

Homework 3

Elliott Pryor

11 November 2023

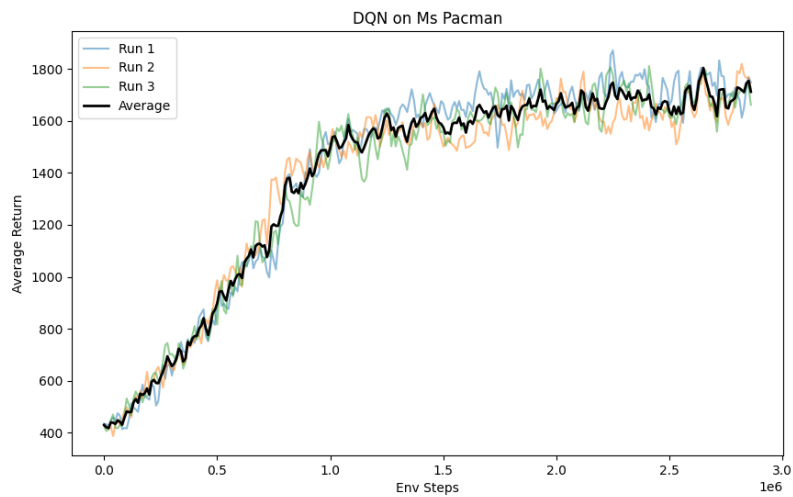
My work can be found on my github: <https://github.com/ElliottP-13/DeepRL-Course>.

Problem 1 Evaluation on Ms Pacman. The results can be seen in Figure 1.

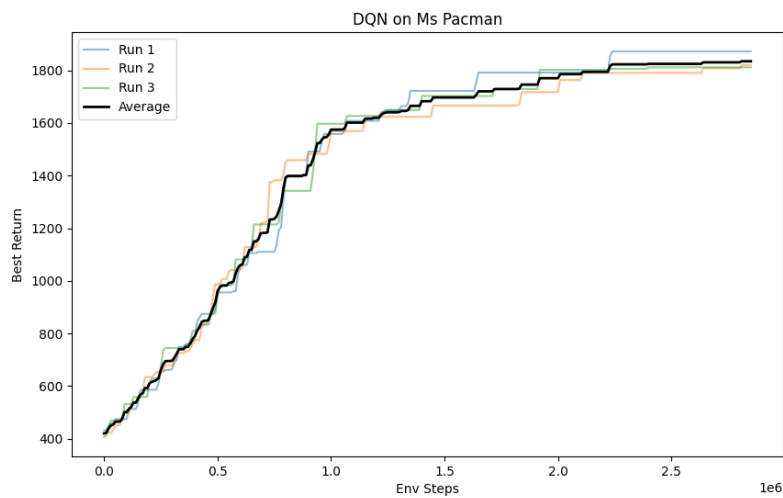
Problem 2 Double Deep Q learning.

We see in Figure 3 that DQN and DDQN perform similarly on Lunar Lander. DDQN does slightly better, but the difference is not significant. We see that both DQN and DDQN converge to a similar value. DQN actually performs slightly better than DDQN, but the difference is not significant. We see that both DQN and DDQN converge to a similar value. DDQN also has lower variance in performance, which may be preferable.

Problem 4 Actor critic results can be seen in Figure 4. We can see that Actor Critic can fully solve the problem in very few iterations. About 40,000 iterations it reaches max performance. However, in homework 2, the vanilla policy gradient did this in about 1000 environment steps. This is probably because vanilla policy gradient is lighter and able to more quickly solve this easy problem.

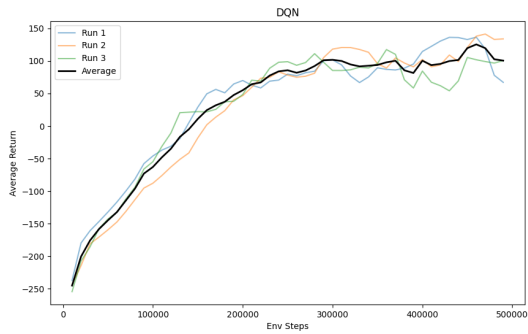


(a) Average reward per episode.

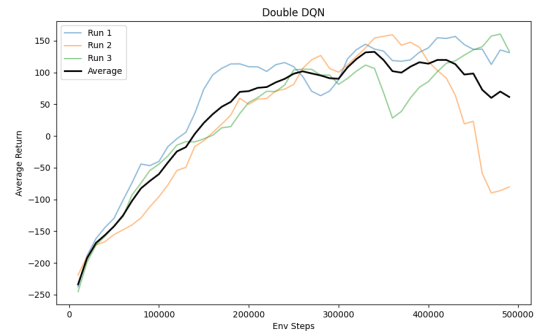


(b) Best reward per episode.

