Assignment 1: Imitation Learning

Andrew ID: mukaiy

1 Behavioral Cloning (65 pt)

1.1 Part 2 (10 pt)

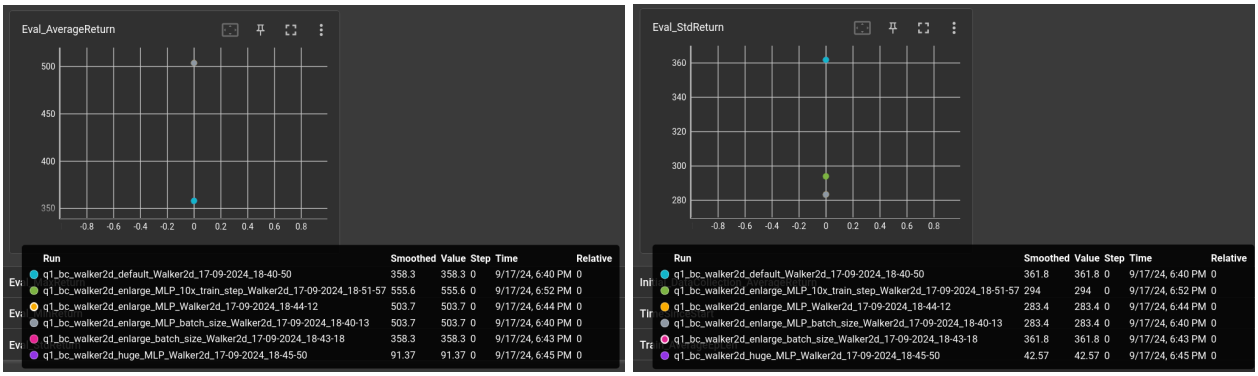
Metric/Env	Ant-v2	Humanoid-v2	Walker2d-v2	Hopper-v2	HalfCheetah-v2
Mean	4713.6533203125	10344.517578125	5566.845703125	3772.67041015625	4205.7783203125
Std.	12.196533203125	20.9814453125	9.237548828125	1.9483642578125	83.038818359375

1.2 Part 3 (35 pt)

Env	Ant-v2		Humanoid-v2	
Metric	Mean	Std.	Mean	Std.
Expert	4713.6533203125	12.196533203125	10344.517578125	20.9814453125
BC	4714.01708984375	93.93411254882812	248.40753173828125 (???)	55.91537094116211

Table 1: Run with *–eval_ batch_ size 5000 –num_ agent_ train_ steps_ per_ iter 10000*

1.3 Part 4 (20 pt)



(a) Average Return. (b) Std Return.

Figure 1: BC agent's performance varies with the value of batch size, training step, MLP size parameters in Walker 2D environment.

Experiment with Humanoid failed because of the complexity of observation (state) space (376 vs Ant's 111), could not improve with noticeable margin. Walker 2D has far less DoF (17 vs Ant's 111), but the dynamics is non-trivial.

All the experiment are conducted with *–eval_ batch_ size 5000*.

Analysis:

The dominant factor is the MLP depth/size and training step, increasing batch size does not help much. The huge MLP failed because of the overfitting issue, and the huge MLP size 512 failed because of the exploding/vanishing gradient issue.

Setup	Default	enlarge_batch_size	enlarge_MLP	huge_MLP
Batch size	1000	10000	1000	1000
Training step	1000	1000	1000	1000
MLP depth	2	2	5	10
MLP size	64	64	128	256 (512 failed)
Average Return	358.3	358.3	503.7	91.37
Std Return	361.8	283.4	283.4	**42.57**

Setup	enlarge_MLP_batch_size	enlarge_MLP_10x_train_step
Batch size	10000	1000
Training step	1000	10000
MLP depth	5	5
MLP size	128	128
Average Return	503.7	**555.6**
Std Return	283.4	294

Table 2: Experiment Setup.

2 DAgger (35 pt)

2.1 Part 2 (35 pt)

Result is shown in Fig. 2.

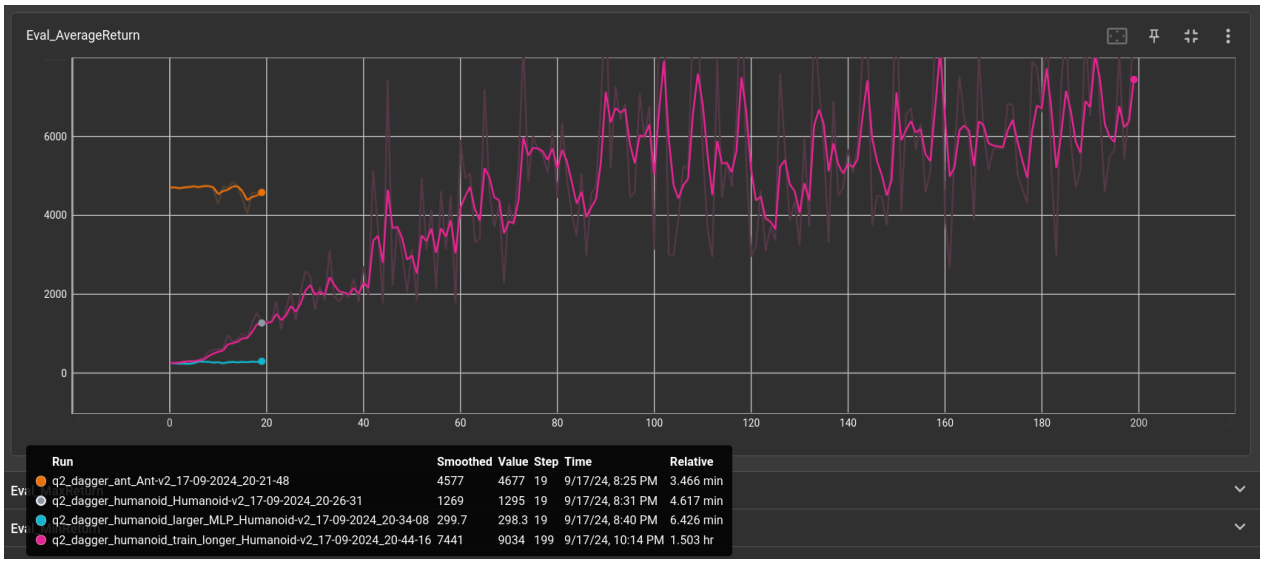
Analysis:

Ant agent in orange curve is already doing a good job in non-DAgger Behavioral Cloning.

In comparison, Humanoid agent's performance gets improved a lot by DAgger as expected. Increasing the depth and size of MLP does not help much compared to the default setting, so the Humanoid agent is trained with the default setting for the 200-iteration run in magenta.

Training DAgger agent with iterations more than 100 does not improve much, possibly due to the network architecture.

The standard deviation of the return is high in the Humanoid environment, which is expected because of the complexity of the environment. The Humanoid expert agent with such a low stda 20.98 must have been trained with a more sophisticated architecture/method.



(a) Average Return.

(b) Std Return.

Figure 2: Learning curve of DAgger agent in Ant-v2 and Humanoid-v2 environment.