CS294 Deep RL Assignment 1: Imitation Learning

Mohamed Khodeir

December 14, 2018

2. Behavior Cloning

2.2 results

I chose to report on the "HalfCheetah" environment as an example where a behavior cloning agent achieves comparable performance to the expert. For the negative example, I chose the "Humanoid" environment.

For my BC agents, I used the same policy architecture as the given experts, which was a 3 layer feed forward network with relu activations in the hidden layer and a linear output layer. In all cases, they were trained for 50 epochs with stochastic gradient descent on 100 rollouts from the relevant expert policy.

	Game	Mean	Std
Expert	HalfCheetah	2374.79	772.96
BC Agent	HalfCheetah	2110.61	947.18
Expert	Humanoid	2908.21	929.65
BC Agent	Humanoid	45.01	13.77

Table 1: Stats reported over 100 rollouts. The HalfCheetah BC agent used a network of 128 and 64 units respectively followed by an output layer over 6 dimensions of the action space. The Humanoid agent 256 and 128 units respectively followed by an output layer over 17 dimensions of the action space.

2.3 Experimentation

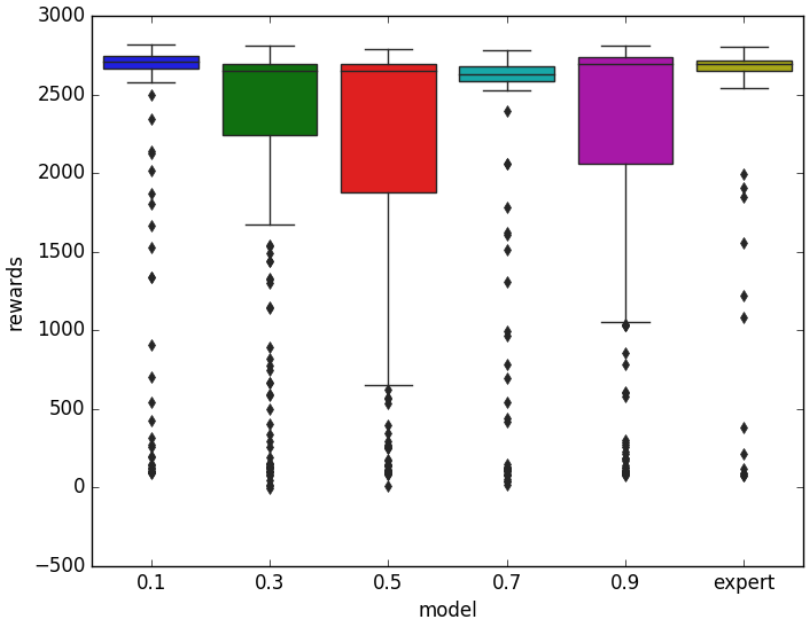

Figure 1: I chose to look at the effect of more training data on the BC Agent's performance. I trained the HalfCheetah agent using subsets of the data between 10% and 100%.

3.2 DAgger

- Dagger Comparison.png

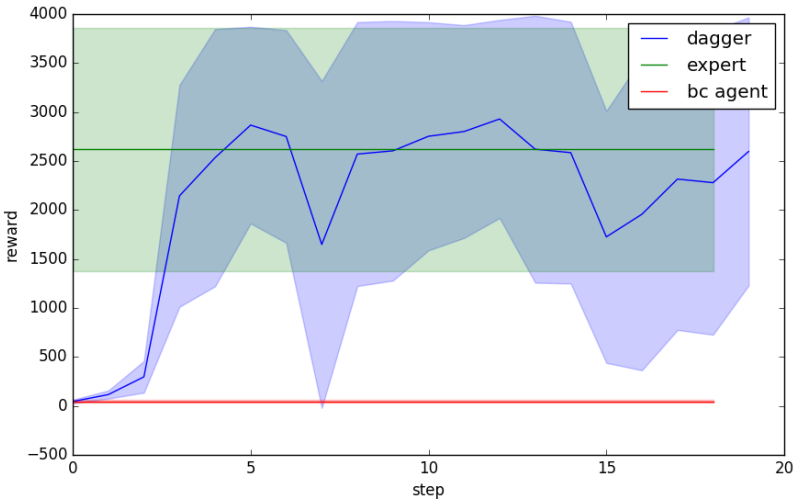

Figure 2: I chose the task on which the behavior cloning agent performed poorest relative to the expert policy - Humanoid. I used the same architecture as in the behavior cloning experiments (i.e. 256 relu ¿ 128 relu -¿ 17 lin). I ran 20 episodes of the algorithm, training the policy for 80000 steps from scratch each episode, and generating 10 rollouts worth of samples for the next iteration.