Figure 1: Performance of DQN on Ms Pacman



(a) DQN on Lunar Lander



(b) DDQN on Lunar Lander

Figure 2: Performance of DQN and DDQN on Lunar Lander

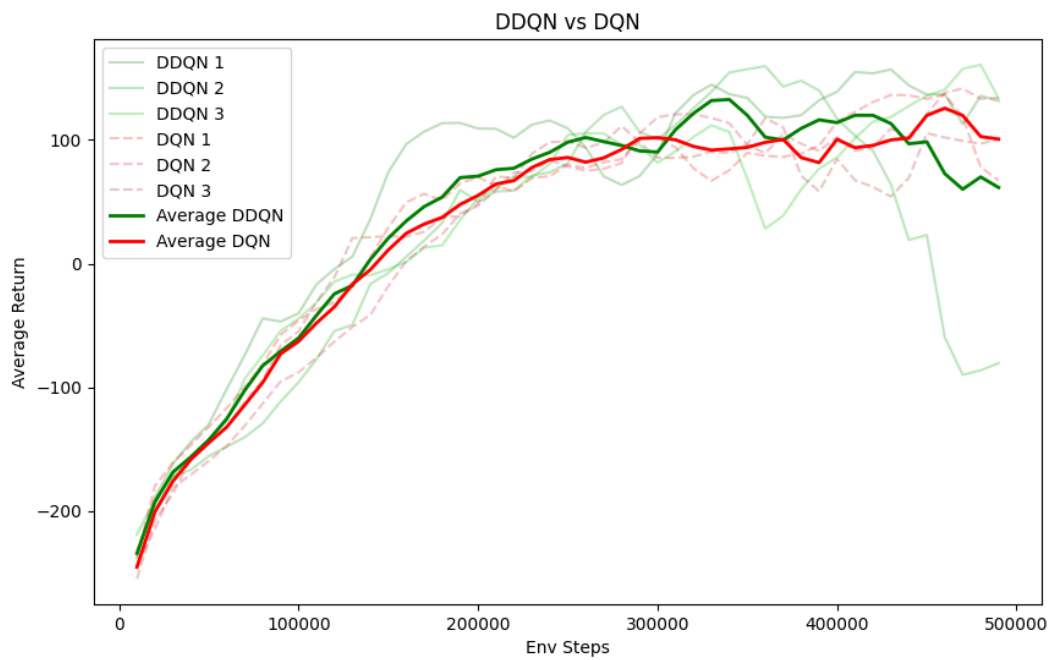


Figure 3: DQN v DDQN on Lunar Lander.

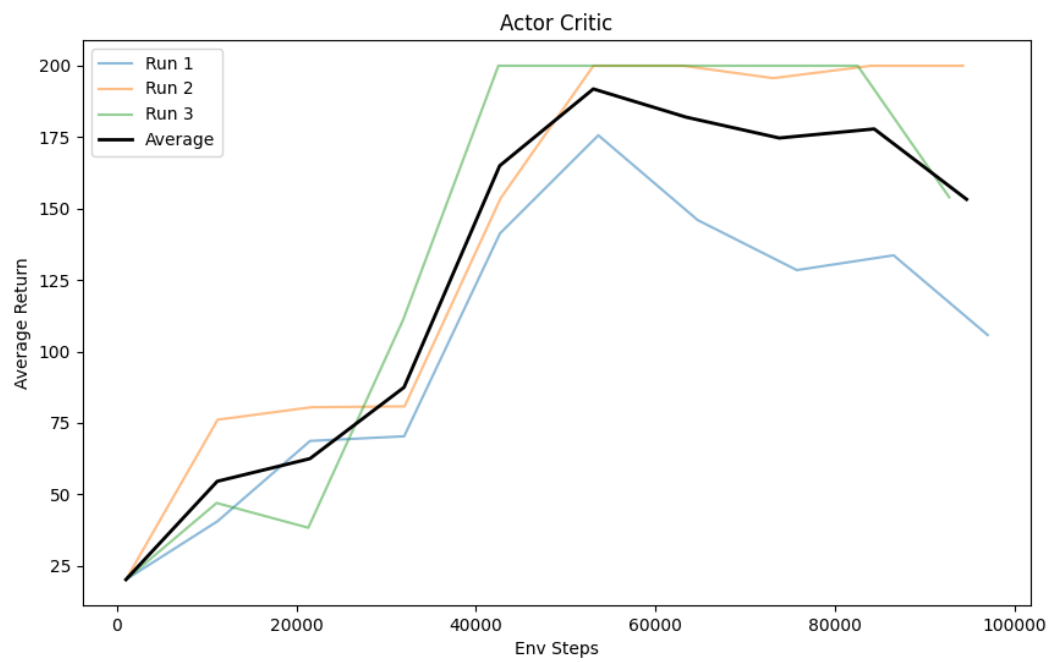


Figure 4: Performance of Actor Critic on cartpole.