Assignment 1: Imitation Learning

Andrew ID: JAROLOVI
Collaborators: MREICH

1 Behavioral Cloning (9.75 pt)

1.1 Part 2 (1.5 pt)

Metric/Env	Ant-v2	Humanoid-v2	Walker2d-v2	Hopper-v2	HalfCheetah-v2
Mean	4713.65	10344.52	5566.85	3772.67	4205.79
Std.	12.20	20.98	9.24	1.95	83.04

1.2 Part 3 (5.25 pt)

Table 1: Behavior cloning (BC) policies were trained with a batch size of 1000, and evaluation batch size of 15000, a training batch size of 100, a 3-layer MLP with a size of 64, a constant learning rate of 5e-3, and 1000 gradient steps per training iteration.

Env	Ant-v2		Humanoid-v2	
Metric	Mean	Std.	Mean	Std.
Expert	4713.65	12.20	10344.52	29.98
BC	4553.22	97.67	306.12	88.61

1.3 Part 4 (3 pt)

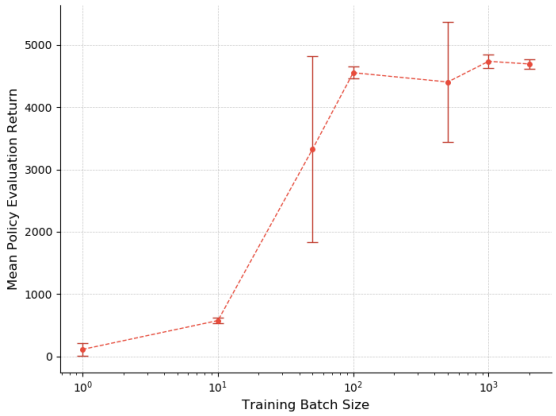

Figure 1: The BC agent's performance in the Ant-v2 environment varies with training batch size. Training batch size is a crucial hyperparameter that impacts performance, determining how many samples are used for gradient update. Smaller batch sizes are more memory-efficient and can encourage generalization due to increased stochasticity. However, this same noise can lead to high variance in gradient updates, destabilizing training and potentially yielding maltrained models. The results indicate a saturation in returns at a training batch size of 100. Larger batch sizes offer little additional performance gain, whereas smaller batch sizes tend to result in maltrained models.

2 DAgger (5.25 pt)

2.1 Part 2 (5.25 pt)

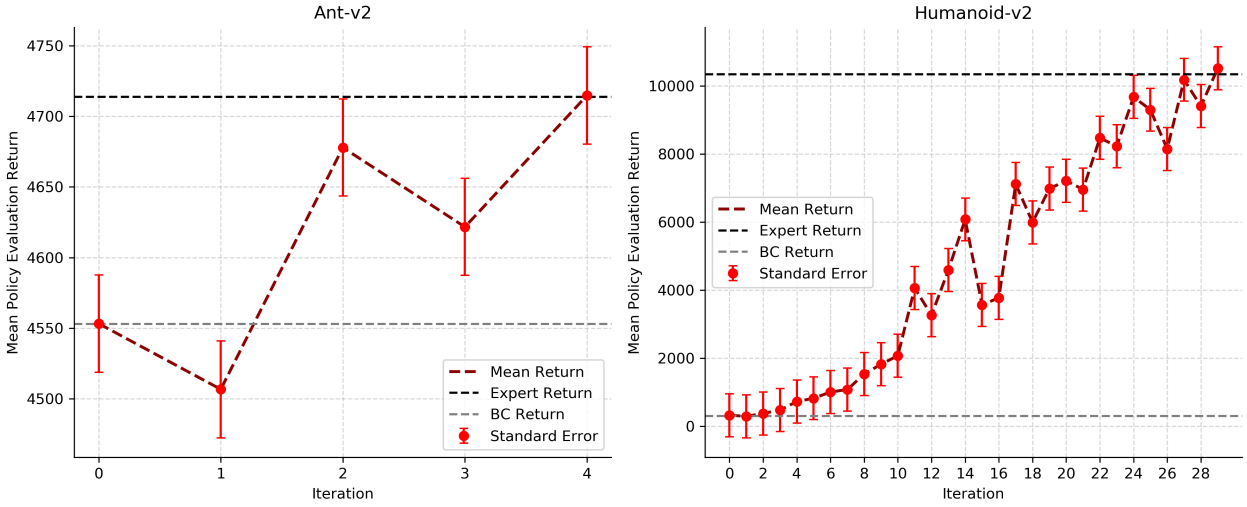

Figure 2: Learning curves depicting the progression of mean evaluation returns for DAgger in the Ant-v2 and Humanoid-v2 environments over training iterations, illustrating convergence toward expert-level performance. Expert mean and behavior cloning (BC) results are also shown for comparison. In the Ant-v2 environment, the policy was trained over 5 iterations using a batch size of 1000, an evaluation batch size of 15,000, a training batch size of 100, a three-layer MLP with 64 hidden units per layer, a constant learning rate of 5e-3, and 1000 gradient steps per training iteration. In the Humanoid-v2 environment, the policy was trained over 30 iterations using a batch size of 1000, an evaluation batch size of 15,000, a training batch size of 100, a four-layer MLP with a 128 hidden units per layer, a constant learning rate of 5e-4, and 1000 gradient steps per training iteration.