

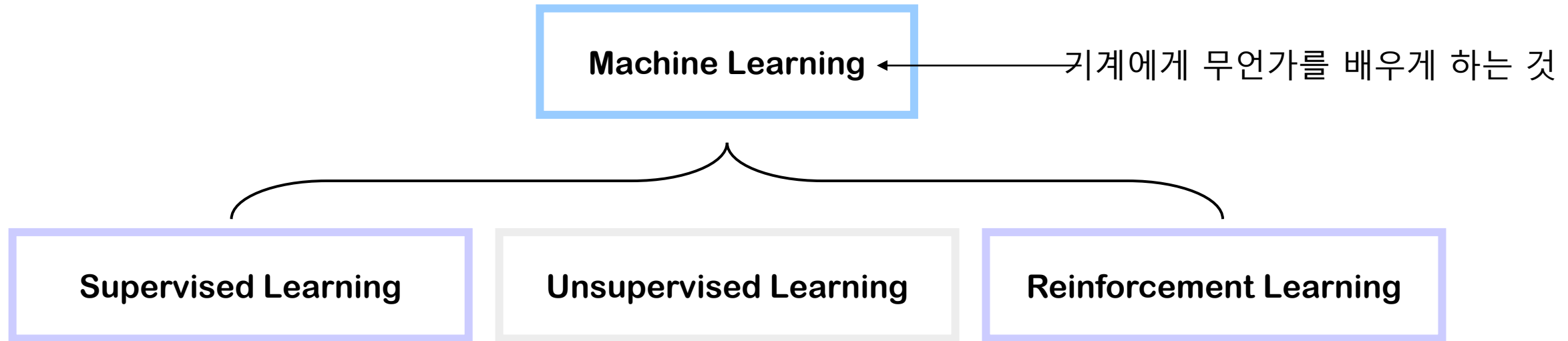
Supervised Learning vs Reinforcement Learning

Hyelin Choi

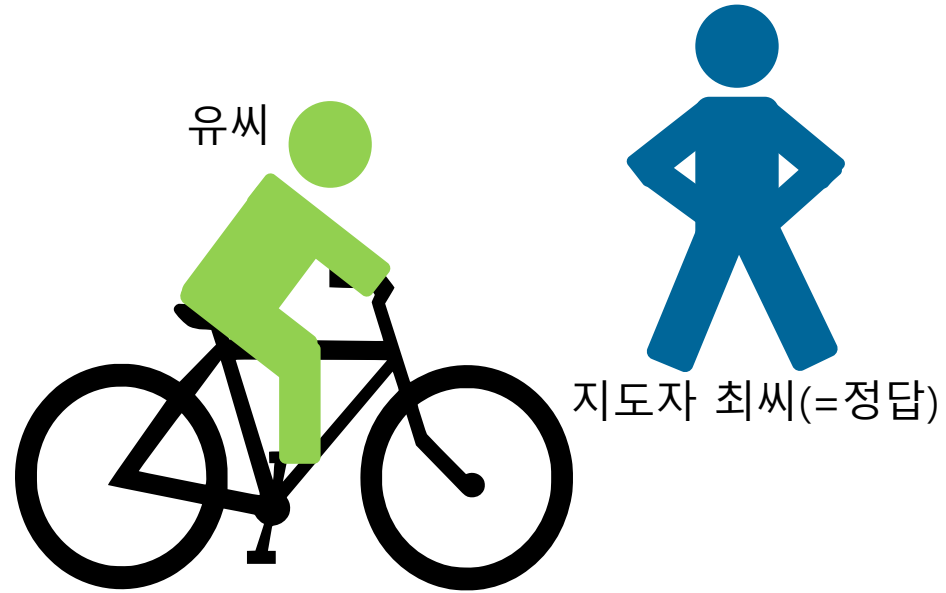
**Department of Mathematics
Sungkyunkwan University**

