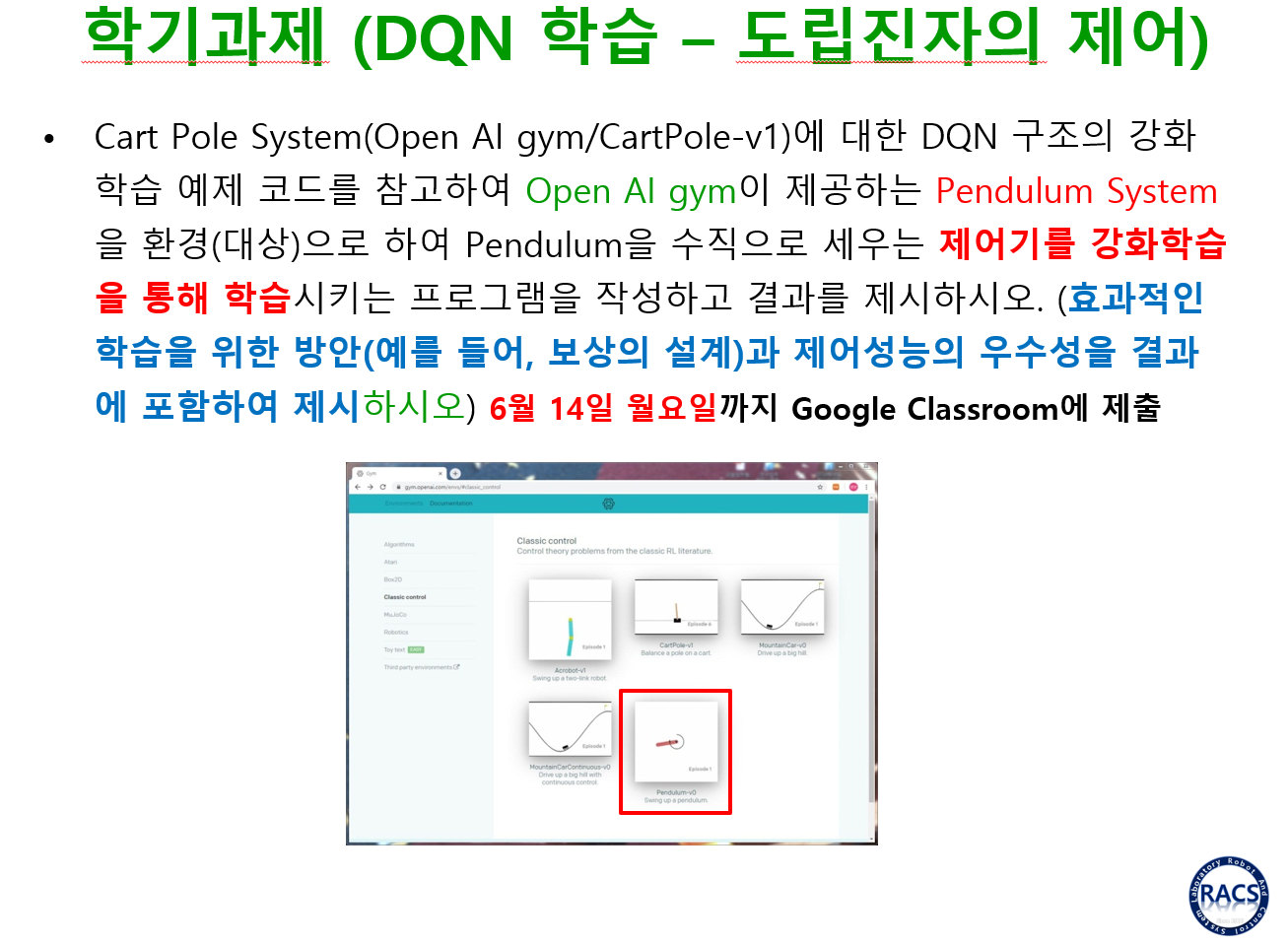
지능시스템

기말대체 과제

2019305059

이현수



작업공간(이산, 연속) 및 관측공간(이산, 연속)을 보고 알고리즘을 분류할 수 있디.

1. Q-learning은 개별 작업 공간과 개별 관측 공간을 처리할 수 있다.
2. DQN은 개별 작업 공간과 이산 및 연속 관측 공간을 모두 처리할 수 있다.
3. DDPG는 연속 작업 공간과 이산 및 연속 작업 공간을 모두 처리할 수 있다.

진자 환경과 같이 지속적인 작업 공간이있는 경우Q-learning만으로는 연속 작업 공간에서 행동하는 방법을 배울 수 없으므로 정책 그라데이션 모델을 ddpg로 사용해야 한다.

Q-learning에서는 Q-값을 극대화하는 액션을 선택하려고 한다. 이때 단순한 방법으로 가능한 각 작업을 하나씩 살펴보고 Q 값이 가장 높은 것을 선택하는 것이다. 그러나 연속 공간에서는 무한한 양의 가능한 작업이므로 이 방법을 사용할 수 없다.

지속적인 작업을 처리하는 방법에는 여러 가지가 있다. 정책 그라데이션 방법, 소프트 액터 비평가, DDPG 등과 같은 많은 알고리즘이 있다.

