

기계학습의 의미와 기계학습 종류

이건명

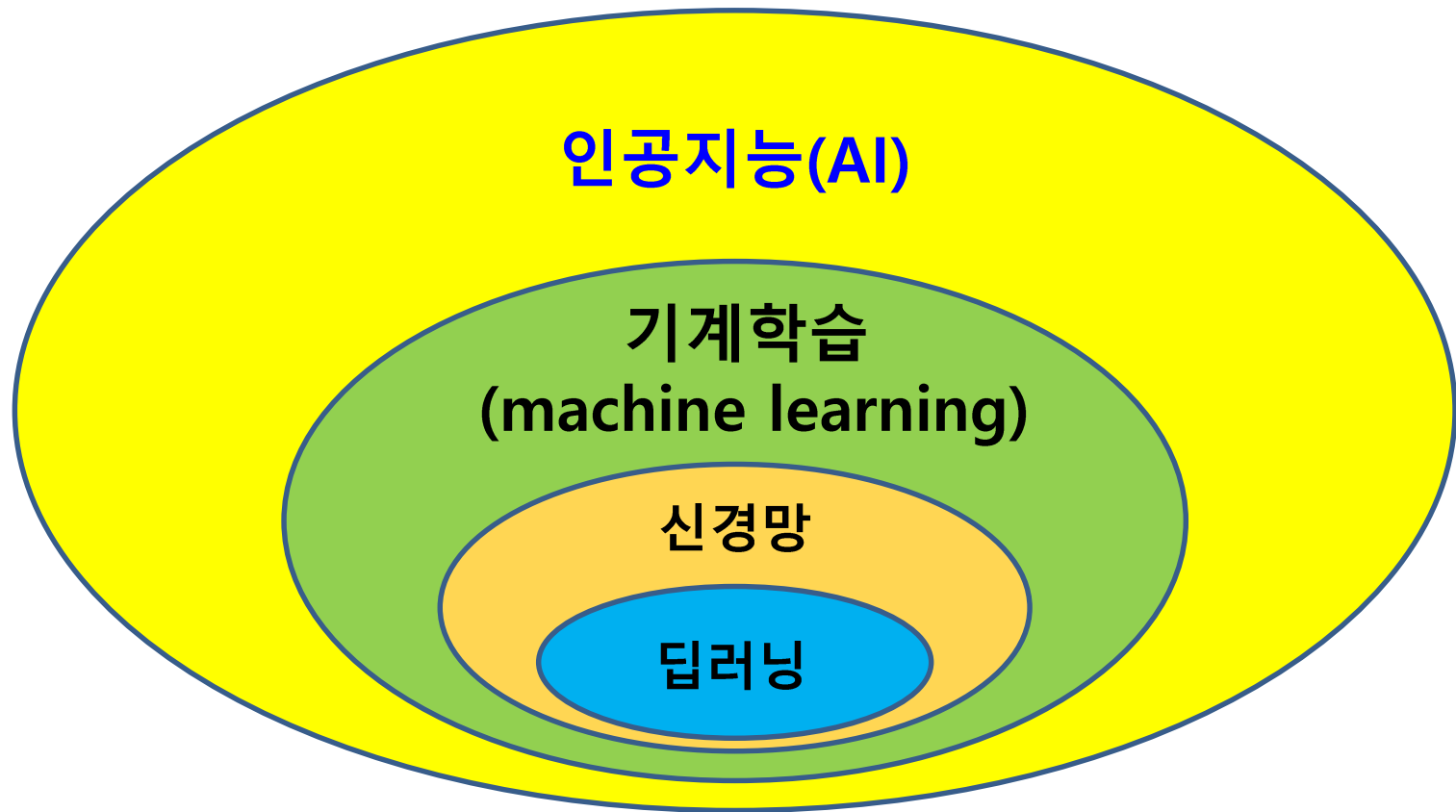
충북대학교 소프트웨어학과

인공지능 : 튜링 테스트에서 딥러닝까지

학습 내용

- 기계학습의 목적을 알아본다.
- 일반적인 프로그래밍과 기계학습의 차이를 알아본다.
- 기계학습의 전략을 알아본다.
- 오컴의 면도날(Occam's Razor)의 의미에 대해서 알아본다.
- 기계학습의 종류에 대해서 살펴본다.

