

기계학습의 의미와 기계학습 종류

이건명

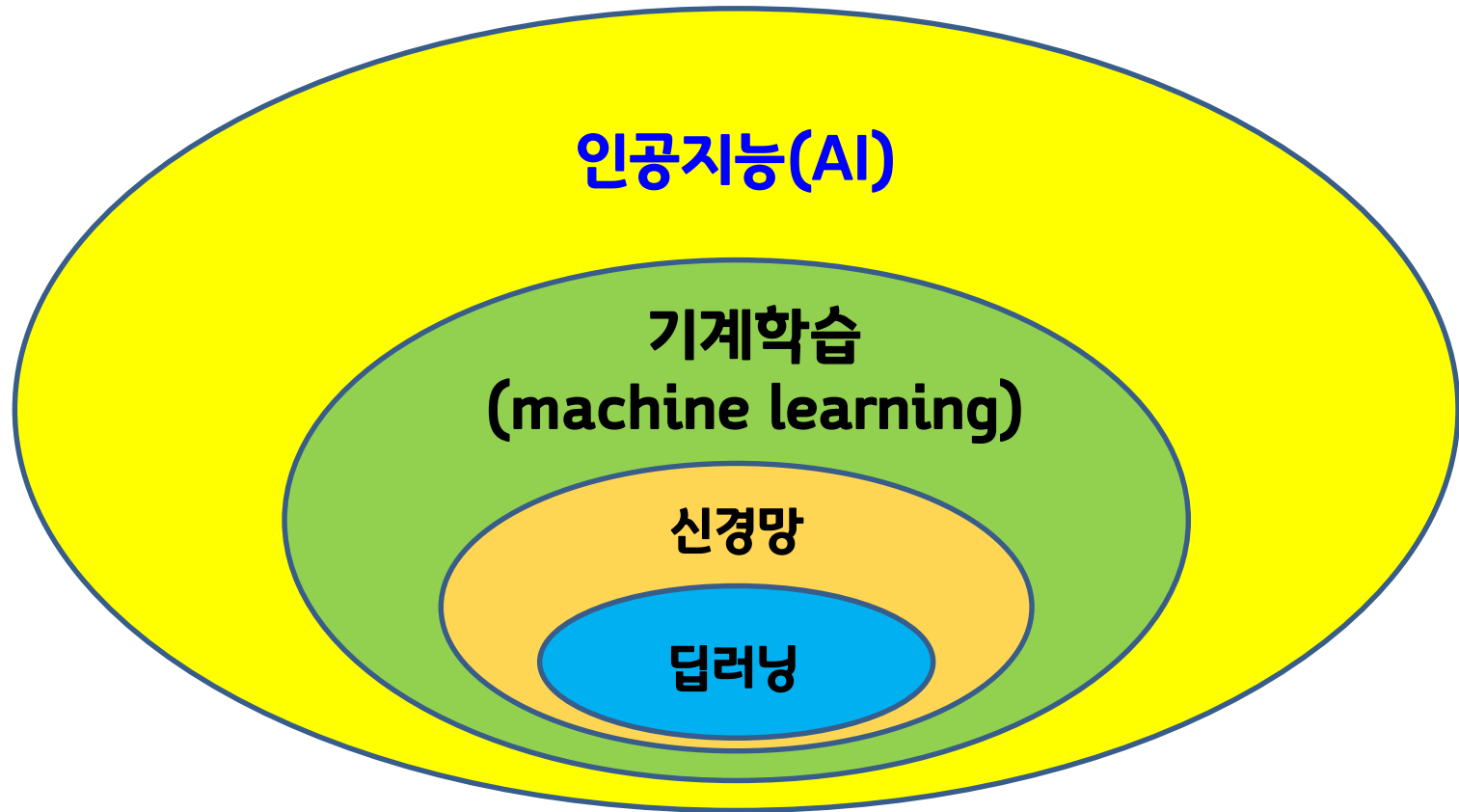
충북대학교 산업인공지능학과

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

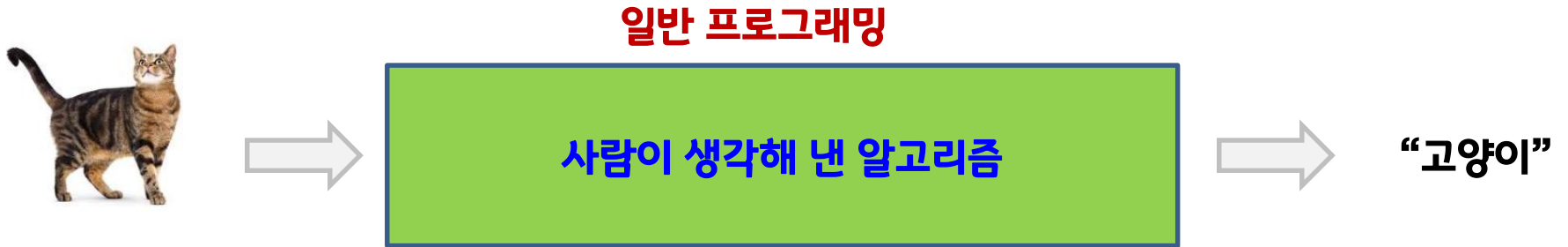
❖ 기계학습(機械學習, machine learning)

- **경험**을 통해서 나중에 유사하거나 같은 **일(task)**를 더 **효율적**으로 처리할 수 있도록 **시스템의 구조나 파라미터를 바꾸는 것**
- 컴퓨터가 데이터로부터 특정 문제해결을 위한 **지식**을 **자동**으로 **추출**해서 **사용**할 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
광고, 사용자 정보, 클릭여부	온라인 광고	정확도
식당리뷰, 긍정/부정	평판도 분석(감성분석)	정확도

기계학습

❖ 일반 프로그래밍 방식



❖ 기계 학습



