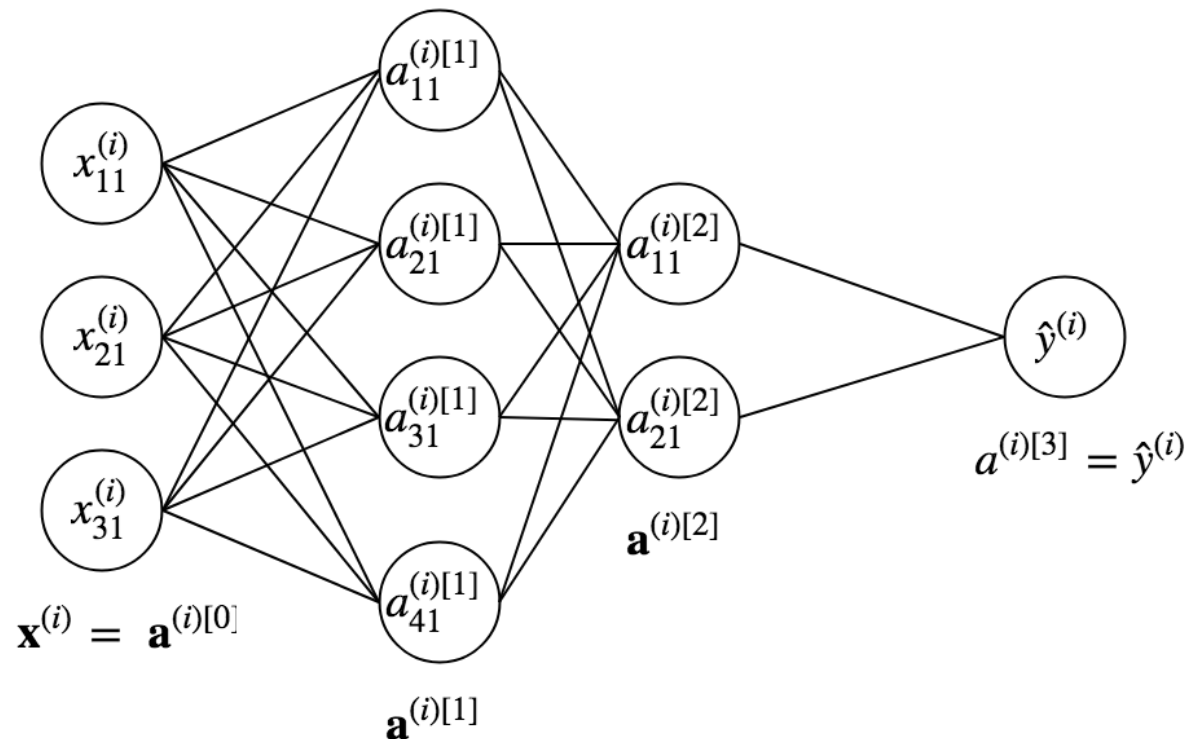


- 입력층(Input Layer) : 입력 데이터 (ex, 수치 값/벡터, 이미지, 신호 등)
- 은닉층(Hidden Layer) : 입력층과 출력층 사이에 존재하는 층
- 출력층(Output Layer) : 출력 결과 (ex, 개/고양이(분류), 수치 값(회귀) 등)

■ 인공신경망 (Artificial Neural Network)

순전파(Forward Propagation) / 추론과정

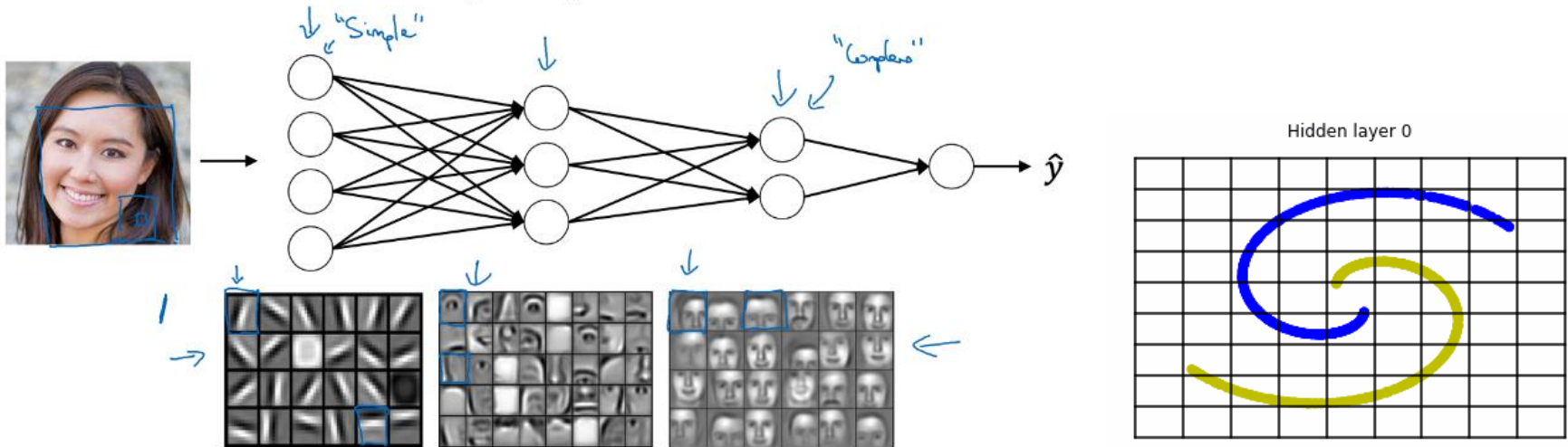
- 인공신경망의 입력층부터 출력층까지 순서대로 변수들을 계산하고 저장



■ 심층신경망

- 왜 깊게 만드나?
 - 층이 깊어질수록 인공신경망의 표현 능력이 강화됨
 - 비선형 활성화 함수를 통해 다양한 함수에 대한 표현 가능

Intuition about deep representation

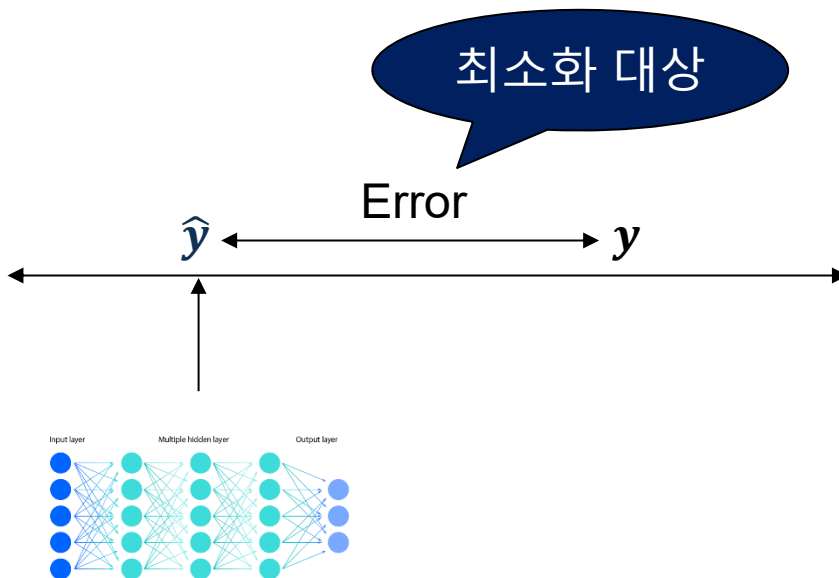


[그림출처: https://www.youtube.com/watch?v=5dWp1mw_XNk/]

