



인공신경망과딥러닝입문

Lecture 05. 머신러닝

동덕여자대학교
데이터사이언스 전공
권 범

목차

❖ 01. 머신러닝 개념



❖ 02. 머신러닝 학습 방법

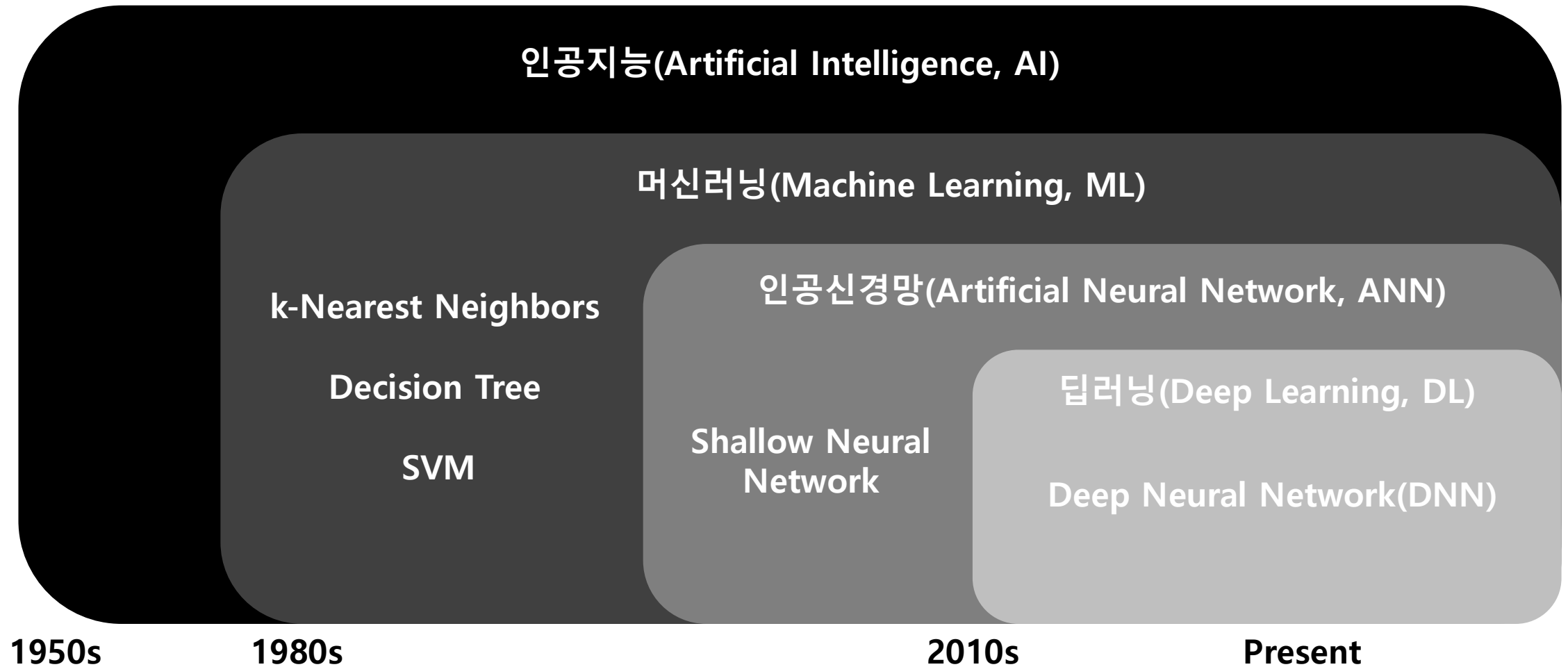
- ① 지도 학습 (x, y)
- ② 비지도 학습 (x)
- ③ 강화 학습
- ④ 준지도 학습 (x, y) (x)

01. 머신러닝 개념

02. 머신러닝 학습 방법

01. 머신러닝 개념

❖ 인공지능, 머신러닝, 딥러닝의 관계



01. 머신러닝 개념

❖ 머신러닝의 정의

- 머신러닝(Machine Learning, ML)이라는 용어는 1959년에 아서 사무엘이 학술지 <IBM Journal of Research and Development>에 기고한 논문에서 처음 사용



“ 머신러닝이란? ”

~~컴퓨터가 명시적으로 프로그램되지 않고도
학습할 수 있도록 하는 연구 분야를 말합니다.~~

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.

아서 사무엘(Arthur Lee Samuel, 컴퓨터 과학자)

01. 머신러닝 개념

❖ 일반적인 프로그래밍 방식과 머신러닝 방식 (1/3)

