1 Q-Learning

1.1 basic Q-learning performance

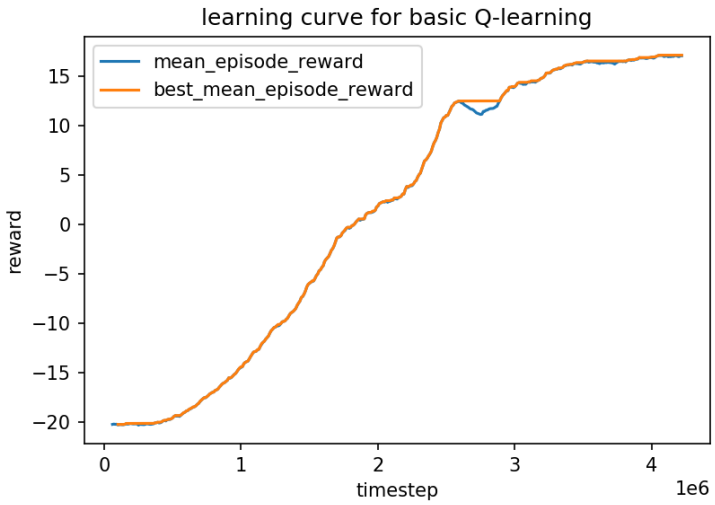

Figure 1: Learning curve for basic Q-learning, 4.2m time steps were trained on *Atari* task.

1.2 double Q-learning

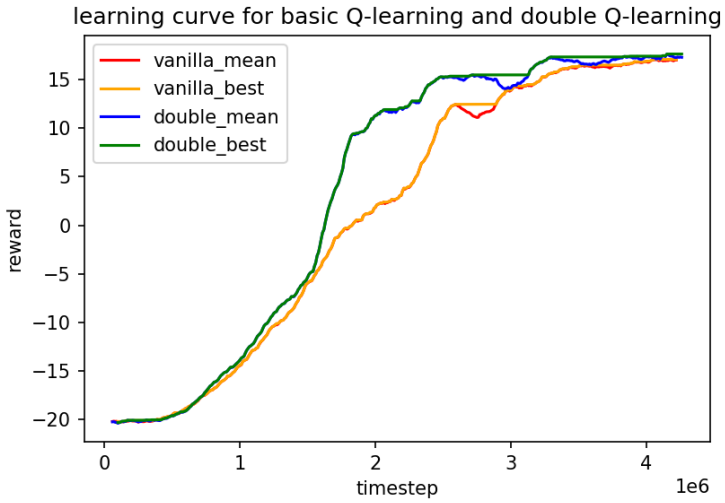

Figure 2: Learning curve for basic Q-learning vs double Q-learning, 4.2m time steps were trained on *Atari* task.