[그림출처: <https://dcn.nat.fau.eu/perceptrons-neural-networks-and-dynamical-systems/>]

■ 선형 회귀 (Linear Regression)의 손실함수

- 딥러닝에서의 **학습**이란?
인공 신경망의 **예측 값(\hat{y})**과 **실제 값(y)** 간의 차이 (오류)를 최소화하는 것
- **손실 함수 (Cost/Loss Function)** 란?
차이(오류)를 정의하는 함수 (ex, Mean Squared Error, Cross Entropy, etc.)



$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

MSE = mean squared error

n = number of data points

Y_i = observed values

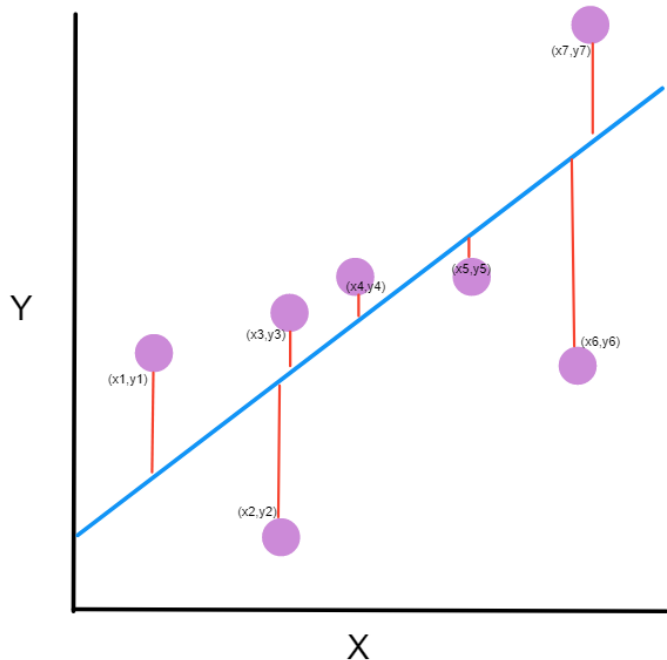
\hat{Y}_i = predicted values

$$\text{CE Loss} = - \sum_{i=1}^{\text{output size}} y_i \cdot \log \hat{y}_i$$

■ 선형 회귀 (Linear Regression)의 손실함수

➤ 손실함수 : Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$



예측한 선과 데이터 간
차이의 제곱을 손실로 정의

■ 선형 회귀 (Linear Regression)의 학습(최적화)

- 선형 회귀에서 최적화란 MSE가 최소화되는 파라미터를 찾는 것
- **Normal Equation** : 수학적으로 한번에 해 계산

$$w^* = (X^T X)^{-1} X^T y$$

→ 데이터가 많거나 $(X^T X)$ 의 역행렬이 존재하지 않으면 계산 불가

- **Gradient Descent** : 경사하강법으로 업데이트

파라미터 업데이트:

$$w \leftarrow w - \eta \frac{\partial L}{\partial w}, \quad b \leftarrow b - \eta \frac{\partial L}{\partial b}$$

기울기 계산:

$$\frac{\partial L}{\partial w} = -\frac{2}{N} \sum_{i=1}^N (y_i - \hat{y}_i) x_i, \quad \frac{\partial L}{\partial b} = -\frac{2}{N} \sum_{i=1}^N (y_i - \hat{y}_i)$$

→ 학습률 η 를 조절하면서 w, b 를 점진적으로 최적값에 수렴시킴.

■ 선형 회귀 (Linear Regression) 학습 실습

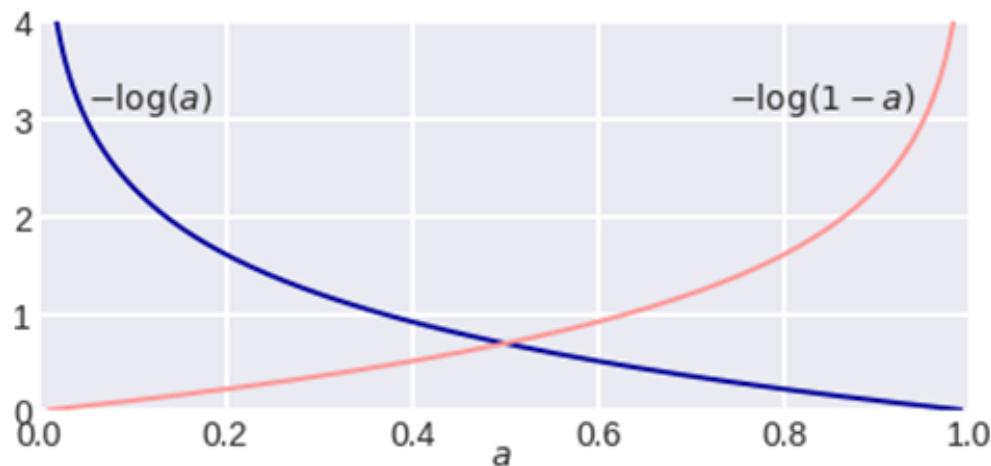
➤ [3주차]\linear_regression.ipynb

■ 로지스틱 회귀 (Logistic Regression)의 손실함수

➤ 손실함수 : Log Loss

$$L(w, b) = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]$$

$$-\begin{cases} \log a_i, & y_i = 1, \\ \log(1 - a_i), & y_i = 0. \end{cases}$$



■ 로지스틱 회귀 (Logistic Regression)의 최적화(학습)

➤ 로지스틱 회귀에서 최적화란 Log Loss가 최소화되는 파라미터를 찾는 것

➤ **Gradient Descent** : 경사하강법으로 업데이트

파라미터 업데이트:

$$w \leftarrow w - \eta \frac{\partial L}{\partial w}, \quad b \leftarrow b - \eta \frac{\partial L}{\partial b}$$