- 일반적인 프로그래밍 방식

- ◆ 입력값에 따라 원하는 결과값이 출력되도록 사람이 내부 동작을 작성함

- 머신러닝 방식

- ◆ 사람이 컴퓨터에게 입력값과 결과값만 충분히 전달해 주면

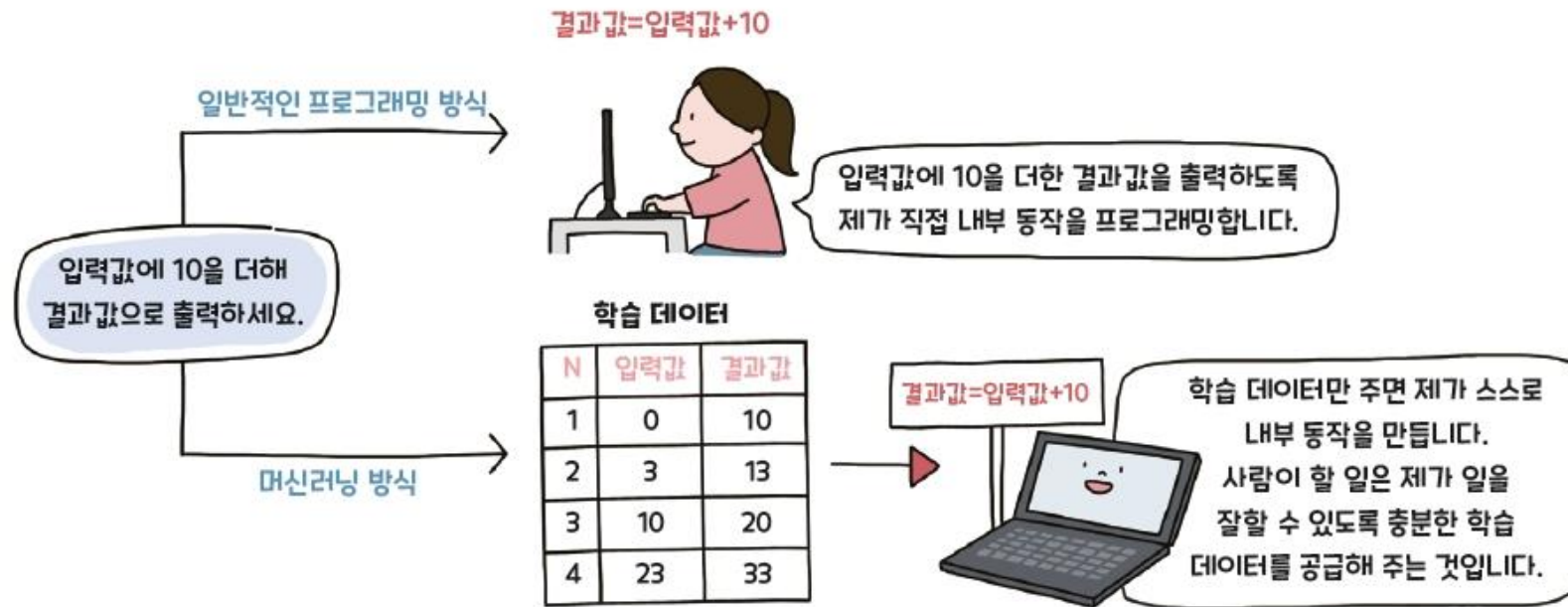
- 컴퓨터가 스스로 입력값과 결과값의 관계를 만족시키는 내부 동작을 찾아냄

01. 머신러닝 개념

❖ 일반적인 프로그래밍 방식과 머신러닝 방식 (2/3)

- 머신러닝은 여러 개의 입력값과 결과값을 컴퓨터에 제공하기만 하면 이 데이터를 바탕으로 컴퓨터가 스스로 내부 동작을 만들어 냄
- 이를 위해서 사람은 양질의 많은 학습 데이터를 공급해야 함

일반적인 프로그래밍 방식과 머신러닝 방식 비교



01. 머신러닝 개념

❖ 일반적인 프로그래밍 방식과 머신러닝 방식 (3/3)

- 인공지능 시대의 소프트웨어 역량

- ◆ 일반적인 프로그래밍 방식

- 우수한 개발자란 논리적인 사고를 통해 프로그래밍 언어를 자유자재로 사용할 수 있는 사람

- ◆ 머신러닝 방식

- 우수한 개발자란 데이터에 대한 통찰력과 수집 능력이 우수하여

- 머신러닝 기계에 양질의 학습 데이터를 충분히 공급할 수 있는 사람

**앞으로는 쓸모 있는 데이터를 누가 더 많이 보유하고
있느냐에 따라 소프트웨어 역량이 결정될 것**

01. 머신러닝 개념

❖ 머신러닝은 언제 주로 사용될까? (1/2)

- 명시적으로 알고리즘을 설계하고 프로그래밍 하는 것이 어렵거나 불가능한 경우에 주로 사용됨
 - ◆ 문제에서 나타날 수 있는 경우의 수가 너무 많은 경우
 - ◆ 규칙 기반 프로그램으로 답을 내기가 어려운 경우

예를 들어

- ① 바둑 경기에서 모든 경우의 수를 찾아서
if-else와 같은 문장으로 처리하는 것은 거의 불가능함
- ② 스팸 메일을 자동으로 걸러내는 작업에도
많은 경우가 있기 때문에 프로그래밍하는 것은 거의 불가능함

01. 머신러닝 개념

❖ 머신러닝은 언제 주로 사용될까? (2/2)

