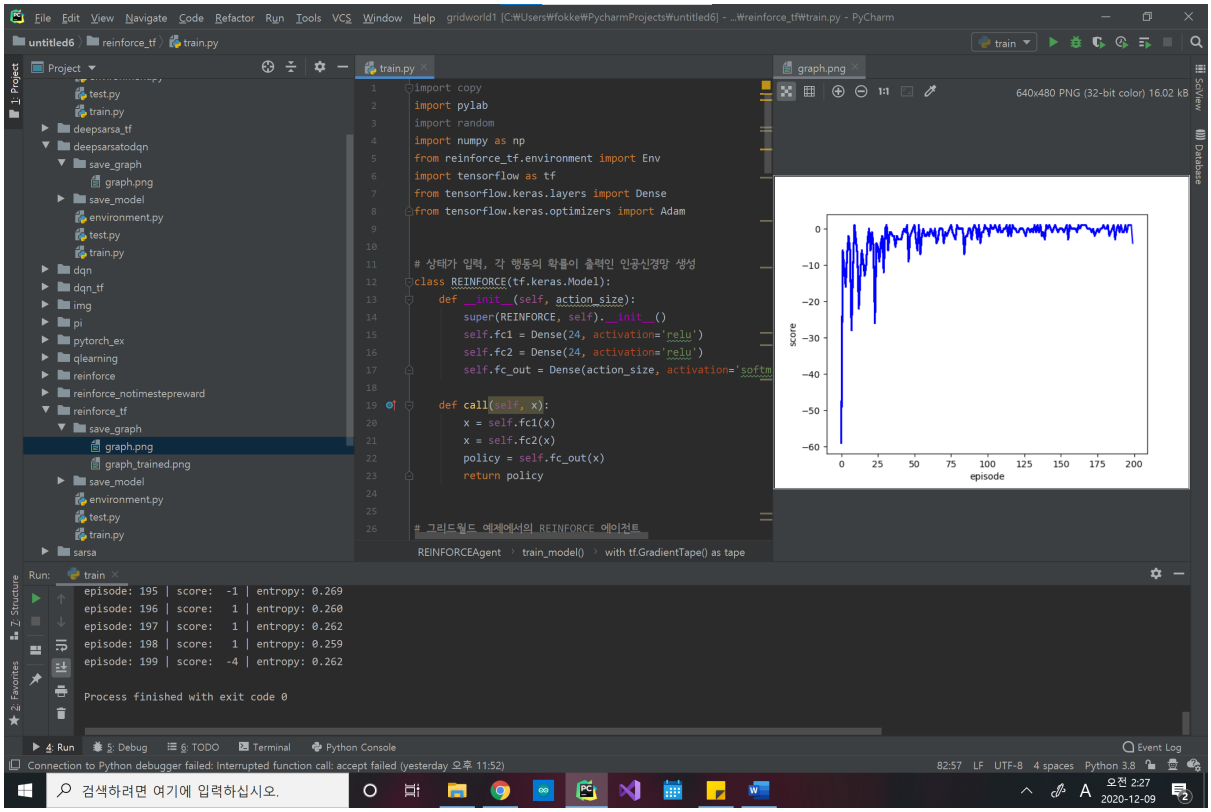
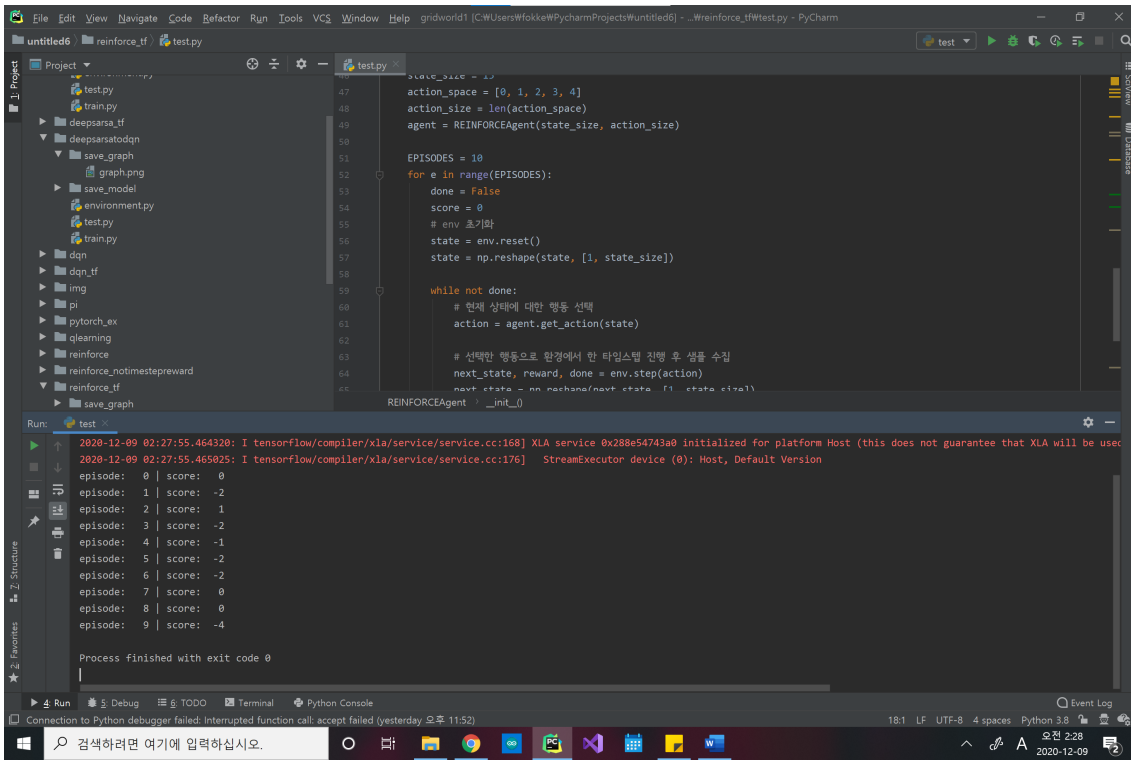
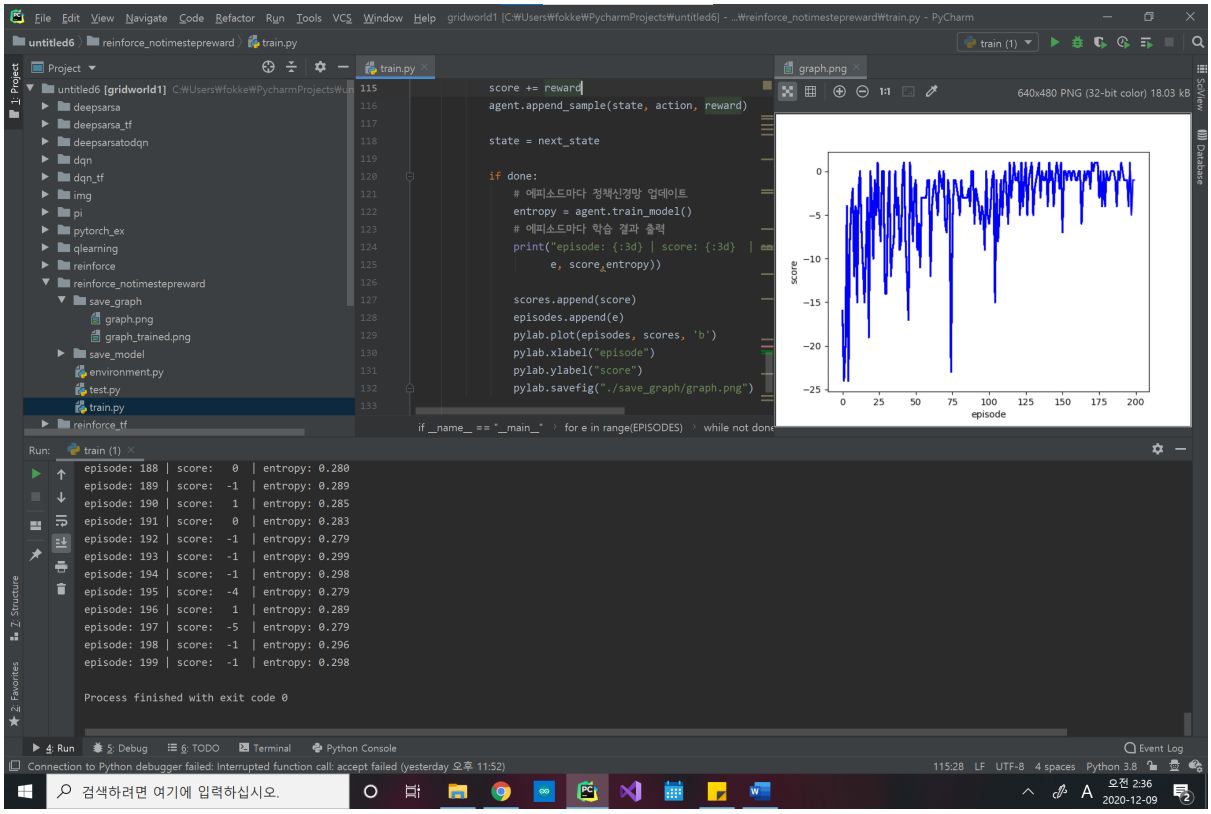
1. Grid world,Reinforce 예제의 tensorflow를 pytorch로 바꾸어보았습니다.



