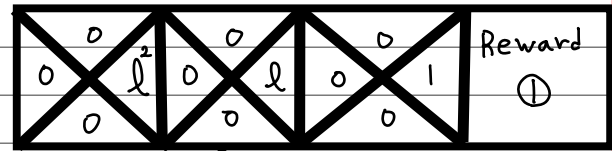


Q-learning

시작			목표

④ Q-Update



$$Q(s_t, a_t) \leftarrow (1-\alpha)Q(s_t, a_t) + \alpha(R_t + \gamma \max_{a'} Q(s_{t+1}, a'))$$

업데이트해라. 기존값 새로운값과 업데이트했을 때 at를 했을 때 받는 목적지 Reward 다음 선택의 최대값

① Greedy action (탐욕적인 행동)

action ①	section ⑤				Reward ①

② Exploration (탐험)

• Greedy

↳ 목적: 새로운 Reward, 새로운 Path 저장 가능

↳ 문제: Exploration 만 계속한다. → 해결: Step이 지날수록 E 값 감소 선택한다.

				Reward ①

③ Discount factor (인자 감소)

↳ 목적: 최적의 경로 찾는다.

				Reward ①