March 26, 2024

Machine Learning



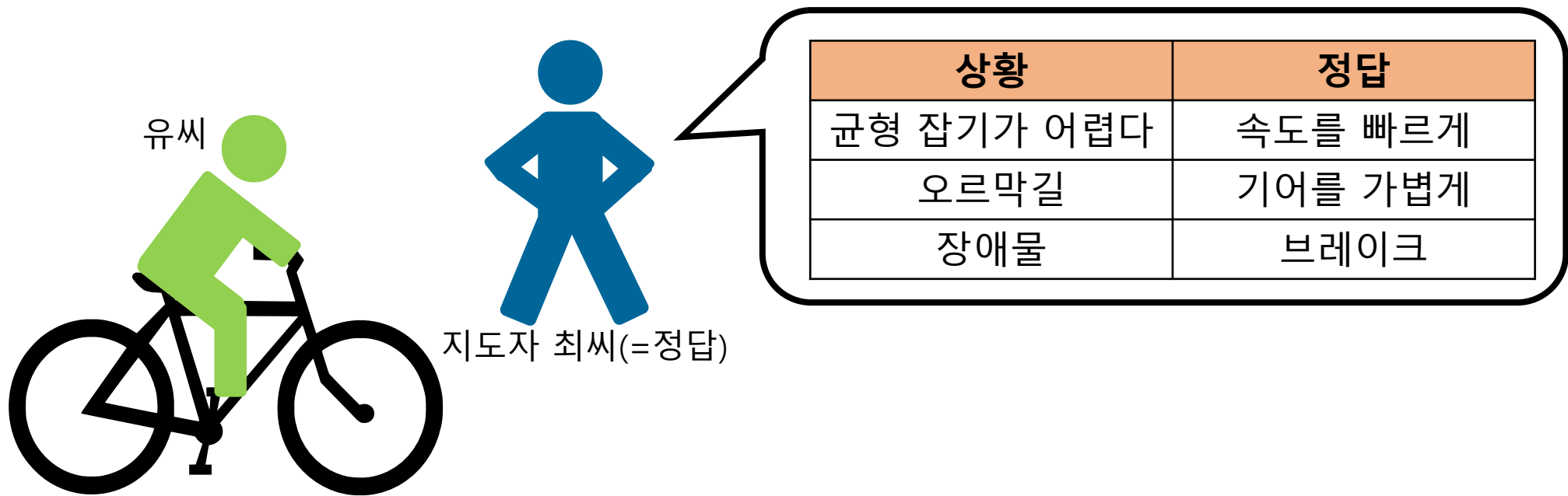
Supervised Learning



Reinforcement Learning



Supervised Learning (유씨)



Supervised Learning (유씨)

속도를 너무 느리게 하면
중심 잡기가 힘들구나

오르막길에서는 기어를
낮추는게 좋구나

장애물이 있으면
브레이크를 밟아야 하는
구나

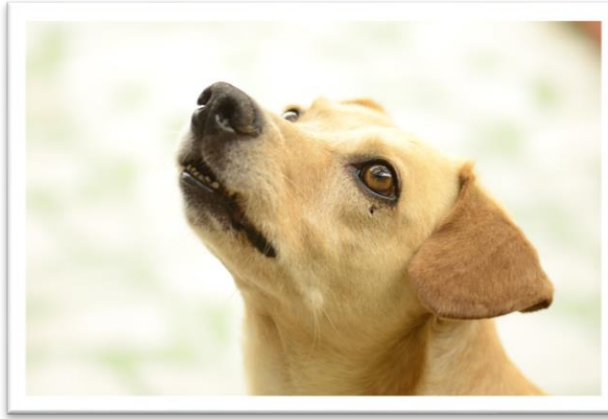


Supervised Learning

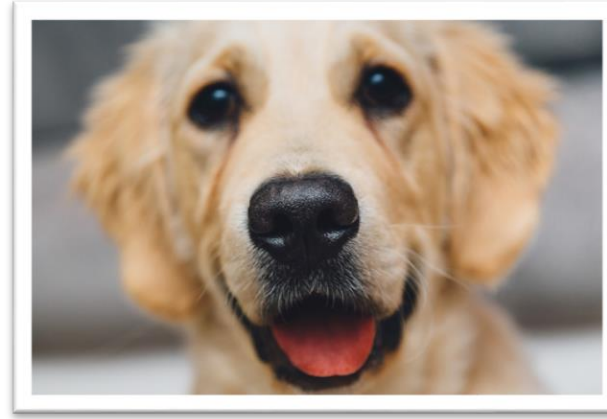
<예시> 사진이 주어지면 해당 사진이 개인지 고양이인지를 출력하는 예측 모델을 학습하는 것



