Assignment 1: Imitation Learning

Name: Jingzhou Liu

Andrew ID: jasonl6

1 Behavioral Cloning (65 pt)

1.1 Part 2 (10 pt)

Table 1: Mean and Standard Deviation of Return of the Expert Data

Metric/Env	Ant-v2	Humanoid-v2	Walker2d-v2	Hopper-v2	HalfCheetah-v2
Mean	4713.65	10344.52	5566.85	3772.67	4205.78
Std.	12.20	20.98	9.24	1.95	83.04

1.2 Part 3 (35 pt)

We compare the behavioral cloning (BC) policy's performance in the Ant-v2 environment against the Hopper-v2 environment. For both environments, the policy is trained using the following hyperparameters:

- `num_agent_train_steps_per_iter`: 1000
- `batch_size`: 1000
- `eval_batch_size`: 5000
- `train_batch_size`: 300
- `ep_len`: 1000
- `n_layers`: 3
- `size`: 128
- `learning_rate`: 10e-3

From Table 2, we observe that BC in the Ant-v2 environment achieves 82% of the expert's performance, whereas BC only achieves 21.7% of the expert's performance in the Hopper-v2 environment.

Table 2: Mean and Standard Deviation of Behavioral Cloning Policies vs. Expert Data

Env	Ant-v2		Hopper-v2	
Metric	Mean	Std.	Mean	Std.
Expert	4713.65	12.20	3772.67	1.95
BC	3870.38	103.15	819.60	249.66

1.3 Part 4 (20 pt)

We analyze the performance of the behavioral cloning agent by varying the learning rate of training on the Ant-v2 environment, illustrated by Figure 1. Note that all other hyperparameters are kept identical to the set described in Section 1.2. We choose to analyze learning rate as we found it empirically having the largest impact on the policy's performance; for instance, increasing the learning rate from 0.005 to 0.01 improves the evaluation mean performance from 29.5% to 82% of the performance of the expert. We note that a learning rate of 0.01 attains both the highest performance and the lowest standard deviation during evaluation.

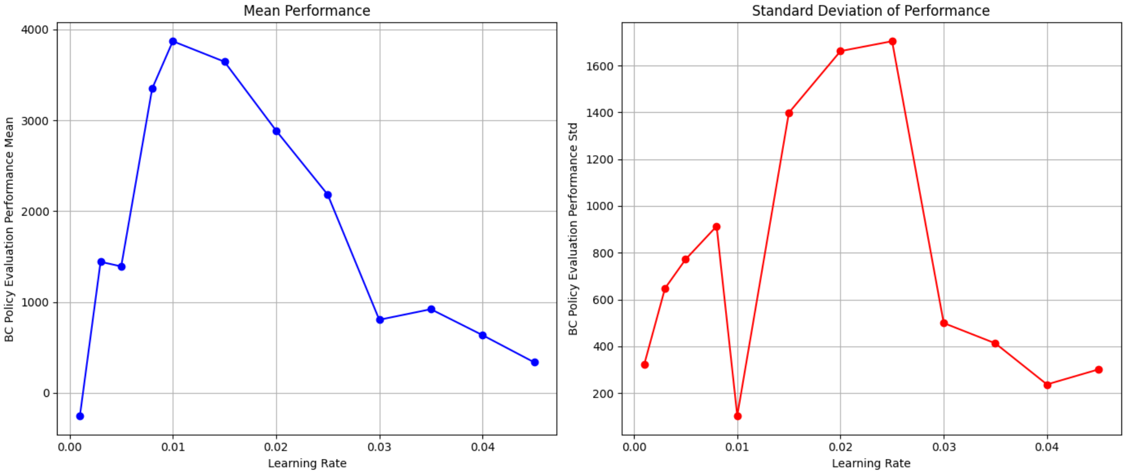

Figure 1: BC agent's performance varies with the value of learning rate parameter in the Ant-v2 environment.

2 DAgger (35 pt)

2.1 Part 2 (35 pt)

Note that these are trained using the same hyperparameters are described in Section 1.3, with the exception of `learning_rate` being lowered to 5e-3 to result in more stable training.

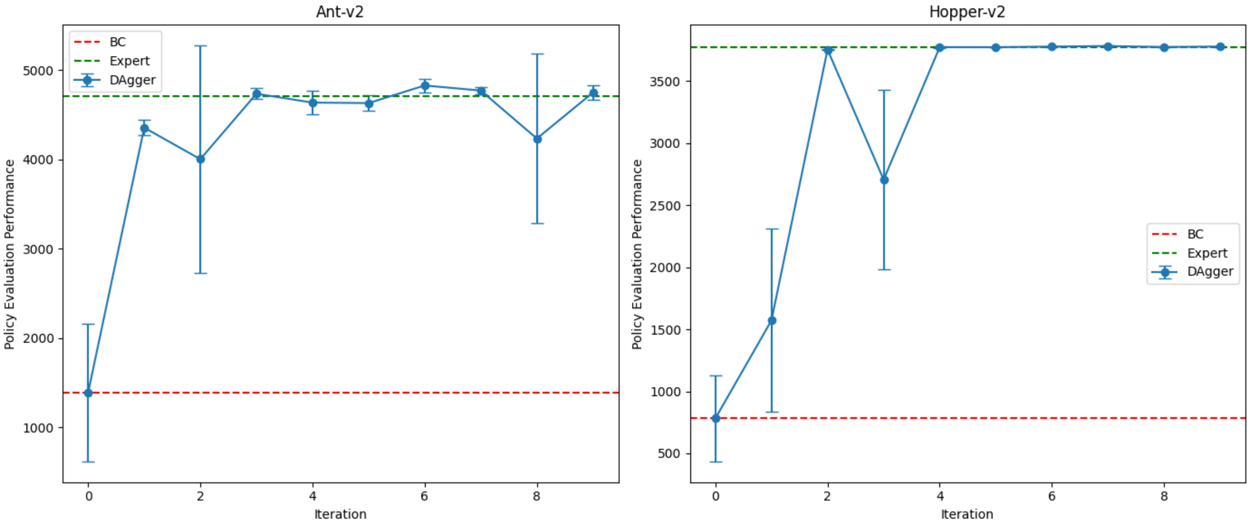

Figure 2: Learning curve, plotting the number of DAgger iterations vs. the policy's mean return, with error bars to show the standard deviation. The left plot shows the performance for the Ant-v2 environment and the right plot shows the performance for the Hopper-v2 environment.