이번 과제에서는 DDPG를 사용해 구현했다.

**DDPG(심도 결정적 정책 그라데이션)는** 연속적인 작업을 학습하기 위한 모델이 없는 오프정책 알고리즘이다.

**DPG(결정적 정책 그라데이션)**와 **DQN(Deep Q-Network)**의 아이디어를 결합.

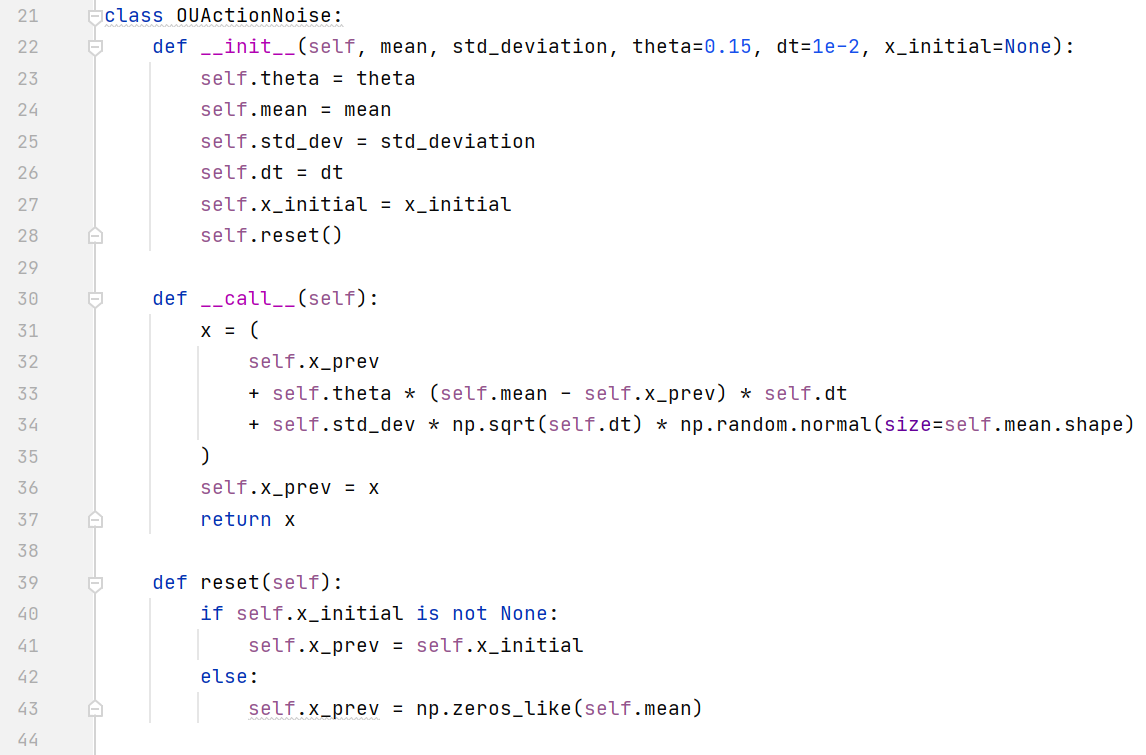
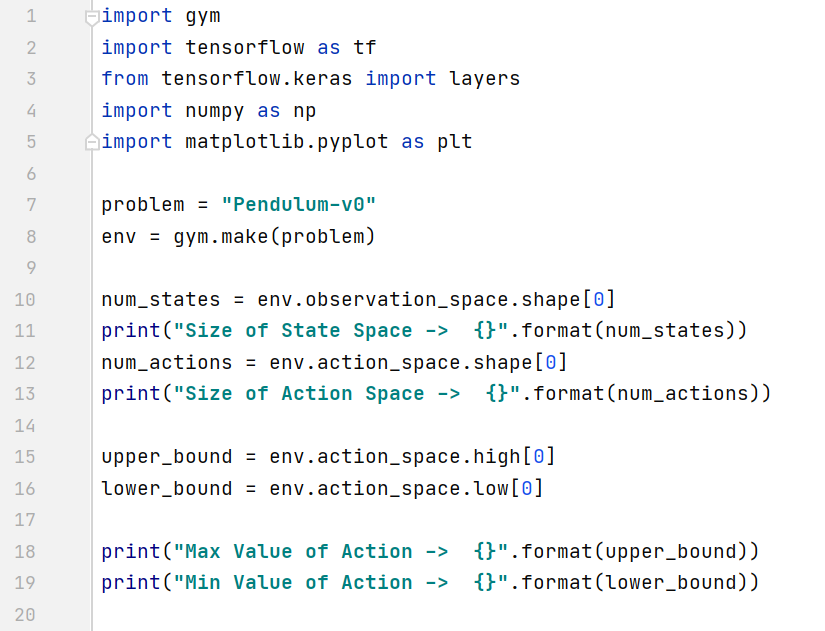
DQN의 익스피리언스 리플레이 및 슬로우 러닝 타겟 네트워크를 사용하며, 연속 작업 공간을 통해 작동할 수 있는 DPG를 기반.