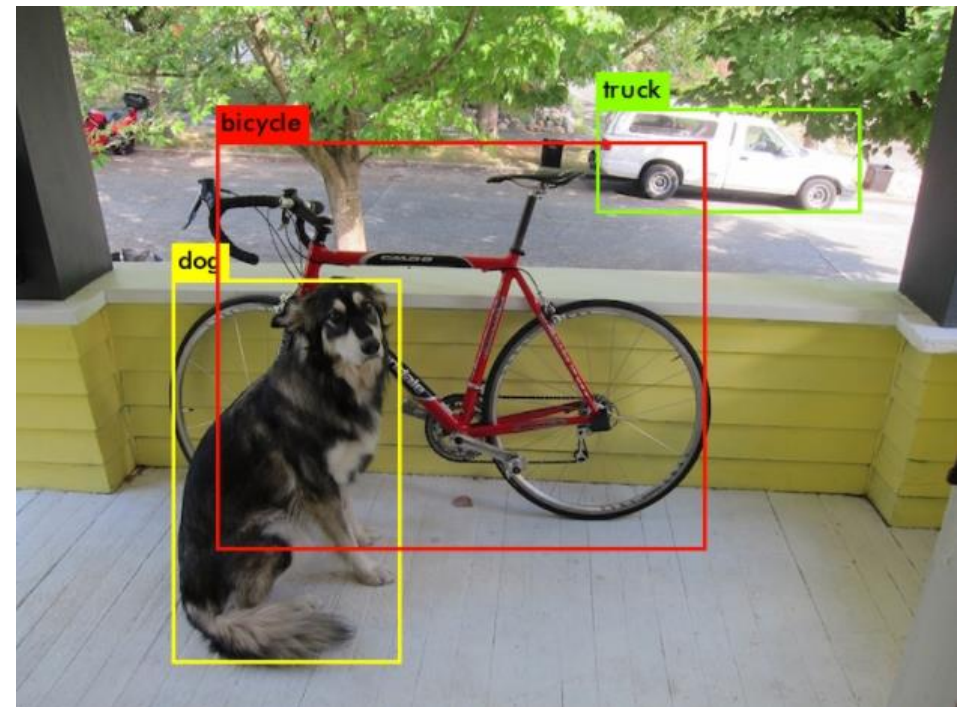
- 머신러닝의 응용 분야

- ◆ 주로, 복잡한 데이터들이 있고 이 데이터들에 기반하여 결정을 내려야 하는 분야
- ◆ 머신러닝 모델을 학습(Learning) 시키려면 많은 데이터가 필요하기 때문에 머신러닝은 빅데이터(Big Data)와 아주 밀접한 관계가 있음

예를 들어

- ① 객체 검출(Object Detection)
- ② 음성 인식(Voice Recognition)
- ③ 글자 인식(Character Recognition)

[영상처리 (컴퓨터 비전)
자연어처리(NLP) ex) Chat GPT]



01. 머신러닝 개념

❖ 머신러닝 모델의 생성 과정 (1/2)

- 머신러닝으로 문제를 해결하기 위해서는 문제 해결에 적합한 머신러닝 모델을 생성해야 함
 - ① 머신러닝 모델 가정
 - 문제 속의 데이터를 잘 설명할 수 있는 머신러닝 모델을 가정
 - ② 머신러닝 모델 학습
 - 모델로부터 학습 데이터에 최적화된 구체적인 함수를 찾음
 - 이 과정을 학습(Learning)이라고 함
 - ③ 머신러닝 모델 적용
 - 학습된 모델을 실제 문제에 적용

01. 머신러닝 개념

❖ 머신러닝 모델의 생성 과정 (2/2)

- 예를 들어, 시험 공부 시간으로 시험 성적을 예상하는 문제를 머신러닝으로 해결하고자 할 때를 가정해 보자

머신러닝
모델 가정

공부 시간으로 성적을 예상할 수 있을까?

공부를 많이 하면 성적도 오를 것
→ 직선과 같은 모델을 가질 것이라 가정

머신러닝
모델 학습

어떤 직선이 공부 시간과 성적 간의 관계를 가장 잘 나타낼까?

친구들의 공부 시간과 점수를 수집
→ 수집한 데이터를 학습 데이터로 사용하여 가장 적합한 직선 찾기

머신러닝
모델 적용

공부 시간을 입력하여 점수를 예측

01. 머신러닝 개념

❖ 사람의 키를 입력했을 때 몸무게를 추측하는 머신러닝의 구현 과정 예제 (1/4)

① 머신러닝 모델 가정

- ◆ 일반적으로 키가 커지면 몸무게도 늘어날 것이라고 가정할 수 있음
- ◆ 따라서 "몸무게= $a \times \text{키} + b$ "와 같은 직선의 방정식을 머신러닝 모델로 가정함

		x	y
No.	이름	키(cm)	몸무게(kg)
1	김*성	100	30
2	박*인	120	40
3	윤*안	130	45
4	최*연	160	60
5	문*승	190	75

[머신러닝 구현 과정 예제의 학습 데이터]

$$\begin{aligned}\text{몸무게} &= a \times \text{키} + b \\ y &= a \times x + b\end{aligned}$$