기계학습

❖ PlayTennis 문제

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

표 4.1 PlayTennis 데이터

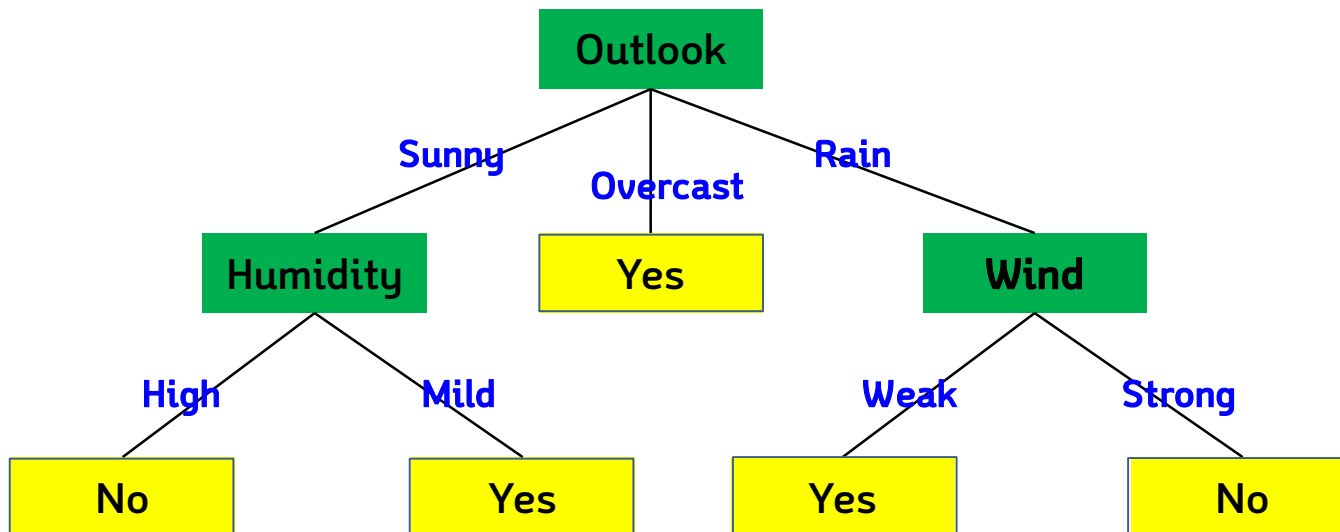
Day 날짜	Outlook 조망	Temperature 기온	Humidity 습도	Wind 바람	PlayTennis 테니스 여부
Day1	Sunny	Hot	High	Weak	No
Day2	Sunny	Hot	High	Strong	No
Day3	Overcast	Hot	High	Weak	Yes
Day4	Rain	Mild	High	Weak	Yes
Day5	Rain	Cool	Normal	Weak	Yes
Day6	Rain	Cool	Normal	Strong	No
Day7	Overcast	Cool	Normal	Strong	Yes
Day8	Sunny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

