Assignment 1: Imitation Learning

Andrew ID: `peiqiy`
Collaborators: `None`

1 Behavioral Cloning (65 pt)

1.1 Part 2 (10 pt)

Table 1 shows the mean and standard deviation of return over five trajectories of expert data for five environments.

Table 1: The 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)

Table 2 shows the eval results on two tasks: Ant-v2 and Hopper-v2. The parameters are set as follows:

- eval_batch_size = 2000
- num_agent_train_steps_per_iter = 1500

Table 2: Eval results on two tasks.

Env	Ant-v2		Hopper-v2	
Metric	Mean	Std.	Mean	Std.
Expert	4713.65	12.20	3772.67	1.95
BC	2824.39	1362.63	582.95	273.92

1.3 Part 4 (20 pt)

Figure 1 shows BC agent performance on the Ant-v2 task, which varies with the value of the parameter "num_agent_train_steps_per_iter". The reason I chose this parameter is because it directly controls how many training steps the agent takes in each iteration, thereby influencing how well the policy improves during the training process. A higher value allows the agent to train for a longer period before each evaluation, potentially leading to more refined policy updates. From this figure, we can see that as the number of training steps increases, the average evaluation performance also increases, indicating that additional training helps the policy to converge toward better behaviors and achieve higher returns on the task. This trend suggests that allowing more steps per iteration gives the agent more time to learn effective control strategies, which is especially important in tasks as complex as Ant-v2.

2 DAgger (35 pt)

2.1 Part 2 (35 pt)

Figure 2 shows the eval results on two tasks: Ant-v2 and Hopper-v2. The parameters are set as follows:

- eval_batch_size = 2000
- num_agent_train_steps_per_iter = 1500

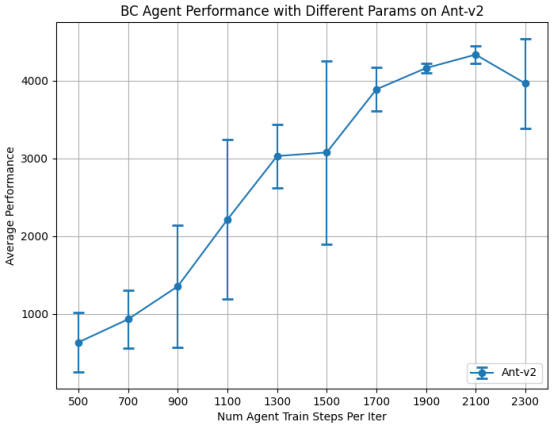
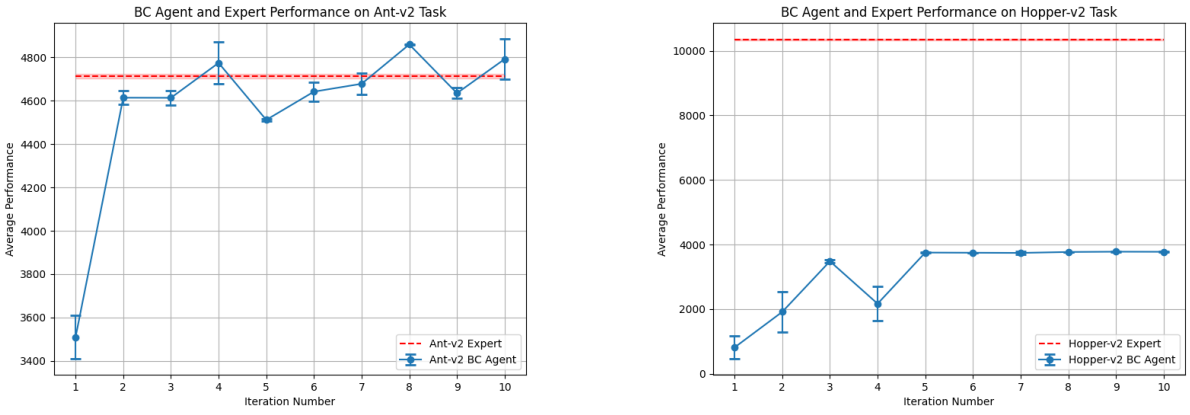


Figure 1: BC agent's performance varies with the value of num steps parameter in Ant-v2 environment.



(a) Ant-v2 environment learning curve.

(b) Hopper-v2 environment learning curve.

Figure 2: Learning curve on two tasks