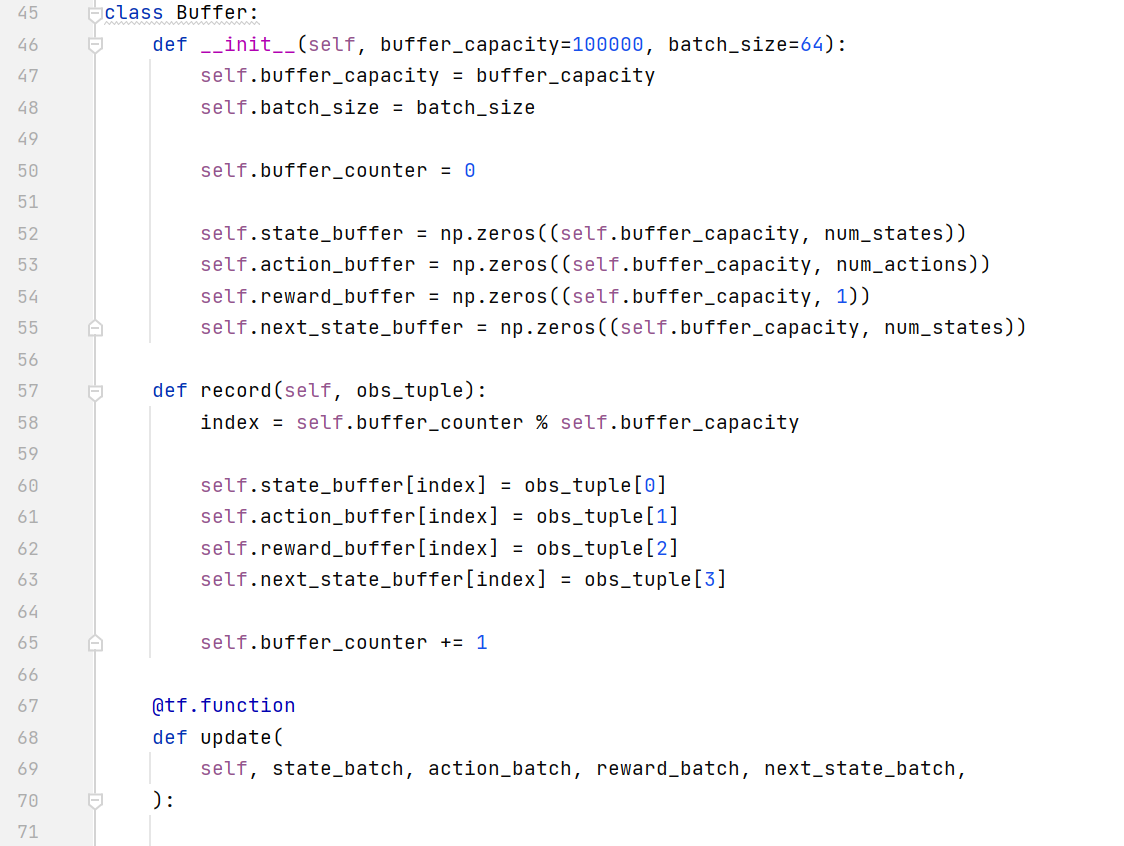
DDPG는 2개의 네트워크를 보유하여 작동한다.