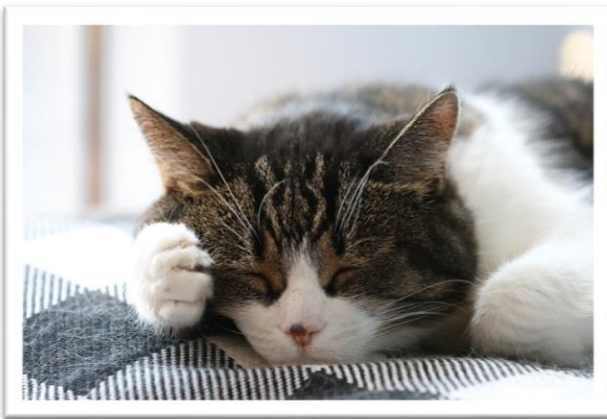
Dog



Dog



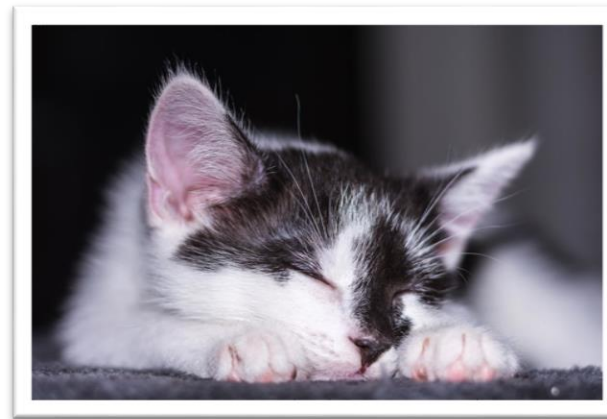
Dog



Cat



Cat



Cat

Training Data

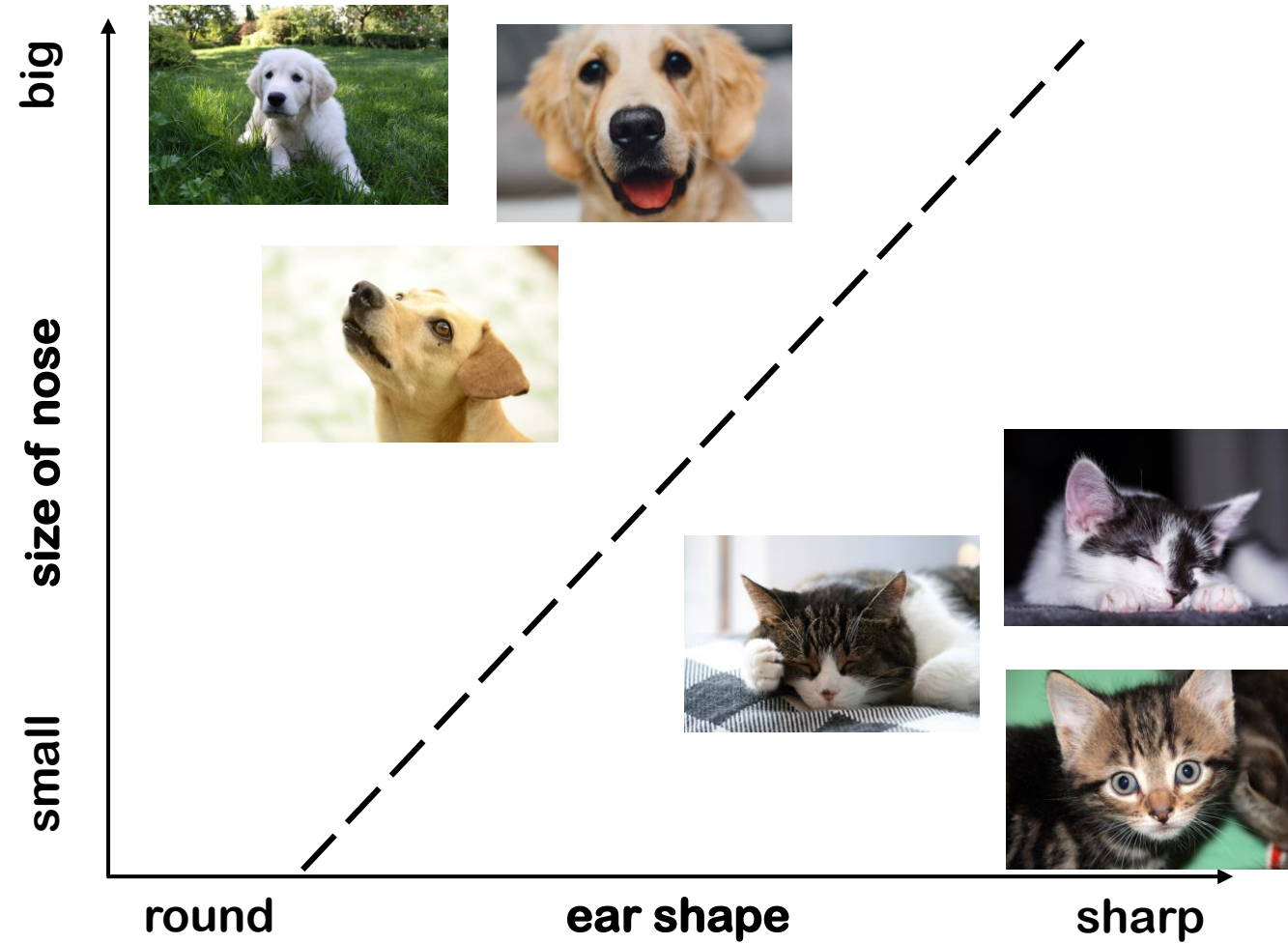
Supervised Learning

- 기준: 귀 모양, 코 크기

	개	고양이
귀 모양	둥글다	뽀족하다
코 크기	크다	작다

이 기준을 자동으로 추출하여 학습하는 방식 => Deep Learning

Supervised Learning



Supervised Learning

Test Data

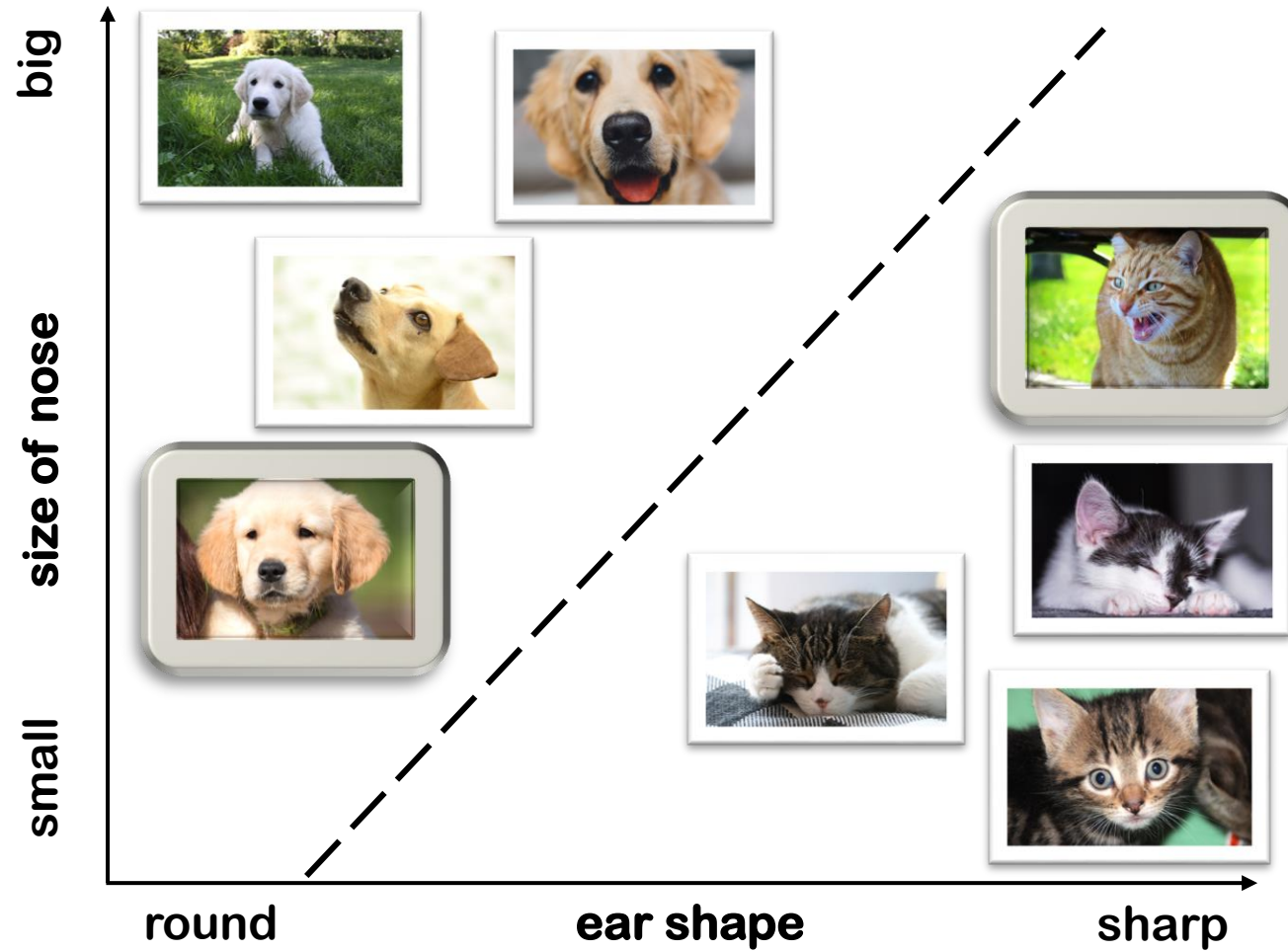


Cat



Dog

Supervised Learning



Reinforcement Learning



지도자
X

Reinforcement Learning

시행착오를 통해 발전해 나가는 학습 과정

순차적 의사결정 문제에서 누적 보상을 최대화하기 위해
시행착오를 통해 행동을 교정하는 학습 과정

Reinforcement Learning

시행착오를 통해 발전해 나가는 학습 과정

순차적 의사결정 문제에서 누적 보상을 최대화하기 위해
시행착오를 통해 행동을 교정하는 학습 과정

Reinforcement Learning (이씨)

<예시> 1m당 1점이 보상 -> 넘어지지않고 최대한 나아가는 것이 목적

페달을 세게 밟아볼까?

방금 내가
보상을 얻기에
나
⇒ 페달을 세
= 보상이 높

두

방금 내가 한 행동은
보상을 얻기에 안 좋은 행동이
구나
⇒ 두 손을 떼고 타는 것
= 보상이 낮은 행동

2점



1m

2m

3m

Reinforcement Learning (이씨)



보상을 높이기 위해서
페달을 항상 세게 밟
으면
되지 않을까?

