

Homework 2

Elliott Pryor

30 October 2023

My work can be found on my github: <https://github.com/ElliottP-13/DeepRL-Course>.

There is not any explicit question given, besides to run

```
$ python ds6559/scripts/run_hw2.py --env_name CartPole-v0 --exp_name test_cartpole  
  --video_log_freq -1  
$ python ds6559/scripts/run_hw2.py --env_name InvertedPendulum-v4 \  
  --exp_name test_pendulum --video_log_freq -1
```

So my results from the first command can be seen in Figure 1 and the second command can be seen in Figure 2.

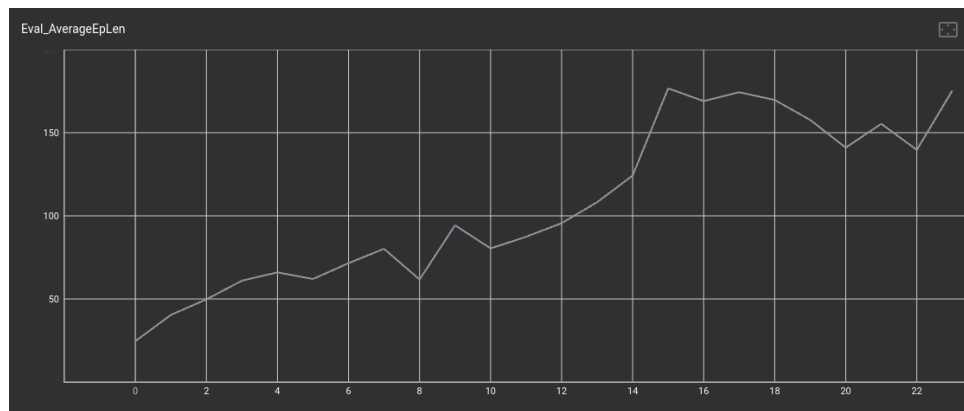


Figure 1: Command 1

Because of the normalization that was asked to do in the code, the program terminates once the optimal policy is reached. This is because max number of environment steps are achieved in all the rollouts, so the standard deviation is zero, thus normalization breaks.

For the first environment, there are only 200 timesteps, and an optimal policy is learned in 24 iterations. The second environment has max timesteps of 1000 which is learned in 103 iterations.

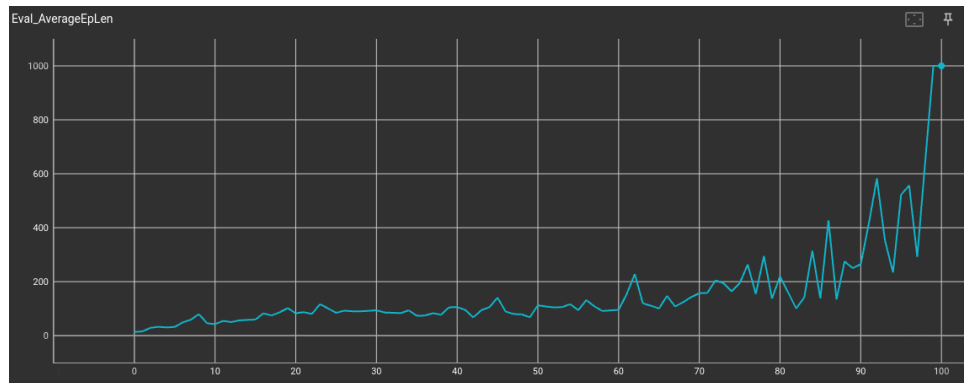


Figure 2: Command 2

A video of the last cartpole rollout can be found at:

<https://github.com/ElliottP-13/DeepRL-Course/blob/main/Writeup/imgs/cartpole.mp4>