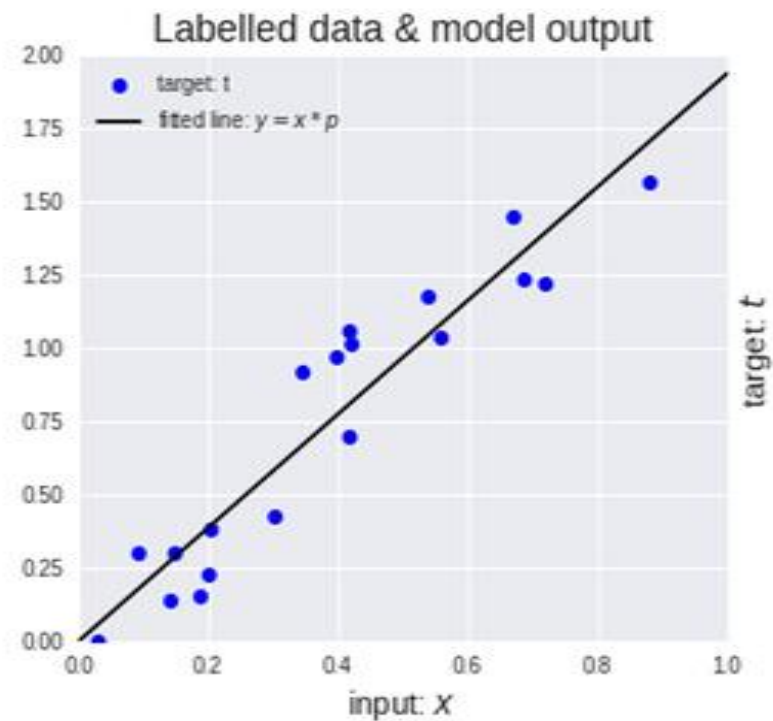
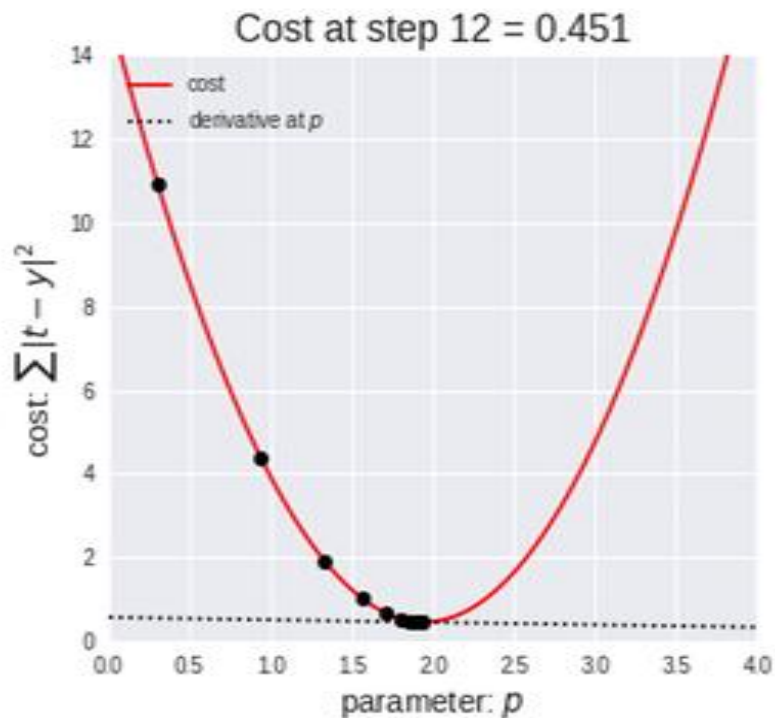
기울기 계산:

$$\frac{\partial L}{\partial w} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)x, \quad \frac{\partial L}{\partial b} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)$$

→기울기에 따라 w와 b를 조절하면서 점진적으로 최적 값에 수렴시킴

■ 로지스틱 회귀 (Logistic Regression)의 최적화(학습)

- 경사 하강법 (Gradient Descent) 은?
손실 함수의 값이 낮아지는 방향으로 파라미터들의 값을 조절하는 학습 방법



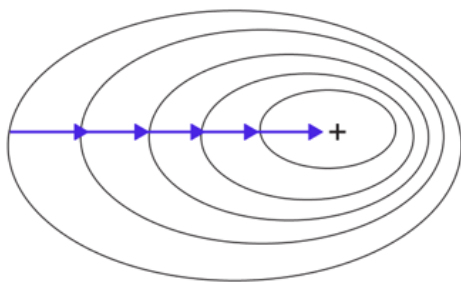
[그림 출처: <https://www.mql5.com/en/articles/11200/>]

■ 로지스틱 회귀 (Logistic Regression)의 최적화(학습)

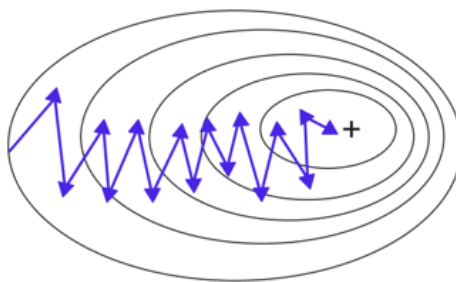
➤ 경사 하강법 알고리즘의 종류 (Batch, Stochastic, Mini-batch)

Batch Gradient Descent	Stochastic Gradient Descent	Mini-batch Gradient Descent
<ul style="list-style-type: none"> 업데이트에 전체 데이터셋 활용 Cost Function 이 매우 Smooth하게 줄어듬 요구되는 컴퓨팅 리소스가 매우 큼 (메모리, 계산량 등) 	<ul style="list-style-type: none"> 업데이트에 하나의 데이터만 활용 Cost Function 에 매우 큰 Variation 존재 학습에 걸리는 시간이 김 	<ul style="list-style-type: none"> 업데이트에 데이터셋의 서브셋 활용 SGD 보다 Smooth한 Cost Function 컴퓨팅 리소스와 학습 시간은 중간