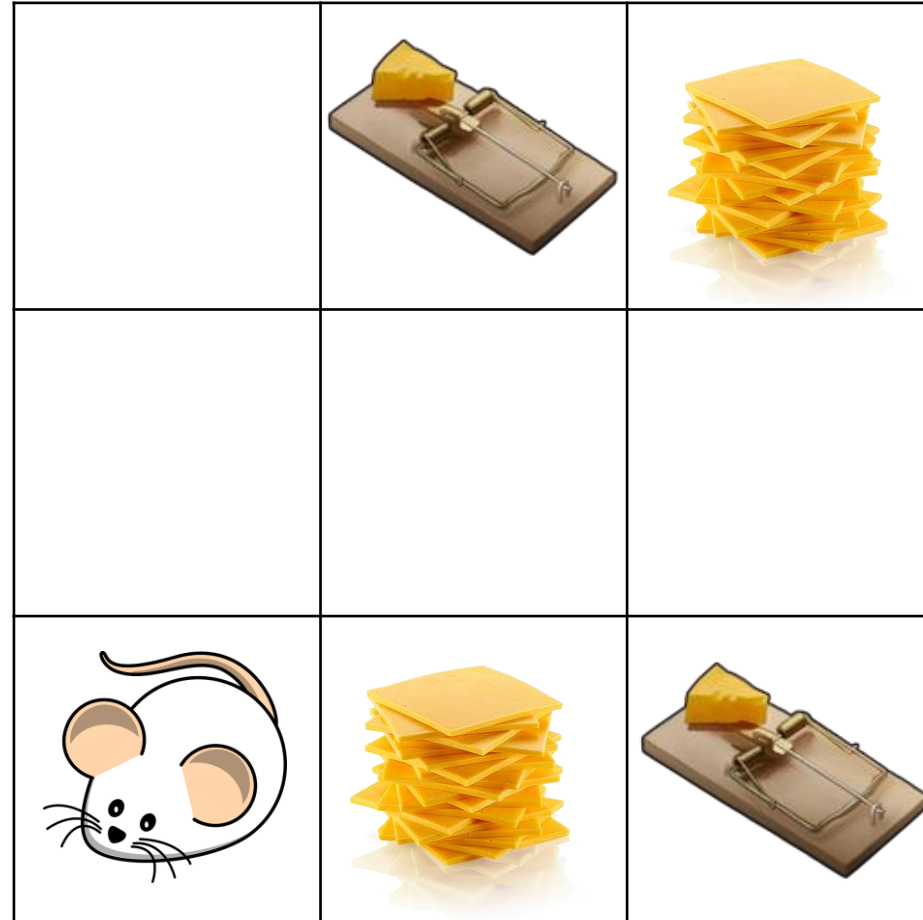
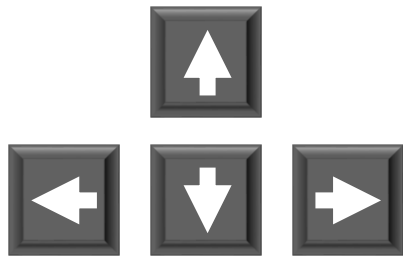
Reinforcement Learning

시행착오를 통해 발전해 나가는 학습 과정

순차적 의사결정 문제에서 누적 보상을 최대화하기 위해
시행착오를 통해 행동을 교정하는 학습 과정

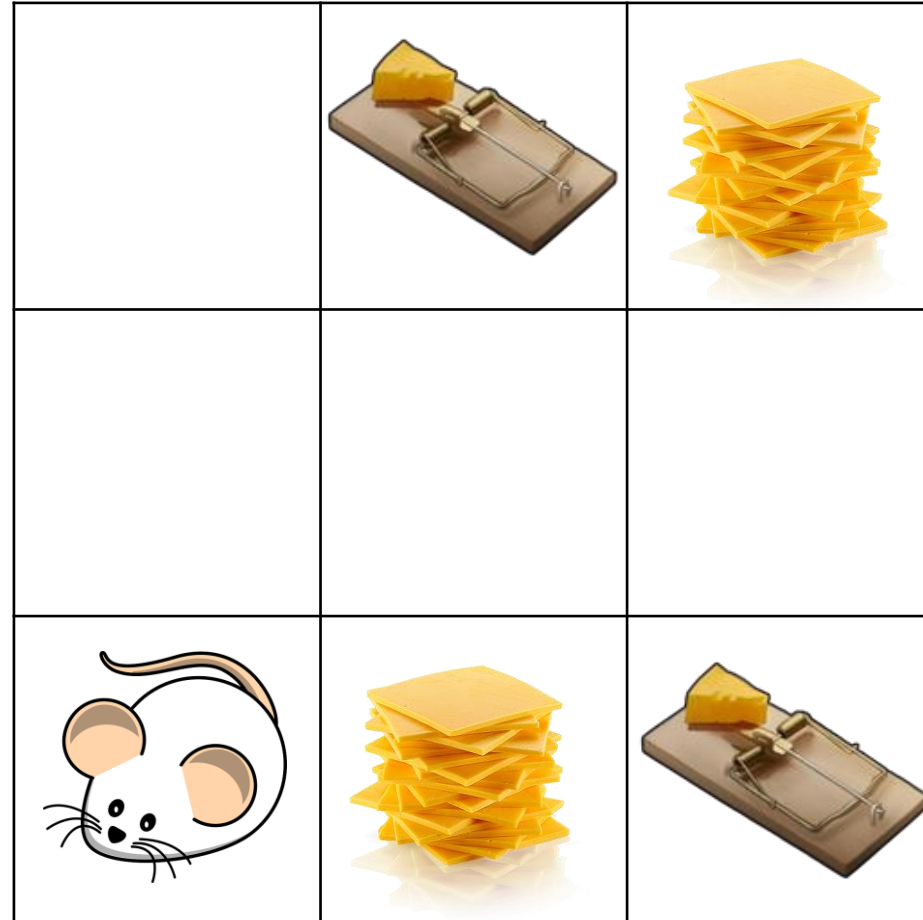
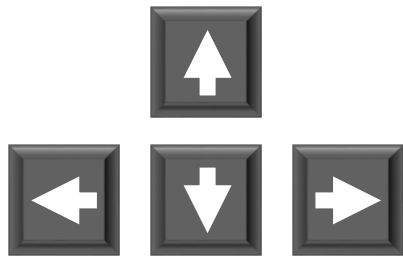
Reinforcement Learning