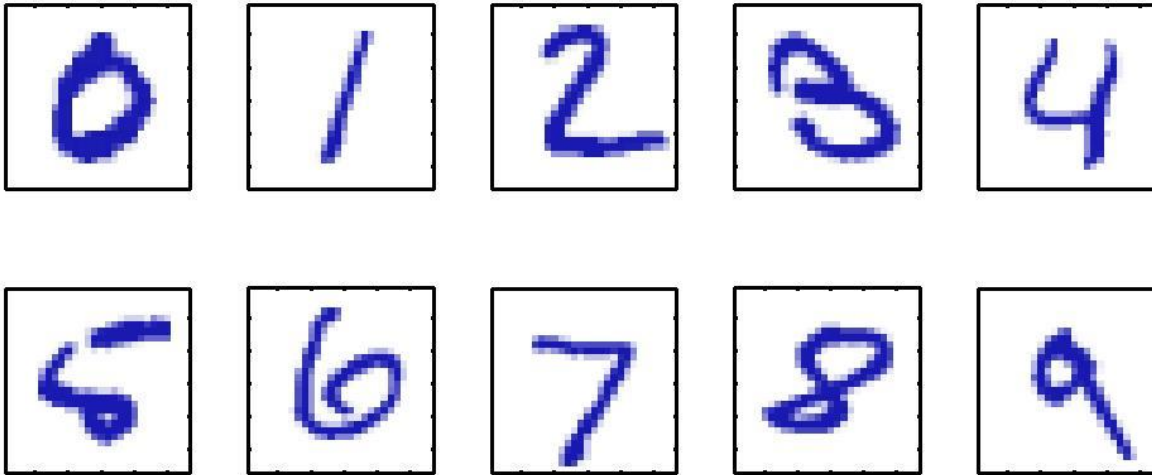
❖ PlayTennis 문제 – cont.



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 설명할 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



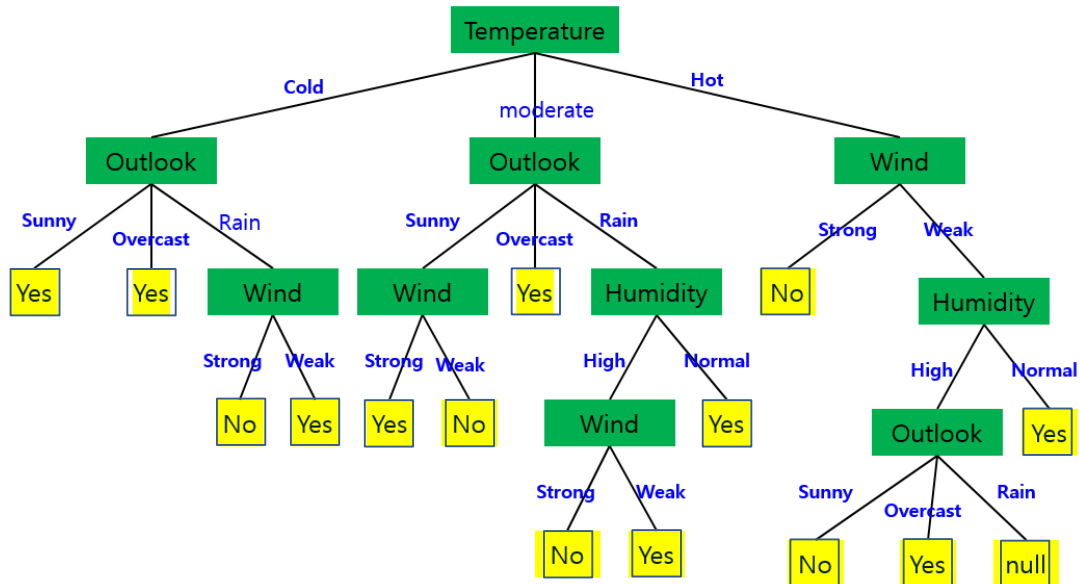
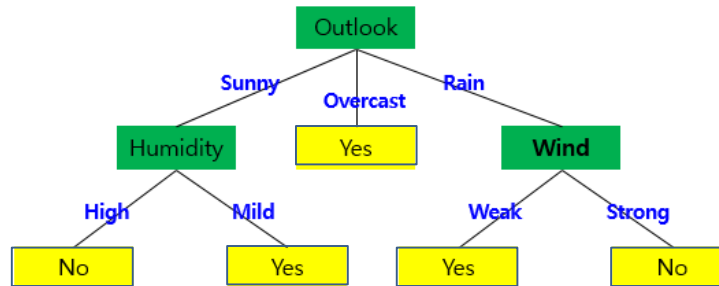
(97,16)
↓
학습된 함수
 $h(x, y)$

