Question 1.1: basic Q-learning performance

Replay buffer size = 500000 (due to the memory limit)
This is the only run over 4m steps (4.36m) using basic DQN. Unfortunately, I only recorded the rewards to episodes (2076 episodes in total) instead of time steps. The x-axis in Fig. 1 is the number of episodes.

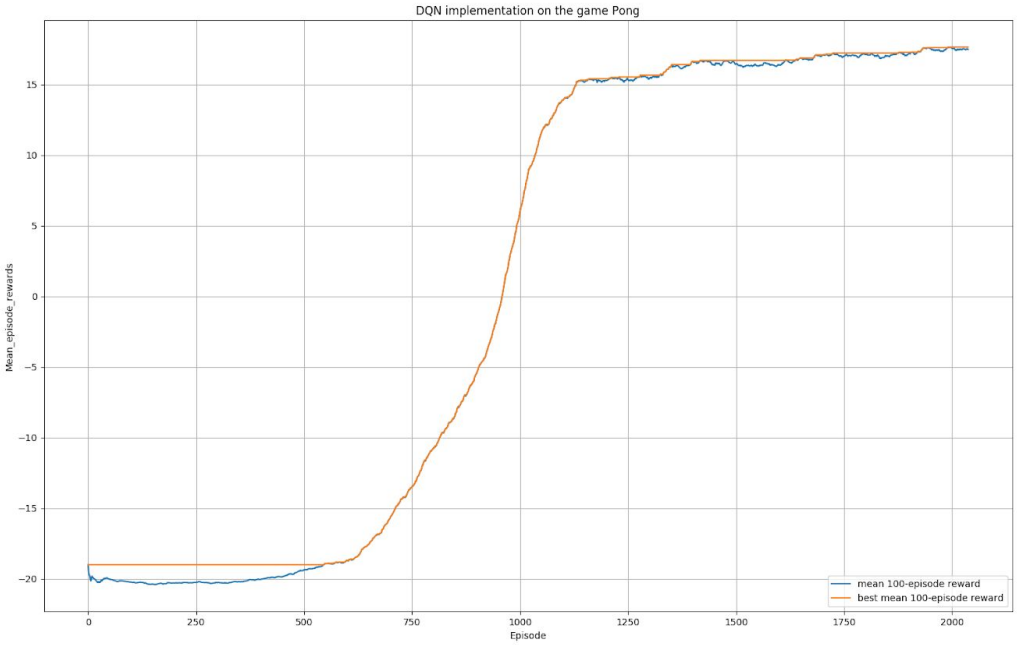

Figure 1. Vanilla DQN

Question 1.2: double Q-learning

Fig. 2 shows the learning curve of vanilla Q-learning with double Q-learning. Vanilla Q-learning achieves the reward around 15 even faster than the double Q-learning. I run both algorithms for 3m steps.

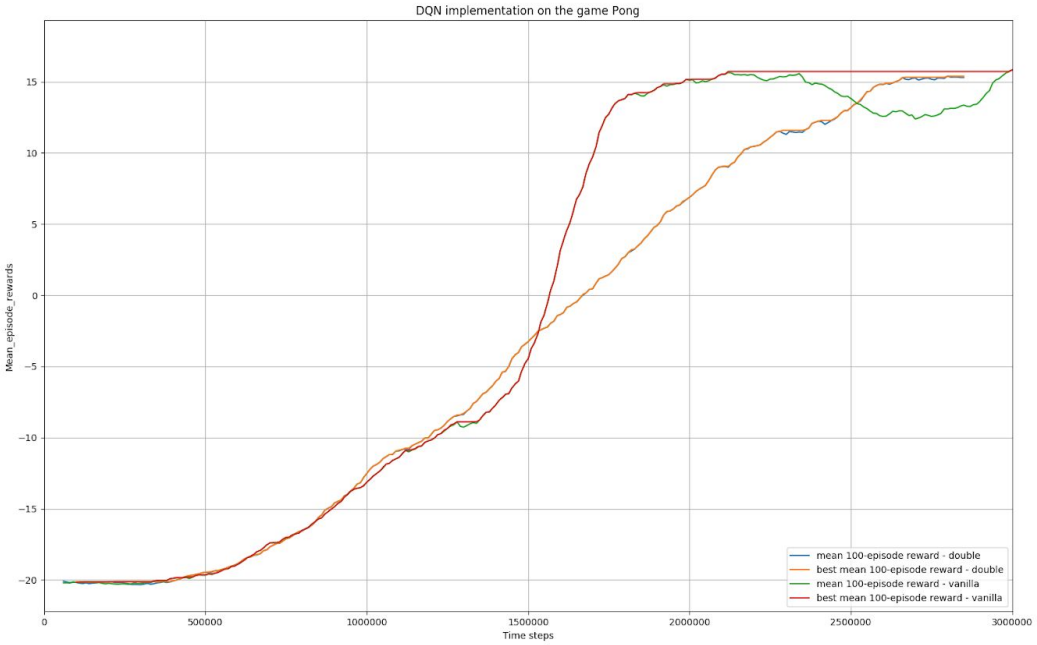

Figure 2. Double Q-learning vs. vanilla Q-learning

Question 1.3: experimenting with hyperparameters

In Question 1.1, I got an memory error using the default replay buffer size of 1000000. Therefore, I ran the basic Q-learning algorithm with different buffer sizes and compared their performances. As the buffer size increases, basic Q-learning achieves better rewards after 2m training steps. However, using a large buffer requires a lot memories.

