최종과제 구현보고서

Simple과 original 메소드를 구현하였다.

모델 설명

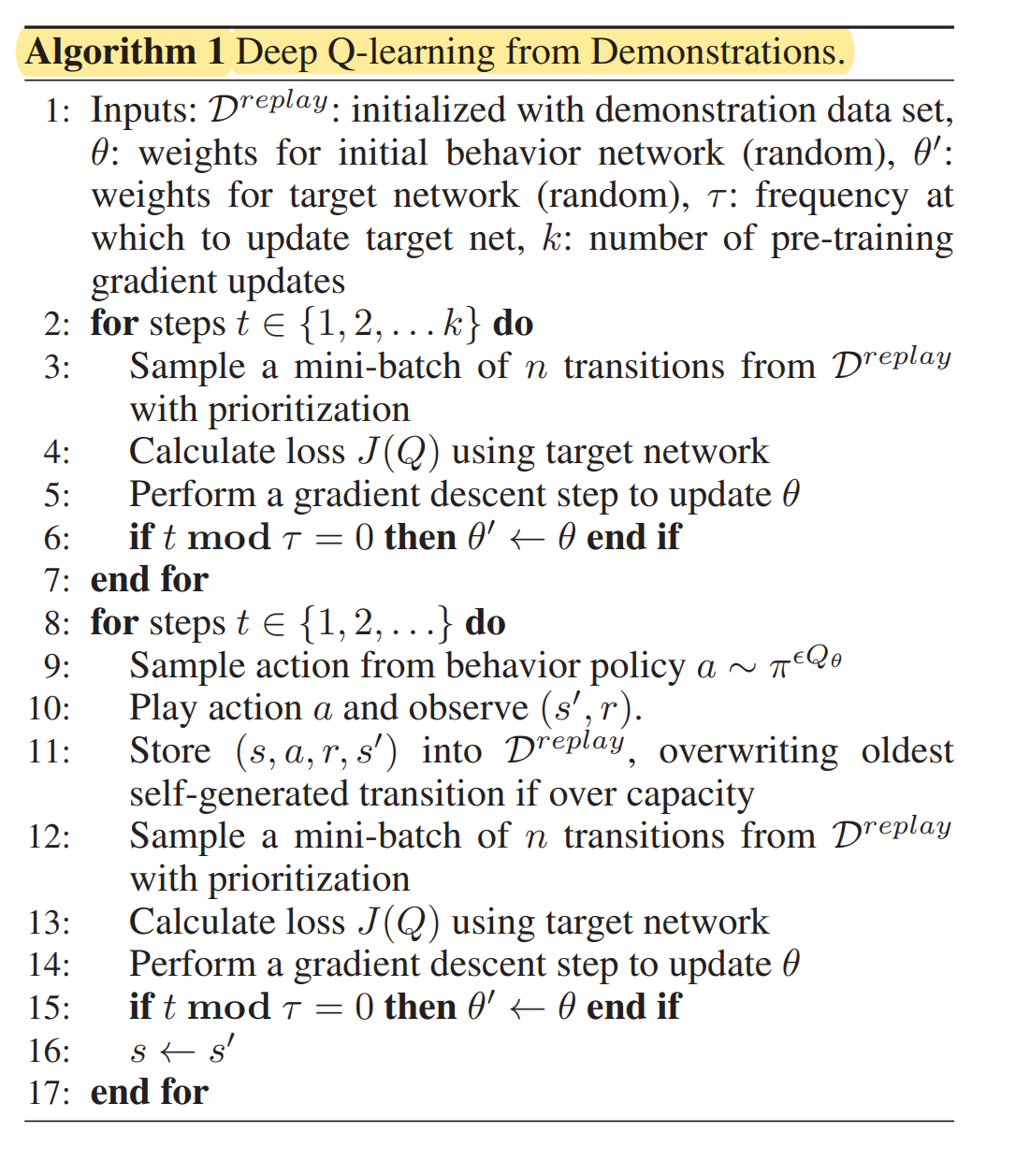
Linear 층 세개로 이루어진 다층 퍼셉트론

은닉층의 뉴런 개수는 30개로 정함

Xavier 균등분포 초기화를 사용하였음

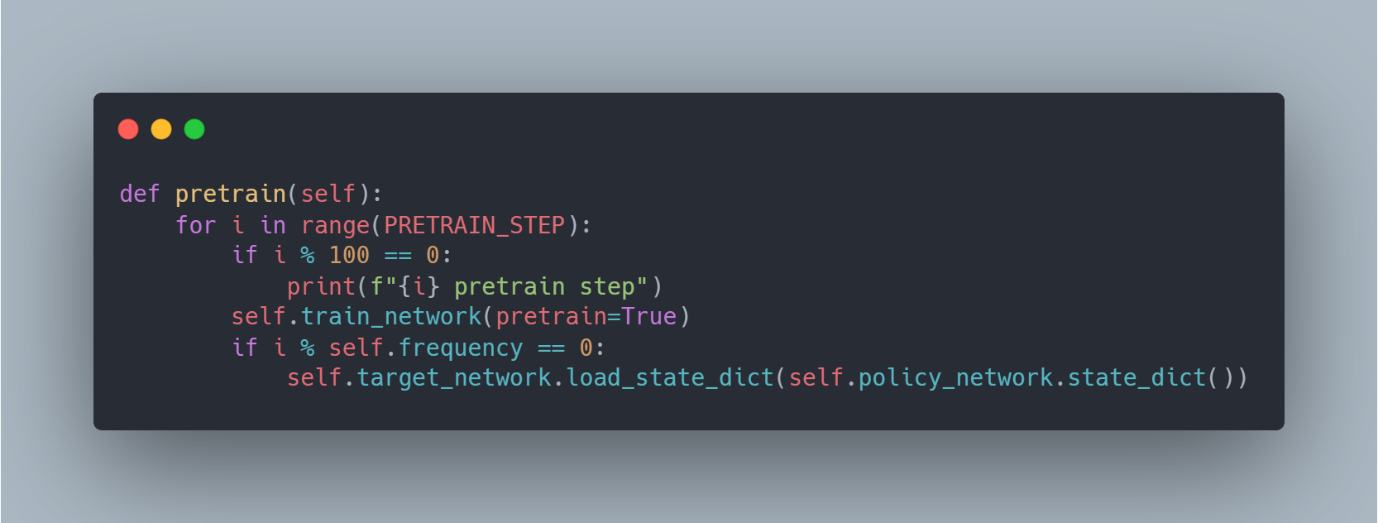
Adam 최적화 함수를 사용하고, L2 손실 대신 MSE 손실 함수를 사용하였음

알고리즘과 코드 비교



2-6줄 pretrain 단계:

100회마다 출력, 정책 네트워크를 훈련시키고 frequency회마다 타겟 네트워크를 업데이트



손실 함수 계산이 포함된 학습 과정

미니배치를 샘플링하고 손실 함수 네가지를 계산한다.

