Assignment 1: Imitation Learning

Andrew ID: `vtambe`
Collaborators: `shriishs`

1 Behavioral Cloning

1.1 Part 2

Mean and Standard deviation for all 5 environments

Environment	Eval_AverageReturn	Eval_StdReturn	Initial_DataCollection_AverageReturn
Ant-v2	4693.75	73.92	4713.65
Hopper-v2	1259.73	143.44	3772.67
HalfCheetah-v2	4102.57	95.56	4205.78
Walker2d-v2	2866.81	2437.66	5566.85
Humanoid-v2	368.77	116.83	10344.52

Table 1: Mean and Standard Deviation for 5 Environments

1.2 Part 3

Table 3 has a comparison of the results of the behavior cloning for the provided expert Ant-v2 and Humanoid-v2 on the respective environments.

While conducting the experiment the parameters listed in table 2 were used.

Parameter	Ant-v2	Humanoid-v2
ep_len	1000	1000
eval_batch_size	5000	5000
num_agent_train_steps_per_iter	1000	1000
number of iterations	1	1
n_layers	2	2
size	64	64
learning_rate	0.005	0.005

Table 2: Parameters used for behaviour cloning on Ant-v2 and Humanoid-v2 Environments

Metric	Ant-v2	Humanoid-v2
Performance	99.57%	3.56%
Eval_AverageReturn	4693.75	368.77
Eval_StdReturn	73.92	116.83
Initial_DataCollection_AverageReturn	4713.65	10344.52

Table 3: Performance Metrics for Ant-v2 and Humanoid-v2 Environments

1.3 Part 4

Hyperparameter tuning for Humanoid2d-v2

The plot in figure 1 shows the results of hyperparameter tuning for behavior cloning on the Humanoid2d-v2 environment. The hyperparameter varied here was the "number of layers".

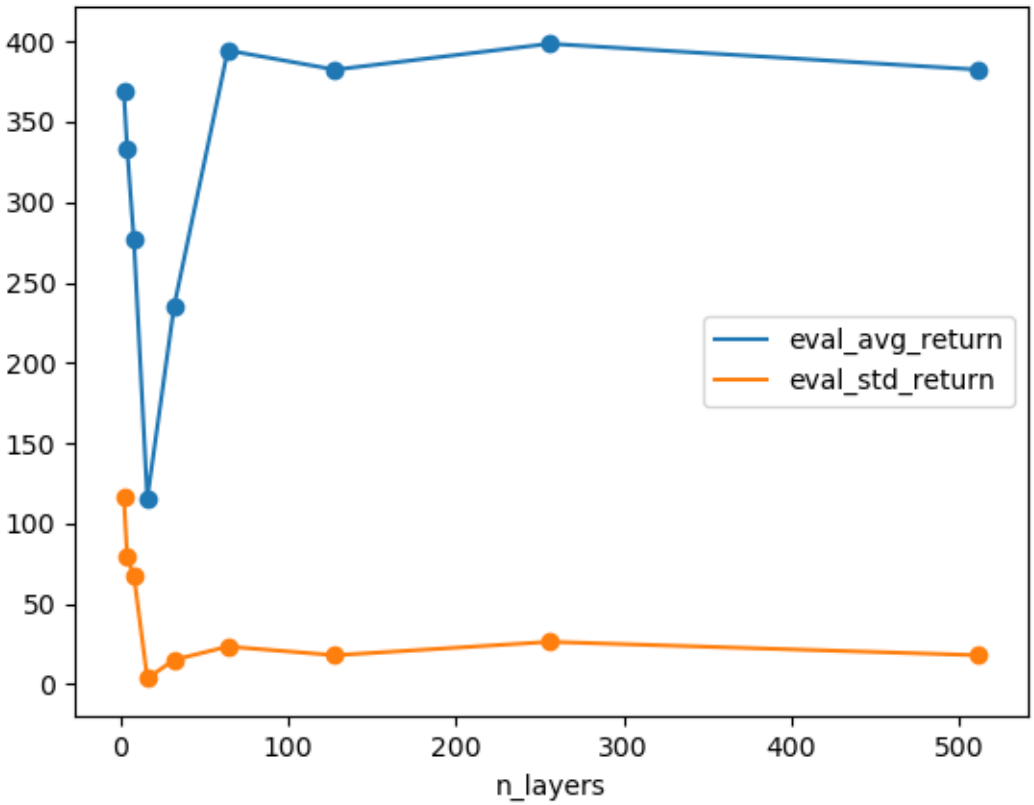

Figure 1: Plot of average returns and std deviation for Behaviour cloning on Humanoid2d-v2 environment vs number of layers. The idea behind choosing this hyperparameter was that the network seems to have a tough time approximating the expert policy as Humanoid2d is a very complex environment - having more layers would allow for more weights which would be helpful in approximating a more complex environment.

2 DAgger

2.1 Part 2

Figure 2 shows an errorbar plot for Dagger on Ant-v4 environment run for 10 itterations. We can see that the trained policy has eval performance better than the expert over 5 trajectories.

Figure 3 shows an errorbar plot for Dagger on Ant-v4 environment run for 20 itterations. We can see that the trained policy has eval performance close to the expert policy over 5 trajectories.