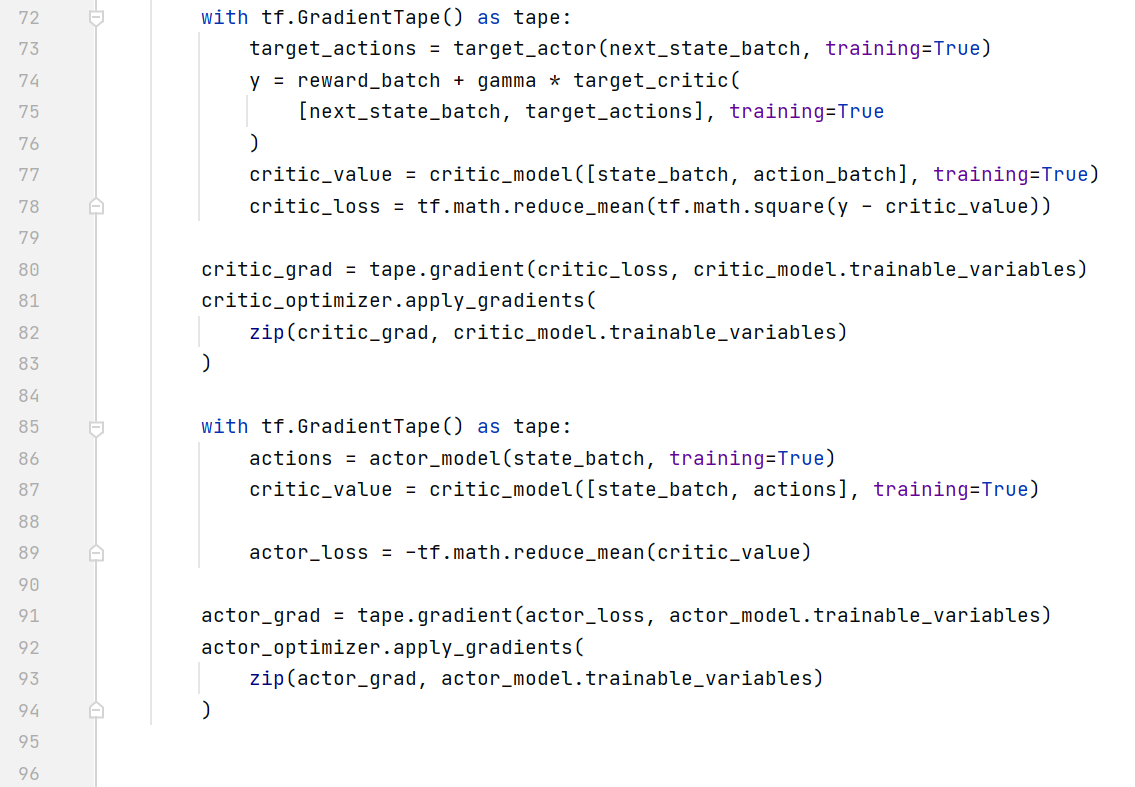
첫 번째 네트워크는 Q-값을 예측하는 법을 학습하고

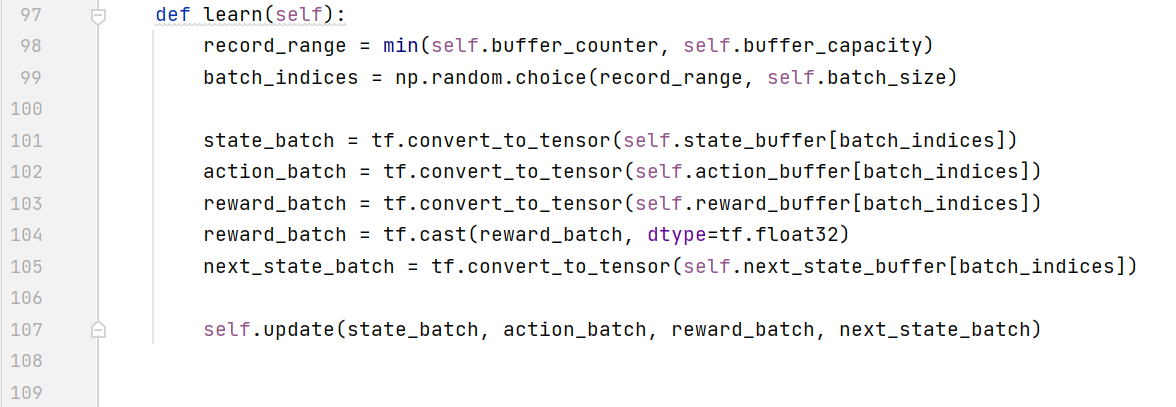
두 번째 네트워크는 주어진 상태 작업을 선택하는 방법을 학습한다.

