Assignment 1: Imitation Learning

Andrew ID: xzhan2
Collaborators: yixuanz4

1 Behavioral Cloning (65 pt)

1.1 Part 2 (10 pt)

TODO

Table 1: Report your result in this table.

Metric/Env	Ant-v2	Humanoid-v2	Walker2d-v2	Hopper-v2	HalfCheetah-v2
Mean	4713.6533203125	10344.517578125	5566.845703125	3772.67041015625	4205.7783203125
Std.	12.196533203125	20.9814453125	9.237548828125	1.9483642578125	83.038818359375

1.2 Part 3 (35 pt)

Table 2: A comparison of results for Ant-v2 and Hopper-v2. Hopper-v2 fails to achieve over 30%. The policy network size is 64-64-8, with training steps of 2000 per iteration, train batch size of 100, batch size of 1000. The comparison is set up with ep lenth of 1000 and eval batch size of 5000

Env	Ant-v2		[Hopper-v2]	
Metric	Mean	Std.	Mean	Std.
Expert	4713.6533203125	12.196533203125	3772.67041015625	1.9483642578125
BC	4555.95751953125	51.607460021972656	1073.1201171875	90.93077850341797

1.3 Part 4 (20 pt)

TODO, fill in the Fig. 1, provide some analysis.

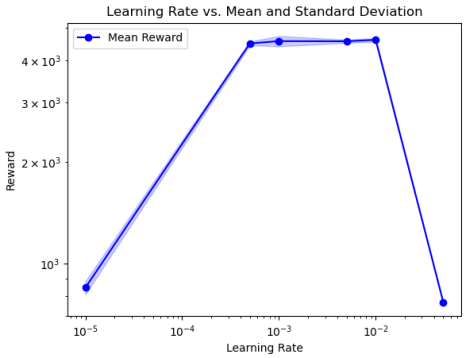

Figure 1: Performance change vs learning rate in ant evnvironment. Learning rate impact the learning results of policy a lot, thus impact the behaviour cloning performance. The blue curve shows the change of mean reward, while the error bar is the standard deviation. Too high or too low learning rate will both make the learning procedure slow or even stuck into local minimum, giving small mean reward

2 DAgger (35 pt)

2.1 Part 2 (35 pt)

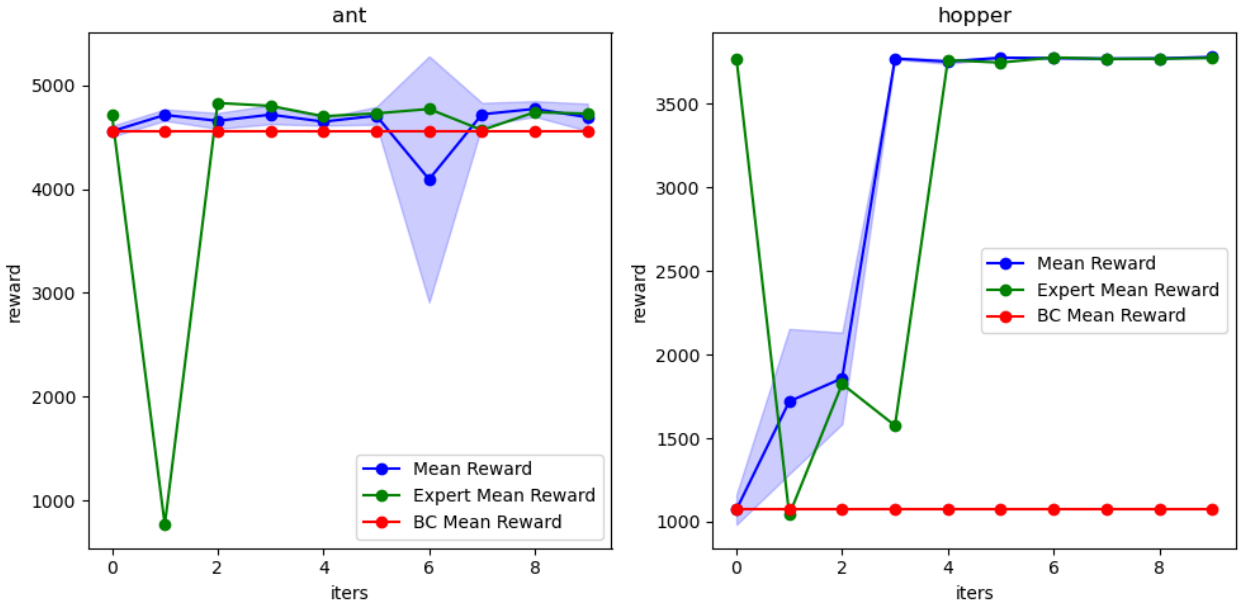

Figure 2: Learning curve, plotting the number of DAgger iterations vs. the policy's mean return, with error bars to show the standard deviation. I choose ant and hopper environments. The policy network size is of 64-64-8, with training steps of 2000 per iteration, train batch size of 100, batch size of 1000. The comparison is set up with ep lenth of 1000 and eval batch size of 5000