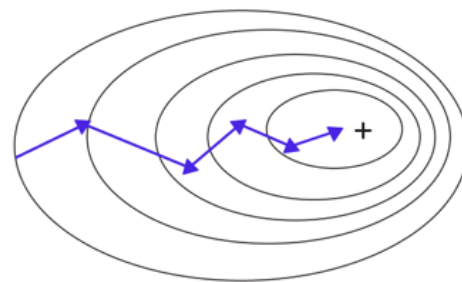
Batch Gradient Descent



Stochastic Gradient Descent



Mini-Batch Gradient Descent

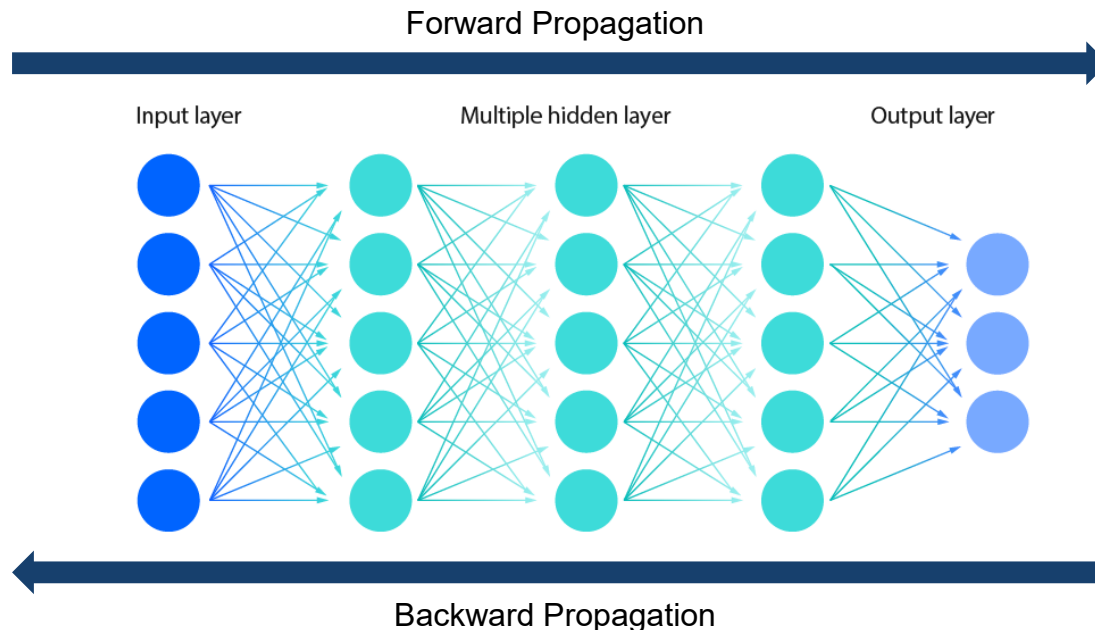


■ 로지스틱 회귀 (Logistic Regression) 학습 실습

➤ [3주차]\logistic_regression.ipynb

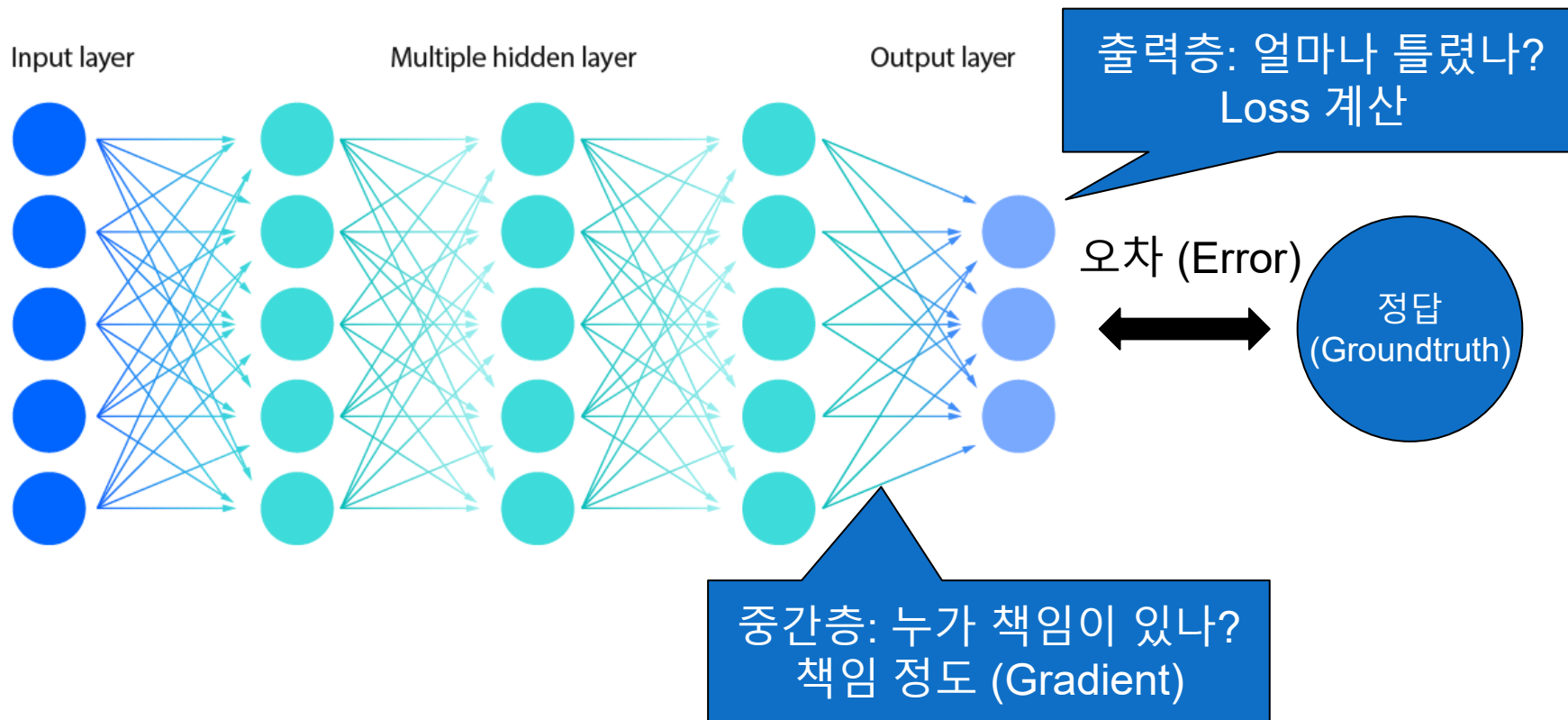
■ 신경망 (Neural Network)의 최적화(학습)

- 순전파(Forward Propagation) / 추론과정
 - 인공신경망의 입력층부터 출력층까지 순서대로 변수들을 계산하고 저장
- 역전파(Backward Propagation) / 학습과정
 - 목적 함수(Objective Function)에 대한 파라미터들의 기울기(Gradient)를 구하는 과정



■ 신경망 (Neural Network)의 최적화(학습)