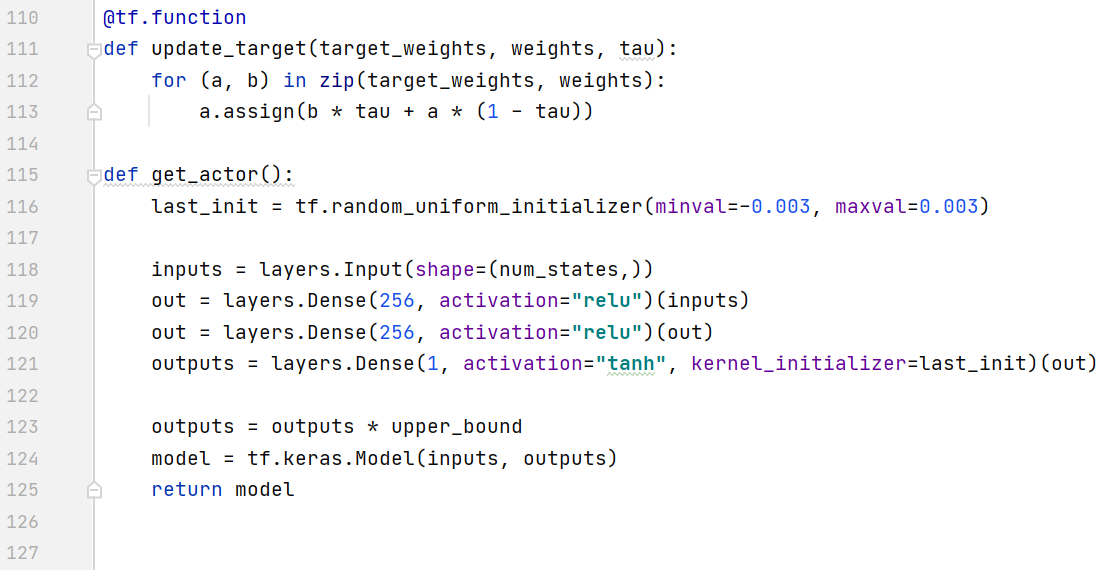
**<코드>**