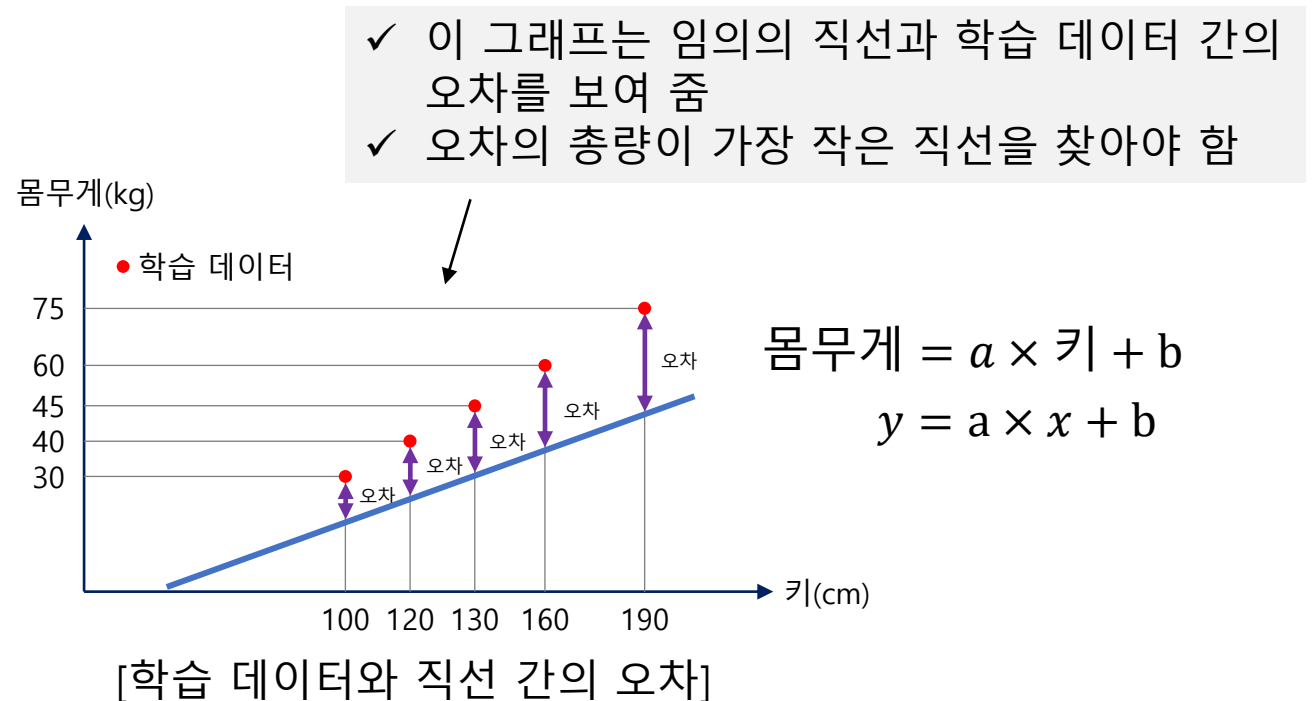
01. 머신러닝 개념

❖ 사람의 키를 입력했을 때 몸무게를 추측하는 머신러닝의 구현 과정 예제 (2/4)

② 머신러닝 모델 학습

- ◆ 가정한 직선의 방정식은 아직 기울기 a 와 y -절편 b 의 값이 결정되지 않은 상태로, 확보한 학습 데이터를 활용하여 최적화된 직선을 구함
- ◆ 최적화된 직선을 구한다는 것은 학습 데이터와 오차가 가장 적은 기울기 a 와 y -절편 b 를 구한다는 의미

No.	이름	x		y	
		키(cm)	몸무게(kg)	키(cm)	몸무게(kg)
1	김*성	100	30		
2	박*인	120	40		
3	윤*안	130	45		
4	최*연	160	60		
5	문*승	190	75		



01. 머신러닝 개념

❖ 사람의 키를 입력했을 때 몸무게를 추측하는 머신러닝의 구현 과정 예제 (3/4)

② 머신러닝 모델 학습

◆ 학습 결과, 직선의 방정식 "몸무게=(0.5×키)-20"이 학습 데이터에 최적화된 함수임

$$\begin{aligned}\text{몸무게} &= a \times \text{키} + b \\ y &= a \times x + b\end{aligned}$$

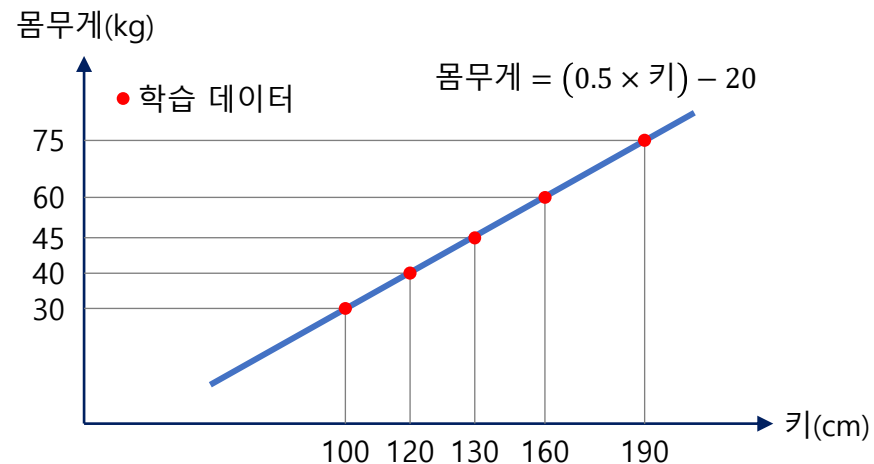


$$\begin{aligned}\text{최적화(Optimization)} \\ a = 0.5, b = -20\end{aligned}$$



$$y = 0.5 \times x - 20$$

		x	y
No.	이름	키(cm)	몸무게(kg)
1	김*성	100	30
2	박*인	120	40
3	윤*안	130	45
4	최*연	160	60
5	문*승	190	75



