Assignment 1: Imitation Learning

Andrew ID: `kanlongy`
Collaborators: `none`

1 Behavioral Cloning (9.75 pt)

1.1 Part 2 (1.5 pt)

TODO

Table 1: Report your result in this table.

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 (5.25 pt)

TODO

Table 2: Fill your results in this table, listing hyperparameters in this caption.

Env	Ant-v2		Humanoid-v2	
Metric	Mean	Std.	Mean	Std.
Expert	4713.65	12.20	10344.52	20.98
BC	4739.81	79.82	306.44	56.53

1.3 Part 4 (3 pt)

As shown in Fig. 1, we varied traning steps from 100 to 10000 while keeping other hyperparameters fixed. Error bars represent standard deviation over 5 rollouts. We chose this hyperparameter because it directly controls how well the policy can fit the expert demonstrations — insufficient training leads to underfitting, while more training allows the model to better approximate the expert's behavior.

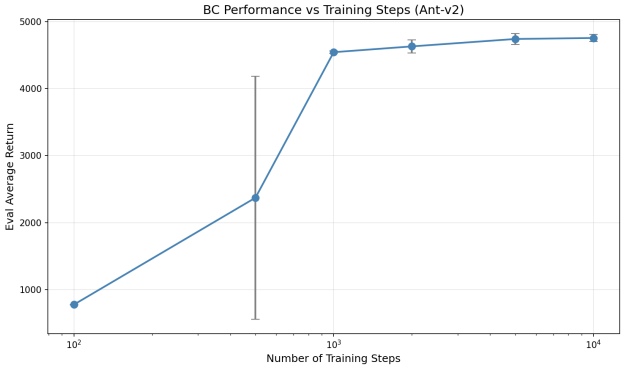

Figure 1: BC agent's performance varies with the value of training steps parameter in Ant-v2 environment.

2 DAgger (5.25 pt)

2.1 Part 2 (5.25 pt)

DAgger iteratively queries the expert to relabel collected observations, allowing the policy to recover from distribution shift that limits pure behavioral cloning.

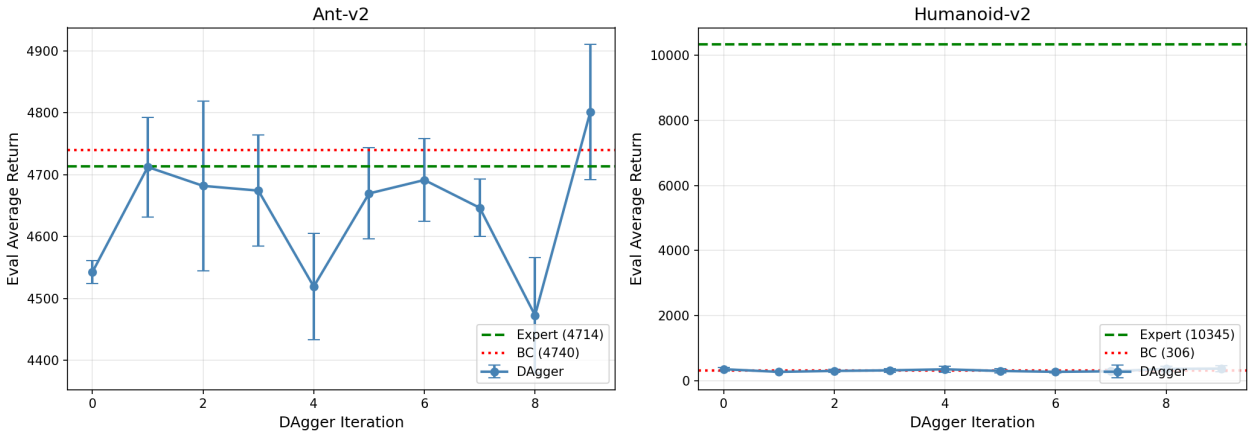

Figure 2: DAgger learning curves on Ant-v2 (left) and Humanoid-v2 (right). The blue line shows DAgger performance over 10 iterations with error bars representing standard deviation over 5 rollouts. Horizontal lines indicate Expert (green, dashed) and BC (red, dotted) performance.