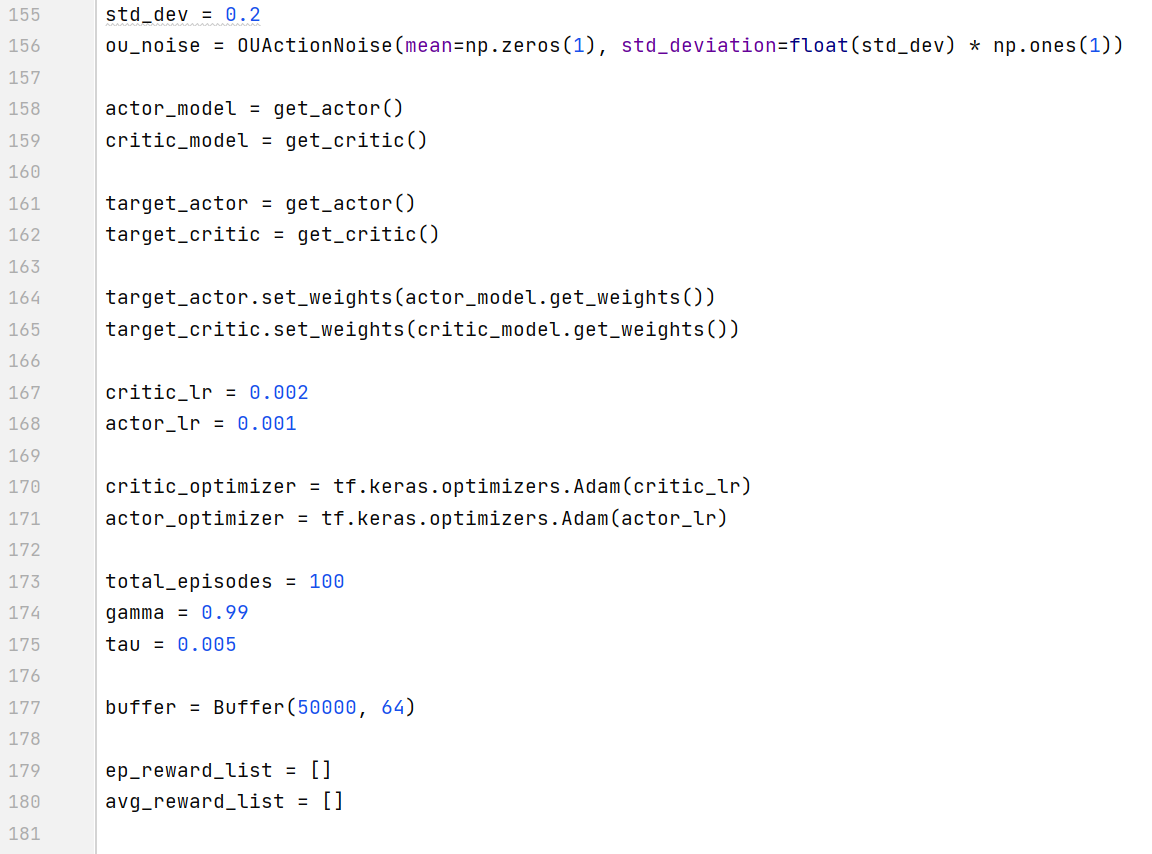
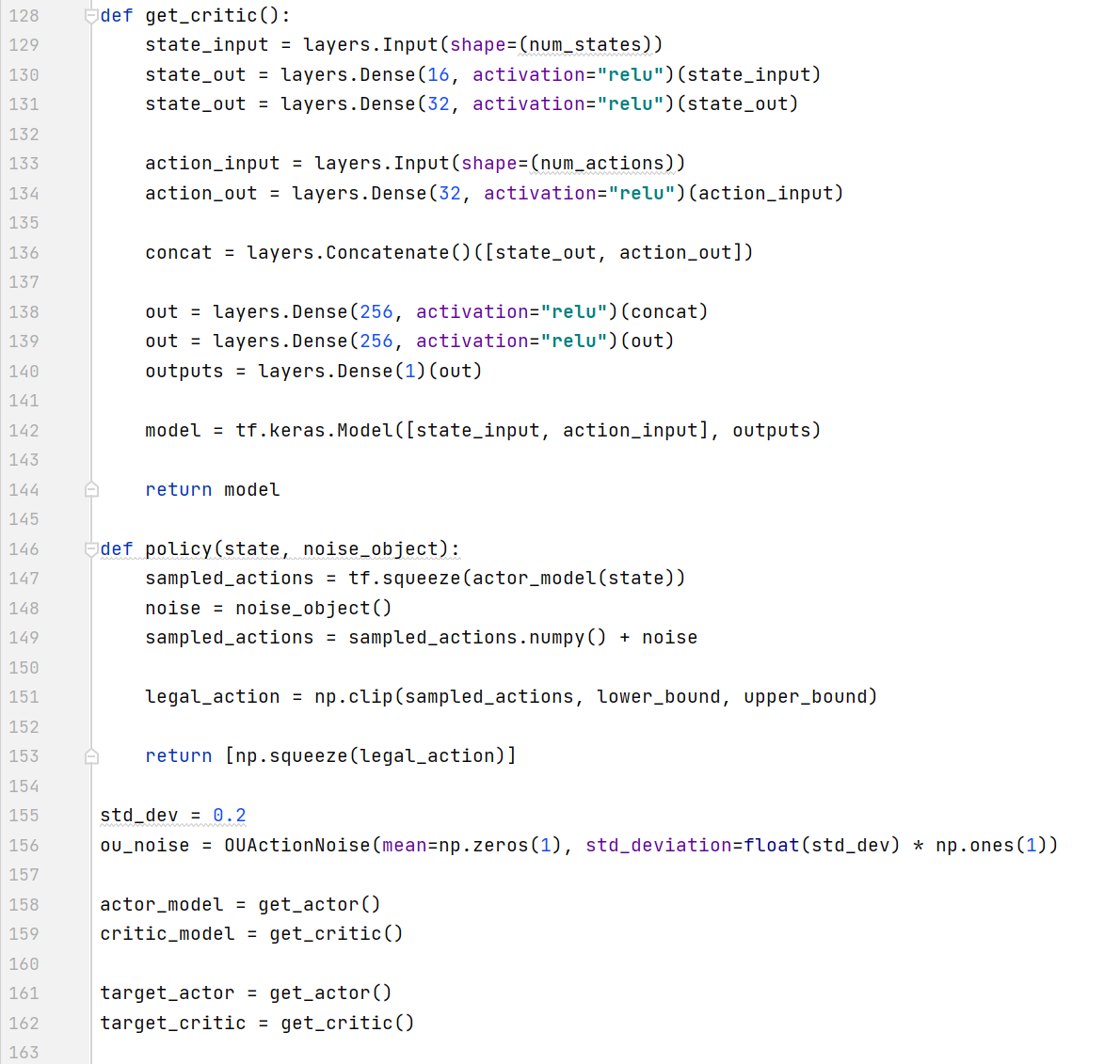


