약간의 맛집 찾기 느낌 reward 가 높은 곳으로 따라가기

- Q-learning

시작해서 맛집까지 찾아나가기

맵이 주어져 있다면 쭉 직진할터인데 이동해가면서 맛집을 찾아내가는 과정 이게 에피소드인 것인데 우예 우예 이동하다가 맛집으로 이동한다.

어떻게 맛집을 들어갔더라 하는 걸 알게되는

Greedy action

이동한 것에 대한 점수를 매길 것이고 reward 가 큰 곳으로 이동할 것

처음에는 다 스코어가 0 이라서 랜덤하게 움직이게 될 것

이 한칸을 state 라고 하는데 이 상태들이 이동하게 될 것이다.

오른쪽으로 간 점수를 쓰고 아래 점수를 쓸 것이다.

처음에는 다 0 이었는데 어쩌다 보니 맛집으로 가서 성공했다.

그럼 그 성공한 직전의 그 방향에 대한 발자취를 남겨놓는다.

맛집 R = 1 이런게 있다면 reward 를 쓰게 된다.

Greedy action

이게 왜 이렇게 움직이는 것에 대한것은 수식으로 증명이 될 것이다.

아직은 다 0인데 다 0인 상황에서 랜덤하게 고르게 될 것

q 러닝은 이동하면서 바로 바로 업데이트를 하게 되는데

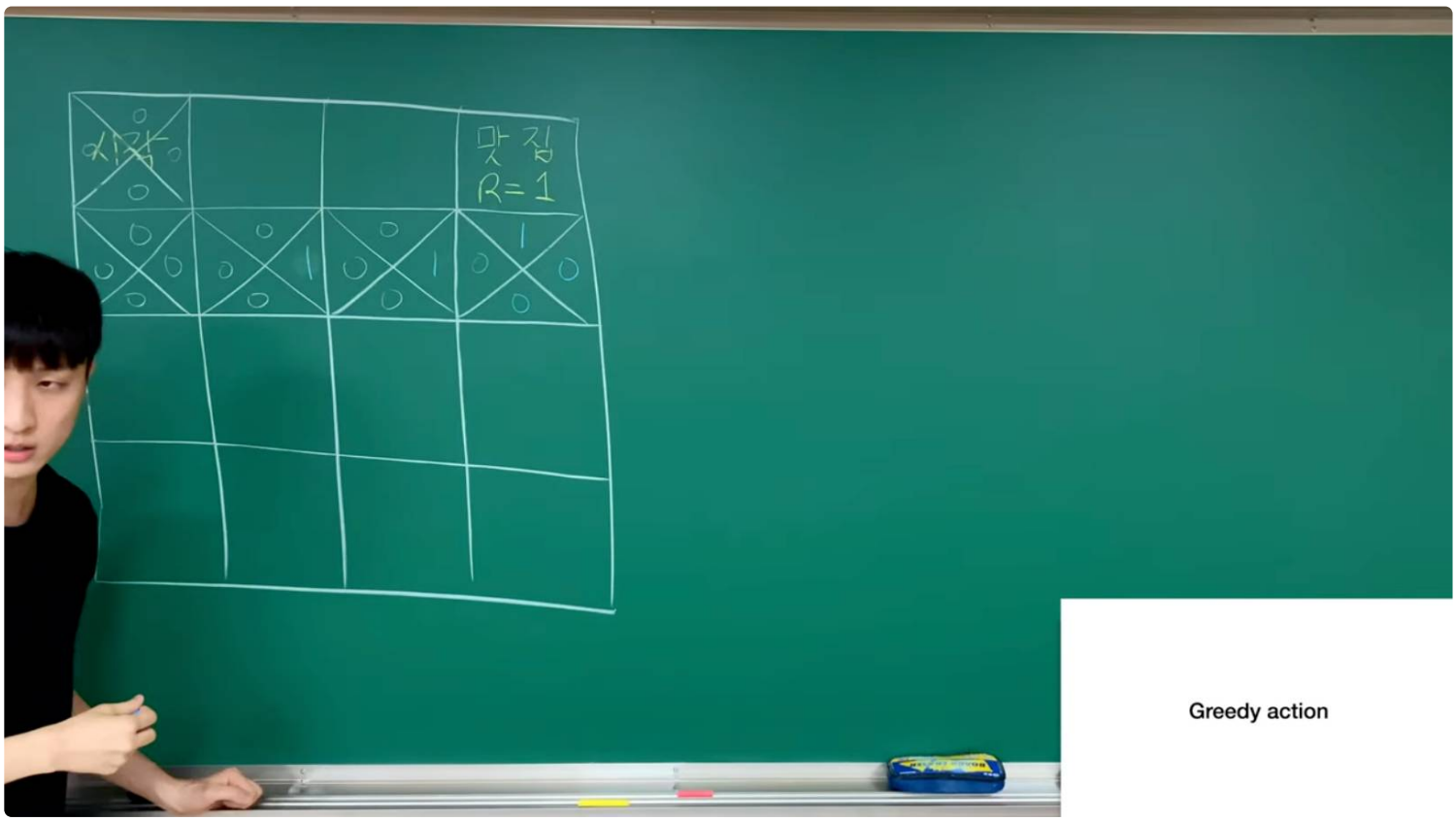
가장 큰 값을 옆에도 새긴다.