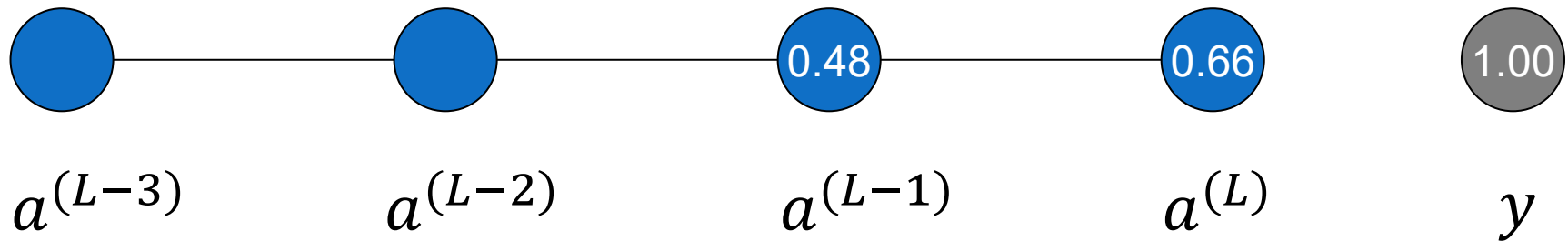
➤ 오류 역전파 (Backpropagation)



■ 신경망 (Neural Network)의 최적화(학습)

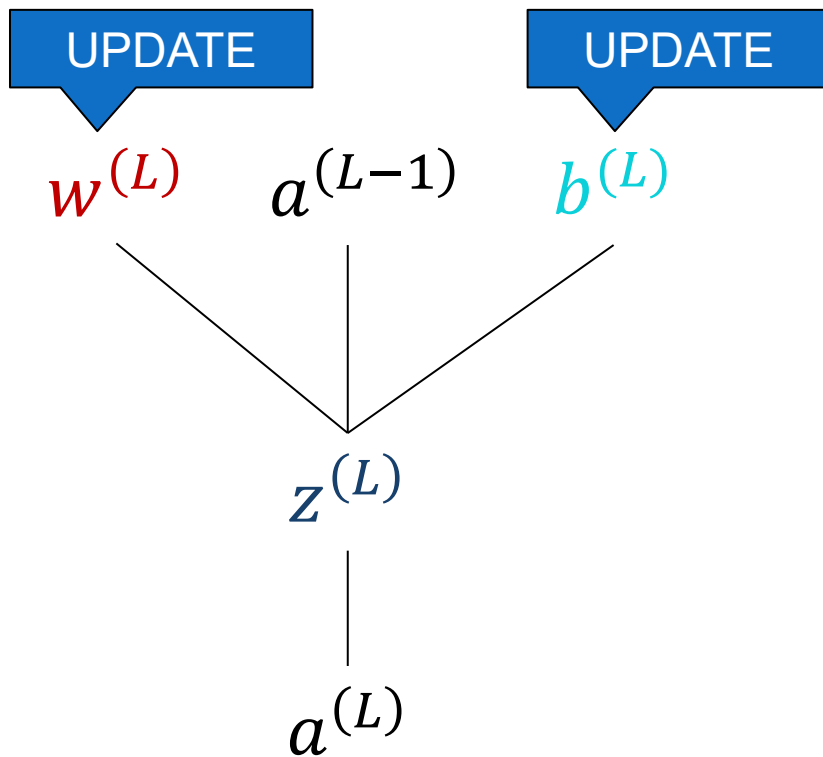
➤ 오류 역전파 (Backpropagation) Simple Example

$$C_0 = (a^{(L)} - y)^2$$



■ 신경망 (Neural Network)의 최적화(학습)

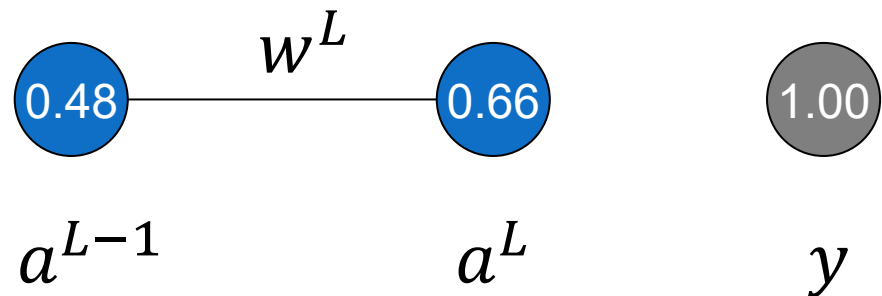
➤ 오류 역전파 (Backpropagation) Simple Example



$$C_0 = (a^{(L)} - y)^2$$

$$z^{(L)} = w^{(L)} a^{(L-1)} + b^{(L)}$$

$$a^{(L)} = \sigma(z^{(L)})$$



■ 신경망 (Neural Network)의 최적화(학습)

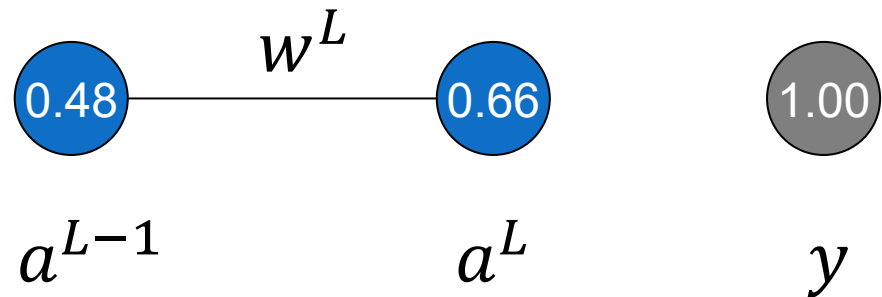
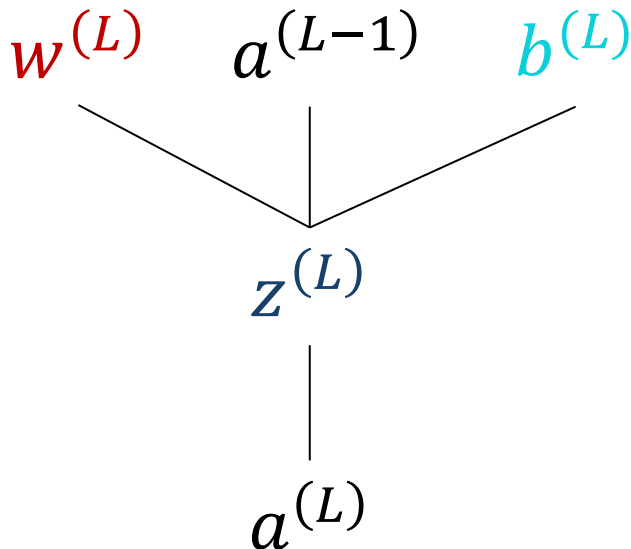
- 오류 역전파 (Backpropagation) Simple Example
- Chain Rule

$$\frac{\partial C_0}{\partial w^{(L)}} = \frac{\partial z^{(L)}}{\partial w^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}} \frac{\partial C_0}{\partial a^{(L)}}$$

$$C_0 = (a^{(L)} - y)^2$$

$$z^{(L)} = w^{(L)} a^{(L-1)} + b^{(L)}$$

$$a^{(L)} = \sigma(z^{(L)})$$



■ 신경망 (Neural Network)의 최적화(학습)