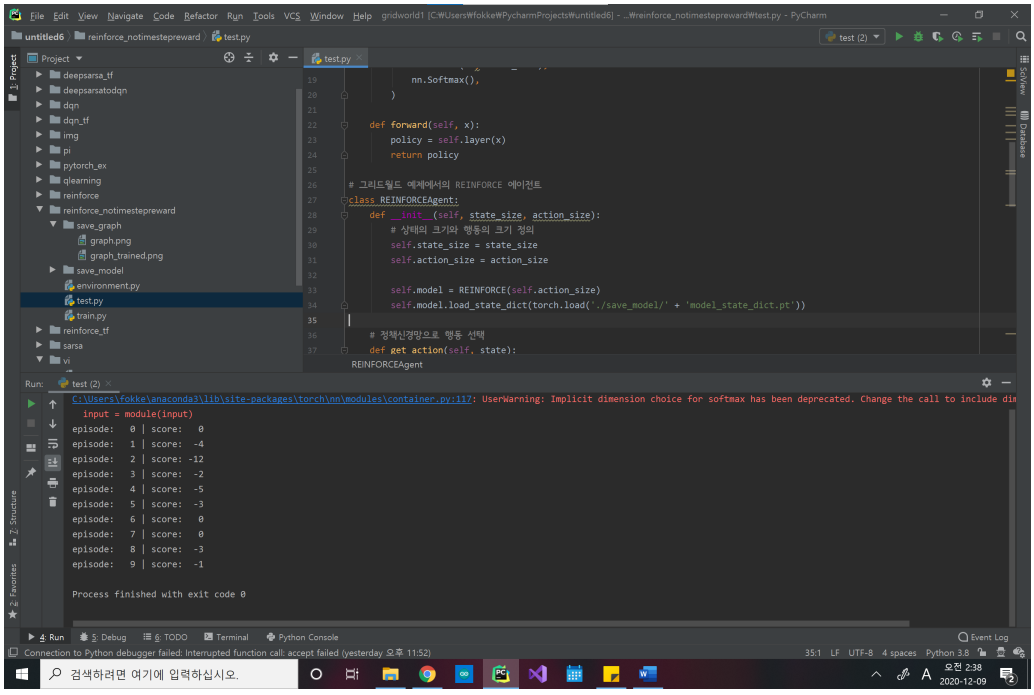
tensorflow로 구현한 train 결과입니다.