인공지능, 기계학습, 신경망, 딥러닝



1. 기계학습

❖ 기계학습(機械學習, machine learning)

- 경험을 통해서 나중에 유사하거나 같은 일(task)를 더 효율적으로 처리할 수 있도록 시스템의 구조나 파라미터를 바꾸는 것
- 컴퓨터가 데이터로부터 특정 문제해결을 위한 지식을 자동으로 추출해서 사용할 수 있게 하는 기술

경험	일	효율(성능)
필기문자 이미지, 글자	문자 판독(인식)	정확도
사진, 얼굴영역	사진에서 얼굴영역 식별	정확도
이메일, 스팸여부	스팸 이메일 판단	정확도
풍경 사진	유사한 풍경 사진 식별	유사도
바둑 대국	바둑두는 방법	승률

기계학습

❖ 일반 프로그래밍 방식

일반 프로그래밍



사람이 생각해 낸 알고리즘



“고양이”

❖ 기계 학습



고양이

기계학습



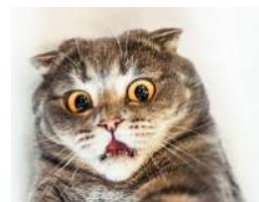
기계학습
알고리즘



고양이 인식
모델



“고양이”



기계 학습

❖ PlayTennis 문제

- 어떤 사람이 테니스를 치는 날의 기상 상황을 조사한 데이터
 - 학습데이터 (training data)

표 4.1 PlayTennis 데이터

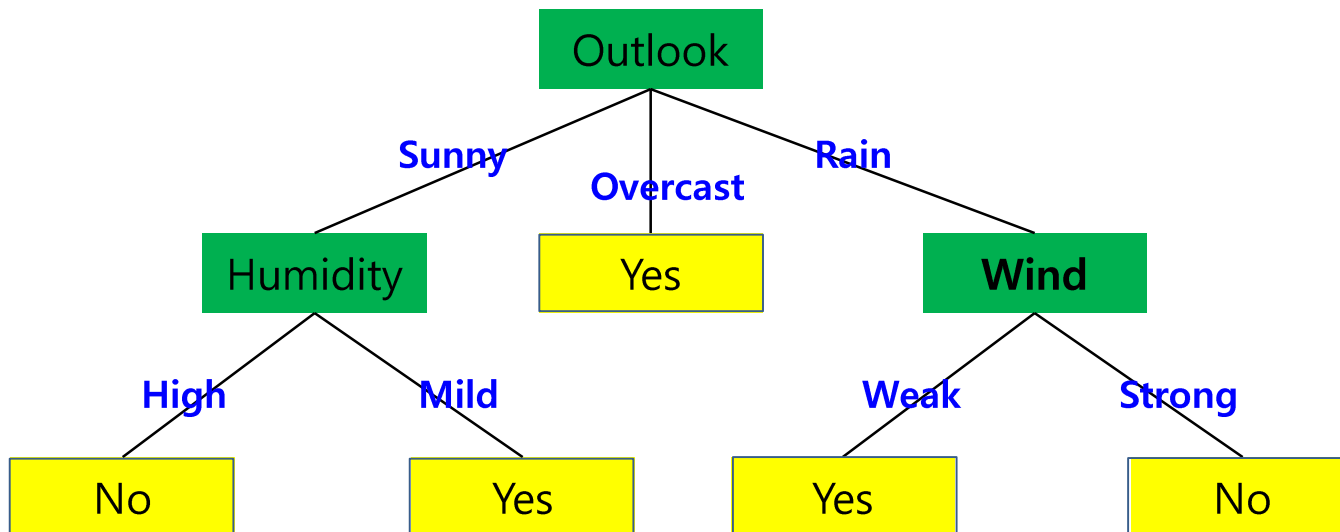
Day 날짜	Outlook 조망	Temperature 기온	Humidity 습도	Wind 바람	PlayTennis 테니스 여부
Day1	Sunny	Hot	High	Weak	No
Day2	Sunny	Hot	High	Strong	No
Day3	Overcast	Hot	High	Weak	Yes
Day4	Rain	Mild	High	Weak	Yes
Day5	Rain	Cool	Normal	Weak	Yes
Day6	Rain	Cool	Normal	Strong	No
Day7	Overcast	Cool	Normal	Strong	Yes
Day8	Sunny	Mild	High	Weak	No
Day9	Sunny	Cool	Normal	Weak	Yes
Day10	Rain	Mild	Normal	Weak	Yes
Day11	Sunny	Mild	Normal	Strong	Yes
Day12	Overcast	Mild	High	Strong	Yes
Day13	Overcast	Hot	Normal	Weak	Yes
Day14	Rain	Mild	High	Strong	No

