Assignment 1: Imitation Learning

Andrew ID: `dameria`
Collaborators: `parthsin, ankitagg, Gen AI`

1 Behavioral Cloning

1.1 Part 2

Environment Name	Eval Avg Return	Eval STD	Expert Avg Return	Accuracy
Ant	2124	1053	4713	45%
HalfCheetah	3418	124	4205	81%
Hopper	541	234	3772	14%
Humanoid	265	4637	10344	2.5%
Walker2D	819	820	5566	6.79%

1.2 Part 3

I'm choosing Hopper as the dataset to make a comparative study with Ant. The result logs are shown below

Table 1: Comparison between Hopper and Ant environments. Hyperparameters- Number of Layer: 4, Learning Rate: 5e-3, Eval Batch Size: 5000, Train Batch Size: 300

Env	Ant-v2		Hopper	
Metric	Mean	Std.	Mean	Std.
Expert	4713	12.2	3772	1.95
BC	2124	1053	541	234

1.3 Part 4

I chose train_batch_size as the hyperparameter to study for this part. The train_batch_size is a crucial hyperparameter in behavioral cloning because it dictates how much data the model sees at once to compute a gradient update. It directly affects the stability and speed of training.

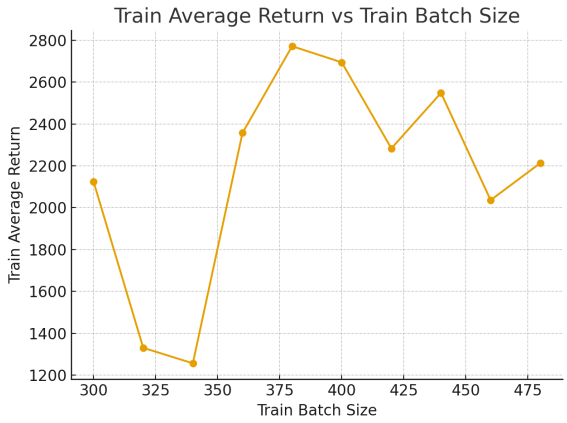


Figure 1: Eval_Average vs Train Batch Size

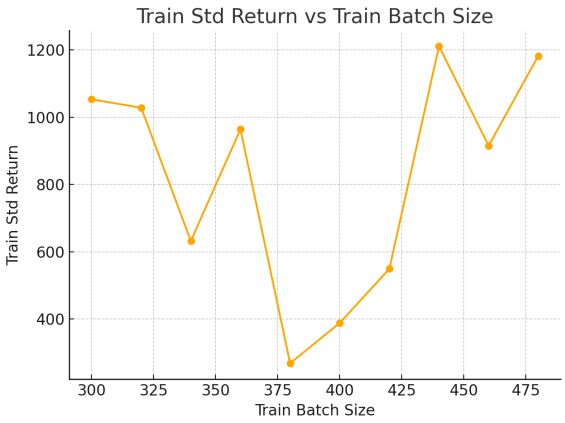
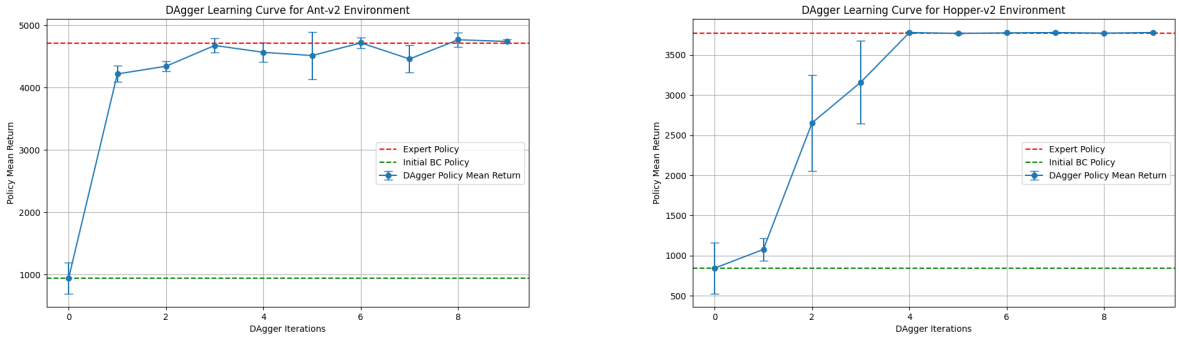


Figure 2: Eval_Std vs Train Batch Size

In my ablation study, the sweet spot appears to be a batch size of 375, where the batch is large enough to provide a stable, efficient gradient but small enough to introduce a little bit of helpful noise. This is why it not only yielded the highest average return (a sign of good performance) but also the lowest standard deviation (a sign of consistent and reliable performance).

2 DAgger

2.1 Part 2



(a) Ant-v2 Environment

(b) Hopper-v2 Environment

Figure 3: Learning curves plotting the number of DAgger iterations vs. the policy's mean return, with error bars showing the standard deviation. (a) Ant-v2 Environment. (b) Hopper-v2 Environment. Both plots include horizontal dashed lines for the expert policy (red) and the initial behavioral cloning policy (green) performance.