- 조건
네 번 움직일 수 있다
벽을 뚫을 수 없다
- 치즈 : 보상 +1
- 쥐덫 : Game Over



Reinforcement Learning

- 조건
네 번 움직일 수 있다
벽을 뚫을 수 없다
- 치즈 : 보상 +1
- 쥐덫 : Game Over



Reinforcement Learning

- 순차적 의사결정 문제



Reinforcement Learning

- 순차적 의사결정 문제의 예시 #1



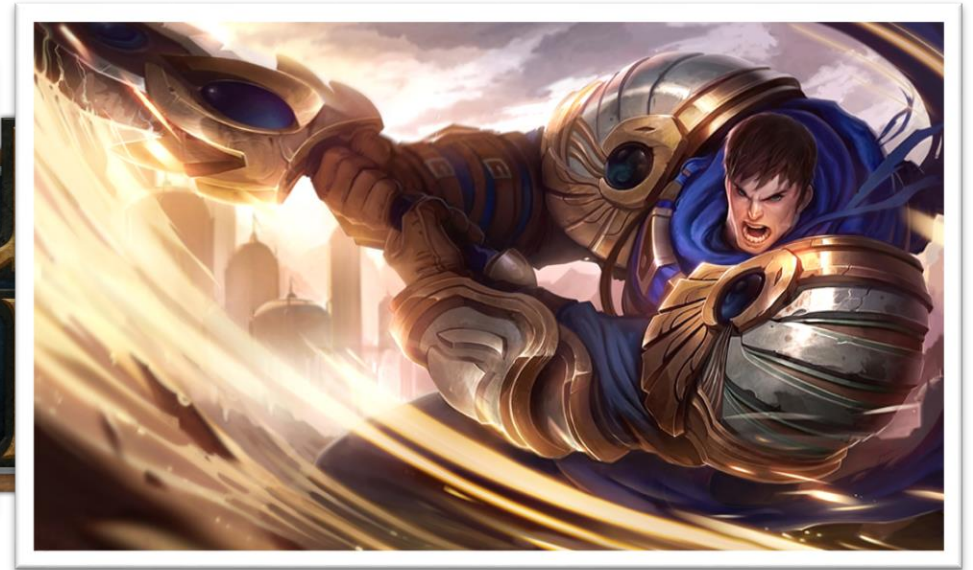
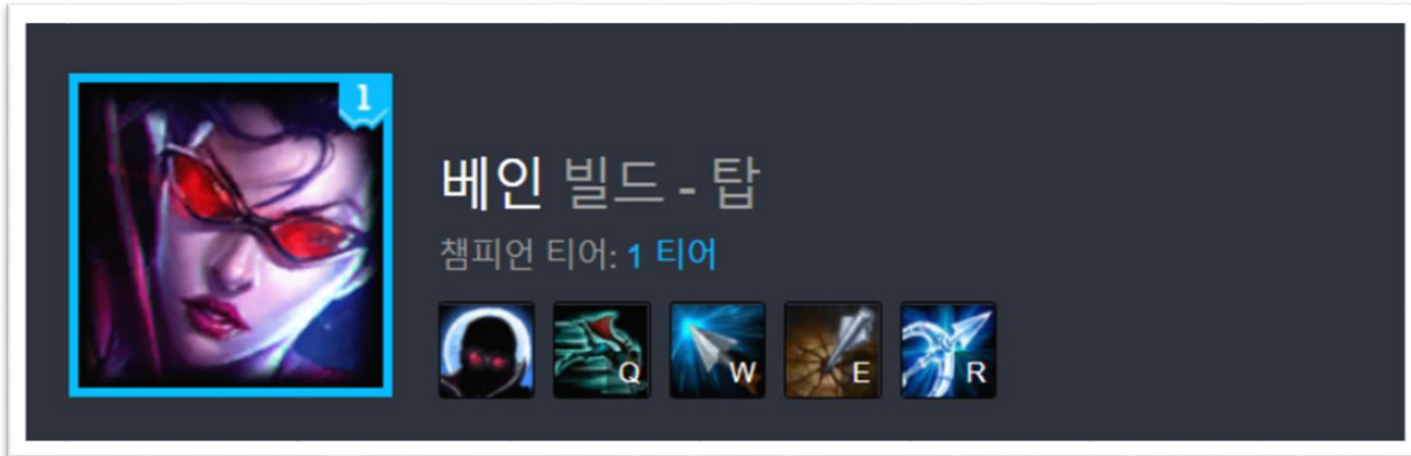
Reinforcement Learning

- 순차적 의사결정 문제의 예시 #2



Reinforcement Learning

- 순차적 의사결정 문제의 예시 #3



Reinforcement Learning(이씨)

<예시> 1m당 1점이 보상 -> 넘어지지않고 최대한 나아가는 것이 목적

페달을 세게 밟아

페달을 세게 밟는 것
= 보상이 높은 행동

행인

페달을 세게 밟는게
항상 높은 보상을 주지는 않는
구나!