- 오류 역전파 (Backpropagation) Simple Example
- Chain Rule

$$\frac{\partial C_0}{\partial w^{(L)}} = \frac{\partial z^{(L)}}{\partial w^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}} \frac{\partial C_0}{\partial a^{(L)}}$$

$$C_0 = (a^{(L)} - y)^2$$

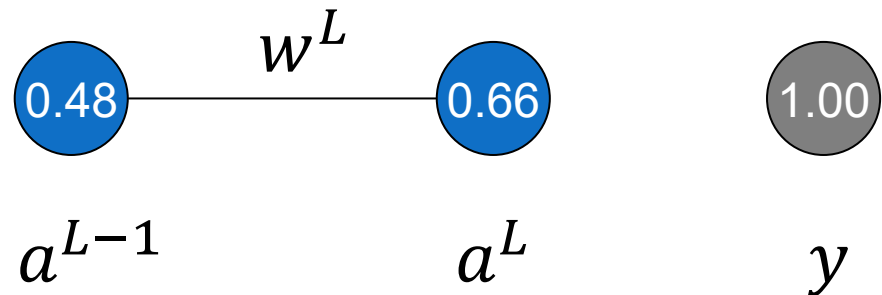
$$z^{(L)} = w^{(L)} a^{(L-1)} + b^{(L)}$$

$$\frac{\partial C_0}{\partial a^{(L)}} = 2(a^{(L)} - y)$$

$$a^{(L)} = \sigma(z^{(L)})$$

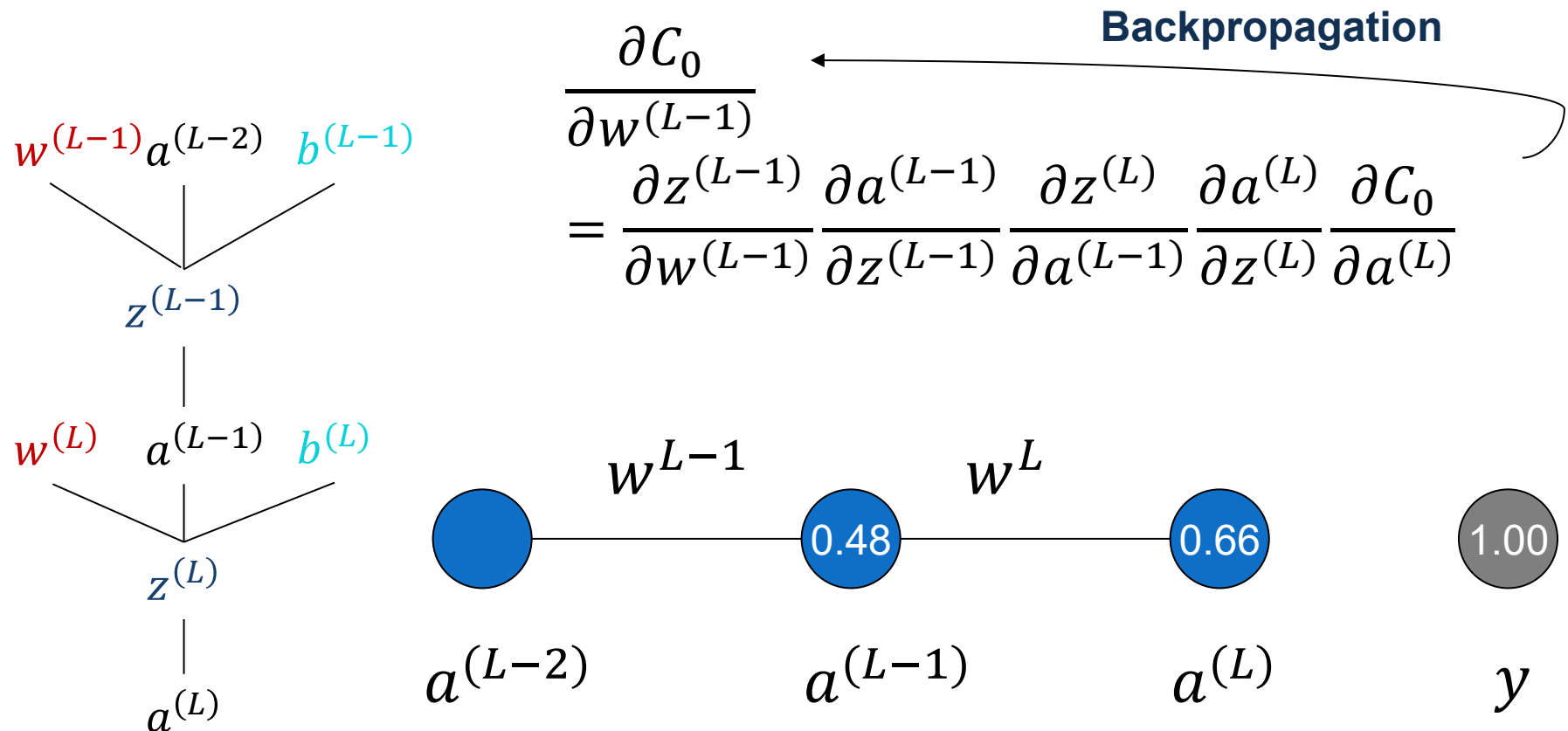
$$\frac{\partial a^{(L)}}{\partial z^{(L)}} = \sigma'(z^{(L)})$$

$$\frac{\partial z^{(L)}}{\partial w^{(L)}} = a^{(L-1)}$$



■ 신경망 (Neural Network)의 최적화(학습)