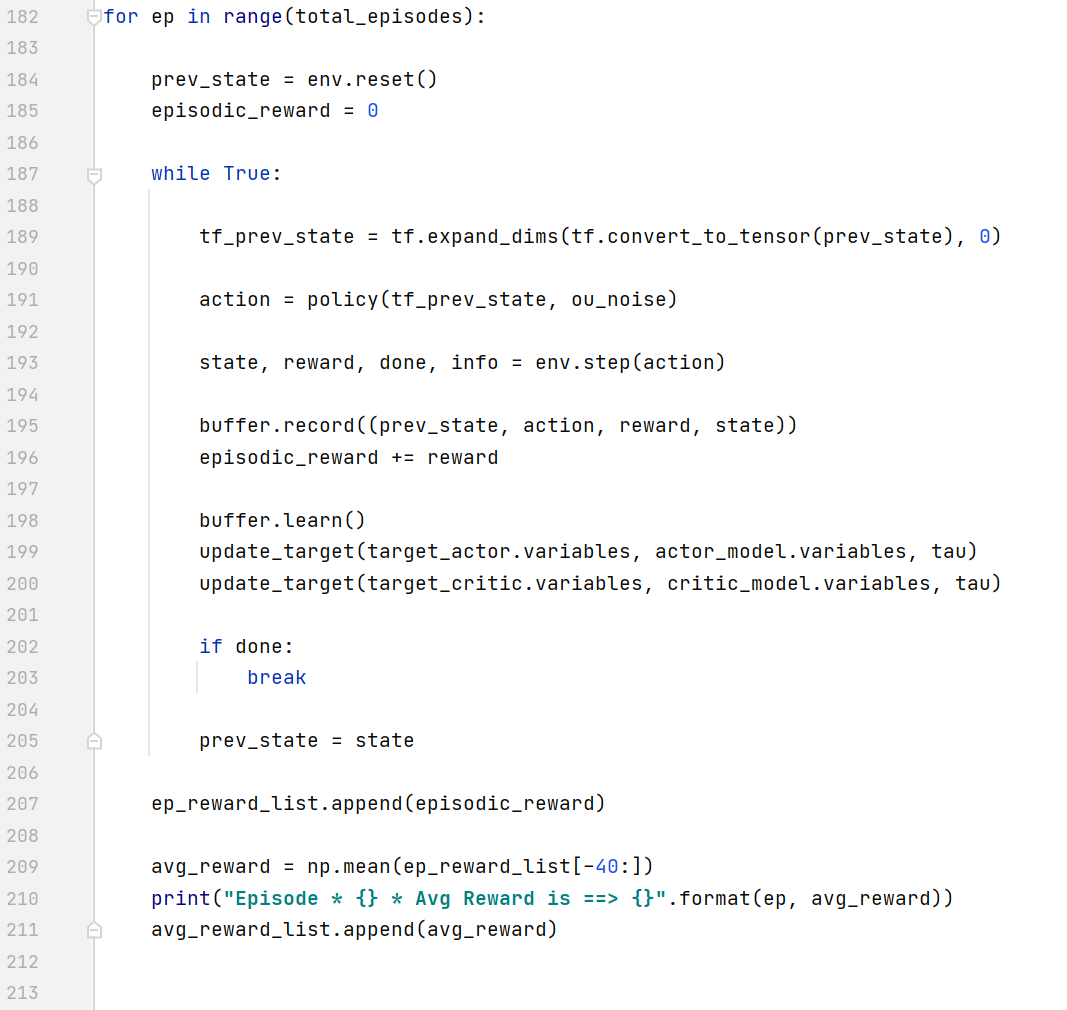


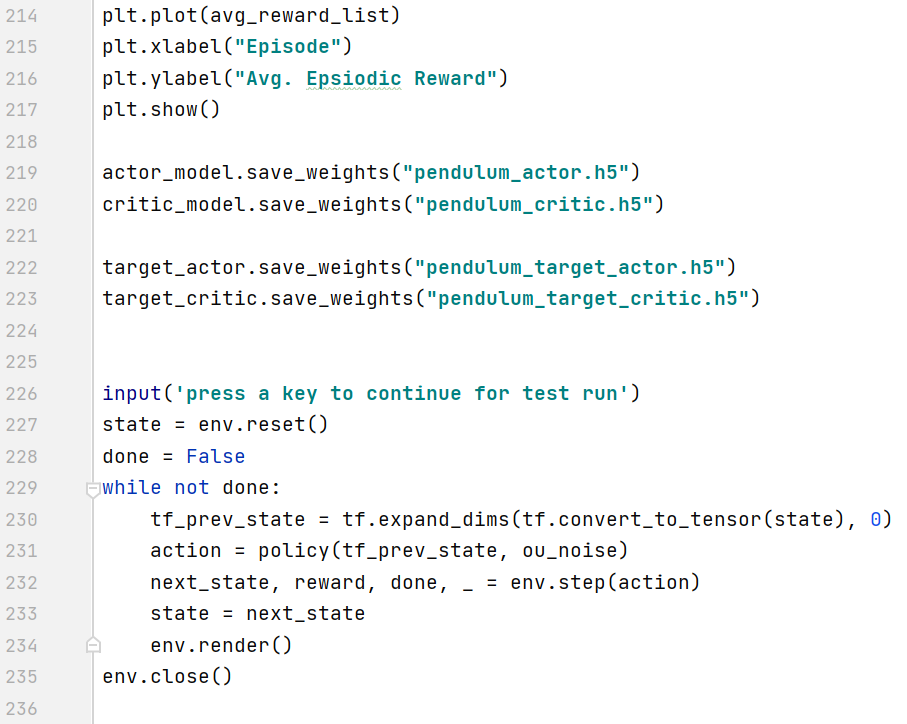








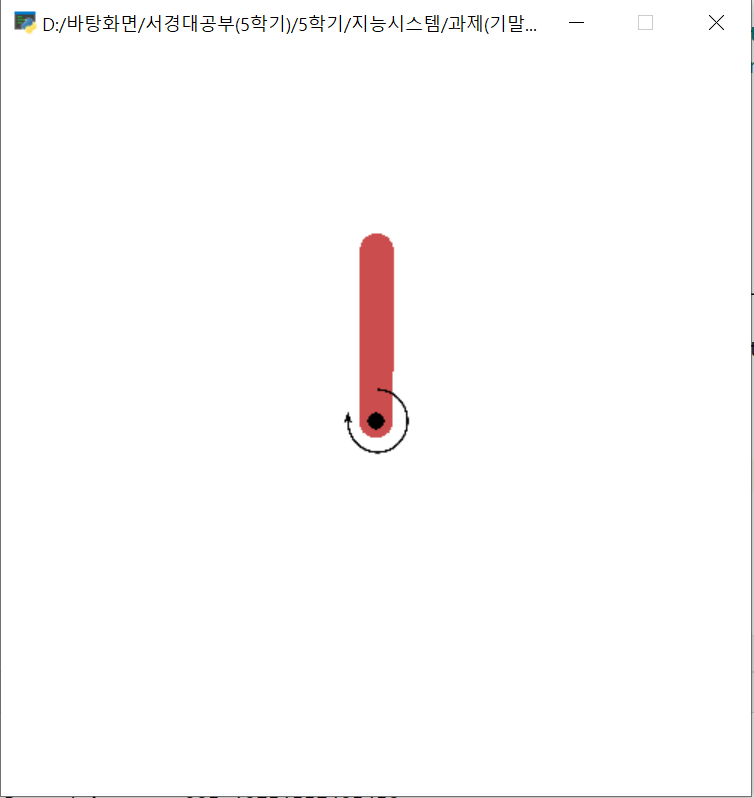




**직접 추가한 코드**

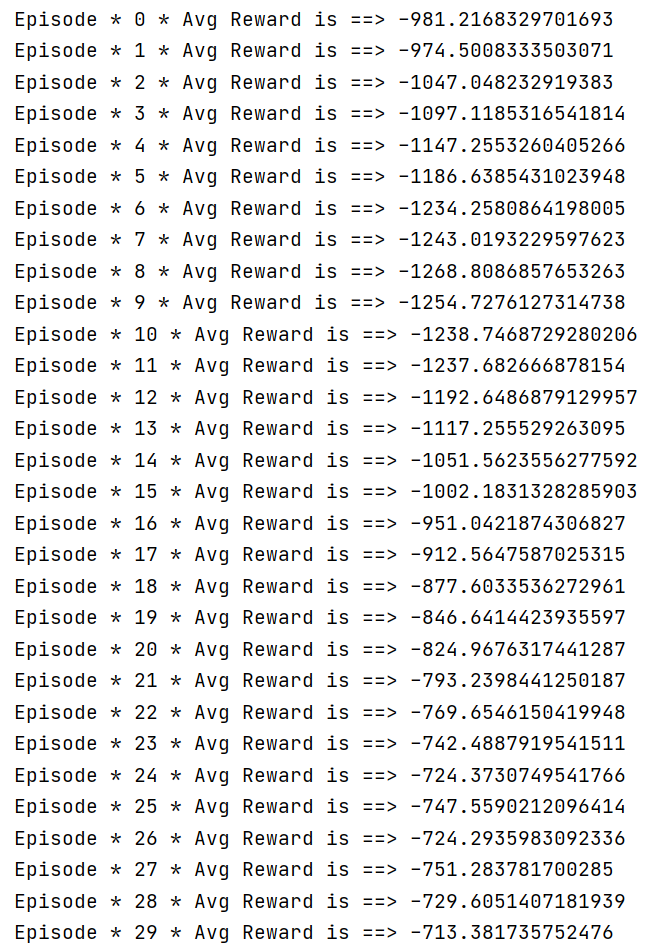
학습결과를 render메소드를 통해 비주얼로 표현한다.