기댓값을 의미하는 Q 함수를 신경망으로 근사하는 DQN 방식을 사용

람다값은 임의로 0.23으로 정하고, 마진 함수의 양수값은 임의로 0.2로 정했음

Q값 계산은 논문의 식 그대로 사용

n-step returns는 미니배치의 다음 값들을 사용

loss 함수는 람다 계수를 취한 모든 손실 함수들의 합

loss 함수를 구한 뒤 오차역전파법으로 손실을 학습



미니배치 샘플링, 메모리 클래스

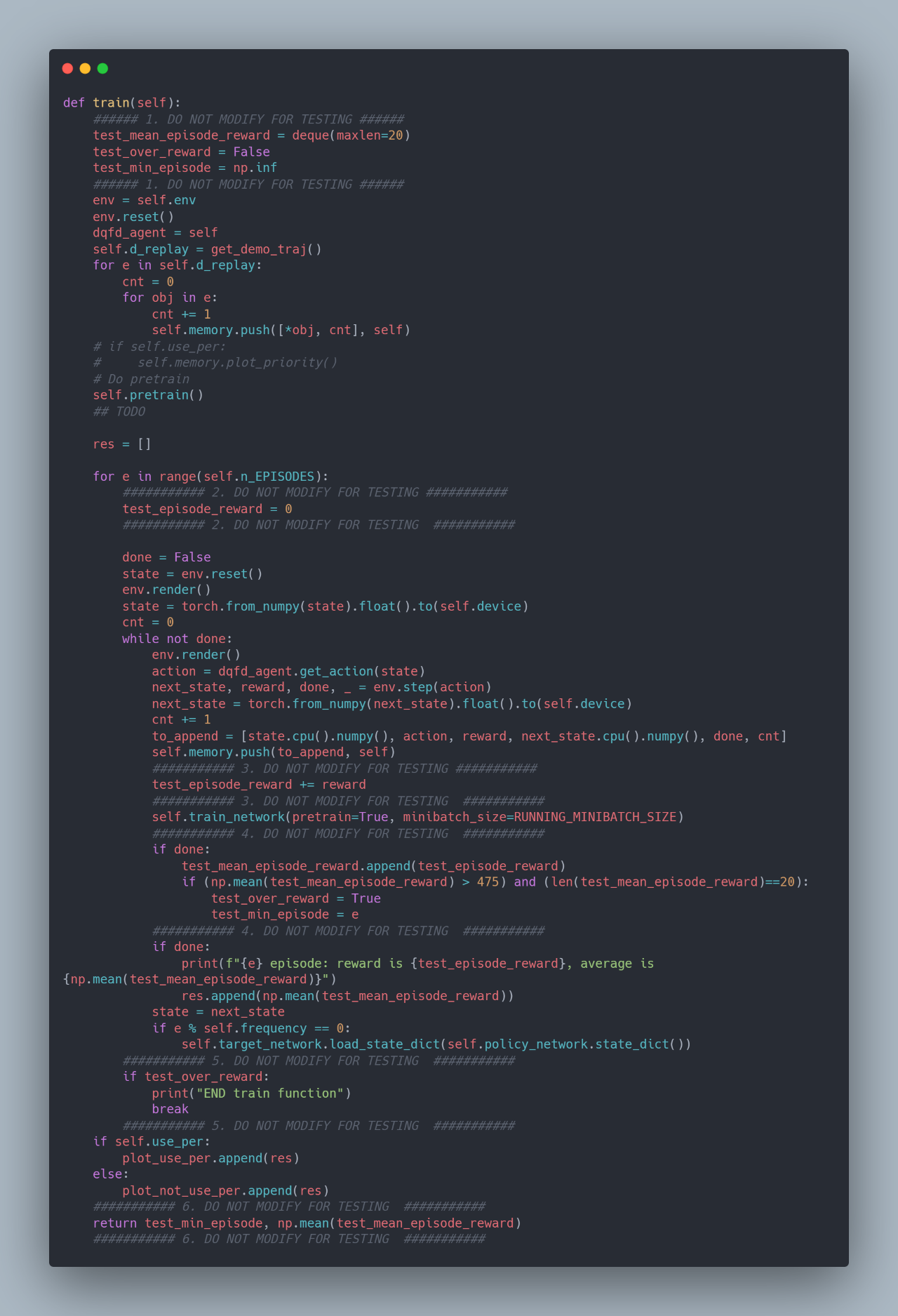
버퍼를 만들고 인덱싱한다

버퍼가 꽉 찼을시에는 처음 인덱스부터 덮어쓰기

push할때 PER를 계산한다.



train 메소드



그래프는 pyplot을 agent\_1에 메인 함수를 생성하고 연동해 작성하였다.