- 오류 역전파 (Backpropagation) Simple Example
- Chain Rule

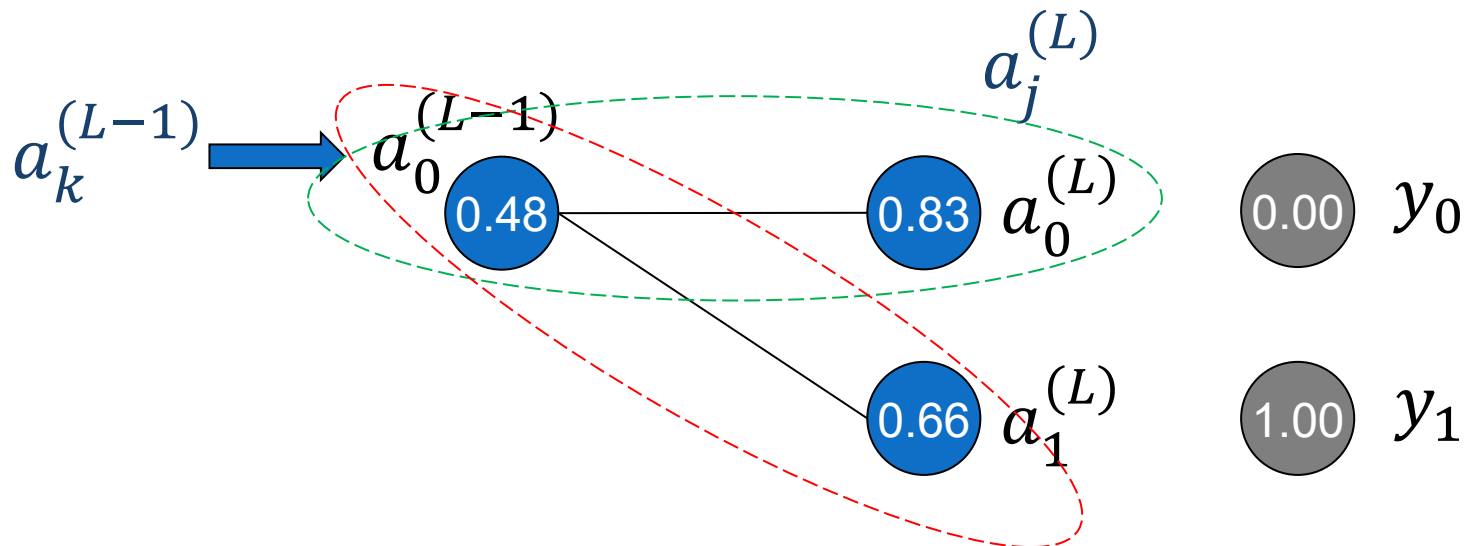


■ 신경망 (Neural Network)의 최적화(학습)

➤ 오류 역전파 (Backpropagation) Complex Example

$$\frac{\partial C_0}{\partial a_k^{(L-1)}} = \sum_{j=0}^{L-1} \frac{\partial z_j^{(L)}}{\partial a_k^{(L-1)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial C_0}{\partial a_j^{(L)}}$$