Tensorflow로 학습한 model을 사용한 test 결과입니다.



이는 pytorch로 바꿔준 train결과입니다

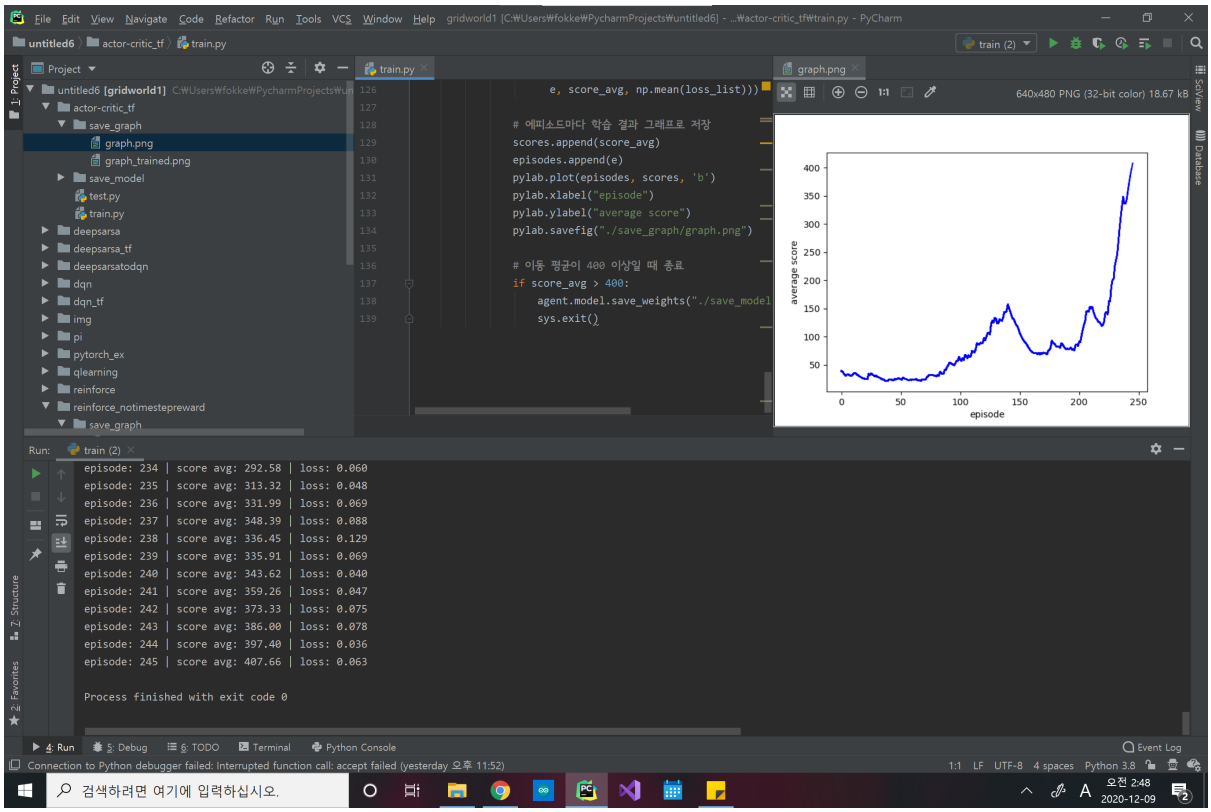


이는 위에서 학습한 model을 사용한 test결과입니다

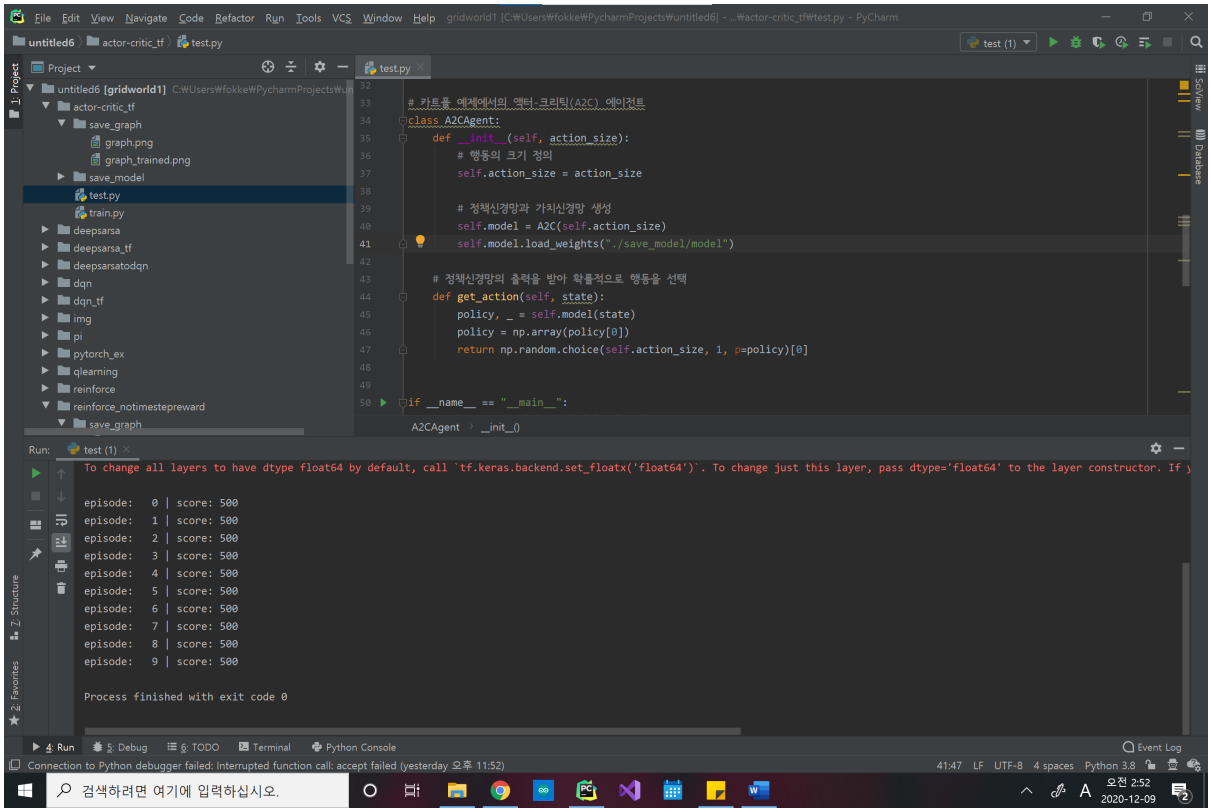
Tensorflow와 pytorch의 train결과와 test결과를 보면 pytorch로 바꿔준 결과가 score가 늦게 올라

