시즌 RL - Lecture 03

노트북:	모두를 위한 머신러닝		
만든 날짜:	2019-01-17 오후 2:21	**수정한 날짜:**	2019-01-17 오후 3:27
작성자:	rr		
태그:	#모두를 위한, .Lecture		

Lecture 03

= Q-function

state, action -Q-> quality(reward)

1. max를 찾는다
2. max의 arg를 따라간다

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

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

optimal policy: 가장 최적화된 정책, *

= Learning Q

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

= Q-learning algorithm

1. 테이블을 0으로 초기화 시킨다
2. 현재 상태를 가져온다 (s)
3. 아래 내용을 무한 반복한다

   어떤 액션을 한다 (a)
   a에 따른 reward를 받는다
   a의 결과로 현재 상태 s에서 s`으로 이동한다
   현재 상태에서 이 액션에 대한 Q를 업데이트 한다
   r과 다음 상태인 s`에서의 max Q 값을 더해서 Q를 계속 업데이트 해 나간다 (Learning Q 식)

4. 무한 반복을 하게 되면 Q가 학습이 된다