1m

2m

3m

Reinforcement Learning

- 특징

미래에 대한 불확실성을 고려한다

현재 보상

만 원

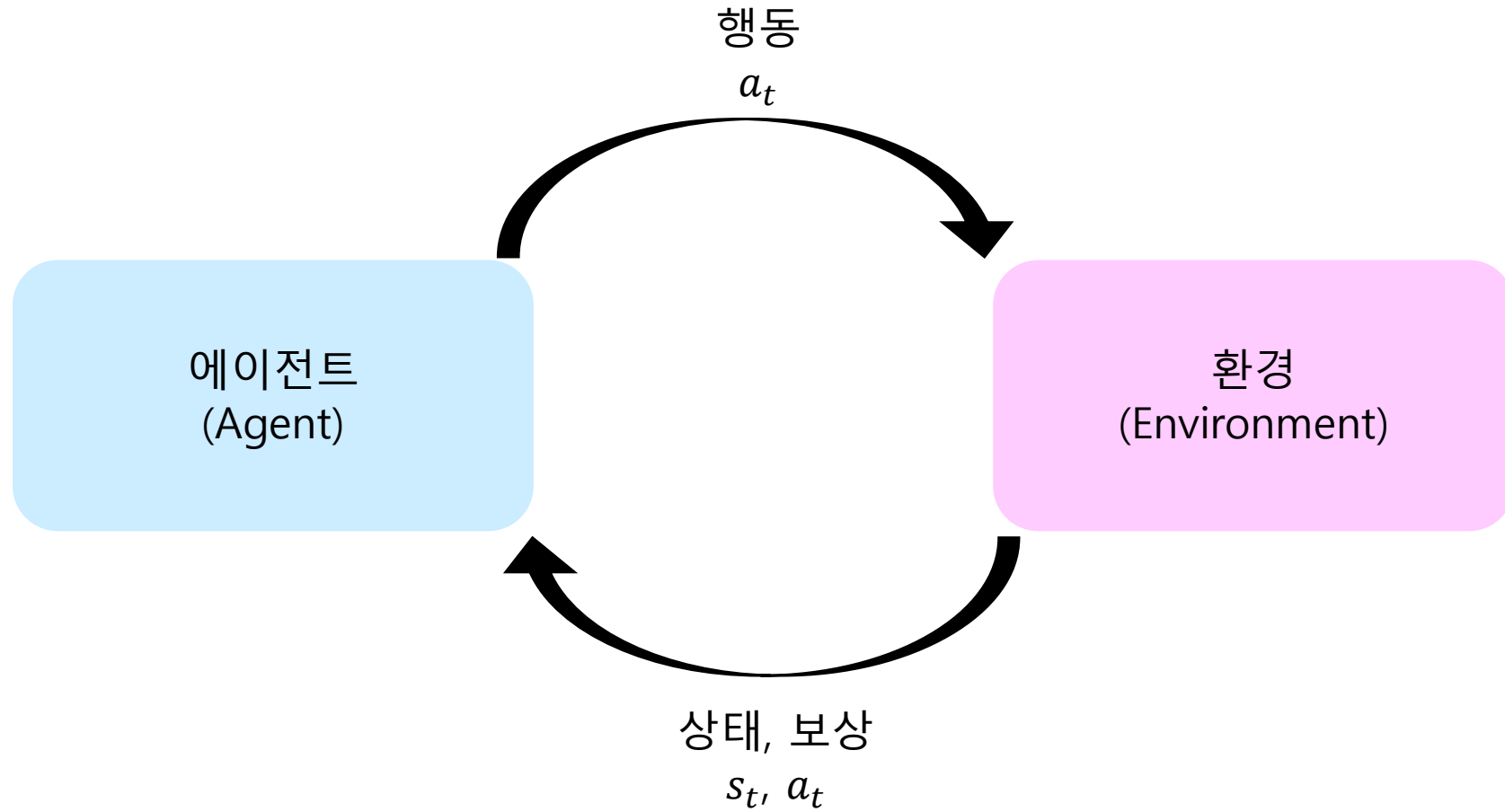


5년 후 보상

3만 원



Reinforcement Learning



Reinforcement Learning

에이전트
(Agent)

- 현재 상태 s_t 에서 어떤 행동(a_t)을 할지 결정
- 결정된 행동 a_t 를 환경으로 보냄

- 환경으로부터 보상과 다음 상태의 정보를 받음

환경
(Environment)