1.3 double Q-learning

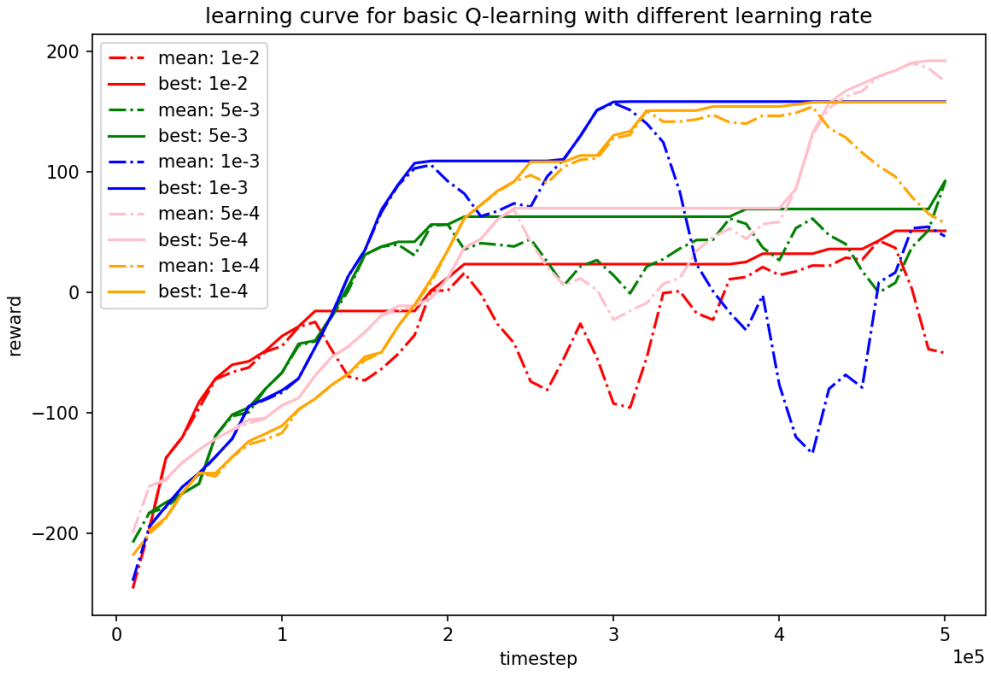

Figure 3: Learning curve for basic Q-learning with different learning rates, 50k time steps were trained on *lunar lander* task. I choose the learning rate parameter to evaluate the sensitivity to hyper-parameters of the algorithm. Typically, larger learning rate fasten the convergence, but it is easy to jump out of a good local minima and finally make the agent fall into a terrible place. Small learning rates requires more iteration to get to a convergence, but the training procedures are more stable, which results in better performance. However, the performance varies in different tasks. So choosing a good learning rate is important in many RL tasks.

2 Actor-Critic

2.1 Sanity check with Cartpole

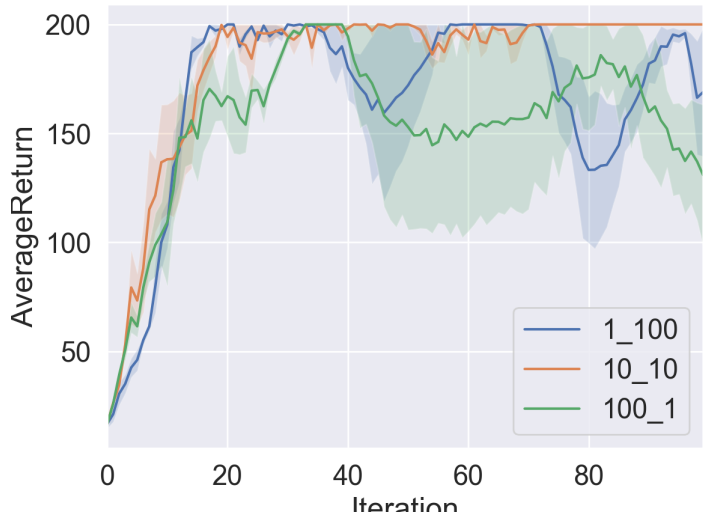

Figure 4: Learning curve for *Cartpole* task with different hyperparameters. From the plot, we can see (10, 10) is the best parameter set for it converges quickly and more stable compared to (1, 100).

2.2 Run actor-critic with more difficult tasks

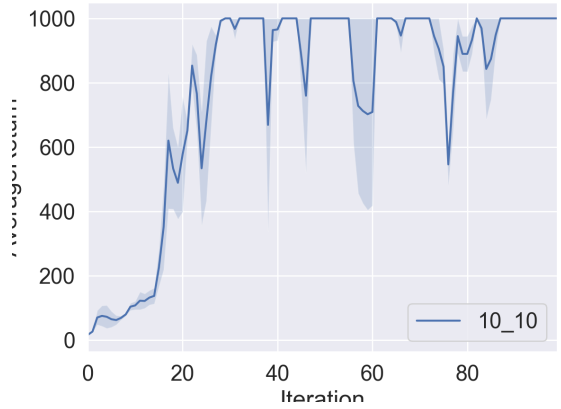

Figure 5: Learning curve for *InvertedPendulum* task.

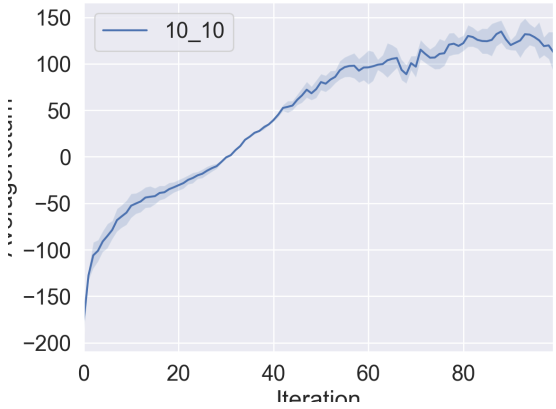

Figure 6: Learning curve for *HalfCheetah* task.