ESE 650, Spring 2020

Homework 5

Yuxiang Qiao [qiaoyx@seas.upenn.edu]

April 24, 2020

Solution 1.

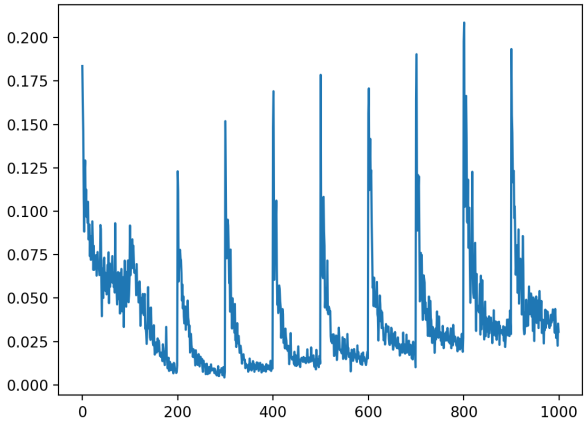

Figure 1: Average loss after each iteration on training set

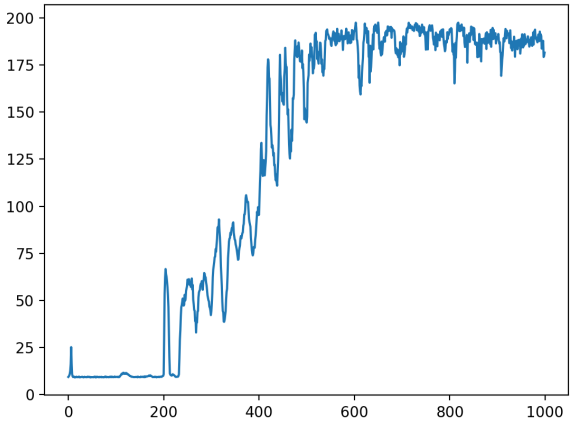

Figure 2: Average cumulative reward after each iteration on training set

For each iteration, the mini-batch size I use is 64. I track loss and average cumulative reward evaluated on 50 trajectories for each iteration. I update target network every 100 iterations. Each time I update the target network, the loss will increase as shown in figure 1. The average reward will increase and become stable over time. The average cumulative reward will converge to 200 eventually.