계속해서 평가를 한다.

옆으로 갔는데 어 뭐가 있네? 하면 값이 업데이트 되는것이다.

여기서 가장 큰값을 스코어링해라

시작을 했다 아래를 랜덤하게 지목했다고 하자 전부다 0이니 업데이트가 안된다. 옆에도 다 0 업데이트 x 마침 옆에 위치한 어떤것에 1이 있다면 1이 적혀짐



근데... 이러면 계속 똑같은 루트를 돌게 될 것이다.

! 대안

Exploration

이러한 탐험이　　앱실론 - greedy

앱실론은 매우 작은 수이다

10 퍼센트 안에 만큼은 랜덤하게 골라서 이동한다.

너무 그리디한 액션만 취해서도 안되고 너무 랜덤하게 액션을 취해서도 안된다...

그럼 어떤 방법이 있을까

! Exploration & Exploitation

익스플로레이션 -> 그냥 탐험

익스플로티에이션 -> 이 판을 이용하자는 것 이 판을 보고 greedy 하게 이동

이 두개에는 trade off 가 존재하는데

1. 새로운 path 를 찾음
2. 새로운 맛집을 찾음

(어떤 방식으로 둬서 바둑을 둬야 이기는지에 대한걸 아는게 목표일 수도 있다.)

바둑의 수가 하나로만 정해져 있는 것은 아니니까

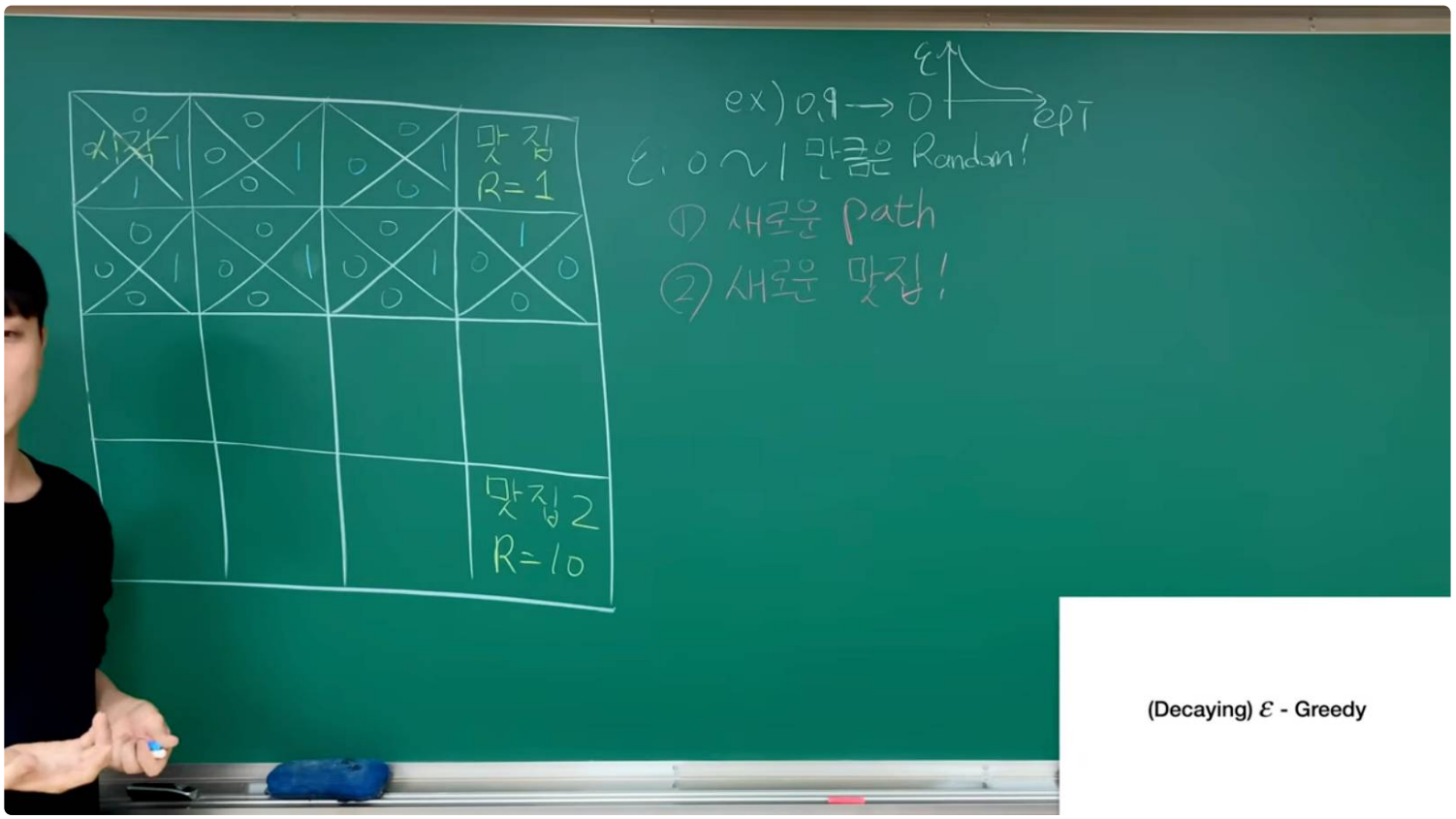
e - greedy (decaying)