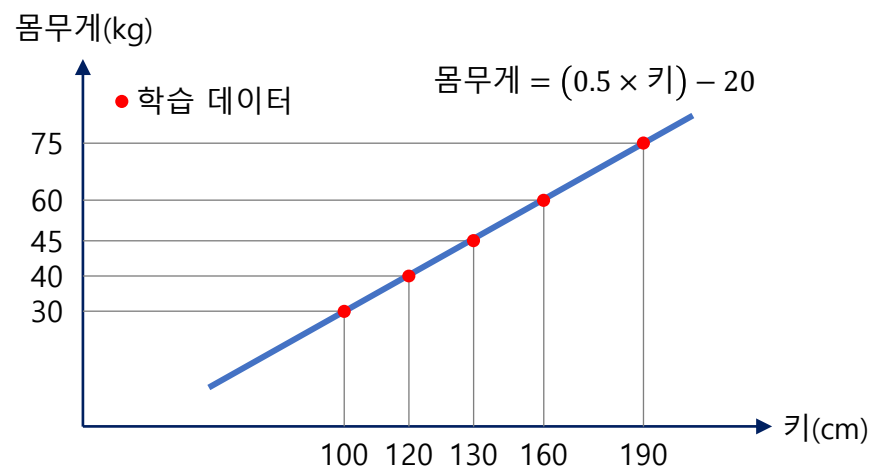
[몸무게와 키의 상관관계에 최적화된 직선의 방정식]

01. 머신러닝 개념

❖ 사람의 키를 입력했을 때 몸무게를 추측하는 머신러닝의 구현 과정 예제 (3/4)

③ 머신러닝 모델 적용

◆ 학습이 완료된 머신러닝 모델을 실제 문제에 적용하여 예측 결과를 확인함



[몸무게와 키의 상관관계에 최적화된 직선의 방정식]

Q) 키가 180cm인 사람의 몸무게는 몇 kg일까요?

$$y = 0.5 \times x - 20$$



$$x = 180 \text{ 대입}$$



$$y = 0.5 \times 180 - 20$$

$$= 90 - 20$$

$$= 70$$

머신러닝 모델의 예측 결과: 70kg

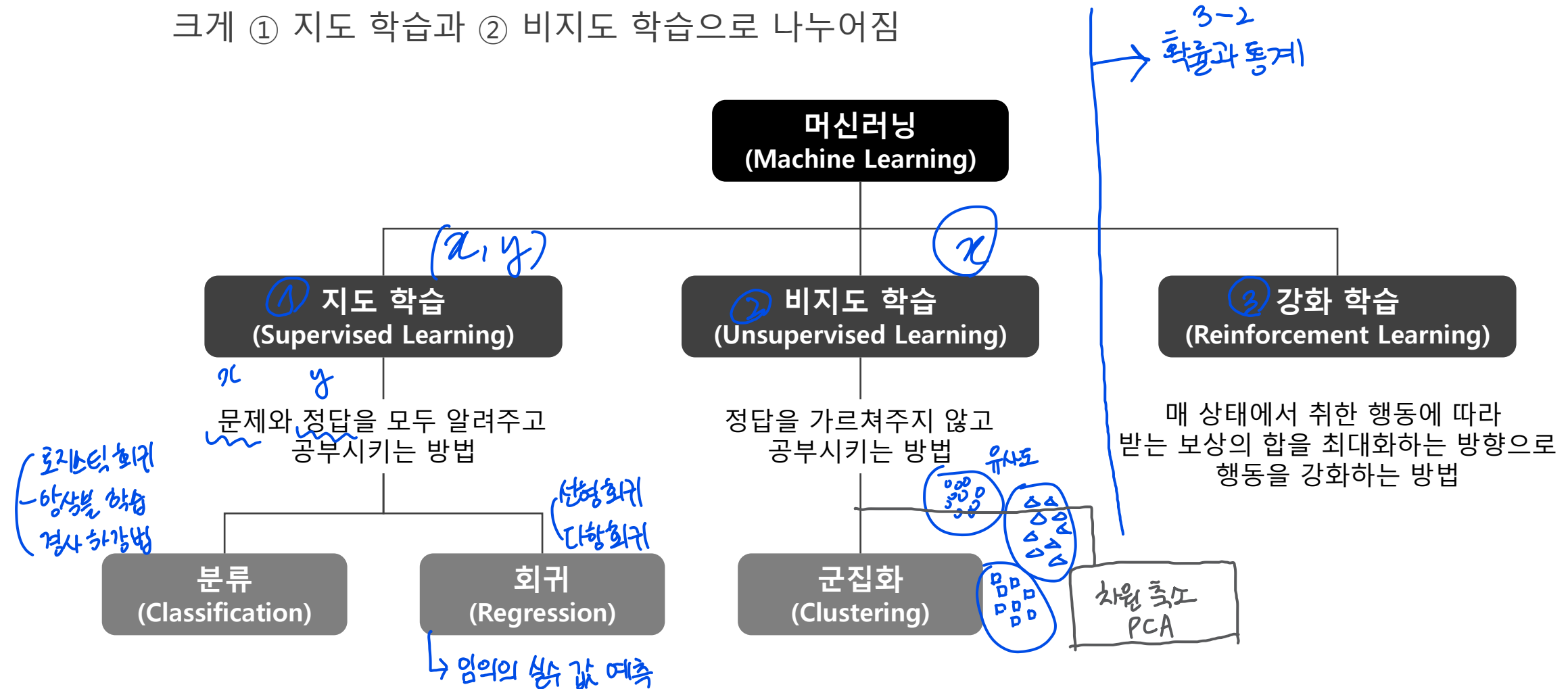
02. 머신러닝 학습 방법

01. 머신러닝 개념

02. 머신러닝 학습 방법

❖ 머신러닝의 학습 방법

- 머신러닝 모델을 학습시킴에 있어서, 입력값에 대한 결괏값을 알고 있는지 여부에 따라 크게 ① 지도 학습과 ② 비지도 학습으로 나누어짐



02. 머신러닝 학습 방법

❖ 지도 학습 (1/7)

