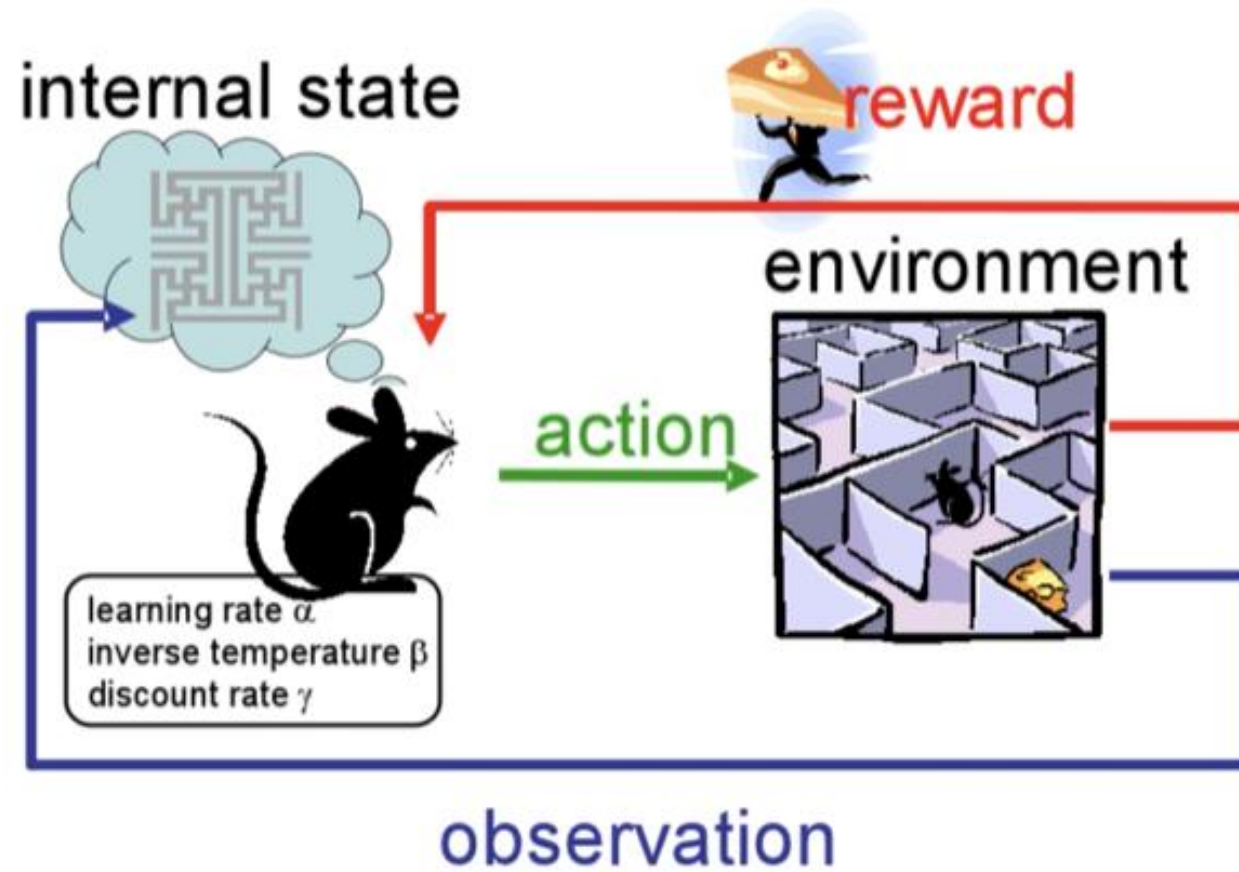


모두를 위한 RL 강좌 : Lecture 1~3

임도연

Lecture 1 : RL 수업 소개

1. 강화학습이란?



Lecture2 : Playing OpenAI GYM Games

Frozen Lake 게임을 통한 실습

S	F	F	F
F	H	F	H
F	F	F	H
H	F	F	G

S : 시작점 F : 얼어있는 땅 H : 구멍 G : 목표 지점(도착점)

Frozen Lake 게임을 통한 실습



Agent

(1) Action (right, left, up, down)