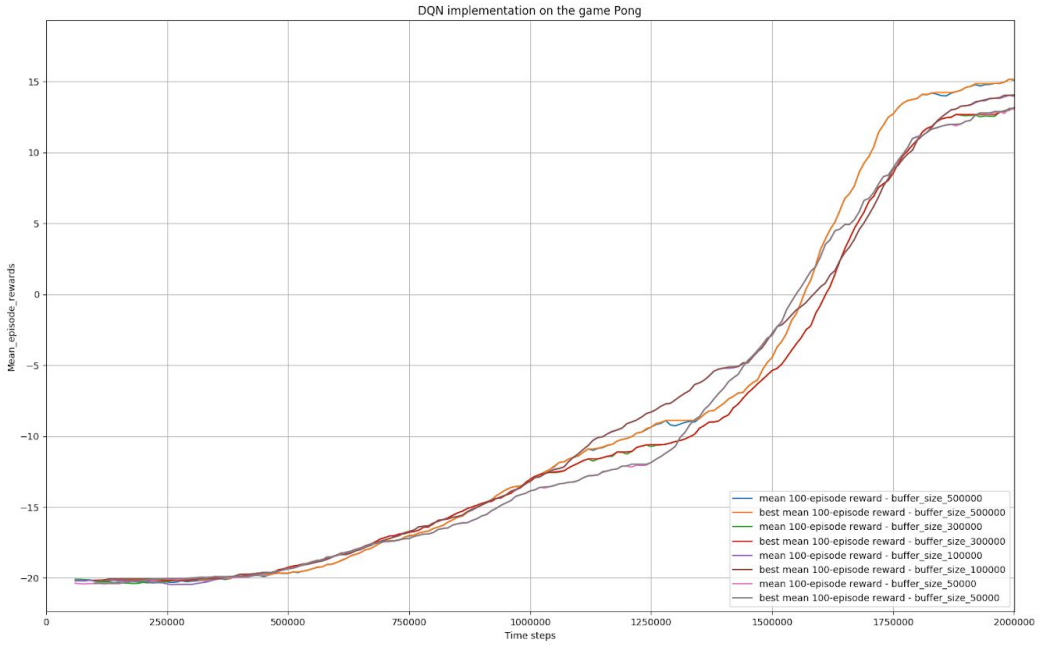

Figure 3. Basic Q-learning with different buffer sizes.

Question 2.1: sanity check with Cartpole

With ntu = 10, ngsptu = 10, actor-critic algorithm achieves the best performance with Cartpole.

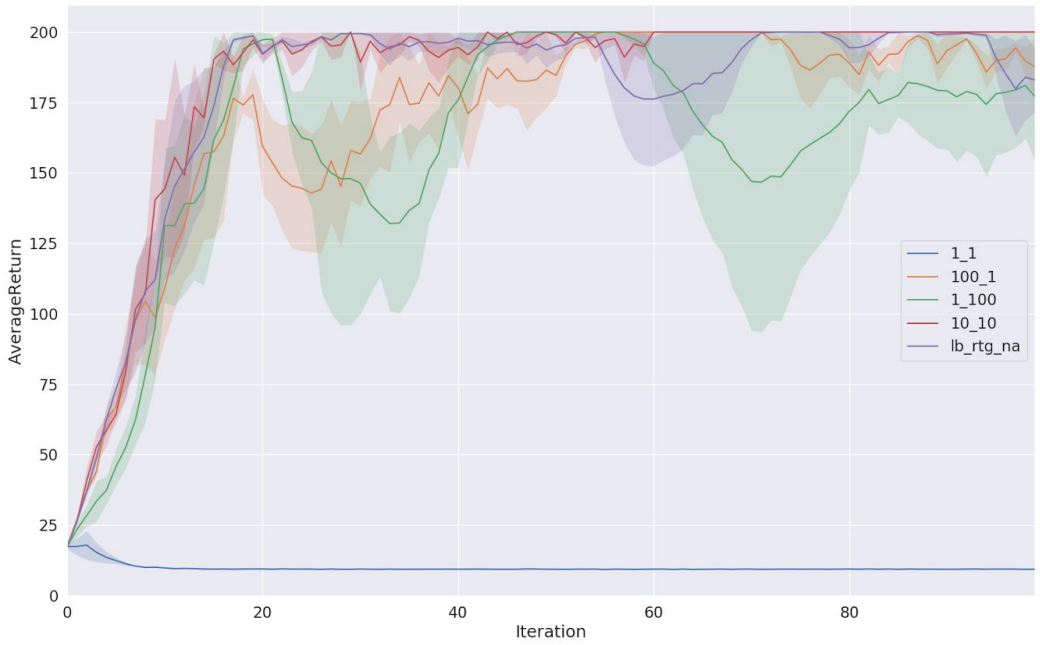

Figure 4. Actor-critic with Cartpole.