The parameters used during training are listed in the table 4

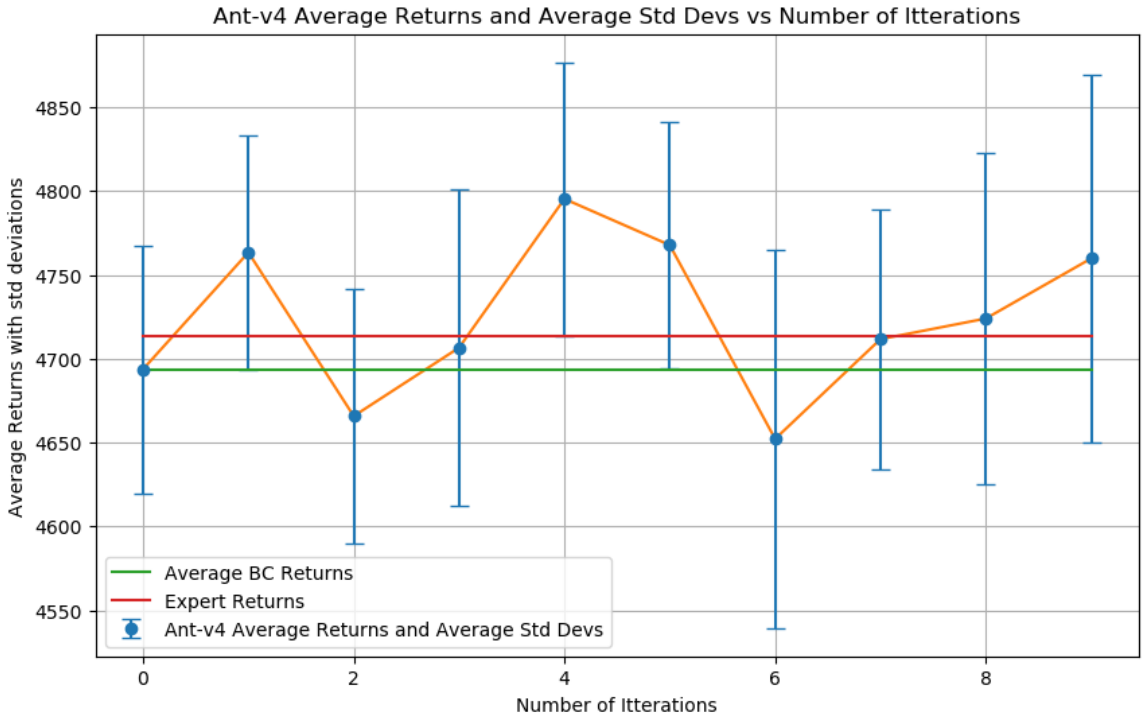

Figure 2: ErrorBar Plot of Average Returns and std deviation for Antv4 using DAGGER over 10 iterations.

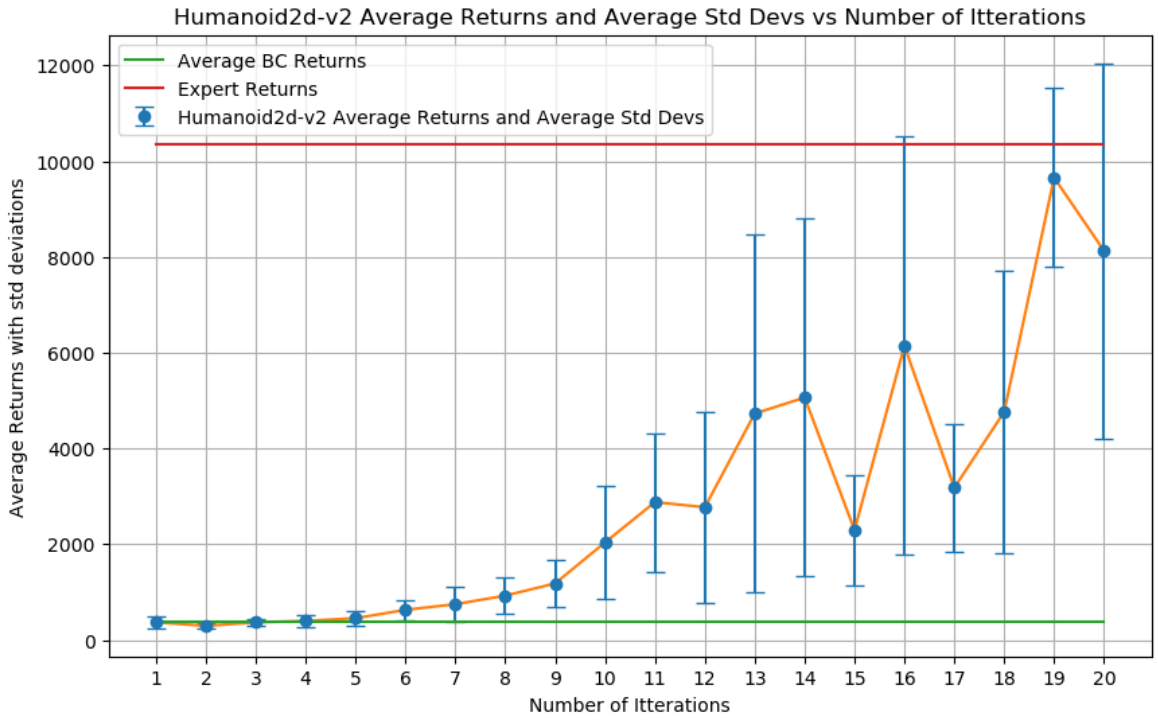

Figure 3: ErrorBar Plot of Average Returns and std deviation for Humanoid2d-v2 using DAGGER over 20 iterations.

Parameter	**Ant-v2**	**Humanoid-v2**
ep_len	1000	1000
eval_batch_size	5000	5000
num_agent_train_steps_per_iter	1000	1000
number of iterations	**10**	**20**
n_layers	2	2
size	64	64
learning_rate	0.005	0.005

Table 4: Parameters used for DAGGER on Ant-v2 and Humanoid-v2 Environments