|  |  |
| --- | --- |
| 과목명 | 강화학습 (4-3) |
| 이름 | 이덕수 |
| 학번 | 20161600 |

이름과 학번을 적고, 답변을 적어주세요.

실습: 4주차 수업 내용을 자유롭게 요약하세요(A4 1 page 분량, 넘어도 상관없음)

- TD학습은 MC학습 처럼 특정 환경의 MDP 정보가 없다고 가정하는

모델프리학습 이며 경험으로부터 직접 학습한다는 점에서 MC학습과 동일함

차이점은 MC 학습은 하나의 에피소드가 끝날 때마다 Q함수를 업데이트 하지만

TD 학습은 상태 변화가 있을 때마다 Q함수를 업데이트함(학습률) 고로 하나의 에피소드가 완전히 끝나지 않아도 단계별 학습 가능 또한

MC 학습은 분산이 크지만,편향이 작고 TD 학습은 분산이 작지만 편향이 큼

가치 반복법: 가치 함수를 산출해서 최적의 정책을 찾는 방법

환경의 MDP 정보를 알고 있다는 가정 하에 진행

각 상태에서 취할수 있는 모든 행동의 가치 함수를 업데이트

TD 학습: 에피소드에서 에이전트가 거치는 상태와 행동에만 Q함수를 업데이트

(전체의 일부를 샘플링)

탐욕 알고리즘: 가치 반복법에서 사용했던 항상 더 큰 가치 함수 또는 Q 함수를 얻기 위한 행동을 선택하는 방식

(행동 공간이 넓은 환경에서는 가치 함수를 최대한 이용한 탐욕 알고리즘 으로 수렴한 정책보다 더 좋은 정책이 나올 가능성 존재 , MC , TD 학습 모두 환경 내 상태 행동을 샘플링 해서 학습에 반영하기에 최적의 정책을 찾아낼 가능성이 낮아짐)

앱실론-탐욕 알고리즘: 에이전트가 가치 함수를 이용하면서 새로운 영역을 탐험할수 있도록 여지를 주는 알고리즘

Q 학습: 행동 정책과 학습 정책을 서로 다르게 하여 특정 상태에서 어떤 결정을 내리는 것이 미래 보상을 가장 높여줄 것인지에 대한 정책 데이터를 지속적으로 업데이트 하는 알고리즘