(2) state, reward



S	F	F	F
F	H	F	H
F	F	F	H
H	F	F	G

Environment

Frozen Lake 게임을 통한 실습



Agent

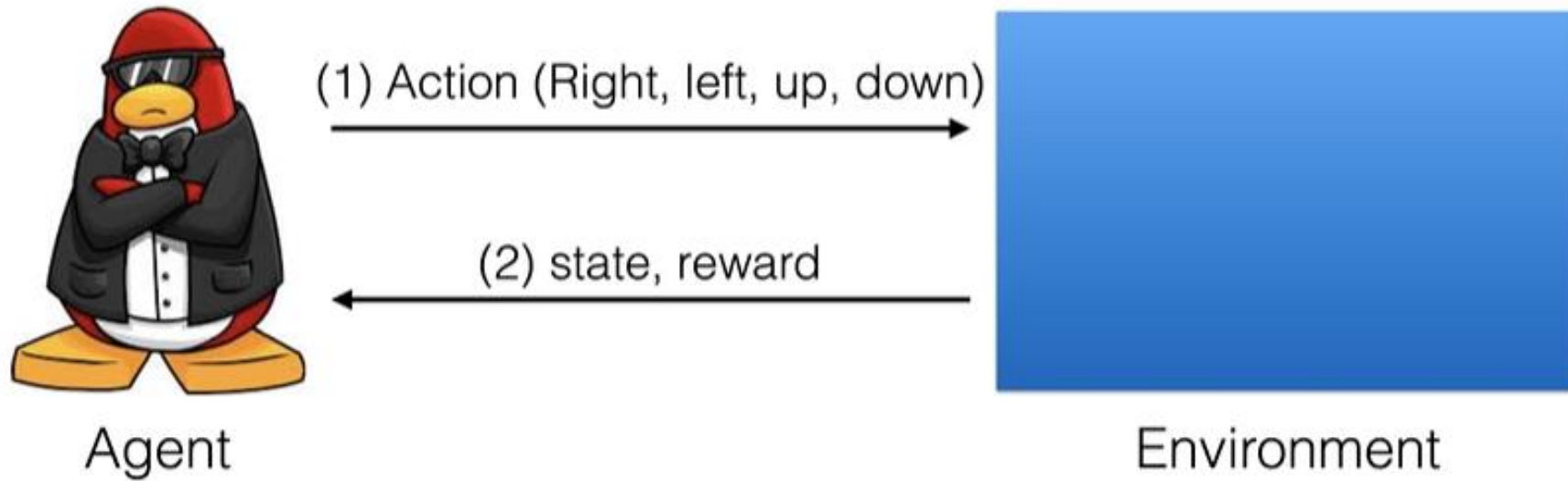
(1) action: RIGHT

(2) state: 1, reward: 0

S	F	F	F
F	H	F	H
F	F	F	H
H	F	F	G

Environment

Frozen Lake 게임을 통한 실습



Lecture3 : Dummy Q-learning

Q-learning



Q (state, action)

Q : 현재 상태에서 취한 행동의 보상에 대한 quality

Q-learning

Q (state, action)

