CS 285 Homework 3

Sagnik Bhattacharya SID: 3033583960

October 21, 2021

Question 1: DQN

Figure 1 shows the learning curve for DQN on Ms. Pac-Man. Default hyperparameters were used along with the command given in the homework specification.

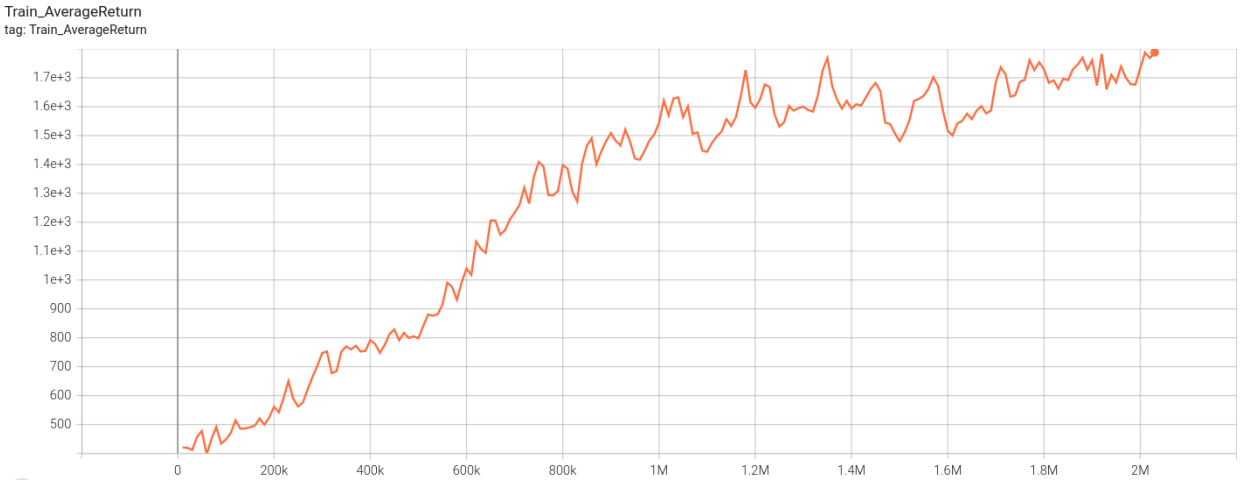

Figure 1: The learning curve for DQN on Ms. Pac-Man.

Question 2: DDQN

Figure 2 shows the learning curves of the average of three runs each of DQN and Double DQN. As we can see, DDQN performs somewhat better due to increased stability.