갔음을 확인 하였습니다. Pytorch가 학습까지 필요한 episode가 좀더 많아보였습니다.

Cartpole example



위는 cartpole 예제를 tensorflow로 구현한 train 결과입니다.

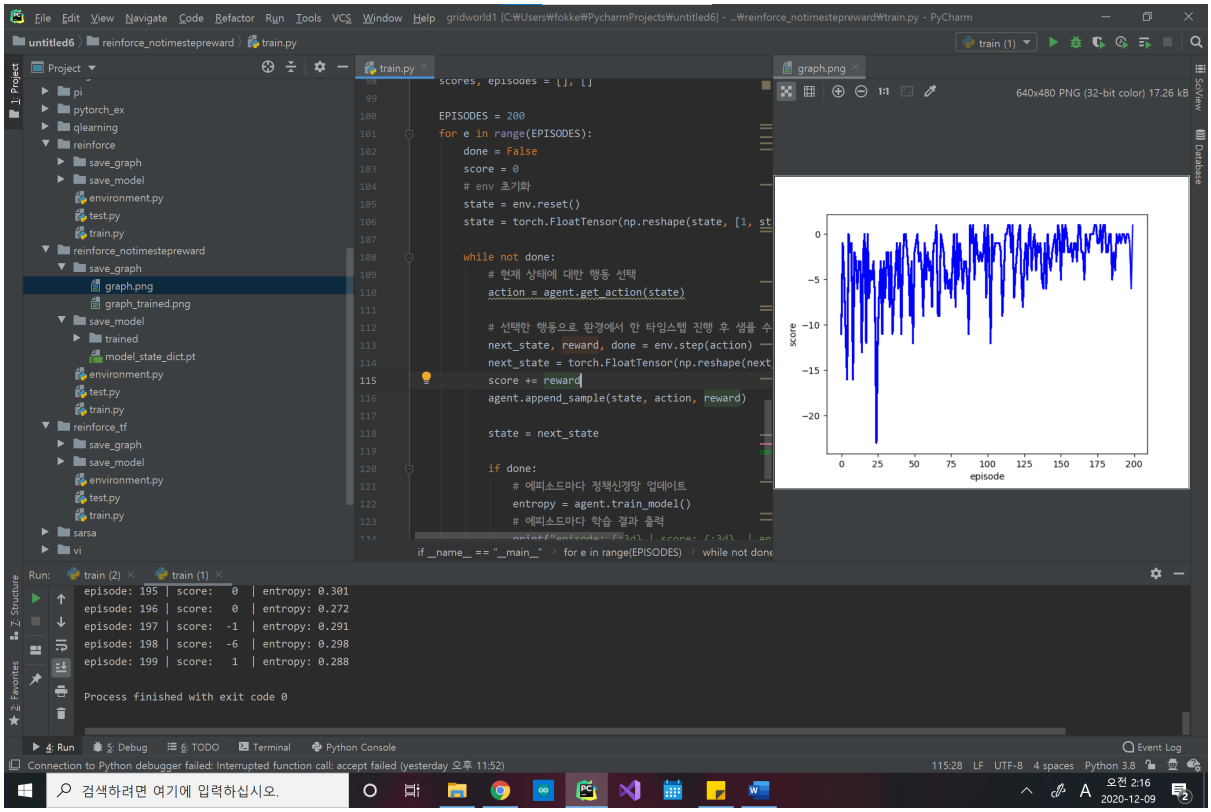


이는 위에서 학습한 model로 test한 결과입니다. 모두 500점을 넘어 잘 학습되었음을 확인하였습니다.