0. 9 -> 0 으로 줄여나가는거

처음에는 아무떼나 수 막 뒀다가 나중에는 점점 이 랜덤함을 줄여나가는 것

epi 소드 별로 e 를 줄여나가는 것 점점 exploration 을 줄여나가는 것

지금 더 낫고 더 안좋다가 없다 경로 마다 덜 좋고 더 좋고가 없음



(Decaying) ε - Greedy

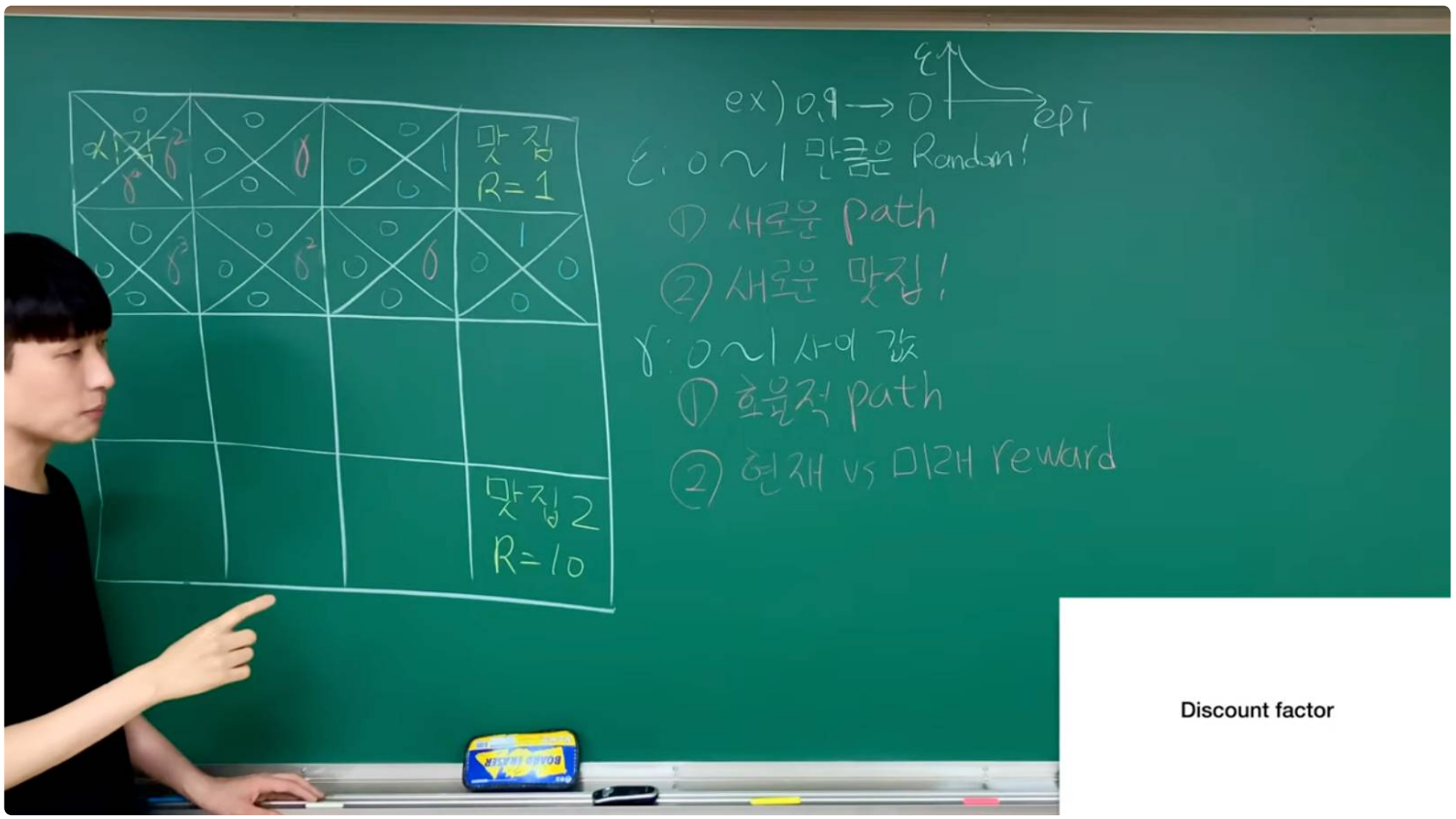