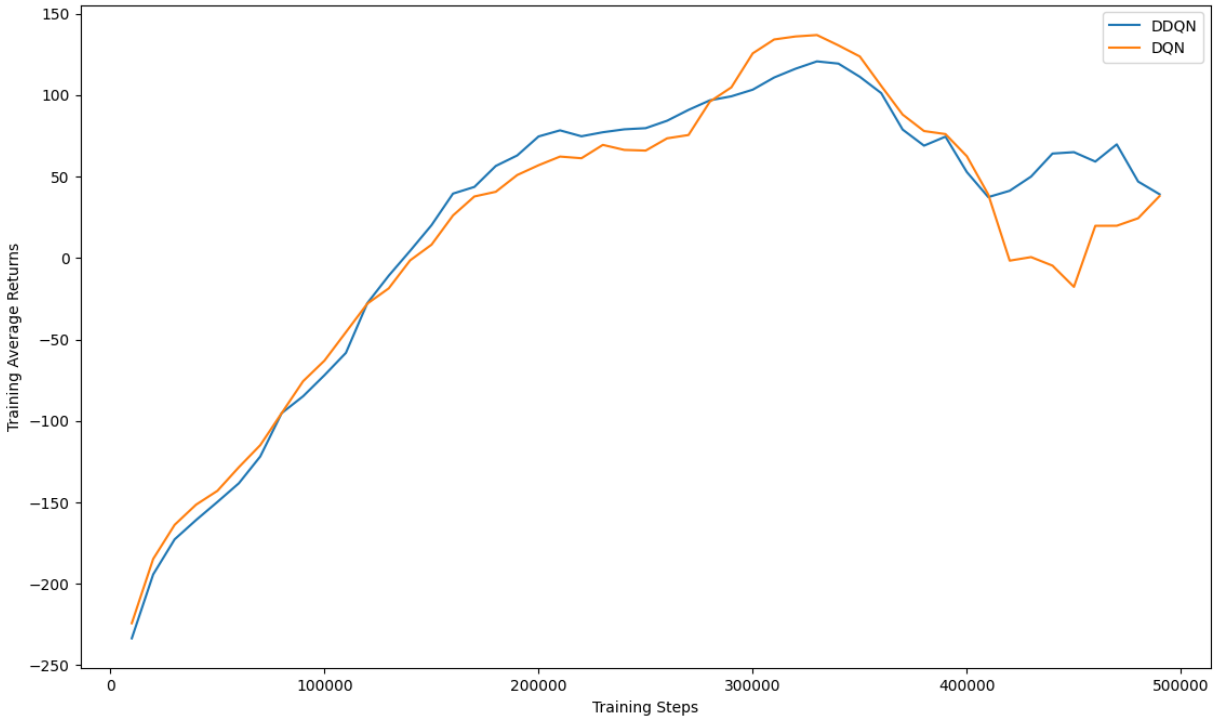

Figure 2: The learning curves comparing DQN and DDQN. As we can see, DDQN is more stable than DQN, even though they end up achieving similar performance most of the time.

Question 3: Hyperparameter Tuning

Figure 3 shows the comparison of changing the learning rate in the Ms. Pac-Man environment. As we can see, if the learning rate is too large or too small, then the algorithm does not end up learning well at all. We need a learning rate that is somewhere in the middle (1e-3 in this case) for significantly improved learning speed. This also shows that DQN is very sensitive to some hyperparameter settings.

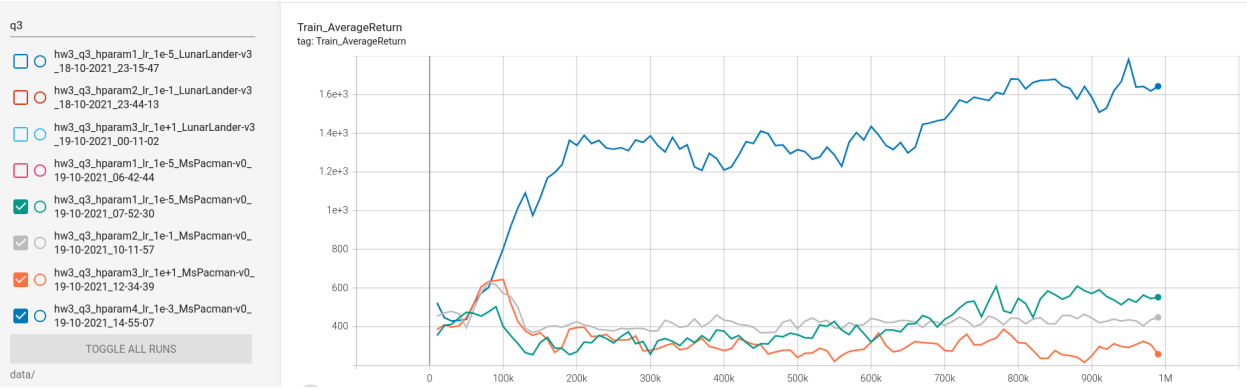

Figure 3: The learning curve for different settings of the learning rate, sweeping from 1e-5 to 1e+1. As we can see, each setting of the learning rate causes training to converge to an optimum with a different value, and choosing the correct learning rate (1e-3 in this case) is essential to finding the best optimum.

Question 4: Actor-Critic

From Figure 4 we can see that we need to perform a sufficient number of gradient and target updates to get the best results in actor-critic. Performing one update each is not sufficient. Performing gradient updates seems more important than performing target updates, as the orange curve is not as good as the blue and red curves. Performing 1 vs 10 target updates only improves the stability but reducing gradient updates from 100 to 10 does not decrease performance at convergence at all if the number of target updates is simultaneously increased from 1 to 10, as we can see by comparing the blue and red curves.

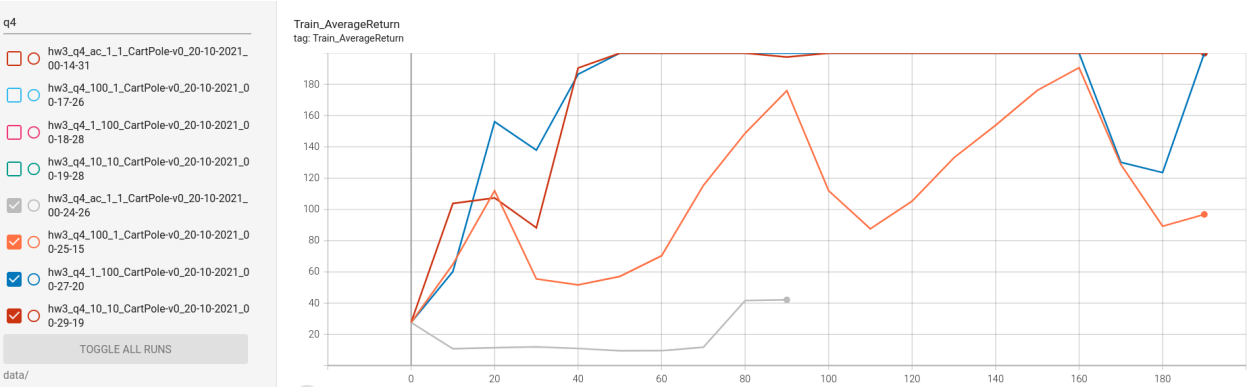

Figure 4: Comparison of various settings of the number of target updates and the number of gradient steps per target update. Results discussed above in detail.

