

Assignment 1: Imitation Learning

Andrew ID: dsreeniv

Collaborators: None

1 Behavioral Cloning

1.1 Part 2

Environment Name	Eval Avg Return	Eval STD	Expert Avg Return	Accuracy
Ant	1578	544	4713	33%
HalfCheetah	2612	121	4205	62%
Hopper	521	179	3772	13.81%
Humanoid	285	46	10344	2.76%
Walker2D	353	276	5566	6.34%

1.2 Part 3

From the above table, I chose Hopper as a dataset close to Ant-v2 due to similar expert costs, training data (2000 observations) and hyperparameters. The full results log is as below:

Output	Ant-v2	Hopper-v2
Eval Avg Return	1578	521
Eval Std Return	544	179
Eval Avg Ep Len	1000	186
Train Avg Return	4713	3772
Train Std Return	12	1.94
Train Avg Ep Len	1000	1000

This demonstrates that Behaviour Cloning does not perform well on Hopper, but does well on Ant since the average evaluation episode length is much shorter for Hopper

1.3 Part 4

I decided to vary the learning rate to understand its impact on the MLP model accuracy for Ant-v2. The chart below illustrates this effect

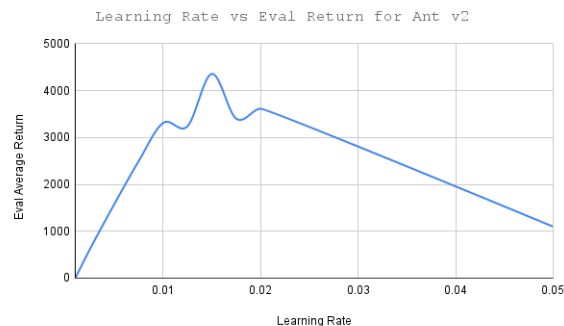


Figure 1: Learning Rate effect on Avg Return

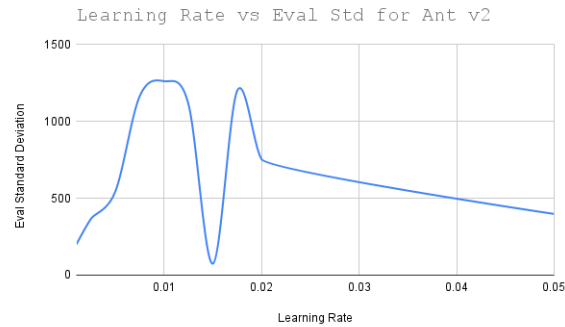


Figure 2: Learning Rate effect on Standard Deviation

I chose this parameter since the average episode length was 200 with the default values which meant the network hadn't really learnt anything - increasing the learning rate would allow for quicker convergence and to hit higher accuracy in the given iterations.

2 DAgger

2.1 Part 2

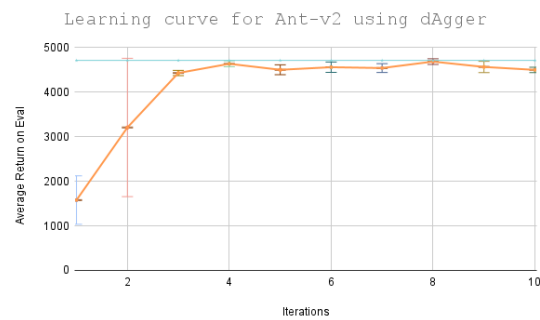


Figure 3: dAgger performance on Ant-v2 over 10 iterations

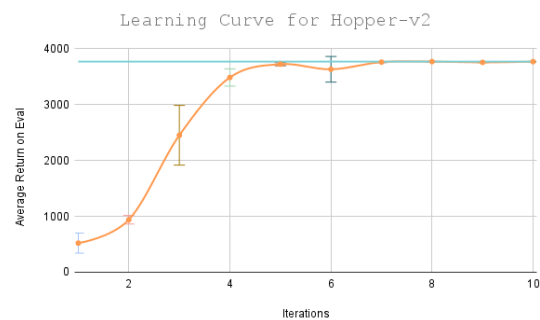


Figure 4: dAgger performance on Ant-v2 over 10 iterations

For the above curves, all the parameters were left at default - (Learning rate = $5e-3$, Number of layers = 2, Layer Width = 64)