Assignment 1: Imitation Learning

Andrew ID: `jihyunki`
Collaborators:

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	4601.81	126.97	254.61	17.43

1.3 Part 4 (3 pt)

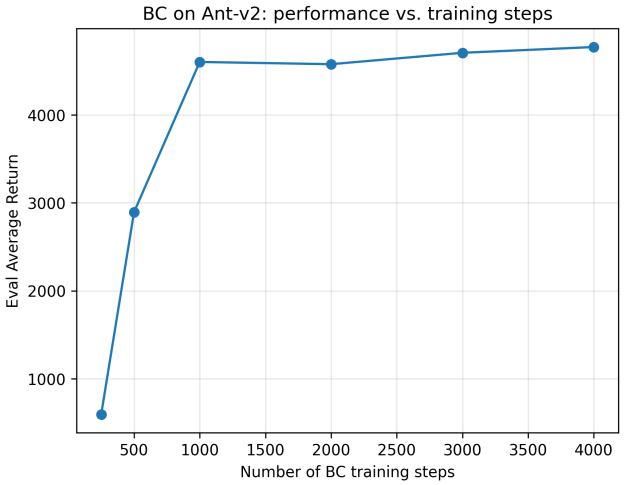

Figure 1: Behavioral cloning (BC) performance on **Ant-v2** as a function of the *number of training steps per iteration*. Points show the evaluation mean return over at least five rollouts with standard-deviation error bars. Increasing the training steps dramatically improves performance from severe underfitting (250 steps) to near-expert performance around 3–4k steps, after which gains taper off.

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

2 DAgger (5.25 pt)

2.1 Part 2 (5.25 pt)

TODO, Report the results for Ant-v2 and [Another Env] in Fig. 2:

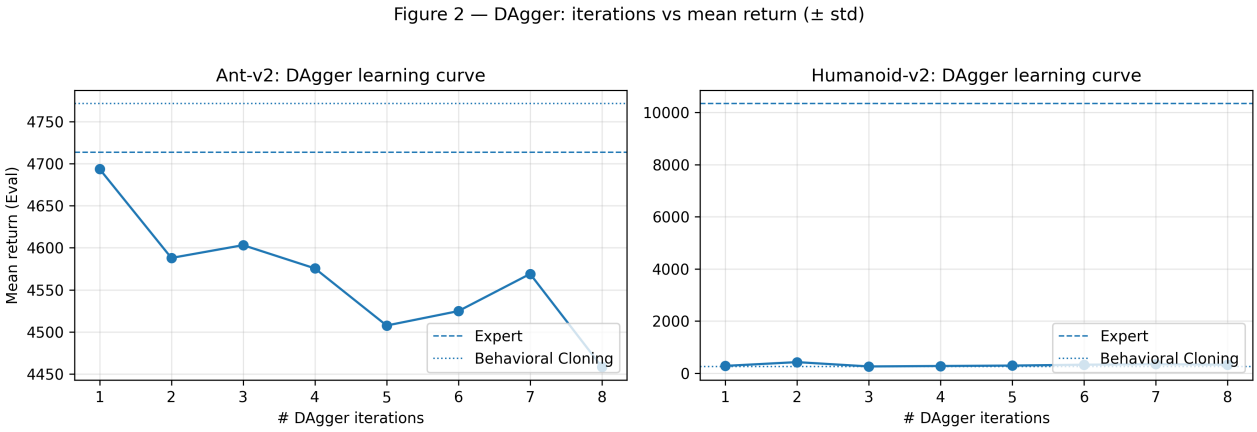

Figure 2: DAgger performance on **Ant-v2** (left) and **Humanoid-v2** (right). Curves plot evaluation mean return versus DAgger iteration, with error bars showing the standard deviation across evaluation rollouts. Horizontal dashed/dotted lines show the expert policy and behavioral cloning (BC) baselines, respectively. Setup: DAgger used a 3-layer MLP (`-n_layers 3`), learning rate 4×10^{-3}, evaluation batch size 5000; expert datasets from `rob831/expert_data` (Ant-v2 and Humanoid-v2). BC baselines were trained with the same learning rate and (for Ant) a 5-layer MLP.