그래서 여기 등장하는게 discount factor 이다

감마라고 함 path 를 좀더 효율적인 것을 찾는 것

0~1 사이의 값을 찾는 것

감마 만큼을 곱해서 복사하는 것 그냥 갖고 오지 말고 r 를 곱해라

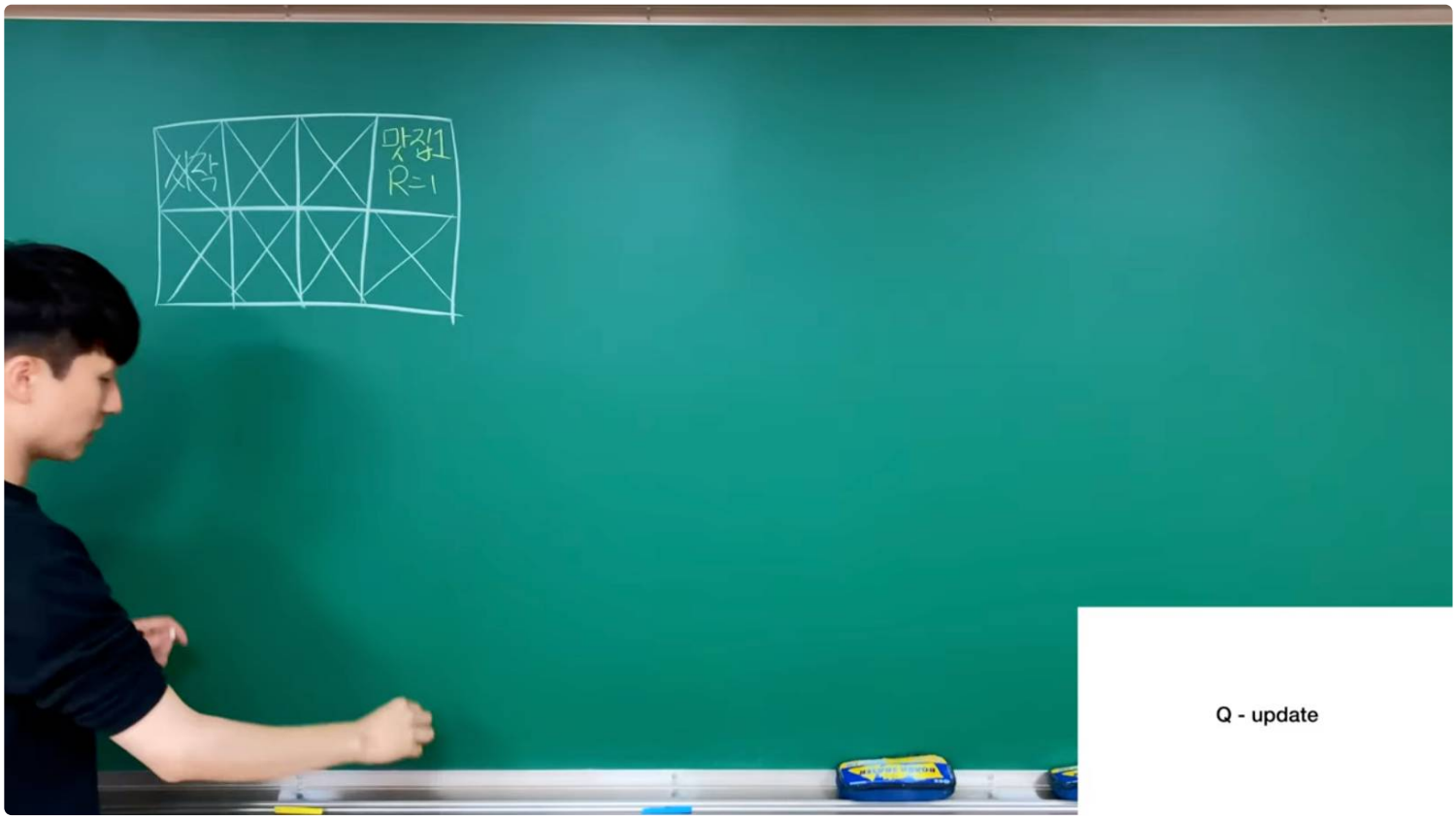
시작 점에서는 그럼 젤 적게 감마값이 곱해진 것을 선택하게 되고 효율적인 path 를 선택할 수 있게 된다.