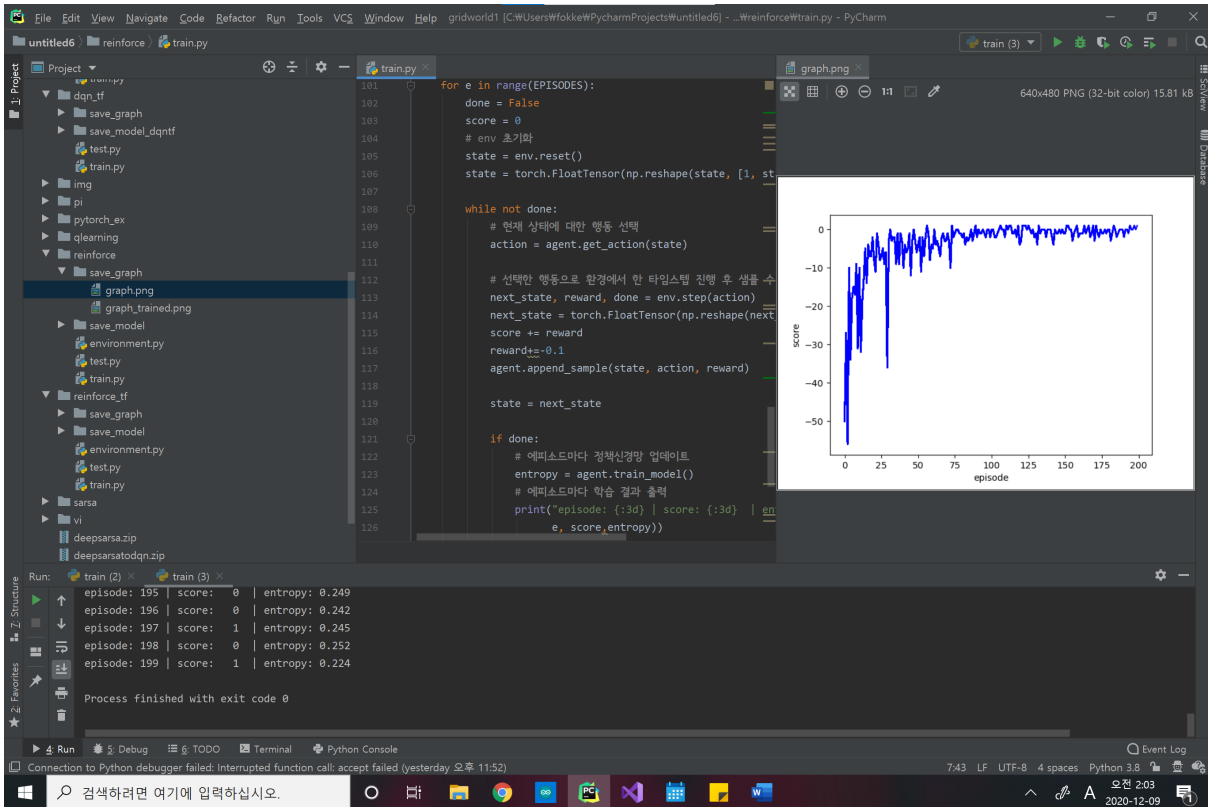
이제 pytorch로 바꾸어준 model을 보겠습니다.

하지만 이를 파이토치로 바꾸어 보았을 때는 어디에선가 문제가 발생해 점수가 안 오르는 현상을 관찰하였다. 밑에 gridworld를 pytorch,actorcritic으로 구현했을 때 학습이 잘되는 것으로 보아, get action, train model, main 함수에는 문제가 없지만 model의 구조만이 달랐기 때문에 model의 구조에 문제가 있는 것 같았다. 하지만 model을 이런저런 방향(Dropout, batchnormalization, lr 바꾸기, model의 깊이 바꿔보기,clip norm 값 수정 등등)으로 바꾸어 보았지만 결국 점수는 올라가지 않았다.

2.

