Assignment 1: Imitation Learning

Andrew ID: haoyus
Collaborators: Myself

1 Behavioral Cloning (65 pt)

1.1 Part 2 (10 pt)

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 (35 pt)

Test Ant-v2 and Humanoid-v2 tasks with n_iter = 1, eval_batch_size = 5000, num_agent_train_steps_per_iter = 1000. Compared to the expert policy with Behavioral Cloning, the BC performance of Ant-v2 reaches 30% of expert policy while Humanoid does not.

Table 2: Fill your results in this table.

Env	Ant-v2		Humannoid-v2	
Metric	Mean	Std.	Mean	Std.
Expert	4713.65	12.20	10344.52	20.98
BC	3782.38	1823.78	283.81	11.96

1.3 Part 4 (20 pt)

1. First test with *num_agent_train_steps_per_iter* in Humanoid-v2, varies from 100-10000.

By comparing the results, it shows that *num_agent_train_steps_per_iter* = 1000 gives the best balance between performance and stability, which returns a good mean (283.81) and lowest standard deviation (11.96). A smaller step size of 100 gives the highest mean return (331.42) with a higher standard deviation (21.95), which shows instability. A larger step size of 10000 leads to more instability and a moderate mean return, and doesn't improve performance much.

The results show that a smaller step size helps the model generalize more without overfitting, but also limits the data the model can explore. A larger step size leads to overfitting with a high standard deviation.

Table 3: Comparison of *num_agent_train_steps_per_iter* in Humanoid-v2

Metric/Para	100	500	1000 (default)	2000	4000	10000
Mean	331.42	276.30	283.81	264.30	272.92	293.14
Std.	21.95	14.88	11.96	20.40	19.02	47.06

2. Second test the train_batch_size in Humanoid-v2, varies from 10-1000.

By comparing the results, it shows that *train_batch_size* = 100 gives a good trade-off between learning speed, stability, and performance. A smaller train batch size of 10 leads to a higher mean return with higher instability due to noise and outlier data. A larger train batch size results in more stable performance but also lacks the

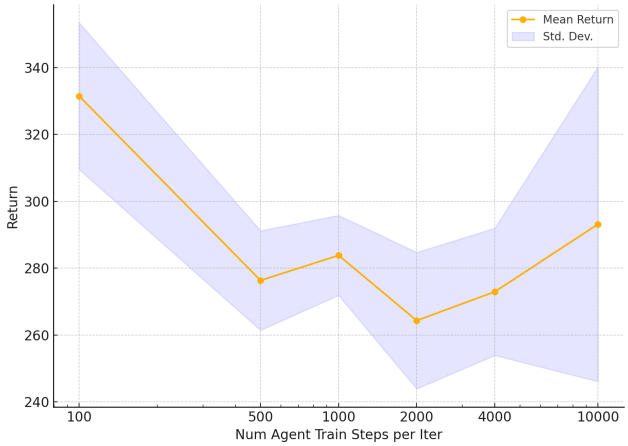

Figure 1: BC agent's performance varies with the value of *train_batch_size* parameter in Humanoid-v2 environment.

generalization from small random sample train data. Also with the batch size varying from 200-1000, the mean return increases slightly due to diminishing.

Table 4: Comparison of *train_batch_size* in Humanoid-v2

Metric/Para	10	50	100 (default)	200	500	1000
Mean	344.13	251.05	283.81	297.33	298.97	291.38
Std.	86.81	13.97	11.96	36.36	40.36	39.83

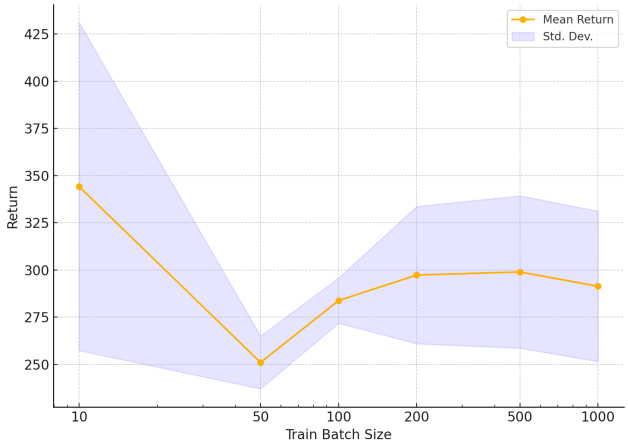

Figure 2: BC agent's performance varies with the value of *train_batch_size* parameter in Humanoid-v2 environment.

2 DAgger (35 pt)

2.1 Part 2 (35 pt)

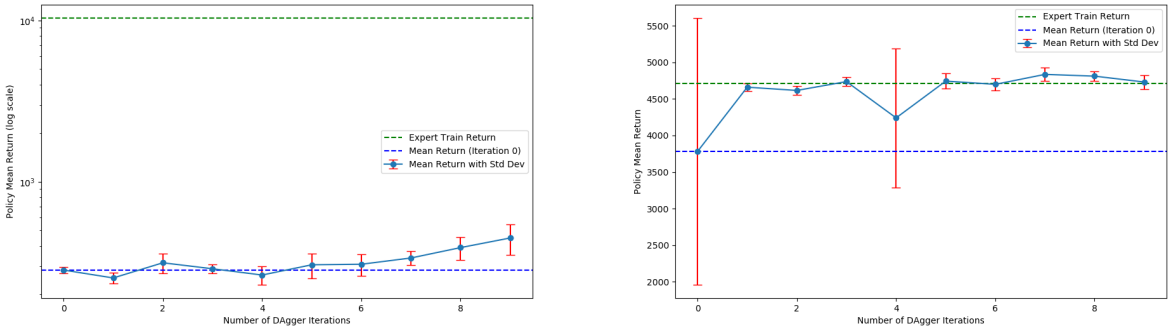

Figure 3: DAgger learning curve (Left: Ant-v2, Right: Humanoid-v2) with default parameters