- 에이전트로부터 받은 행동 a_t 를 통해서 상태 변화를 일으킴
- 그 결과 상태는 $s_t \rightarrow s_{t+1}$ 로 바뀜
- Agent에게 줄 보상 r_{t+1} 도 함께 계산
- s_{t+1} 와 r_{t+1} 을 에이전트에게 전달

Reinforcement Learning

현재 상태: 0m

에이전트
(Agent)

환경
(Environment)

- 행동: 페달을 세게 밟는다

- 상태 : 2m

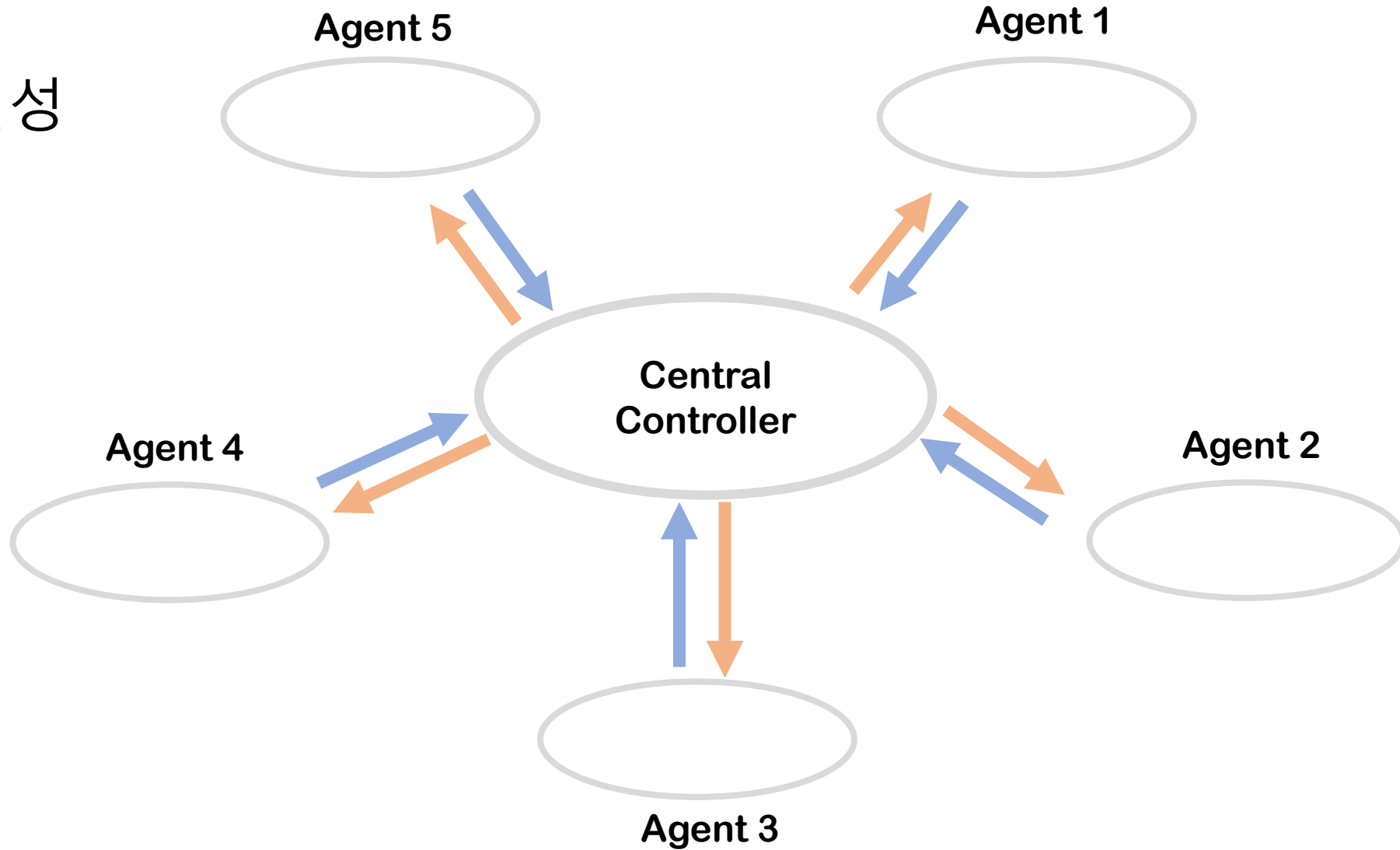
- 상태 변화 : 앞으로 2m 나아감
(0m→2m)
- 보상 : +2

Reinforcement Learning

시행착오 { (상태 1, 행동 11, 상태 21, 행동 21, 상태 31, 행동 31, ...) 정책 1 누적 보상 : 3
(상태 1, 행동 12, 상태 22, 행동 22, 상태 32, 행동 32, ...) 정책 2 누적 보상 : 7
(상태 1, 행동 13, 상태 23, 행동 23, 상태 33, 행동 33, ...) 정책 3 누적 보상 : 5

장점 #1

- 병렬성



장점 #2

- 자가 학습(Self-learning)



AlphaGo

Thank you for listening

Reinforcement Learning

- 특징 #1

직접적으로 정답을 제시하지는 않는다