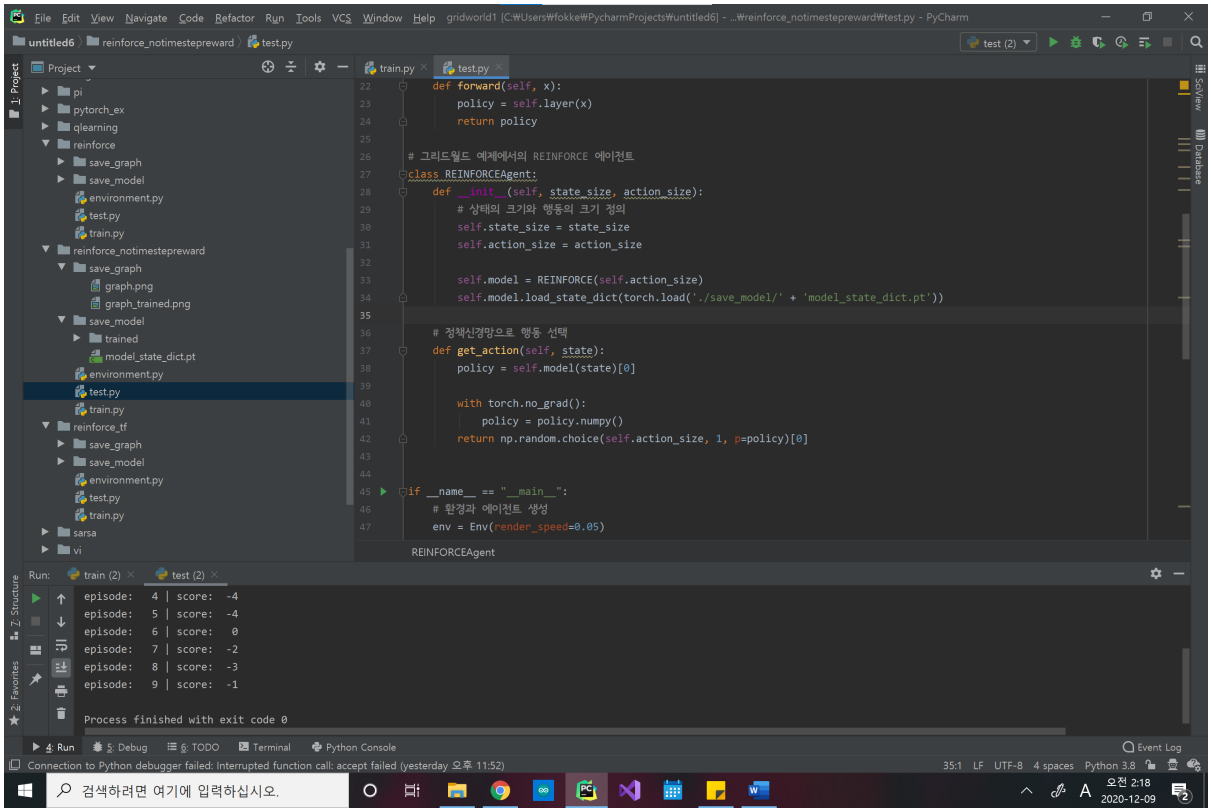


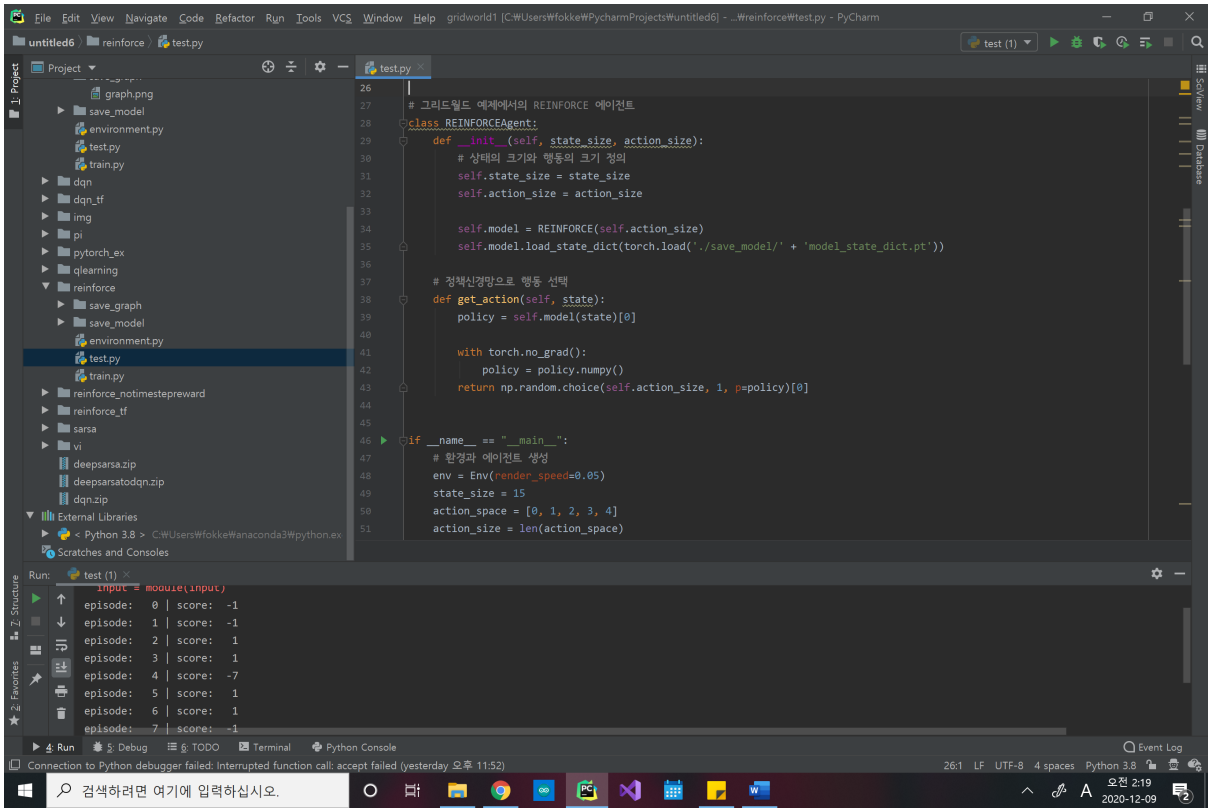
이는 time step마다 -0.1의 reward를 주지 않은 train결과이다



위는 pytorch로 바꾼이후 time step마다 -0.1의 reward를 준 model의 train 결과이다.

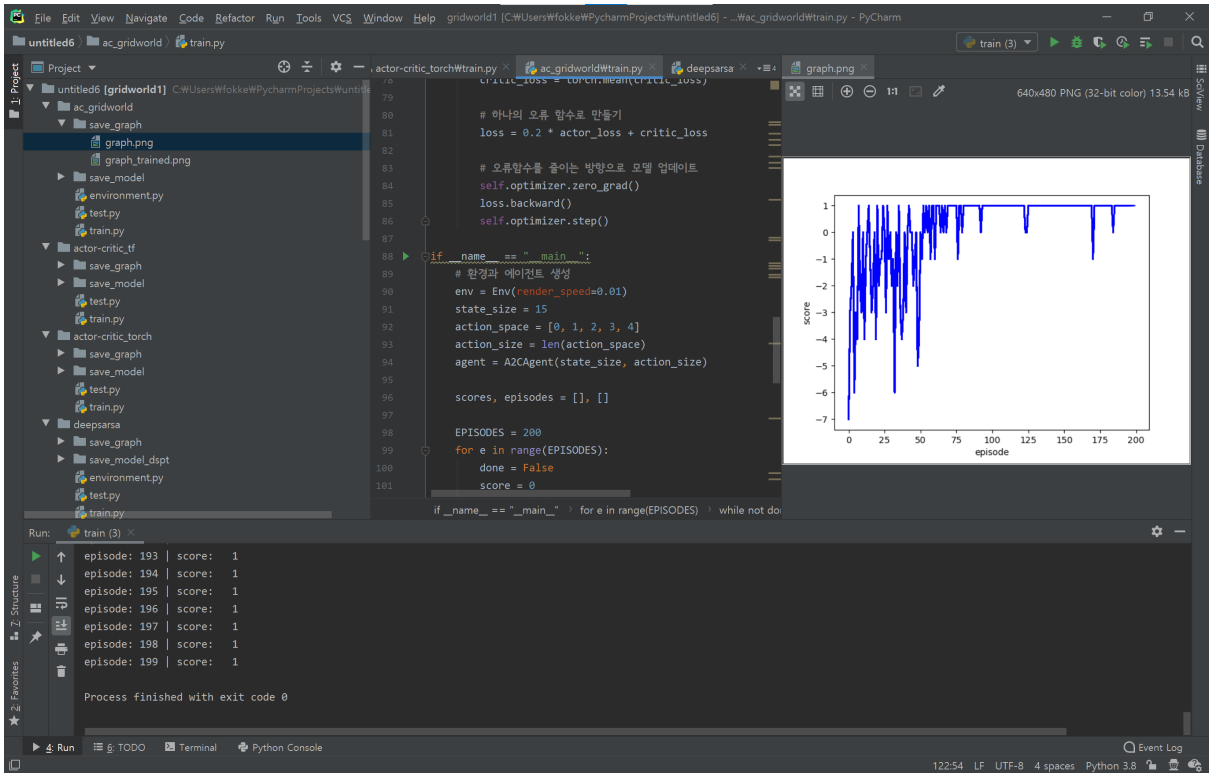
그래프를 비교해보면 아래의 모델이 훨씬 학습이 잘되며 정확도가 높은 것을 육안으로 확인할 수 있다.





