Homework 4 Report

Behavioral Cloning

For both tasks, default args were used.

```
class Args:

 #@markdown expert data
 expert_policy_file = 'cs285/policies/experts/Ant.pkl' #@param
 expert_data = 'cs285/expert_data/expert_data_Ant-v4.pkl' #@param
 env_name = 'Ant-v4' #@param
 exp_name = 'bc_ant' #@param
 do_dagger = True #@param {type: "boolean"}
 ep_len = 1000 #@param {type: "integer"}
 save_params = False #@param {type: "boolean"}

 num_agent_train_steps_per_iter = 1000 #@param {type: "integer"})
 n_iter = 5 #@param {type: "integer"})

 #@markdown batches & buffers
 batch_size_initial = 2000 #@param {type: "integer"})
 batch_size = 1000 #@param {type: "integer"})
 eval_batch_size = 10_000 #@param {type: "integer"}
 train_batch_size = 100 #@param {type: "integer"}
 max_replay_buffer_size = 1_000_000 #@param {type: "integer"}

 #@markdown network
 n_layers = 2 #@param {type: "integer"}
 size = 64 #@param {type: "integer"}
 learning_rate = 5e-3 #@param {type: "number"}

 #@markdown logging
 video_log_freq = -1 #@param {type: "integer"}
 scalar_log_freq = 1 #@param {type: "integer"}

 # some properties omitted for brevity
```

Rollouts

I'm using `ep_len=1000` and `eval_batch_size=10_000` to produce 10 rollouts.

Part 1a: Half Cheetah

Collecting data for eval...
Eval_AverageReturn : 3149.32080078125
Eval_StdReturn : 154.32643127441406
Eval_MaxReturn : 3382.9130859375
Eval_MinReturn : 2815.89794921875
Eval_AverageEpLen : 1000.0
Train_AverageReturn : 4034.7999834965067
Train_StdReturn : 32.8677631311341
Train_MaxReturn : 4067.6677466276406
Train_MinReturn : 4001.9322203653724
Train_AverageEpLen : 1000.0
Training Loss : 0.043427277356386185
Train_EnvstepsSoFar : 0
TimeSinceStart : 10.225802421569824
Initial_DataCollection_AverageReturn : 4034.7999834965067

Part 1b: Hopper

Collecting data for eval...
Eval_AverageReturn : 979.667724609375
Eval_StdReturn : 349.65399169921875
Eval_MaxReturn : 2125.21337890625
Eval_MinReturn : 357.337158203125
Eval_AverageEpLen : 295.0882352941176
Train_AverageReturn : 3717.5129936182307
Train_StdReturn : 0.3530361779417035
Train_MaxReturn : 3717.8660297961724
Train_MinReturn : 3717.159957440289
Train_AverageEpLen : 1000.0
Training Loss : 0.04094276949763298
Train_EnvstepsSoFar : 0
TimeSinceStart : 11.587031364440918
Initial_DataCollection_AverageReturn : 3717.5129936182307

Part 2: Varying `train_steps`

For this part, I chose to increase `num_agent_train_steps_per_iter` in hopes that the agent would learn more from the expert before being evaluated. I changed it from `1000` to `5000`.
From the results below, it looks like it worked in increasing the `Eval_AverageReturn` score as compared to with `1000` training steps.

Part 2a: HalfCheetah

Collecting data for eval...
Eval_AverageReturn : 3972.861328125
Eval_StdReturn : 60.224185943603516
Eval_MaxReturn : 4120.2353515625
Eval_MinReturn : 3896.466796875
Eval_AverageEpLen : 1000.0
Train_AverageReturn : 4034.7999834965067
Train_StdReturn : 32.8677631311341
Train_MaxReturn : 4067.6677466276406
Train_MinReturn : 4001.9322203653724
Train_AverageEpLen : 1000.0
Training Loss : 0.0032688395585864782
Train_EnvstepsSoFar : 0
TimeSinceStart : 19.65921449661255
Initial_DataCollection_AverageReturn : 4034.7999834965067

Part 2b: Hopper with 5000

Collecting data for eval...
Eval_AverageReturn : 1678.6815185546875
Eval_StdReturn : 621.1459350585938
Eval_MaxReturn : 3709.64892578125
Eval_MinReturn : 963.0943603515625
Eval_AverageEpLen : 491.4761904761905
Train_AverageReturn : 3717.5129936182307
Train_StdReturn : 0.3530361779417035
Train_MaxReturn : 3717.8660297961724
Train_MinReturn : 3717.159957440289
Train_AverageEpLen : 1000.0
Training Loss : 0.004322831518948078
Train_EnvstepsSoFar : 0
TimeSinceStart : 21.426495790481567
Initial_DataCollection_AverageReturn : 3717.5129936182307