- 문제에 대한 정답을 알고 있는 학습 데이터를 활용하여 머신러닝 모델을 학습시키는 방식
- 문제를 입력값, 입력 데이터라고 함
- 정답을 결괏값, 목표값, 또는 레이블(Label)이라고 함
- 지도 학습은 결괏값에 따라 2가지 유형으로 나눌 수 있음
 - ① 분류(Classification)
 - ② 회귀(Regression)

02. 머신러닝 학습 방법

❖ 지도 학습 (2/7)

- 분류(Classification)

- ◆ 어떤 입력 데이터가 들어오더라도 학습에 사용한 레이블 중에 하나로 결괏값을 결정
- ◆ 레이블이 **이산**적인(Discrete) 경우, 즉 0, 1, 2, 3, ...과 같이 **유한**한 경우

예를 들어

- ✓ 사진이 주어졌을 때, 고양이(Label: 0) 사진인지 또는 강아지(Label: 1) 사진인지 분류
- ✓ 필기체로 쓴 숫자가 주어졌을 때, 숫자 인식(0~9까지의 레이블 중에 하나로 결정)

02. 머신러닝 학습 방법

❖ 지도 학습 (3/7)

- 분류의 종류

이진 분류(Binary Classification)

- 레이블이 2개의 그룹(Class)으로만 구성

0, 1

Label: 고양이



Label: 강아지



'고양이', '강아지' 그룹 중에 하나로 분류

다중 분류(Multiclass Classification)

- 레이블이 3개의 그룹(Class) 이상으로 구성

Label: 0



Label: 1



⋮

Label: 9



0부터 9까지의 그룹 중에 하나로 분류

02. 머신러닝 학습 방법

❖ 지도 학습 (4/7)

- 분류 문제를 해결하기 위한 기법

- ① k -최근접 이웃(k -Nearest Neighbors, k -NN)

- ② 의사 결정 트리(Decision Tree)

- ③ 로지스틱 회귀(Logistic Regression)

- ④ 서포트 벡터 머신(Support Vector Machine, SVM)

- ⑤ 나이브 베이즈(Naïve Bayes)

- ⑥ 앙상블 학습(Ensemble Learning)

02. 머신러닝 학습 방법

❖ 지도 학습 (5/7)

- 회귀(Regression)

- ◆ 입력 데이터에 대한 결괏값으로 학습에 사용한 레이블 이외의 값이 나올 수 있음
- ◆ 레이블이 실수(Real Number)인 경우

예를 들어

- ✓ 키(Height) 정보가 주어졌을 때, 몸무게를 예측
- ✓ 공부한 시간 정보가 주어졌을 때, 시험 성적 예측
- ✓ 커피를 몇 잔 마셨는지에 대한 정보가 주어졌을 때, 수면 시간 예측

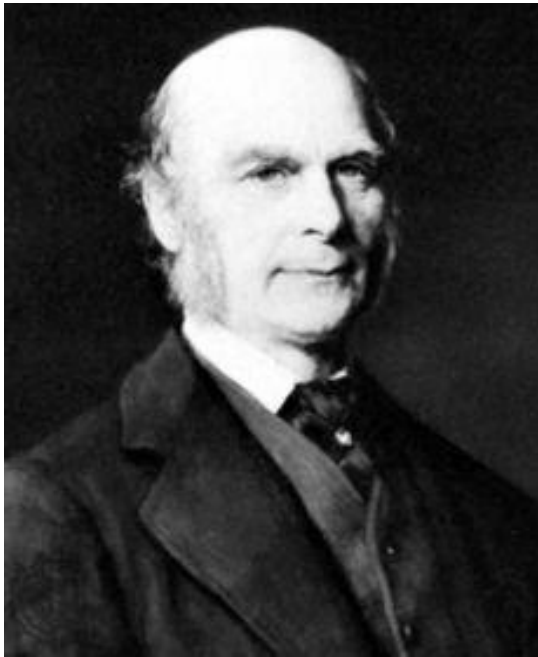
02. 머신러닝 학습 방법

❖ 지도 학습 (6/7)

- 회귀(Regression, 回歸)

- 回: 돌아올 회
- 歸: 돌아갈 귀

◆ 19세기, 통계학자이자 인류학자인 프랜시스 골턴이 처음 사용했다고 알려져 있음



프랜시스 골턴(Francis Galton)

- ✓ 아버지와 자식의 키를 분석함
- ✓ 사람의 키는 세대를 거듭할 수록 평균에 가까워지는 경향이 있다는 것을 발견
- ✓ 키가 큰 아버지의 자식은 아버지보다 키가 작고,
키가 작은 아버지의 자식은 아버지보다 키가 크다
- ✓ 즉 세대를 거듭할 수록 큰 키는 작아지고,
작은 키는 커져서 결국 평균에 수렴한다
- ✓ 이를 프랜시스 골턴은 "평균으로 돌아간다(=회귀)"라고 표현함

변수 사이의 관계를 분석하는 방법을 역사적인 이유 때문에
"회귀(Regression, 回歸)"라고 부르게 되었음

02. 머신러닝 학습 방법

❖ 지도 학습 (7/7)

- 회귀 문제를 해결하기 위한 기법
 - ◆ k -최근접 이웃(k -Nearest Neighbors, k -NN)
 - ◆ 의사 결정 트리(Decision Tree)
 - ◆ 선형 회귀(Linear Regression)
 - ◆ 다항 회귀(Polynomial Regression)