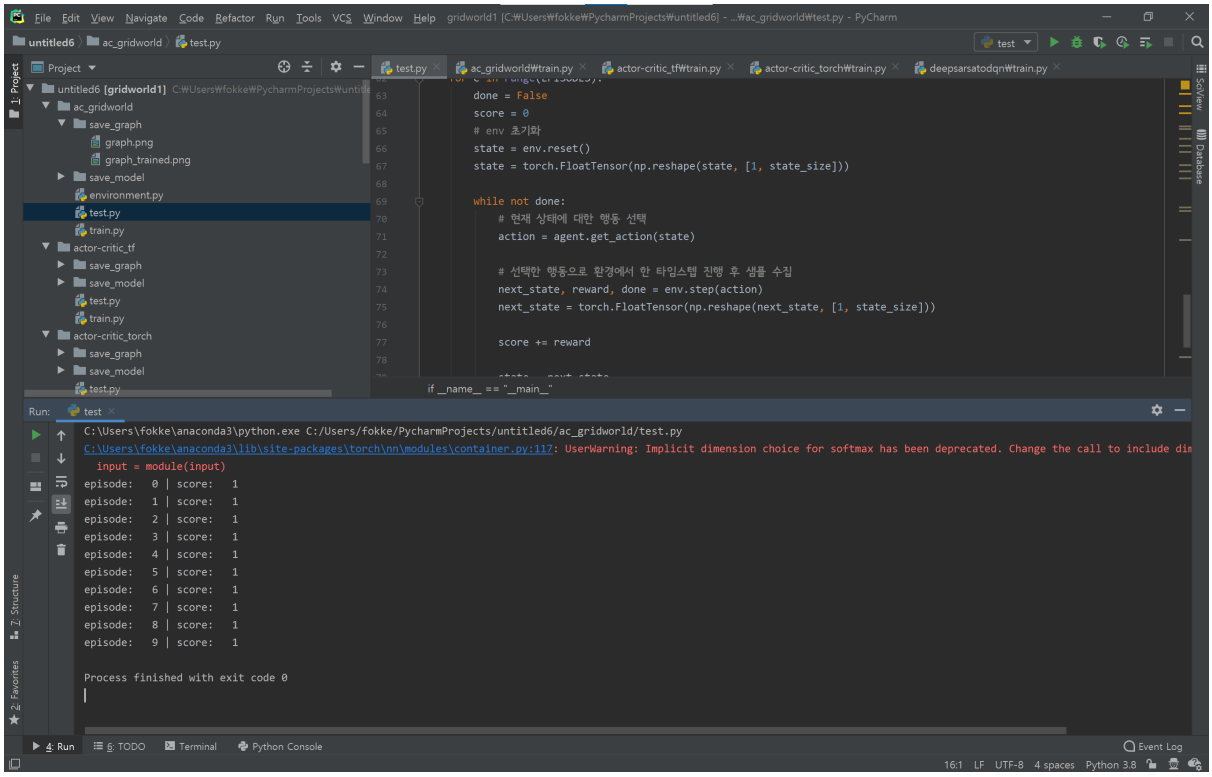
test결과도 아래의 모델을 보면 1의 score로 높지만 위의 모델은 아직 부족한 것을 확인하였다.

3 . grid world를 actor critic을 이용한 model로 바꾸어준 train 결과이다

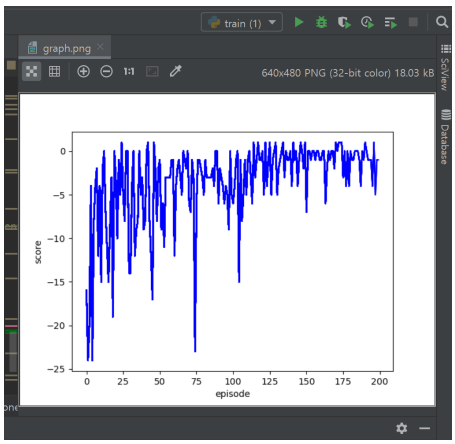


보면 매우 매우 좋은 성능으로 동작함을 확인하였다.

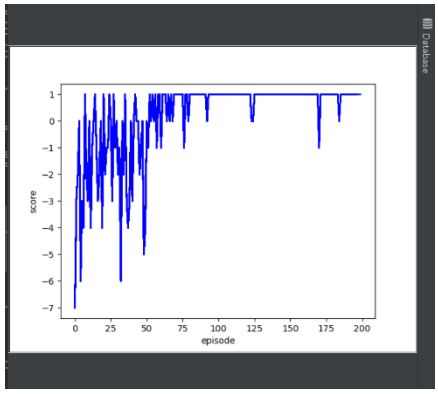
Cartpole과 달리 kernel initialize기능을 빼주었다.



test결과도 매우 훌륭하게 나왔다.



