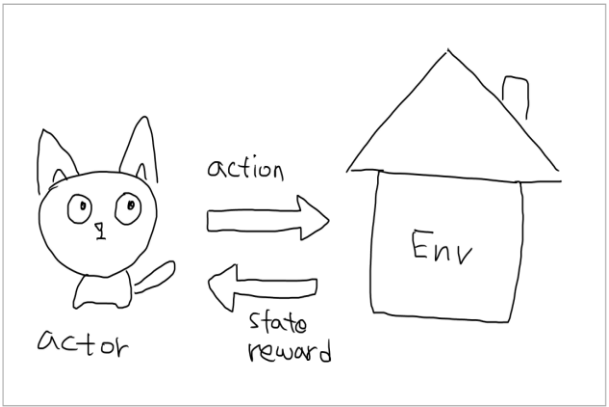
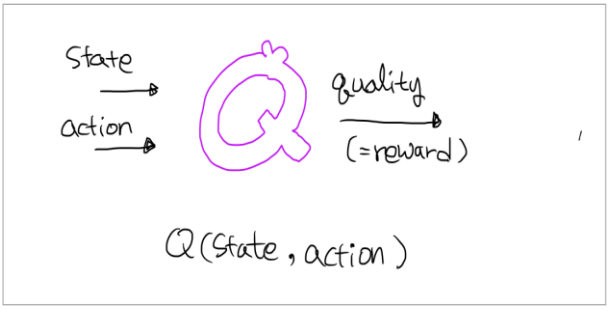
Reinforce Learning

- 강화학습은 state 와 reward 를 통해 학습한다.
- **actor** 가 주어진 **Environment** 에서 어떤 **action** 을 취할 때마다 변화된 **state** 와 **reward** 를 얻게 된다.

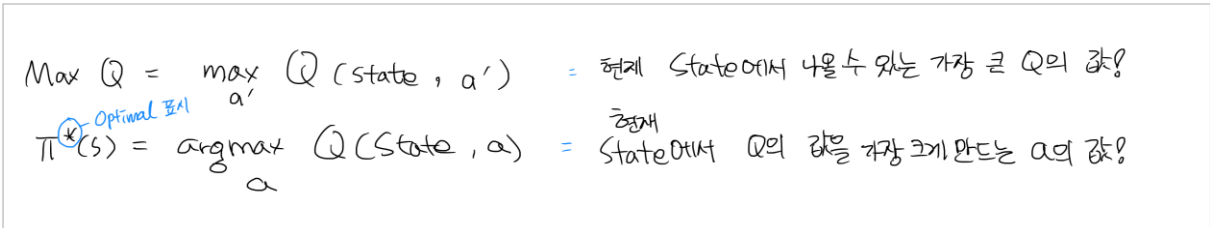


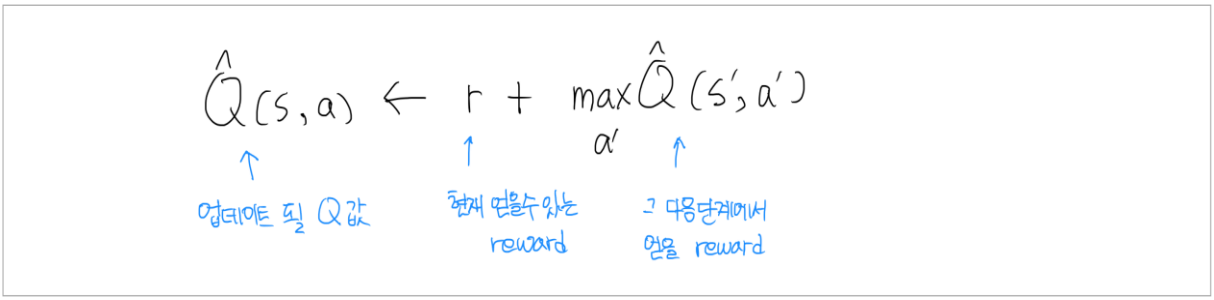
Q-learning



- Q 에 **State** 와 **action 을** 입력하면 **quality (reward)** 를 출력한다

Optimal policy (π), and Max Q





Q-algorithm

1 Q (s, a) ← 0 : Q 의 모든 값을 0 으로 초기화한다.
2 환경을 만들고 상태 (state) 를 가져온다.

(③ ~ ⑥ 을 무한 반복한다.)

1 어떤 행동 (action) 을 취한다.
2 행동에 대한 보상 (reward) 과 변화된 상태 (state') 를 받는다.
3 $\hat{Q}(s,a) \leftarrow r + \max_{a'} Q(s',a')$
: 보상 (reward) 과 다음 상태 (state') 의 최대 Q 값을 더해서 Q 를 업데이트 한다.
4 상태 (state) 를 변화된 상태 (state') 로 바꾼다.