Q (s1, LEFT): 0

Q (s1, RIGHT): 0.5

Q (s1, UP): 0

Q (s1, DOWN): 0.3

max값

Q가 가지는 최대값을 의미

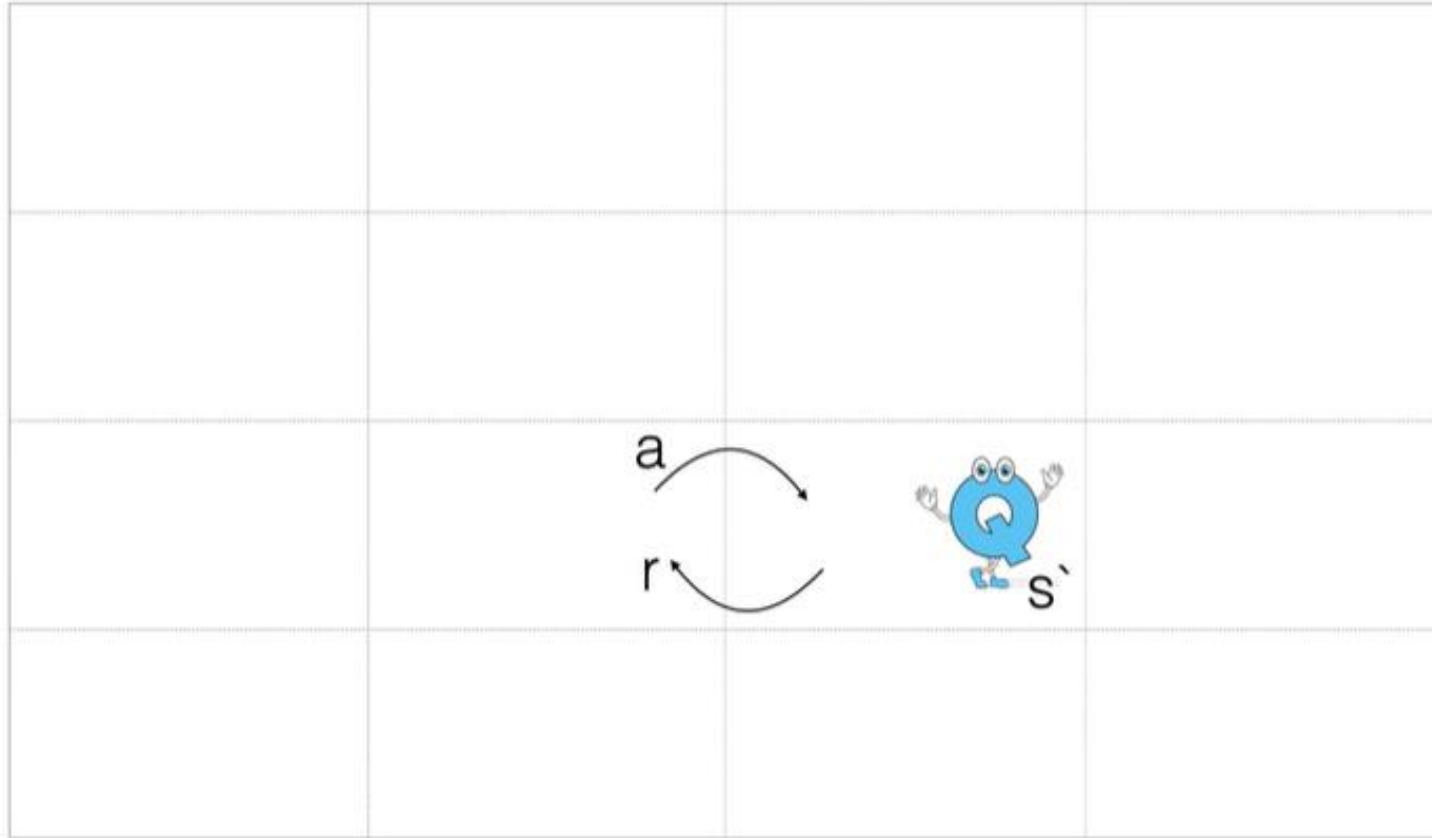
$$\text{Max } Q = \max_{a'} Q(s, a')$$

$$\pi^*(s) = \operatorname{argmax}_a Q(s, a)$$

최대값일 때 가지게 되는 변수 값

* : 최적의 값을 의미

Q-learning



Q-learning

S	F	F	F
F	H	F	H
F	F	F	H
H	F	F	G

$s_0, a_0, r_1, s_1, a_1, r_2, \dots, s_{n-1}, a_{n-1}, r_n, s_n$

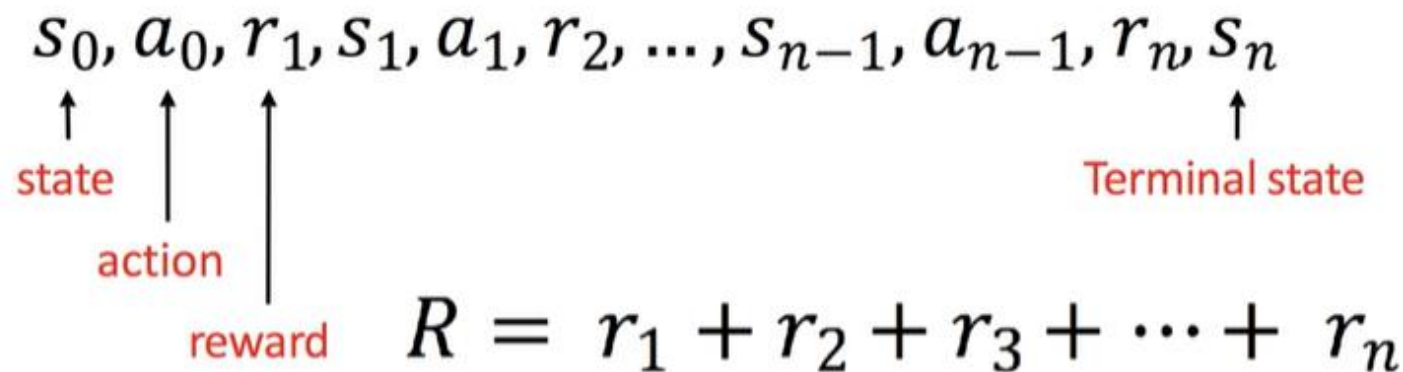
↑
state

↑
action

↑
reward

↑
Terminal state

Q-learning



$$R_t = r_t + r_{t+1} + r_{t+2} + \dots + r_n$$

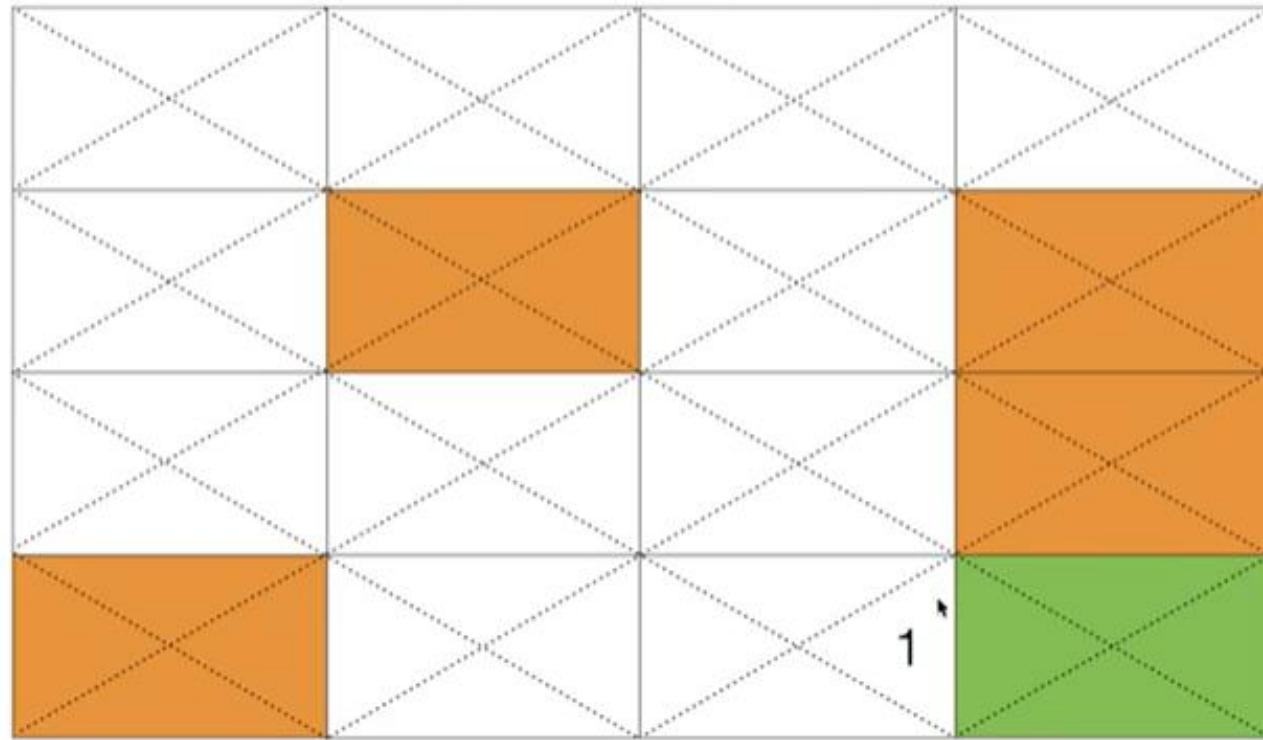
$$R(t+1) = r_{t+1} + r_{t+2} + \dots + r_n$$

$$R(t) = r_t + R(t+1)$$

$$Q(s,a) = r + \max_{a'} Q(s', a')$$

$$R(t)^* = r_t + \max R(t+1)$$

Q-learning



$$Q(s_{14}, a_{\text{right}}) = r = 1$$

$$Q(s_{13}, a_{\text{right}}) = r + \max(Q(s_{14}, a)) = 0 + \max(0, 0, 1, 0) = 1$$

Q-learning

For each s, a initialize table entry $\hat{Q}(s, a) \leftarrow 0$

Observe current state s

Do forever:

- Select an action a and execute it
- Receive immediate reward r
- Observe the new state s'
- Update the table entry for $\hat{Q}(s, a)$ as follows:

$$\hat{Q}(s, a) \leftarrow r + \max_{a'} \hat{Q}(s', a')$$

- $s \leftarrow s'$

Q&A