Assignment 1: Imitation Learning

Andrew ID: `shriishs`
Collaborators: `None`

1 Behavioral Cloning

1.1 Part 2

Environment	Expert's average return	Expert's std return
Ant-v2	4713.6533203125	12.196533203125
Humanoid-v2	10344.517578125	20.9814453125
Walker2d-v2	5566.845703125	9.237548828125
Hopper-v2	3772.67041015625	1.9483642578125
HalfCheetah-v2	4205.7783203125	83.038818359375

Table 1: Expert data

1.2 Part 3

Environment	Expert's average return	Eval average return	Eval std return	Percentage
Ant-v2	4713.6533203125	3287.592041015625	2010.4427490234375	69.74%
Humanoid-v2	10344.517578125	304.390625	38.27568054199219	2.94%
Walker2d-v2	5566.845703125	2088.454345703125	2104.83837890625	37.51%
Hopper-v2	3772.67041015625	947.5502319335938	168.98275756835938	25.11%
HalfCheetah-v2	4205.7783203125	3951.74755859375	62.33942413330078	93.95%

Table 2: Default parameters with changes eval_batch_size = 5000 & n_layers = 1

1.3 Part 4

The hyperparameter chosen for tuning is the number of gradient steps in an iteration. Increasing the number of gradient steps will allow the policy or the neural network to converge better. This will improve the policy's evaluation performance as seen below.
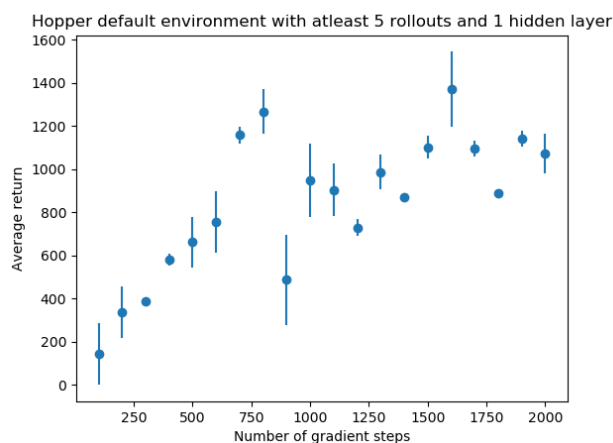

Figure 1: Hopper-v2 default parameters with changes eval_batch_size = 5000 & n_layers = 1
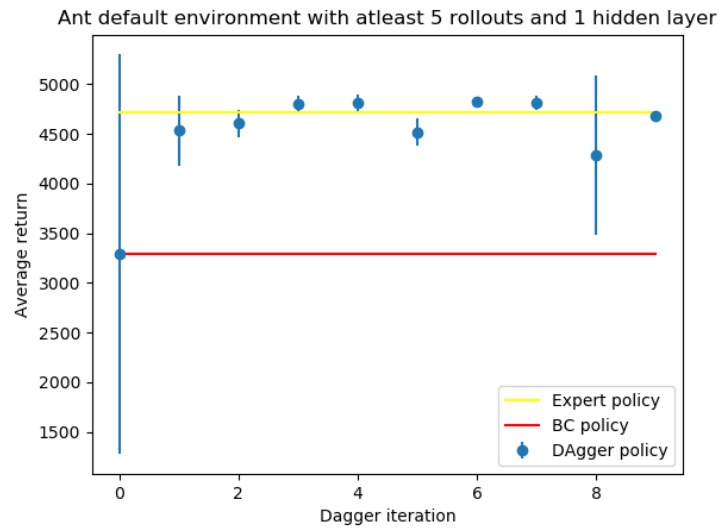
2   DAgger

2.1   Part 2

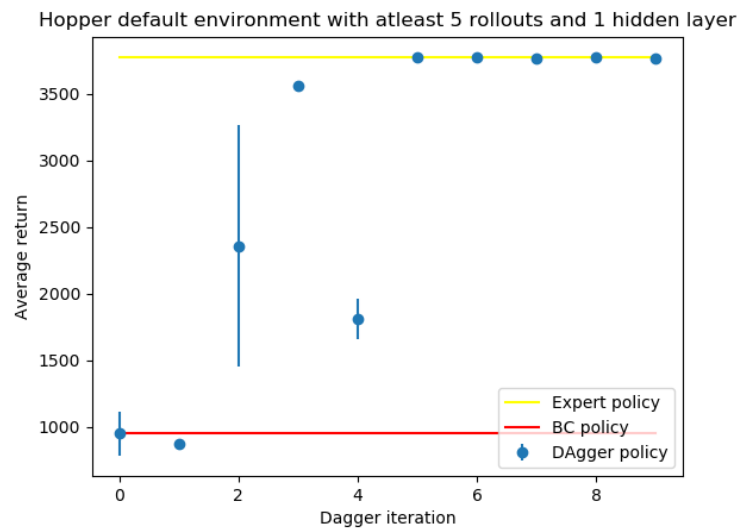Figure 2: Ant-v2 default parameters with changes n_iter = 10, eval_batch_size = 5000 & n_layers = 1

Figure 3: Hopper-v2 default parameters with changes n_iter = 10, eval_batch_size = 5000 & n_layers = 1