(출처: Machine Learning, Tom Mitchell, 1995)

- 테니스를 치는 날은?
- '흐리고 적당한 온도에 습도는 높고 바람이 센 날' 테니스를 칠까?

기계학습

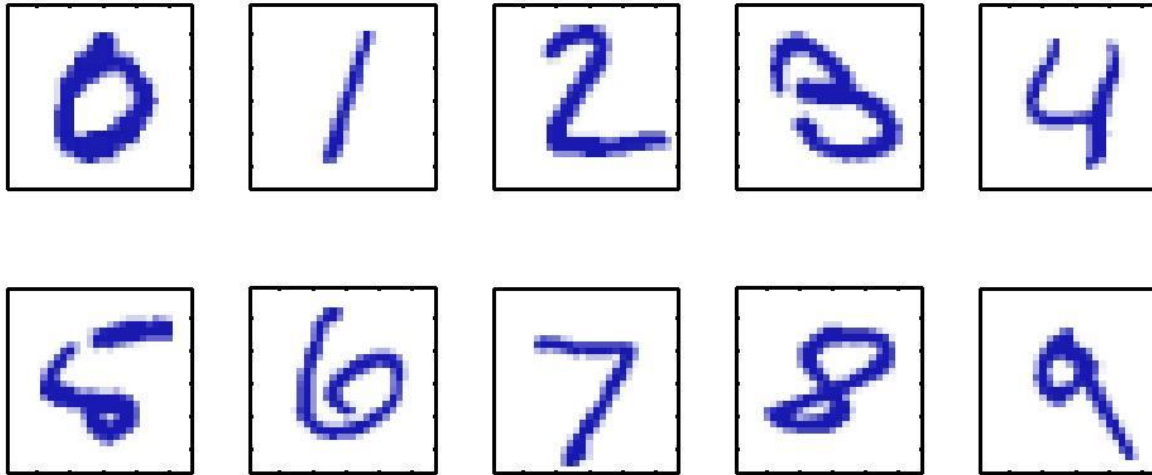
❖ PlayTennis 문제 – cont.



Outlook 조망	Temperature 기온	Humidity 습도	Wind 바람	PlayTennis 테니스 여부
Sunny	Hot	Mild	Weak	?
Rain	Hot	High	Weak	?

기계학습

❖ 필기문자 인식



- 직접 만든 규칙이나 휴리스틱(heuristics)
 - 복잡
 - 불충분한 성능
- 기계학습 방법
 - 자동으로 분류 규칙이나 프로그램 생성
 - 괄목할 만한 성능

기계학습

❖ 연역적 학습 (deductive learning)

- 연역적 추론(deductive inference)을 통한 학습

❖ 귀납적 학습 (inductive learning)

- 사례들(examples)을 **일반화(generalization)**하여 **패턴(pattern)** 또는 **모델(model)**을 추출하는 것
- 일반적인 기계학습의 대상
- 학습 데이터를 **잘 설명**할 수 있는 **패턴**을 찾는 것
 - **오컴의 면도날(Occam's razor)**
 - 어떤 현상의 인과관계를 설명할 때 불필요한 가정을 삼가야 한다
 - 가능하면 학습 결과를 간단한 형태로 표현하는 것이 좋다



