CS 285 HW 1 Report
Yang Lyu

Q1. Behavior Cloning

Q1.2: compare different environments

Ant-v4: achieved high Eval_AverageReturn!
Initial_DataCollection_AverageReturn : 4725.849609375

Trial	Eval_AverageReturn	Eval_StdReturn
1	4538.8916015625	79.895881652832
2	4690.03515625	171.648788452148
3	4566.3876953125	54.9478950500488

Walker2d: poor performance (less than 30%)!
Initial_DataCollection_AverageReturn : 5557.6083984375

Trial	Eval_AverageReturn	Eval_StdReturn
1	832.942932128906	464.512115478516
2	493.213806152344	253.150985717773
3	107.093048095703	197.114562988281

Table 1. for each environment, 3 trial rollouts are performed using different random seeds (from 1 to 3). All hyperparameters during the training are the same, such as the MLP architecture, eval_batch_size (5000) and ep_len (1000).

Q1.3: try different hyperparameters

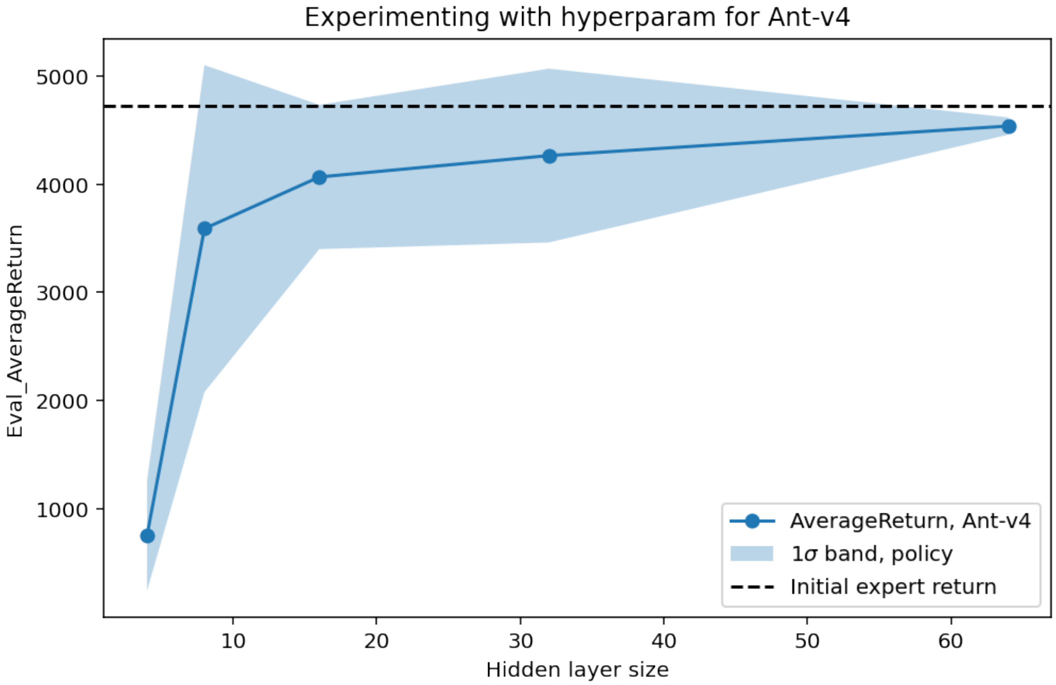

Figure 1. For Ant-v4, we set all hyperparameters to the same values as in Q1.1 (in this case, random seed is set to 1) except **hidden layer size**, where we increased the MLP hidden layers size gradually from 4 to 64. We observe that the policy performance is hugely impacted by the hidden layer size when the size is small (i.e. less than 20). This is expected because the observation space of Ant-v4 environment is 27 and there are 8 degrees of freedoms in the action, which requires a high expressivity of the neural network.