Ex)

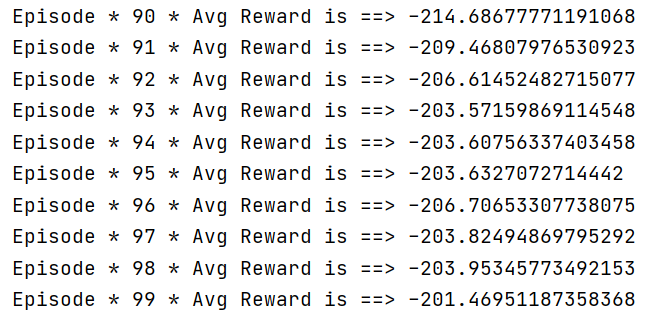


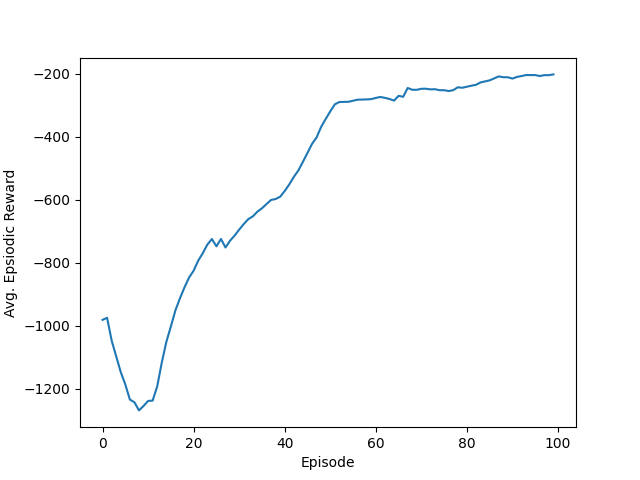
**<실행결과>**

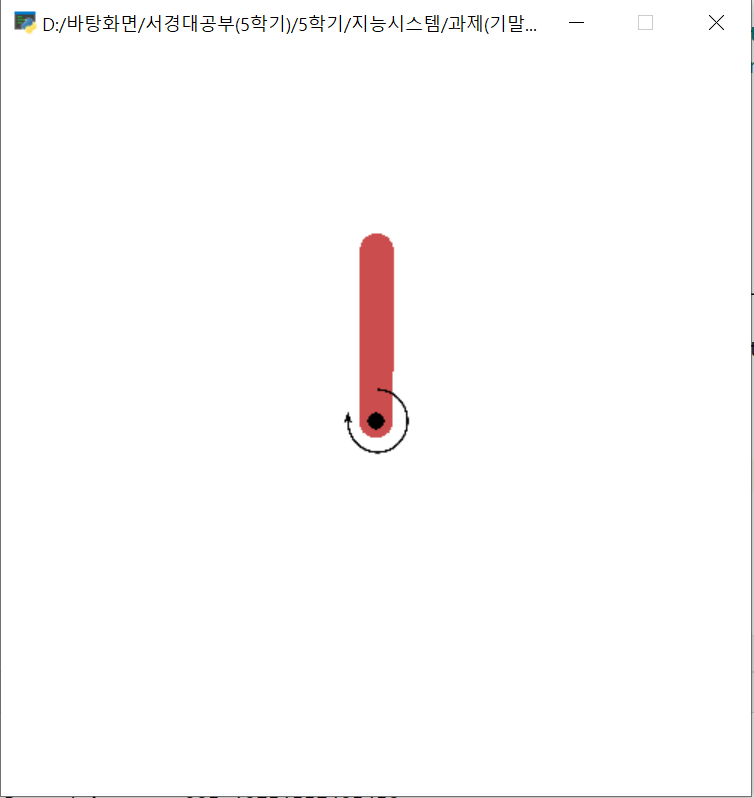




**(중간 생략)**







결과를 보면 학습이 정상적으로 되어 목표하는 결과가 나왔다.