(1,2) -> 3
(4,8) -> 12
(10,13) -> 23
(23,52) -> 75
(43,101) -> 144
(14,31) -> 45
(123,15) -> 138



학습알고리즘



학습된 함수
 $h(x, y)$

(97,16)



113

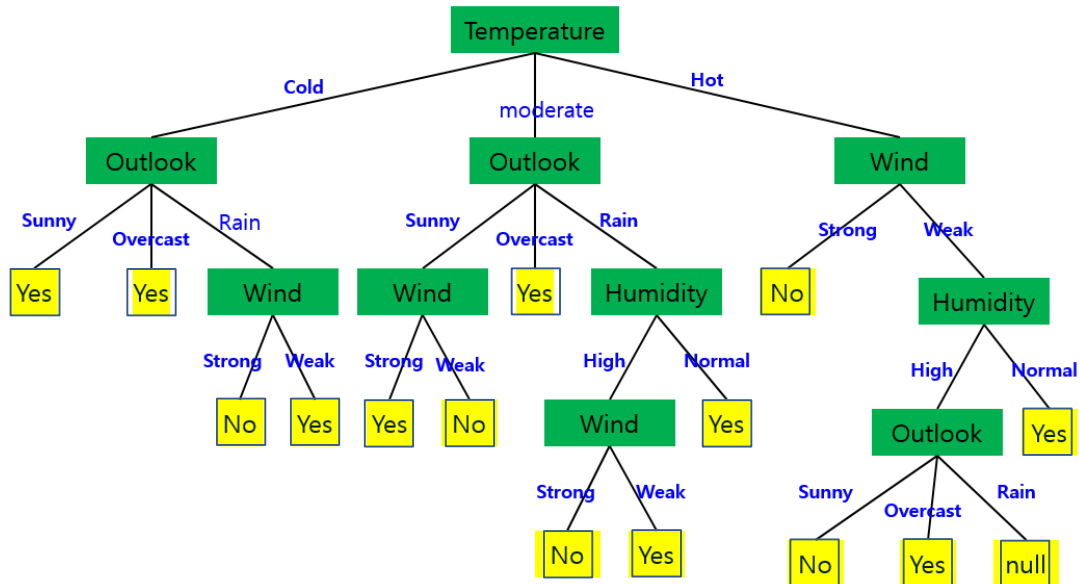
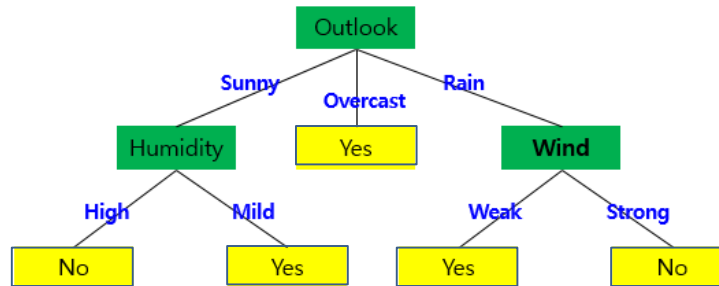


