

Assignment 1: Imitation Learning

Andrew ID: bharathh
Collaborators: None

1 Behavioral Cloning (9.75 pt)

1.1 Part 2 (1.5 pt)

Table 1: Mean and standard deviation of return over two trajectories of the expert data for each environment.

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)

Table 2: Hyperparameters: n_layers=5, learning_rate=3e-3, eval_batch_size=5000; rest are default values. BC Agent reaches 47.85% and 7.99% of the expert performance for Ant-v2 and Hopper-v2 environments respectively.

Env	Ant-v2		Hopper-v2		
	Metric	Mean	Std.	Mean	Std.
Expert	4713.65	12.20	3772.67	1.95	
BC	2255.59	1699.11	301.59	359.24	

1.3 Part 4 (3 pt)

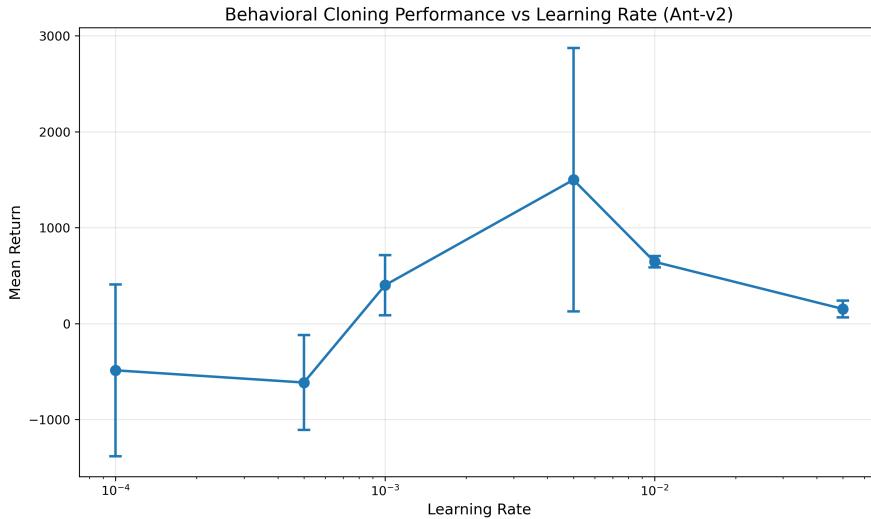


Figure 1: The graph shows mean return \pm standard deviation of the BC Agent in Ant-v2 environment over approximately five rollouts for each learning rate value: $10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}$ and 5×10^{-2} . Rest of the hyperparameters are same as in Table 2. Learning rate was chosen because it controls the gradient step size and can impact both convergence speed and final performance, as can be seen from the graph.

2 DAgger (5.25 pt)

2.1 Part 2 (5.25 pt)

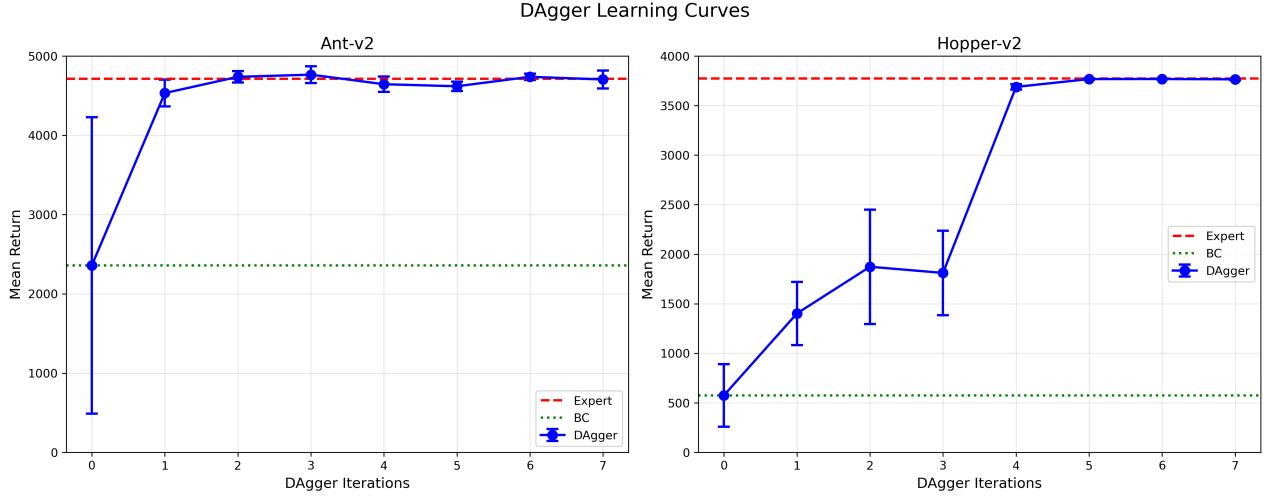


Figure 2: DAgger learning curves for Ant-v2 (left) and Hopper-v2 (right) environments. The graphs show mean return \pm standard deviation over approximately five rollouts for each DAgger iteration. Hyperparameters: n_layers=3, n_iter=8, eval_batch_size=5000; rest are default values. As seen, DAgger outperforms BC agent, reaching 99.8% of the expert performance in both environments.