Question 5: Difficult Actor-Critic

Figures 5 and 6 shows the learning curves for HalfCheetah and InvertedPendulum respectively using actor-critic.

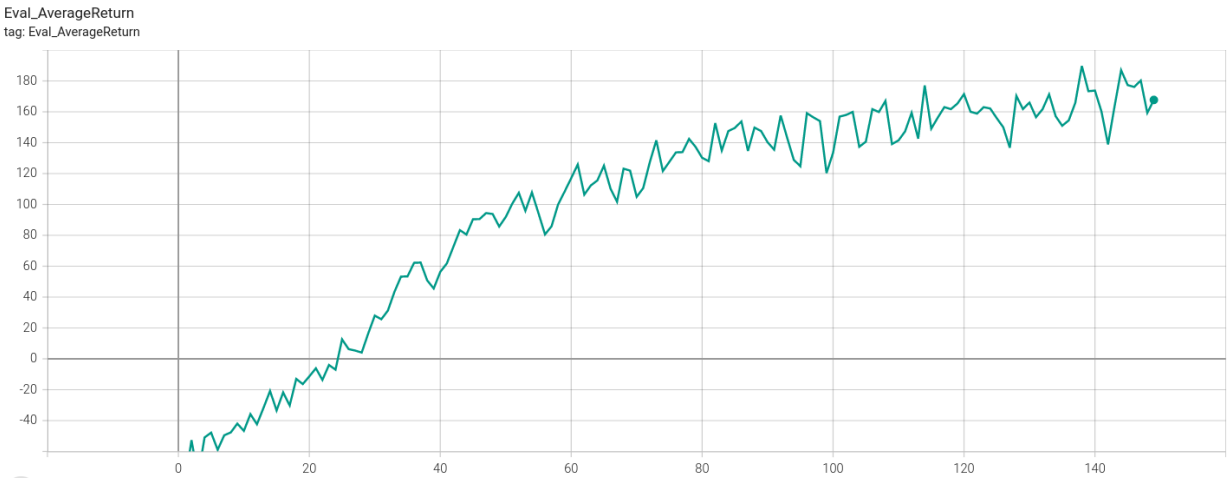

Figure 5: The learning curve for HalfCheetah using actor-critic.

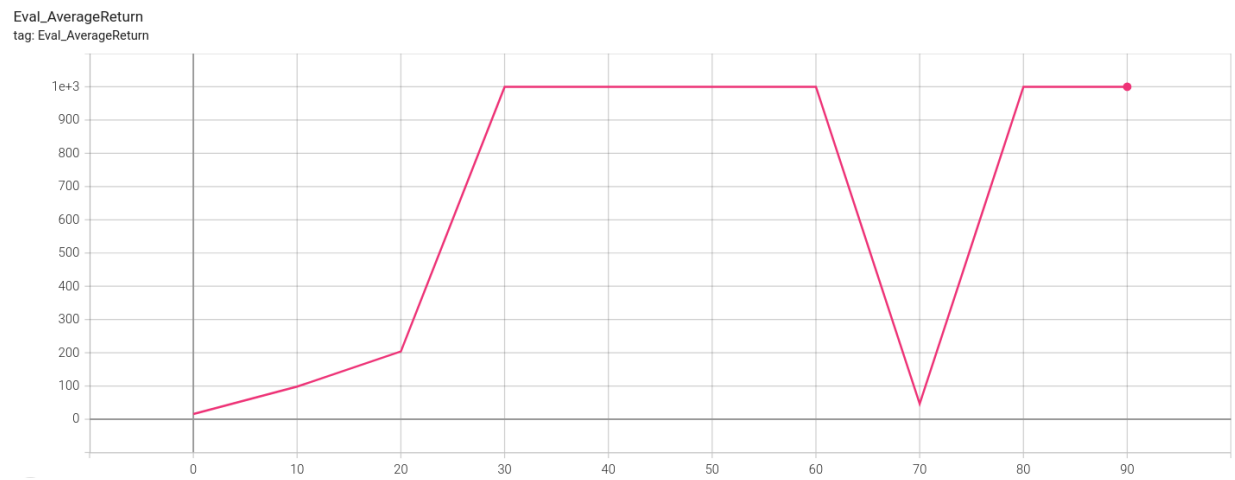

Figure 6: The learning curve for InvertedPendulum using actor-critic.