DAgger

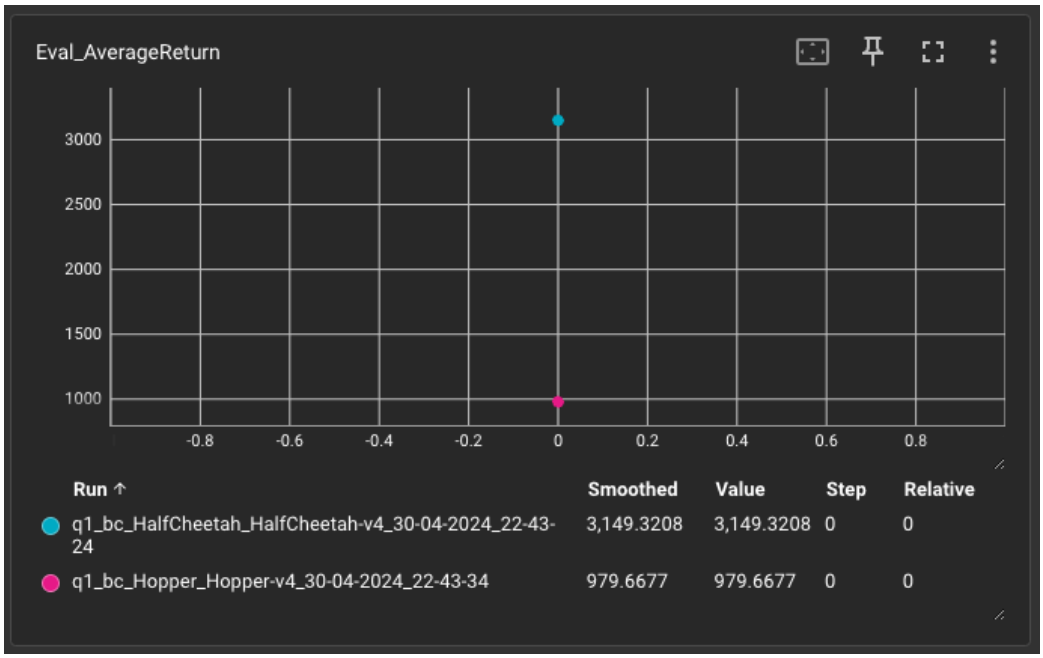
Part 1: HalfCheetah

Final Iteration
Eval_AverageReturn : 4055.351318359375
Eval_StdReturn : 0.0
Eval_MaxReturn : 4055.351318359375
Eval_MinReturn : 4055.351318359375
Eval_AverageEpLen : 1000.0
Train_AverageReturn : 4093.72802734375
Train_StdReturn : 0.0
Train_MaxReturn : 4093.72802734375
Train_MinReturn : 4093.72802734375
Train_AverageEpLen : 1000.0
Training Loss : 0.0015648063272237778
Train_EnvstepsSoFar : 9000
TimeSinceStart : 185.56412601470947