$$C_0 = \sum_{j=0}^{L-1} \left(a_j^{(L)} - y_j \right)^2$$

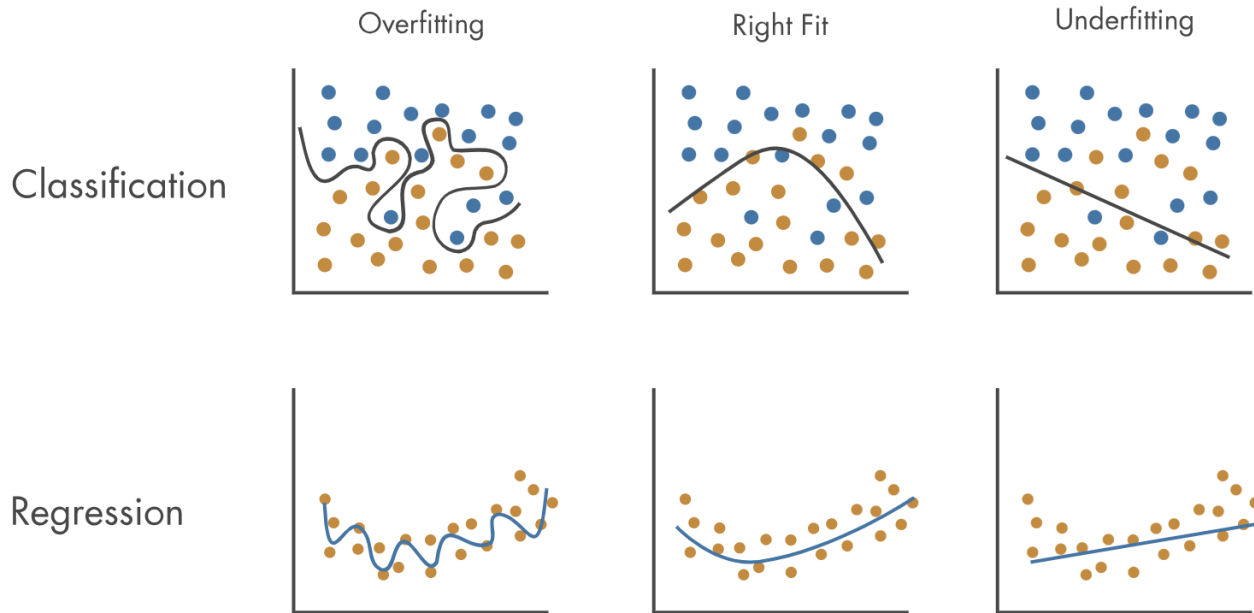


■ 신경망 (Neural Network) 학습 실습

➤ [3주차]\neural_network.ipynb

■ 신경망에서의 과적합 문제

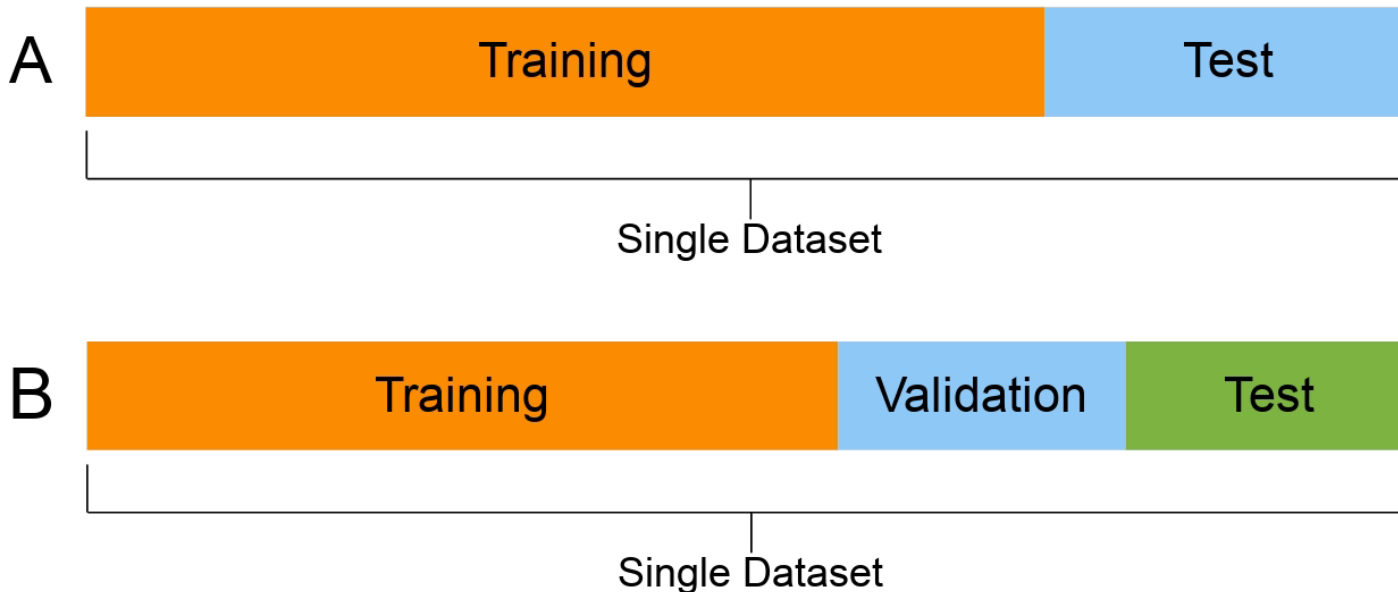
- 딥러닝에서 학습이란 $y = f(x)$ 를 만족하는 $f()$ 를 찾는 것
- 이때, 학습된 모델 $f()$ 가 새로운 데이터(학습데이터x)로 일반화되지 않는 경우 모델이 과적합되었다고 함



- 과적합 문제가 발생하는 이유1: 학습 데이터의 다양성 부족
- 과적합 문제가 발생하는 이유2: 모델의 복잡도 (파라미터의 양)

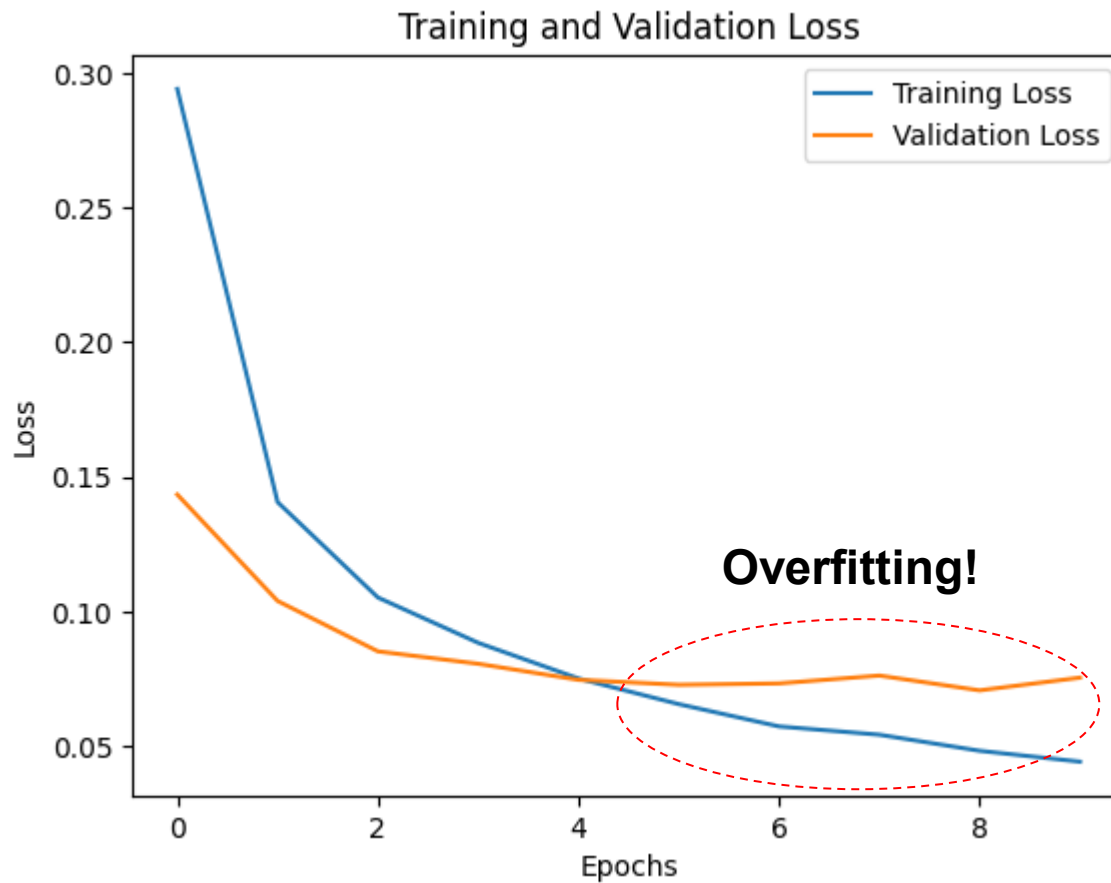
■ 신경망에서의 과적합 문제

- Training Data : 모델의 학습에 사용
- Validation Data : 하이퍼파라미터 튜닝 등 좋은 모델을 산출하는데 사용
- Test Data : 모델의 최종 성능 평가에 활용



■ 신경망에서의 과적합 문제

➤ Training / Validation Loss Graph



■ 차후 수업 진행 안내

주차	수업내용	수업방식
1	Lecture Introduction	대면수업
2	Introduction to Pattern Recognition	대면수업
3	Basic of Neural Network	대면수업
4	Learning Strategies: Supervised Learning	대면수업
5	Learning Strategies: Unsupervised Learning	대면수업
6	Model Architecture: Convolutional Neural Networks	대면수업
7	Model Architecture: Sequence Models	대면수업
8	Midterm exam	대면수업
9	Pattern Recognition Application: Visual Anomaly Detection	대면수업
10	Pattern Recognition Application: Temporal Anomaly Detection	대면수업
11	Pattern Recognition Application: Predictive Maintenance with Time Series	대면수업
12	Advanced Pattern Recognition	대면수업
13	Final Project: Proposal	대면수업
14	Final Project: Discussion	대면수업
15	Final Project: Presentation	대면수업

감사합니다

Q&A