02. 머신러닝 학습 방법

❖ 비지도 학습 (1/3)

- 입력값에 대한 정답이 없는, 즉 레이블이 없는 데이터를 활용하여 머신러닝 모델을 학습시키는 방식
- 정답이 없기 때문에, 데이터에서 그 데이터의 특징을 찾아서 스스로 구분해 나감
- 이렇게 구분해 나가면 새로운 데이터가 들어왔을 때 그 데이터가 어떤 그룹에 속하는지 판단할 수 있음
- 즉, 정답이 없는 데이터를 사용해서 스스로 판단할 수 있는 지능을 가지게 됨
- 비지도 학습에는 대표적으로 **군집화**(Clustering)라는 방법이 있음

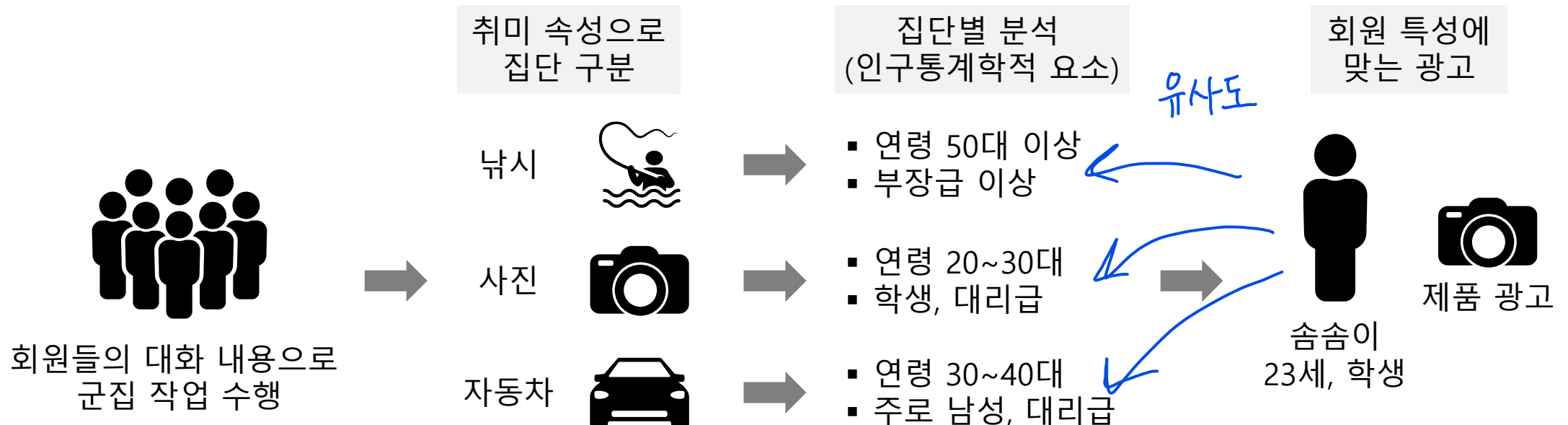
02. 머신러닝 학습 방법

❖ 비지도 학습 (2/3)

- 군집화(Clustering) → 추천 시스템

- ◆ 레이블 없이, 데이터 간에 존재하는 비슷한 속성을 기준으로 데이터를 묶는 작업임
- ◆ 군집화에 필요한 기준 속성은 데이터를 기반으로 머신러닝 모델이 스스로 찾아냄
- ◆ 군집화는 마케팅 분야에서 많이 활용되고 있음

마케팅 분야의 군집 작업 활용 예



02. 머신러닝 학습 방법

❖ 비지도 학습 (3/3)

- 군집화 문제를 해결하기 위한 기법

- ① k -평균 군집화(k -Means Clustering)

- ② 밀도기반 클러스터링(Density Based Spatial Clustering of Applications with Noise, DBSCAN)

02. 머신러닝 학습 방법

❖ 강화 학습 (1/3)

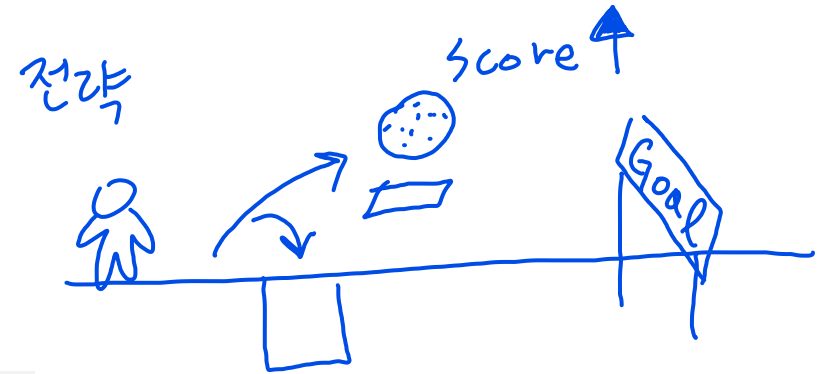
- 입력값에 대한 정답 대신 어떤 일을 잘했을 때 보상(Reward)을 주는 것으로 머신러닝 모델을 학습시키는 방식

핵심 목표

- ✓ 시행착오를 통해 보상의 총합이 최대가 되는 일련의 행동을 찾는 것
- ✓ 나중에 더 큰 이익을 기대하며 지금의 손해를 감수하는 식의 전략적 행동을 탐색할 수 있다는 의미와 같음