$$h(x, y) = x + y$$

$$h(x, y) = (2x + 3y)(2x + 2y)y / (4xy + 6y^2)$$

기계학습

- 오컴의 면도날(Occam's razor) 원리에 따른 선택



2. 기계학습의 종류

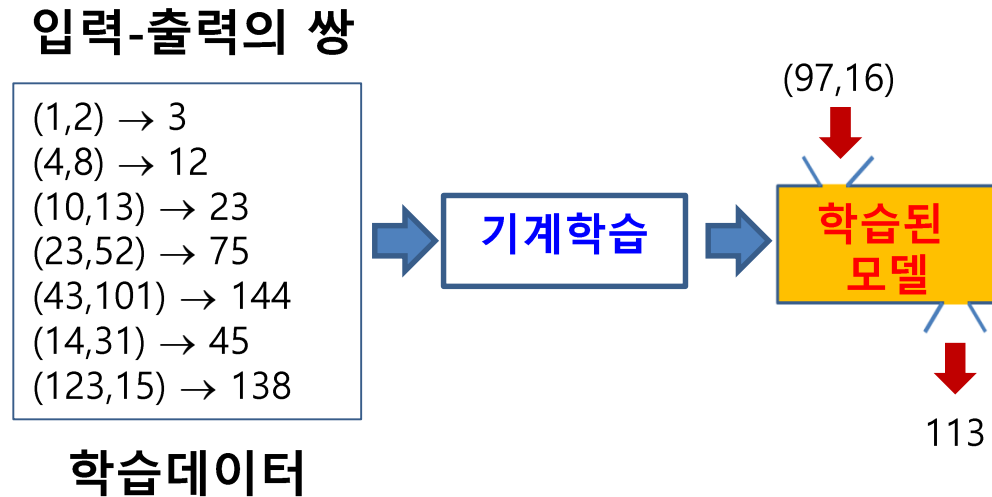
❖ 학습데이터의 형태와 학습 지식의 형태에 따른 분류

- 지도학습(supervised learning)
- 비지도학습(unsupervised learning, 자율학습)
- 강화학습(reinforcement learning)

기계학습의 종류

❖ 지도학습(supervised learning)

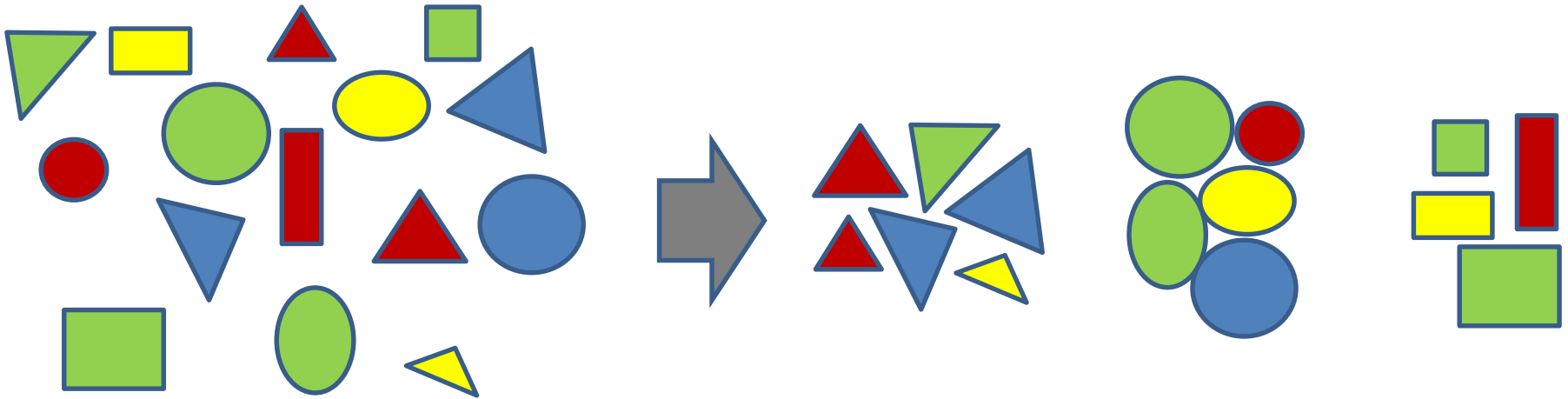
- **입력(문제)-출력(답)**의 데이터들로 부터 새로운 입력에 대한 출력을 결정할 수 있는 패턴 추출



기계학습의 종류

❖ 비지도학습 (unsupervised learning, 자율학습)

