

Homework 1

Lizhi Guo

October 19, 2018

Section 2: Behavioral Cloning

2. Run behavioral cloning (BC) and report results on two tasks – one task where a behavioral cloning agent achieves comparable performance to the expert, and one task where it does not. When providing results, report the mean and standard deviation of the return over multiple rollouts in a table, and state which task was used. Be sure to set up a fair comparison, in terms of network size, amount of data, and number of training iterations, and provide these details (and any others you feel are appropriate) in the table caption.

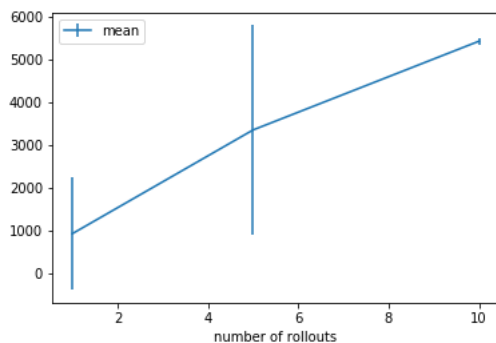
I ran behavior cloning on Walker2d and HalfCheetah and have the results shown in the table. The behavioral cloning agent achieves comparable result in HalfCheetah and slightly worse result in Walker2d.

	mean	variance
Expert: HalfCheetah	4152.05	79.54
BC: HalfCheetah	4138.37	70.71
Expert: Walker2d	5490.35	78.12
BC: Walker2d	5294.85	860.96

For both tasks, the training data have 20000 pairs of outputs and inputs with each pair consisting of 17 input data and 6 output data. I used the same network to learn both tasks. The network has 4 fully-connected layers and each hidden layer has 400 rectified linear units. I trained the network 20000 times for each task and the final mean squared errors of both tasks are less than 0.001.

3. Experiment with one hyperparameter that affects the performance of the behavioral cloning agent, such as the number of demonstrations, the number of training epochs, the variance of the expert policy, or something that you come up with yourself. For one of the tasks used in the previous question,

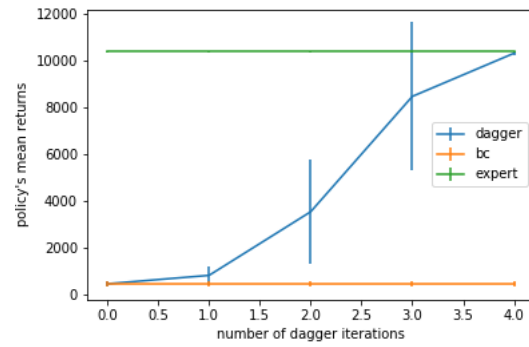
show a graph of how the BC agent’s performance varies with the value of this hyperparameter, and state the hyperparameter and a brief rationale for why you chose it in the caption for the graph.



The hyperparameter is the number of expert demonstrations. I varied the number of demonstrations by changing the number of rollouts. I chose this parameter in order to show that more demonstrations lead to better performance.

Section3 DAgger

2. Run DAgger and report results on one task in which DAgger can learn a better policy than behavioral cloning. Report your results in the form of a learning curve, plotting the number of DAgger iterations vs. the policy’s mean return, with error bars to show the standard deviation. Include the performance of the expert policy and the behavioral cloning agent on the same plot. In the caption, state which task you used, and any details regarding network architecture, amount of data, etc. (as in the previous section).



I trained dagger algorithm to perform Humanoid-v2 task. My neural net consists of 3 layers, with each inner layer consisting of 200 rectifier units. I used 10 rollouts as the initial training data, which are 10000 input-output pairs.