응용 분야

- ✓ 사업 분석 분야: 게임 전략, 금융시장의 투자 전략, 광고 노출 전략 등
- ✓ 로봇틱스 분야: 로봇 팔 제어, 이족보행 제어, 게임
- ✓ 자율주행 분야

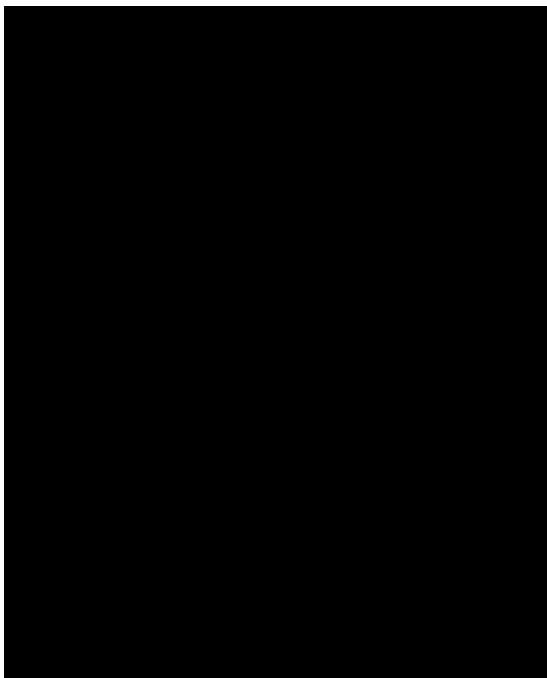


02. 머신러닝 학습 방법

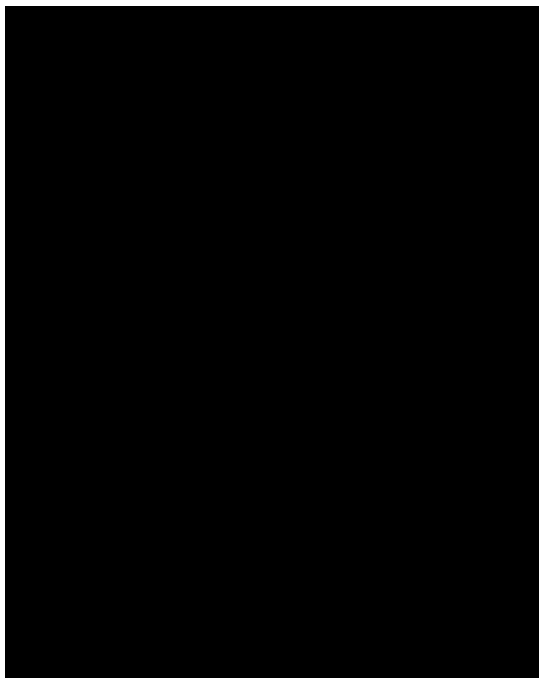
❖ 강화 학습 (2/3)

- 딥마인드(DeepMind)의 벽돌 깨기 게임

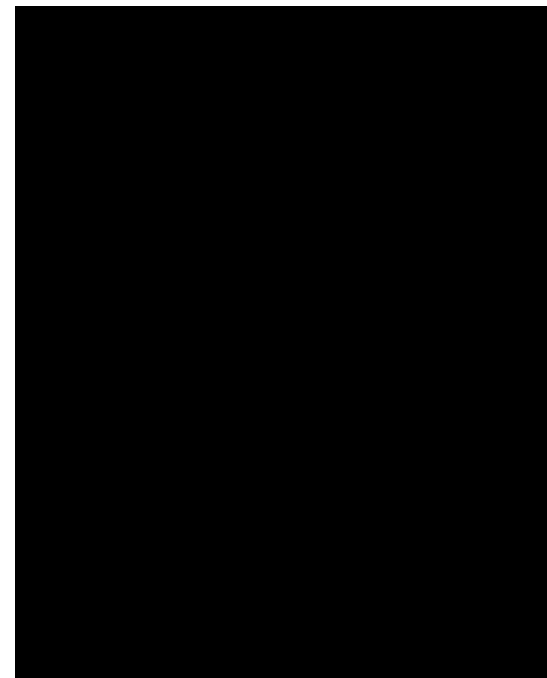
- ◆ 처음 인공지능을 훈련시킬 때만 하더라도 인공지능이 공을 따라가지 못하여 금방 게임이 끝났으나, 학습한 인공지능이 사람은 도저히 따라갈 수 없는 속도와 정확도로 게임을 진행하게 됨
- ◆ 강화 학습으로 벽돌 깨기를 학습시킨 인공지능은 게임 고수들이 하는 것과 같은 방식을 사용함



학습 초창기



중반



끝

02. 머신러닝 학습 방법

❖ 강화 학습 (3/3)

- 딥마인드의 인공지능 모델 더미가 걸어 다니는 모습
 - ◆ 더미의 무게 중심이 어떨 때 넘어지는지에 대한 정보를 입력한 후, 더미에게 넘어지지 말라는 명령과 다양한 환경을 제시
 - ◆ 연구진들은 이 실험을 통해 사람의 모습을 닮은 로봇인 휴머노이드가 어떻게 걷고 뛰어야 하는지 강화 학습으로 스스로 학습하는 모습을 보여줌



[영상출처] <https://youtu.be/gn4nRCC9TwQ>

❖ 01. 머신러닝 개념

❖ 02. 머신러닝 학습 방법

THANK YOU!

Q & A

- Name: 권범
- Office: 동덕여자대학교 인문관 B821호
- Phone: 02-940-4752
- E-mail: bkwon@dongduk.ac.kr