ESE 650 - Learning in Robotics
Homework 4

Akash Sundar

May 2023

1 Problem 1

a) In the code, the stochastic controller $u_\theta(\cdot|x)$ is implemented using a neural network with a single hidden layer. The neural network takes the state x as input and produces a mean and std. deviation for the normal distribution that generates the action u. The mean and standard deviation are then used to sample an action from the normal distribution. The log-likelihood log $u_\theta(u|x)$ is computed using the probability density function of the normal distribution with the sampled action u, mean and standard deviation obtained from the neural network.

The constraint $|u| \leq 1$ is imposed in the code by applying the tanh function to the output of the neural network. This scales the output of the neural network to the range [-1, 1], ensuring that the absolute value of the action is always less than or equal to 1.

b) Code implemented and submitted under code section

c)

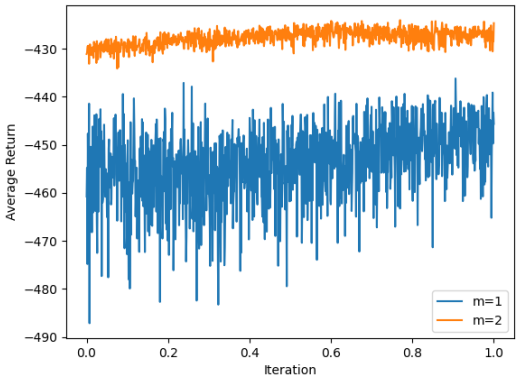

Figure 1: Comparison of Cumulative reward for change in mass

Since the pendulum has a higher mass, it will require more force to be applied to move it. This means that the optimal policy learned with a mass of 1 might not perform well with a mass of 2. As a result, the magnitude of the cumulative reward may decrease when the trained policy is evaluated on the pendulum with a mass of 2. Hence, the result of the experiment collaborates with theory.

2 Problem 2

1. Average return of training environment for every 1000 weight updates:

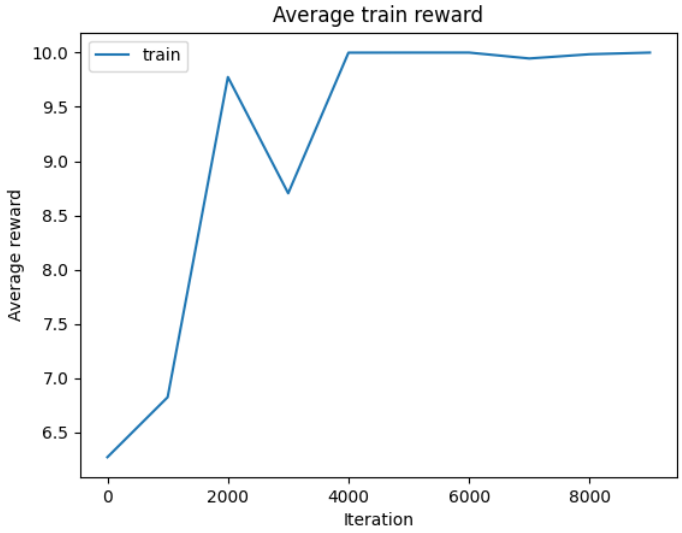

Figure 2: Average Reward - Training

2. Average return of evaluation environment for every 1000 weight updates:

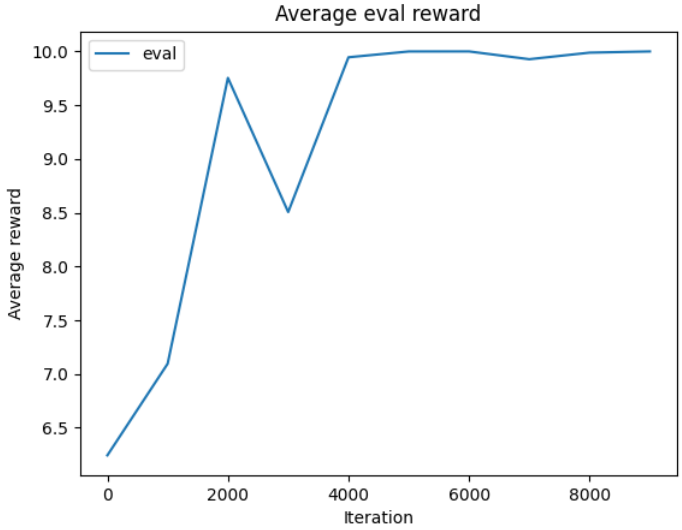

Figure 3: Average Reward - Evaluation

Running the code for a larger number of weight updates leads to a stabilization of the parameters as seen below:

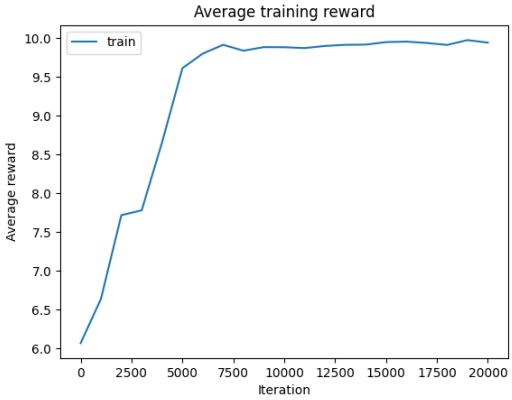

Figure 4: Average Reward - Training

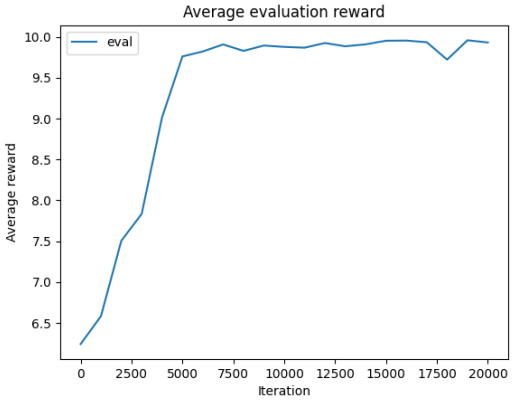

Figure 5: Average Reward - Evaluation