↓
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

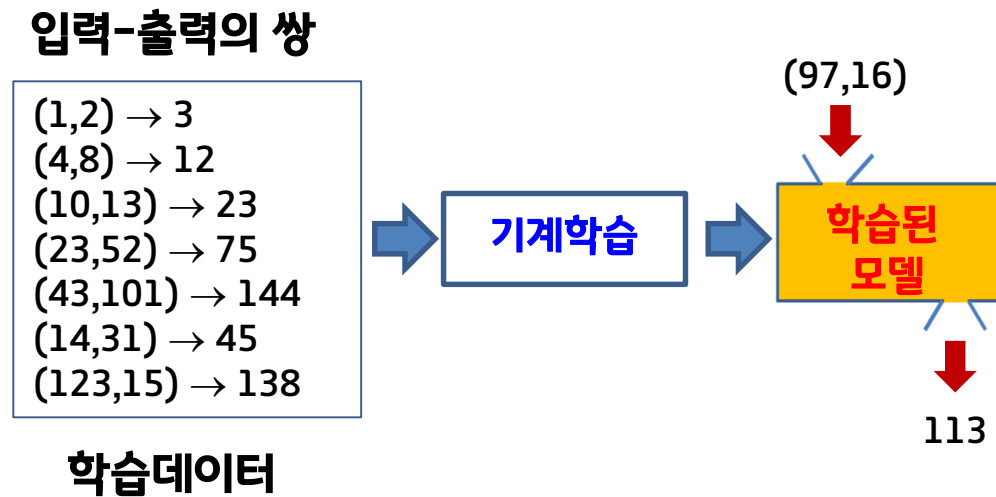
❖ 학습데이터의 형태와 학습 지식의 형태에 따른 분류

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

기계학습의 종류

❖ 지도학습(supervised learning)

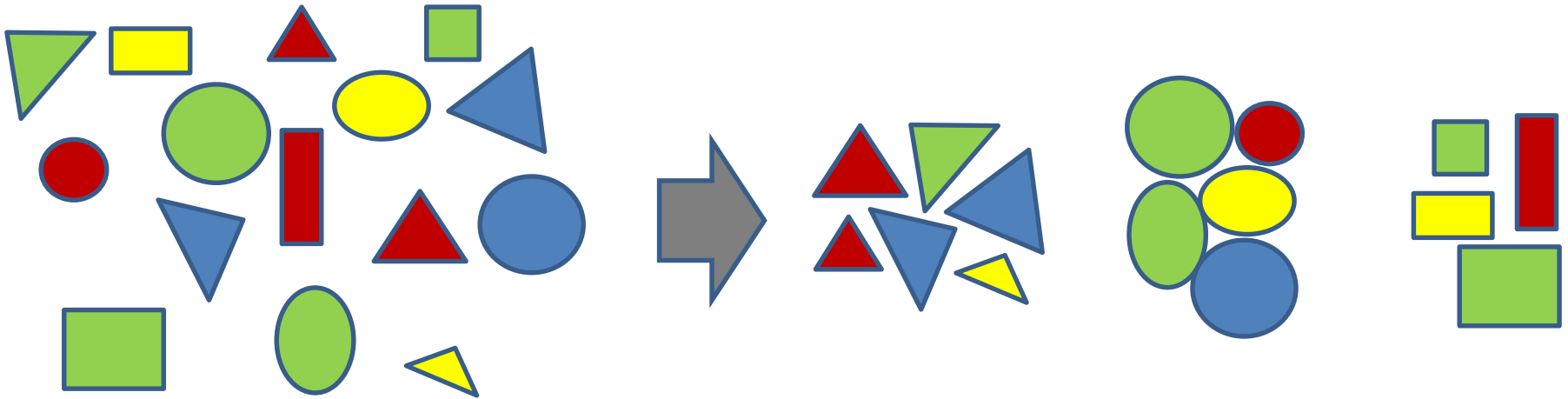
- **입력(문제)-출력(답)**의 데이터들로부터 새로운 입력에 대한 출력을 결정할 수 있는 패턴 추출