Question 2.2: run actor-critic with more difficult tasks

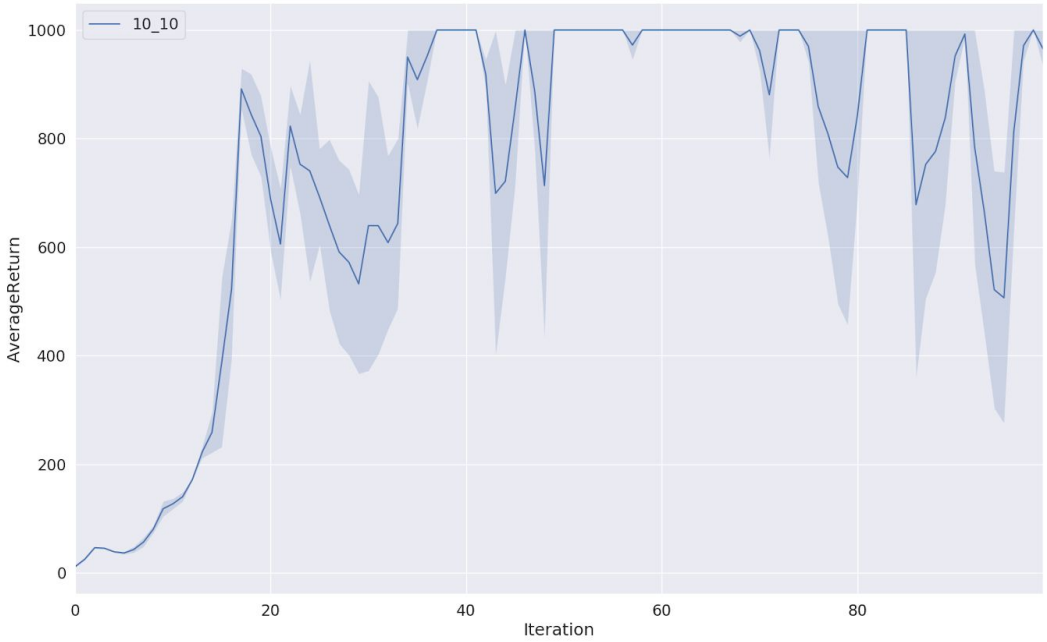

Figure 5. Actor-critic with InvertedPendulum.

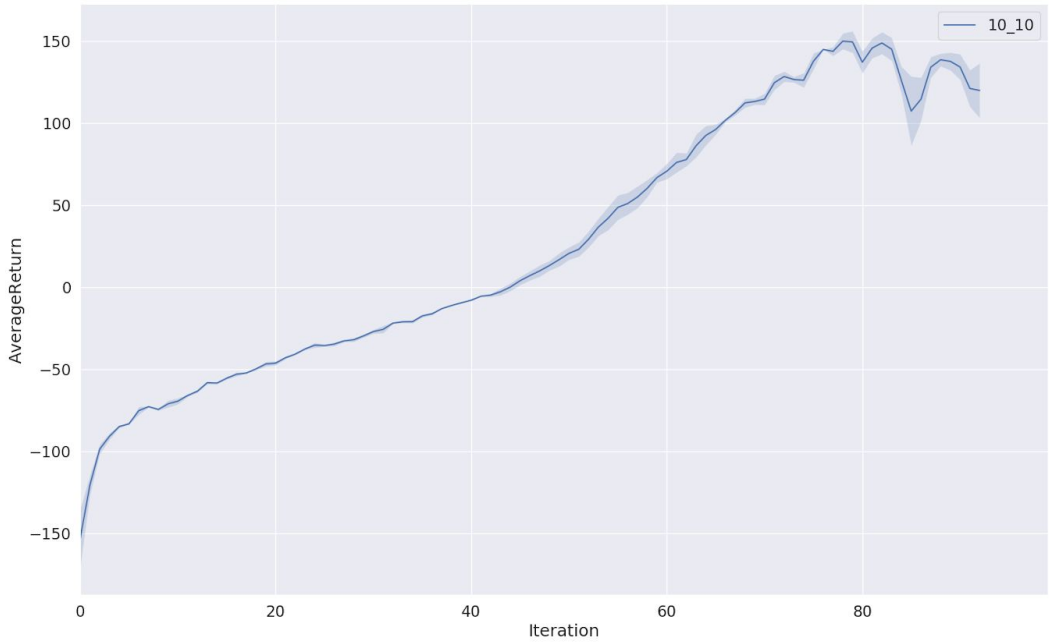

Figure 6. Actor-critic with Halfcheetah.