Q2. DAgger

Q2.2: run DAgger on two environments

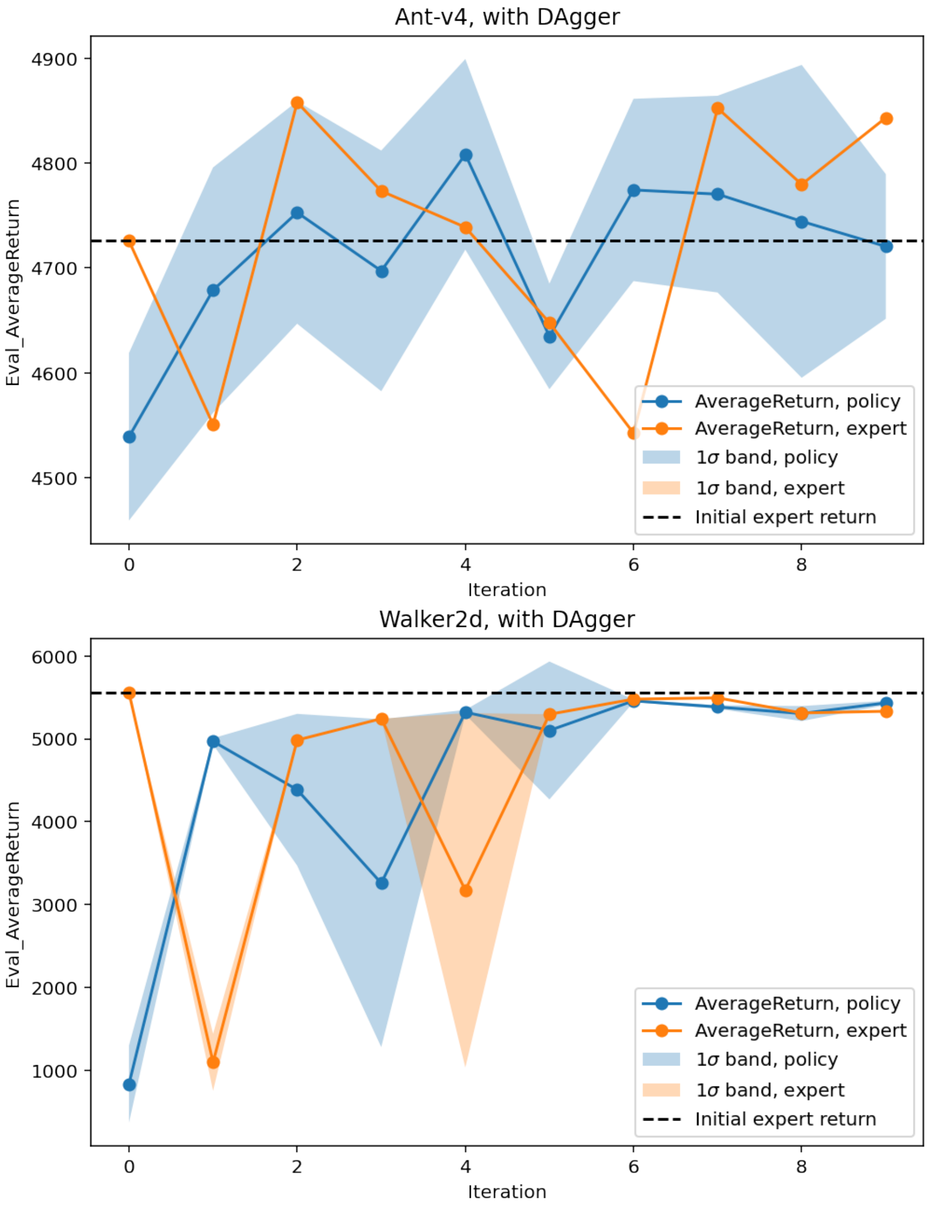

Figure 2. DAgger was ran for Ant-v4 and Walker2d. For both experiments, the same sets of hyperparameters are used, e.g. ep_len= 1000, eval_batch_size = 5000. Other parameters (including hidden layer size) are set to default values in run_hw1.py.