Assignment 1: Imitation Learning

Andrew ID: liweiy
Collaborators: azou

1 Behavioral Cloning (9.75 pt)

1.1 Part 2 (1.5 pt)

Rounded to two decimals.

Table 1: Report your result in this table.

Metric/Env	Ant-v2	Humanoid-v2	Walker2d-v2	Hopper-v2	HalfCheetah-v2
Mean	4713.65	10344.52	5566.85	3772.67	4205.78
Std.	12.20	20.98	9.24	1.95	83.04

1.2 Part 3 (5.25 pt)

Rounded to two decimals.
Hyperparameters remains the same for both environment: ep_len: 1000, num_agent_train_steps_per_iter: 1000, batch_size: 1000, train_batch_size: 100, eval_batch_size: 5000, n_layers: 2, hidden_layer_size: 64, learning_rate: 5e-3.

Table 2: Fill your results in this table.

Env	Ant-v2		Humanoid-v2	
Metric	Mean	Std.	Mean	Std.
Expert	4713.65	12.20	10344.52	20.98
BC	4683.13	83.31	276.18	27.72

1.3 Part 4 (3 pt)

Rounded to two decimals.
I choose Humanoid-v2 environment as my experiment environment.
I modified three parameters, train_batch_size: 100 -> 500, n_layers: 2 -> 5, size: 64 -> 128.
The rationale behind this is to use a deeper and wider neural network to boost the performance. We can see average, max, and min returns all improved. However, the drawback of a deeper network may cause overfitting. The effect of overfitting is shown as the increase in StdReturn.

Table 3: Report your result in this table.

Metric/Hyperparameters	Original	Tuned
AverageReturn	276.18	301.42
StdReturn	27.72	48.58
MaxReturn	375.52	456.49
MinReturn	189.61	237.89
AverageEpLen	51.93	56.06

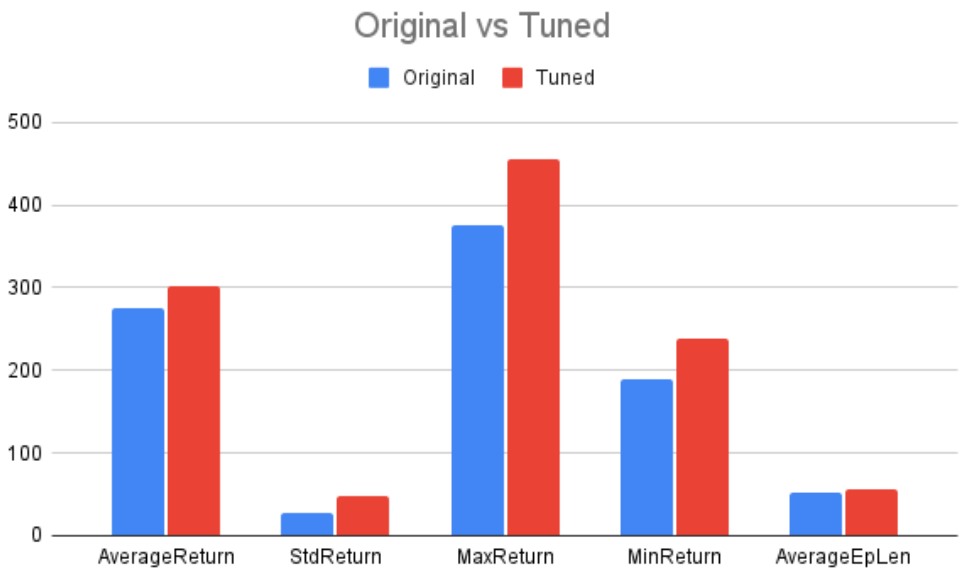

Figure 1: BC agent's performance varies with the value of n_layers, train_batch_size, and size parameters in Humanoid-v2 environment.

2 DAgger (5.25 pt)

2.1 Part 2 (5.25 pt)

For the Ant-v2 environment, I only trained the DAgger for 10 iters, and it's already at the same performance as the expert. However, for the Humanoid-v2 environment, I increased n_layers to 5, hidden neurons to size 128 to capture more features, increased iters to 100, and increased train_batch_size to 1000 to boost the efficiency in training.

Table 4: Fill your results in this table.

Env	Ant-v2		Humanoid-v2	
Metric	Mean	Std.	Mean	Std.
Expert	4713.65	12.20	10344.52	20.98
DAgger	4827.50	48.59	10506.08	48.08

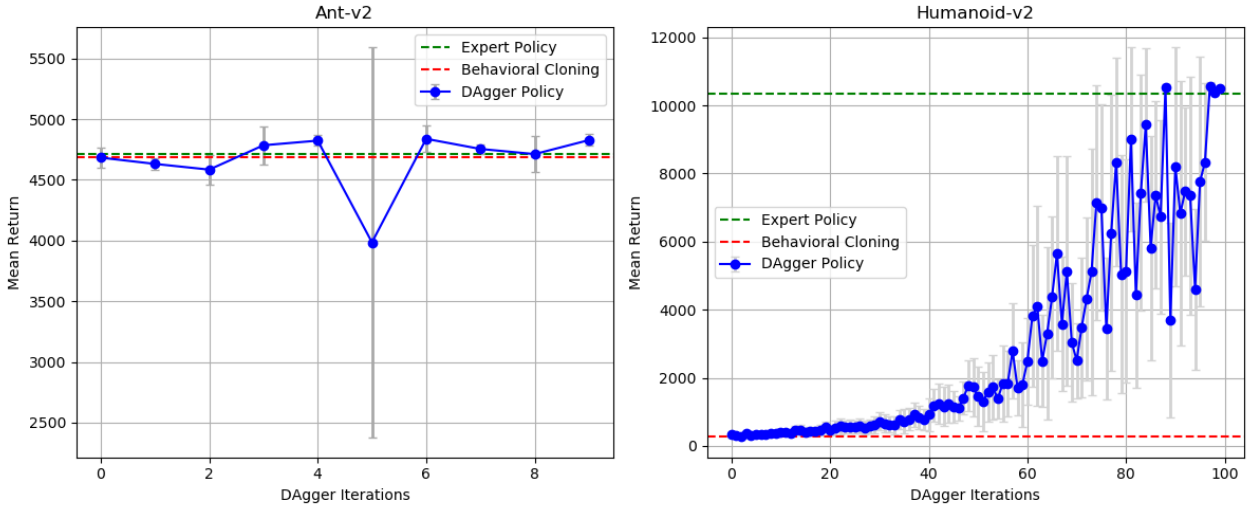

Figure 2: Learning curve, plotting the number of DAgger iterations vs. the policy's mean return, with error bars to show the standard deviation.