기계학습의 종류

❖ 비지도학습 (unsupervised learning, 자율학습)

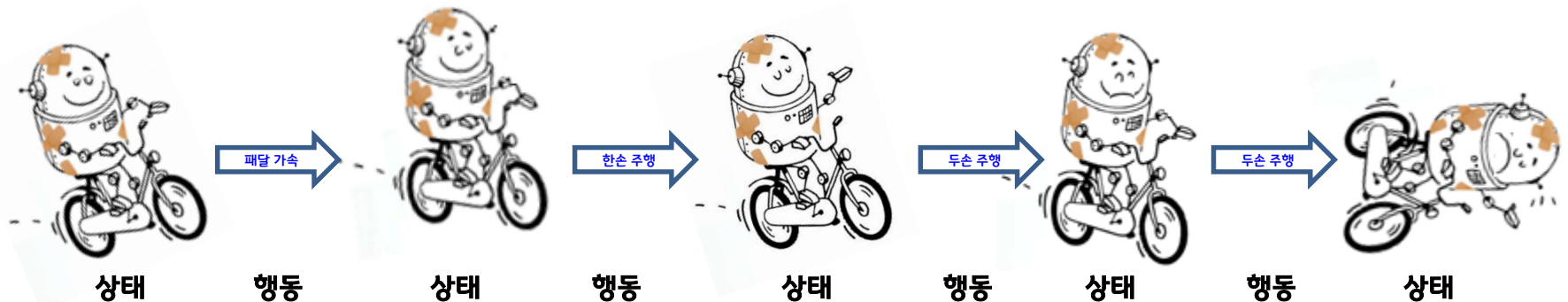
- 출력에 대한 정보가 없는 데이터로 부터 패턴 추출



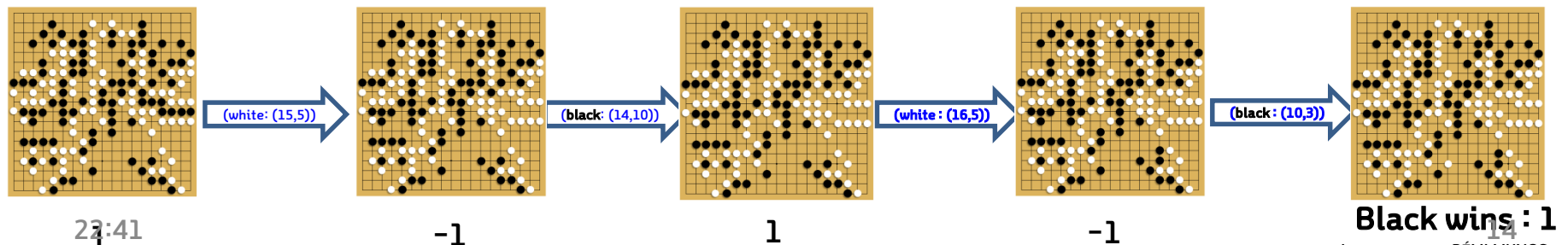
기계학습의 종류

❖ 강화학습(reinforcement learning)

- 출력에 대한 정확한 정보를 제공하지는 않지만, **평가정보**(reward)는 주어지는 문제에 대해 각 **상태**에서의 **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.

Quiz

❖ 지도학습이란 무엇인가?

- ① 비지도 학습으로 모델을 훈련시키는 방법
- ② 데이터에 레이블이 없는 학습 방식
- ③ 입력 데이터에 대한 적절한 출력을 예측하도록 모델을 훈련시키는 방법
- ④ 보상 시스템을 통해 모델을 학습시키는 방법

❖ 지도학습의 가장 큰 특징은 무엇인가?