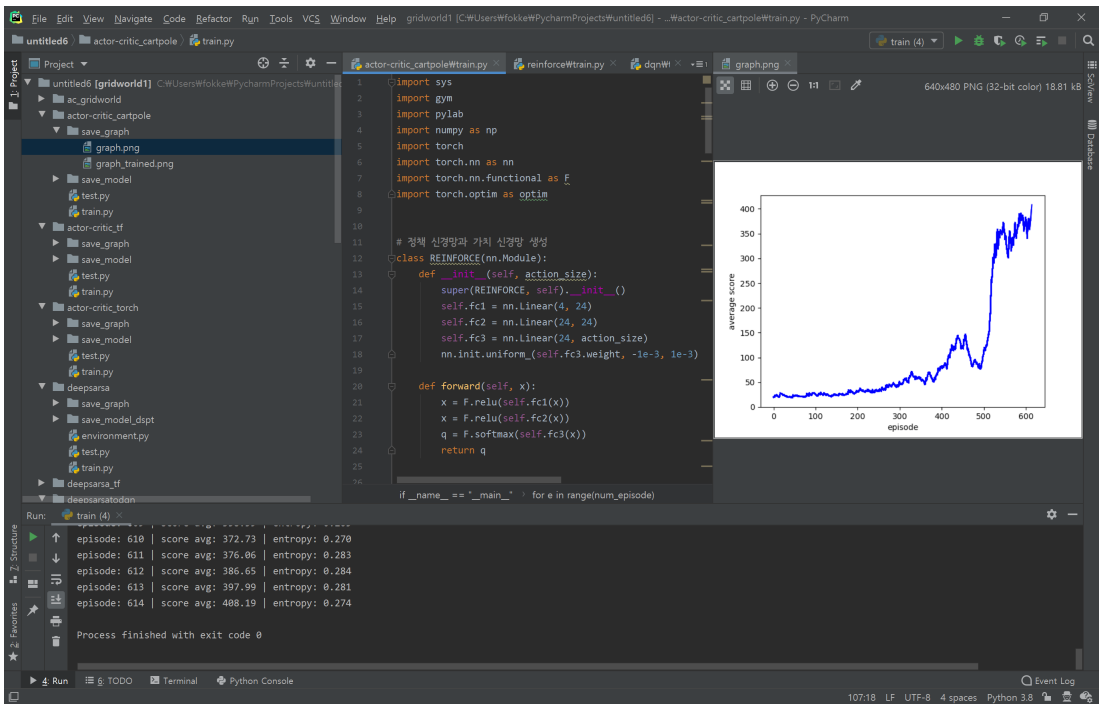
그래프 1



그래프2

그래프 1은 grid world를 reinforce로 학습시켰을 때이며 actor critic으로 학습시킨 그래프2와 비교해 보면 actor critic의 점수 그래프가 훨씬 안정적이게 1값에 수렴함을 확인하였다.

4. cart pole 예제를 reinforce를 활용한 알고리즘으로 바꾸어보았다.

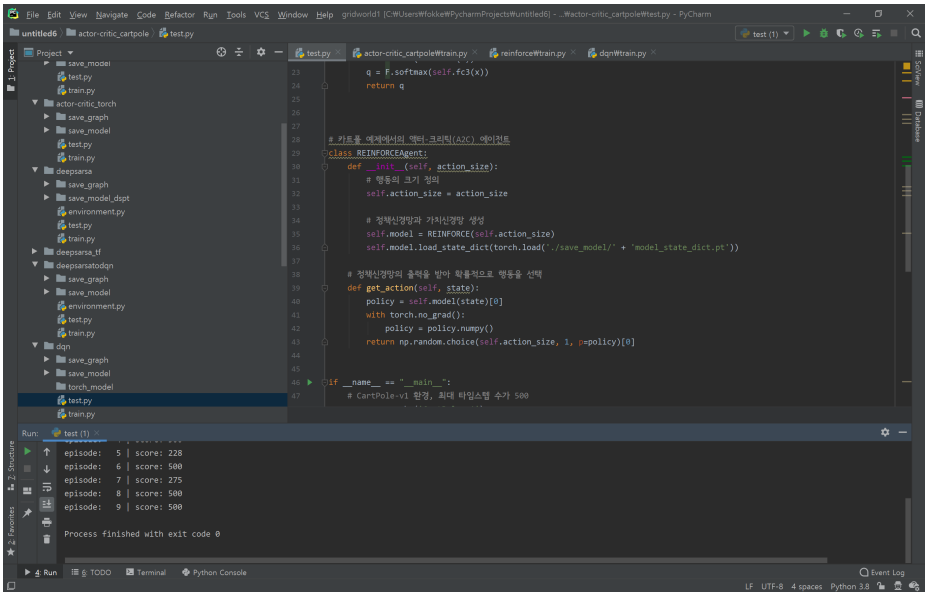


train결과를 보면 614 episode만에 avg score 400을 넘어 성공적으로 학습되었다.

위에 cartpole을 Actor critic, pytorch로 구현 했을때는 점수가 오르지 않았었는데 reinforce, pytorch로 구현한 결과 성공적으로 구현이 되었음을 확인하였다. 결국 Actor critic의 model에 문제가 있음을 간접적으로 추론할수 있었다. 이전에 gridworld에서 쓴 train_model함수를 그

대로 사용하였으며 model만을 약간 바꾸어주었다.

교과서의 그래프와 비교해봤을 때 그래프의 형태가 굉장히 유사함을 확인할수 있었다.



위 model을 사용한 test결과도 성공적으로 500점이 나오는 것을 확인하였다.