Discount factor

감마가 작을 수록 미래의 리워드를 생각을 하게됨

곱해질수록 더 작아지니까 설레발 치는 경우가 적어짐

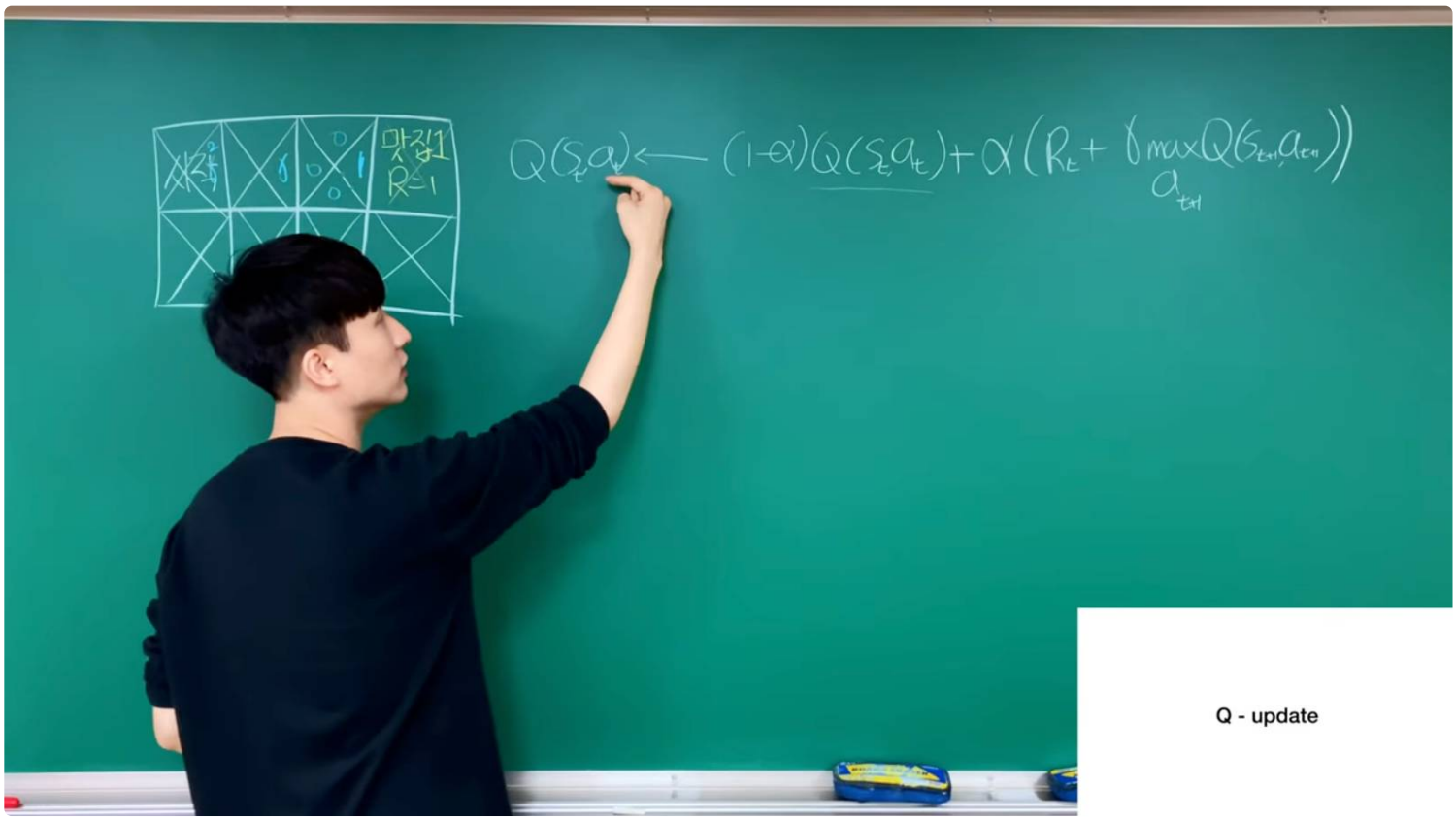


Q - update

업데이트 할 때 q 값들을 업데이트 해나가는데 그대로 가져 오는 것이다.

하지만 업데이트가 좀 더 부드럽게 실행된다.

r 가 아니라 좀 다르게 업데이트를 함



q 값을 이 뒤에 값으로 업데이트를 해라

오른쪽의 값을 여기로 업데이트 해라

알파는 0~1 의 값인데

식을 보면 a 가 크면 오른쪽을 더 신경 쓰겠단 것이고

a가 작으면 왼쪽을 더 신경쓰겠다는 것이다.

우리는 at 를 했을 때 받는 reward 를 rt 라고 표현한다.

만약에 해당 계산 후 이동할 위치에 reward 가 속하지 않는다면 rt 값은 0이 된다.

action 들 중에 젤 큰것을 보면 1이라서 그 다음 저 q 값은 1이 되고 감마를 곱한 후 a를 곱하게 된다.

다음 판에서 a 들 중에서 제일 큰거 그래서 max 가 붙음

(a 가 0.5 이라고 한다)

저 앞에는 원래 가지고 있는 값인데

Q(St, At) 는 처음에는 0일 것이다

만약 똑같이 감마인데 a 가 껴들었으니...

원래 값에다가 더해서 넣어주는 것! 직역하면 새로운것을 얼마나 많이 받아들이냐에 대한 것이다.

사실 이 수식이 나오는 것은 다른 이유가 있긴 한데

자기를 유지하고 새로운것을 받아들인다는 것 그 의미

지그믄 의미만 받아들여라