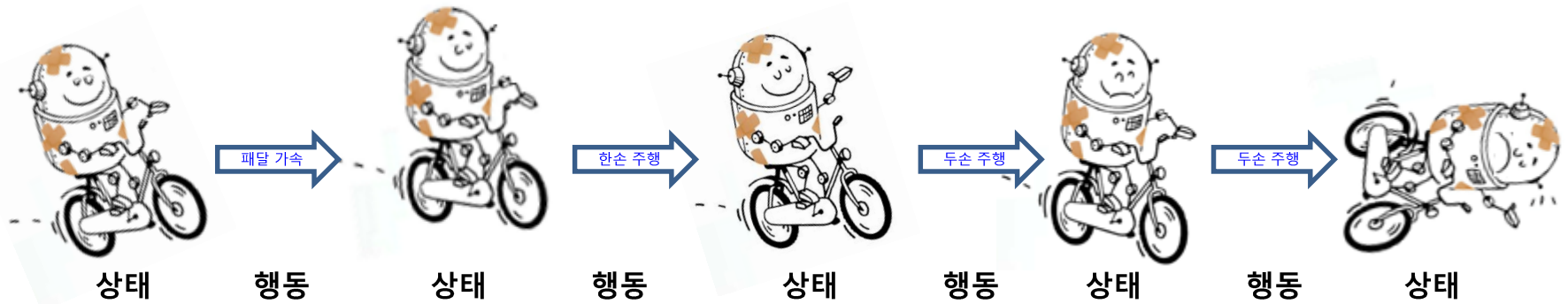
- 출력에 대한 정보가 없는 데이터로부터 패턴 추출



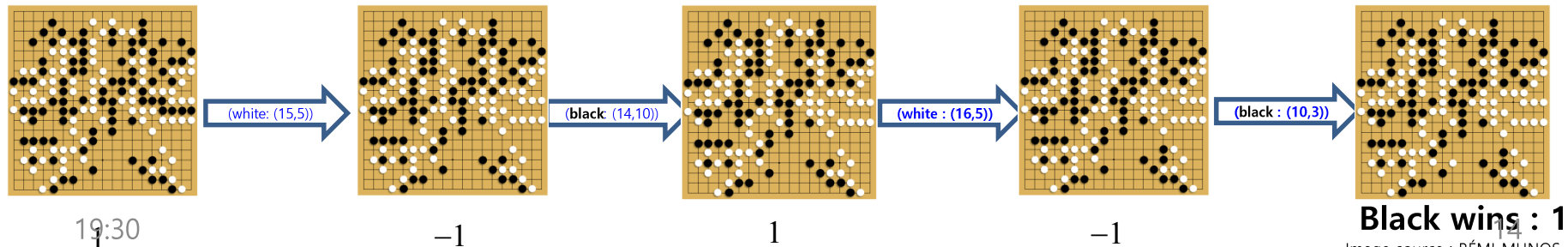
기계학습의 종류

❖ 강화학습(reinforcement learning)

- 출력에 대한 정확한 정보를 제공하지는 않지만, **평가정보(reward)**는 주어지는 문제에 대해 **각 상태에서의 행동(action)**을 결정
- 문제에 대한 직접적인 답을 주지는 않지만 경험을 통해 **기대 보상(expected reward)**이 최대가 되는 **정책(policy)**을 찾는 학습
 - 정책** : 각 상태 별로 취할 행동을 정해 놓은 것



정책(policy)



Quiz

❖ 다음 기계학습에 대한 설명으로 옳지 않은 것을 선택하시오.

- ① 기계학습은 1980년대부터 시작된 인공지능의 분야이다.
- ② 기계학습은 학습을 통해서 특정 일을 수행할 수 있도록 하는 방법을 찾는 분야이다.
- ③ 기계학습은 학습 데이터들을 일반화하는 패턴이나 규칙성을 찾는 것으로 귀납적 학습인 경우가 많다.
- ④ 기계학습을 할 때는 데이터에서 적합한 특징이 추출되어 사용되어야 높은 성능을 얻을 수 있다.

❖ 다음 기계학습의 종류에 대한 설명으로 옳지 않은 것을 선택하시오.

- ① 지도 학습에서 학습 데이터의 출력은 수치형 속성이어야 한다.
- ② 비지도 학습에서는 학습 데이터에 출력 정보가 없다.
- ③ 반지도 학습에서는 학습 데이터의 출력이 주어지는 것과 그렇지 않는 것이 함께 사용된다.
- ④ 강화 학습에서는 상황별로 어떤 행동을 해야 할지 결정하는데, 학습 데이터에는 상황에 대해 수행한 행동에 대한 보상 정보가 주어진다.