Part 2: Hopper

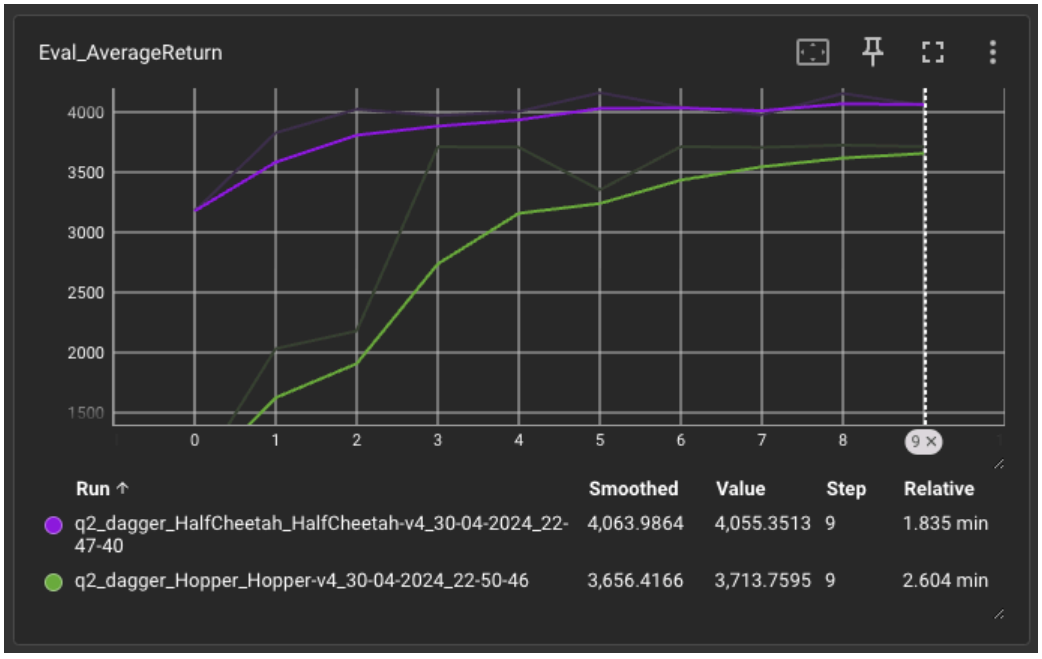
Final Iteration
Eval_AverageReturn : 3713.759521484375
Eval_StdReturn : 0.0
Eval_MaxReturn : 3713.759521484375
Eval_MinReturn : 3713.759521484375
Eval_AverageEpLen : 1000.0
Train_AverageReturn : 3723.124755859375
Train_StdReturn : 0.0
Train_MaxReturn : 3723.124755859375
Train_MinReturn : 3723.124755859375
Train_AverageEpLen : 1000.0
Training Loss : 0.0024139871820807457
Train_EnvstepsSoFar : 9278
TimeSinceStart : 197.94536757469177

Behavioral Cloning



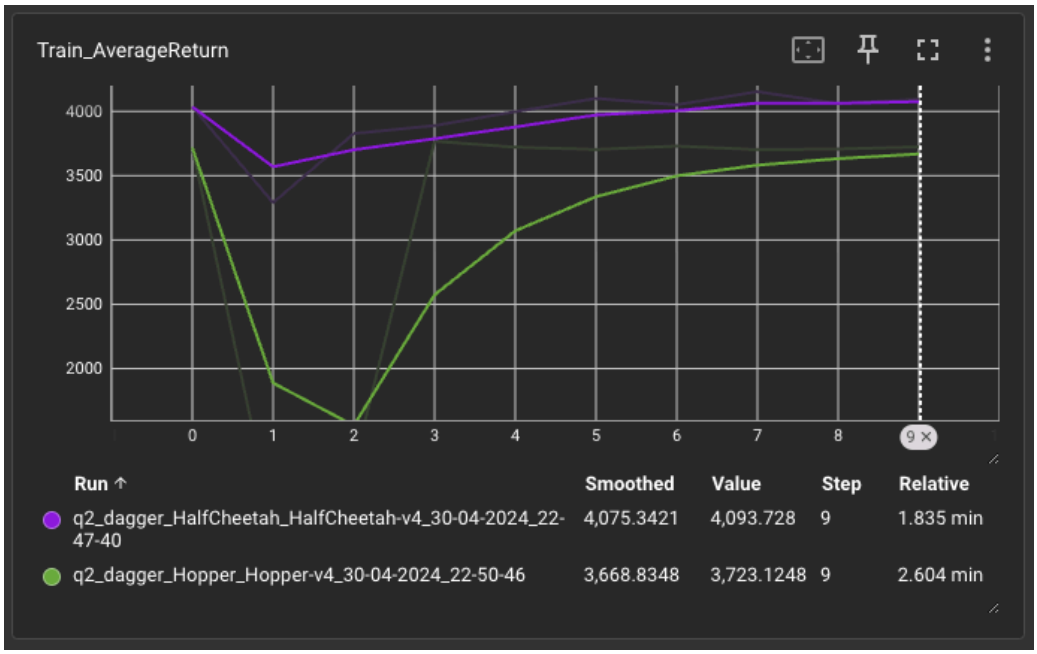
Fixed at 3149 for HalfCheetah and 979 for Hopper (should be a straight horizontal line).

DAgger



HalfCheetah vs Hopper comparison.
Unfortunately I don't have the plots for the standard deviation error bars.

Expert Performance



Unfortunately I don't have this on the same plot as the BC agent.

Final Thanks 🙏

Thank you for an amazing semester for CS545! Lowkey one of the classes I learned the most in 🚀