- ① 레이블없는 데이터만 사용하여 학습한다.
- ② 항상 정확한 예측을 보장한다.
- ③ 레이블된 학습 데이터를 사용하여 모델을 학습시킨다.
- ④ 데이터 없이 학습이 가능하다.

❖ 지도학습의 예가 아닌가?

- ① 분류
- ② 회귀
- ③ 군집화
- ④ 순위매기기

Quiz

❖ 오컴의 면도날의 원리는 무엇인가?

- ① 가장 복잡한 해설을 우선적으로 고려한다.
- ② 불필요하게 가정을 늘리지 않는다.
- ③ 항상 가장 간단한 해설이 옳다.
- ④ 모든 가설은 반드시 검증되어야 한다.

❖ 오컴의 면도날은 어떤 종류의 학문적 원칙으로 가장 잘 알려져 있나?

- ① 경제학 원칙
- ② 심리학 이론
- ③ 철학적, 과학적 방법론 원칙
- ④ 사회학적 관점

❖ 오컴의 면도날을 사용할 때 주의해야 할 점은 무엇인가?

- ① 항상 단순한 해설이 옳다는 것을 의미한다.
- ② 복잡한 해설이 틀렸다는 것을 의미한다.
- ③ 단순성만을 추구하면 항상 옳다.
- ④ 단순한 해설을 선호하지만, 그것이 항상 옳은 해설이라는 것을 보장하지 않는다.

Quiz

❖ 귀납적 학습 방법은 어떤 접근 방식을 사용하는가?

- ① 주어진 정보만을 기반으로 결론을 도출한다.
- ② 일반적인 원칙에서 출발해 특정 사례를 예측한다.
- ③ 특정 사례들을 관찰하고 이를 바탕으로 일반적인 원칙이나 규칙을 도출한다.
- ④ 무조건적인 규칙을 따른다.

❖ 연역적 학습 방법은 _____ 에서 특정 사례로의 추론 방식을 취한다.

- ① 특정 사례
- ② 일반적인 원칙
- ③ 같은 특정 사례들
- ④ 관련 없는 사례들

❖ 어떤 연구자가 여러 사례를 관찰하고 그 결과를 바탕으로 일반적인 규칙을 세운다면, 그 연구자는 _____ 학습 방법을 사용하고 있다.

- ① 연역적 ② 귀납적 ③ 추론적 ④ 비교적

Quiz

❖ 비지도 학습의 주요 특징 중 하나는?

- ① 학습 과정 중에 라벨이 주어진다.
- ② 각 데이터 포인트에 라벨을 할당하는 것이 주 목적이다.
- ③ 라벨이 없는 데이터를 사용하여 패턴을 찾는다.
- ④ 지도 학습보다 정확도가 높다.

❖ 비지도 학습의 주요 목적 중 하나는 무엇인가?

- ① 분류된 라벨의 오류율을 최소화한다.
- ② 데이터의 숨겨진 구조나 패턴을 발견한다.
- ③ 미래의 데이터 포인트를 예측한다.
- ④ 데이터의 라벨을 빠르게 추정한다.

❖ 강화학습에서 '에이전트'는 무엇을 의미하는가?

- ① 학습을 수행하는 환경
- ② 보상을 결정하는 규칙
- ③ 학습 과정에서 행동을 선택하고 실행하는 주체
- ④ 모든 가능한 행동의 집합

Quiz

❖ 강화학습의 목표는 무엇인가?

- ① 모든 상황에서 최적의 행동을 선택한다.
- ② 주어진 데이터를 가장 잘 설명하는 모델을 찾는다.
- ③ 과거 데이터를 기반으로 미래를 예측한다.
- ④ 입력 데이터에 대한 라벨을 정확하게 예측한다.

❖ 강화학습에서 '보상'의 역할은 무엇인가?

- ① 에이전트의 행동을 제한하는 역할
- ② 학습 데이터의 라벨을 제공하는 역할
- ③ 에이전트의 행동에 따른 피드백을 제공하는 역할
- ④ 학습의 속도를 조절하는 역할