CS294 Homework 1

Philippe Poulin

September 19, 2017

Warmup

We first show the evolution of the loss through training a single layer model with 128 units and a relu activation layer, trained on 20 expert rollouts of 1000 steps.

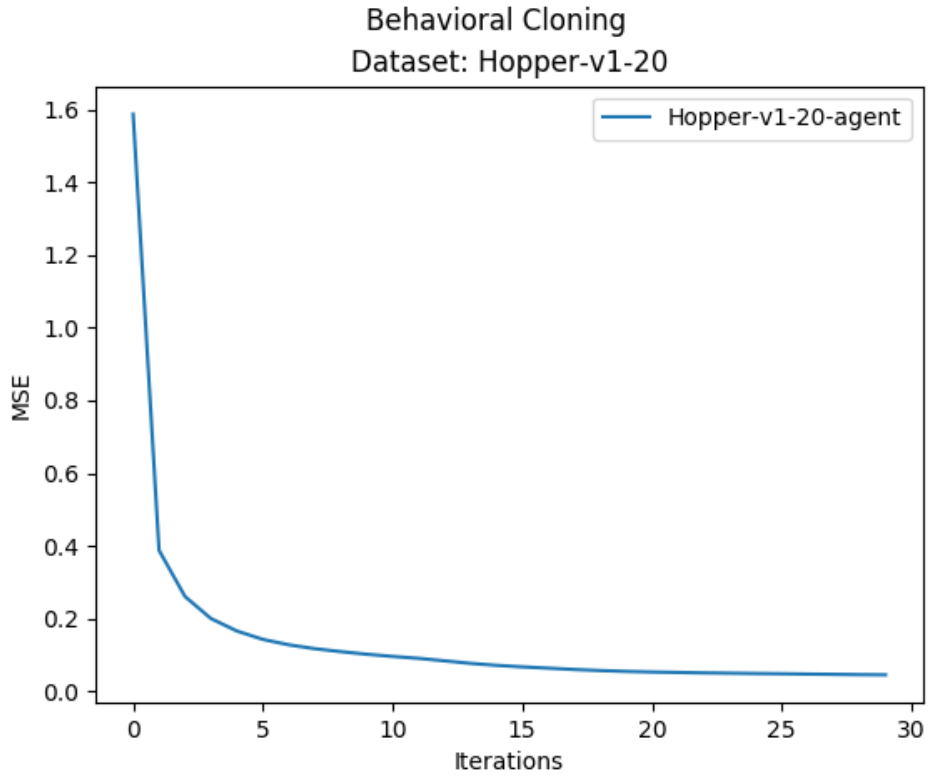

Figure 1: Training loss for behavioral Cloning on Hopper-v1 environment using 20 rollouts

Behavioral Cloning

Using the same architecture as before, the performance of the trained on Ant-1 task is comparable to the expert's.

Table 1: Performance comparison on Ant-v1 task using 20 rollouts

Policy	**Returns mean**	**Returns std**
128-relu (20 rollouts)	4815	88
Expert	4828	101

However, it is significantly worse than the expert on the Humanoid-v1 task.

Table 2: Performance comparison on Humanoid-v1 task using 20 rollouts

Policy	**Returns mean**	**Returns std**
128-relu (20 rollouts)	1467	987
Expert	10417	48

Adding more rollouts to the training set seems to improve performance.

Table 3: Performance comparison on Humanoid-v1 task using 20 rollouts

Policy	**Returns mean**	**Returns std**
128-relu (20 rollouts)	1467	987
128-relu (100 rollouts)	2438	773
Expert	10417	48

DAgger

Using the DAgger training process, the trained model is competitive with the expert (when using 10 iterations).

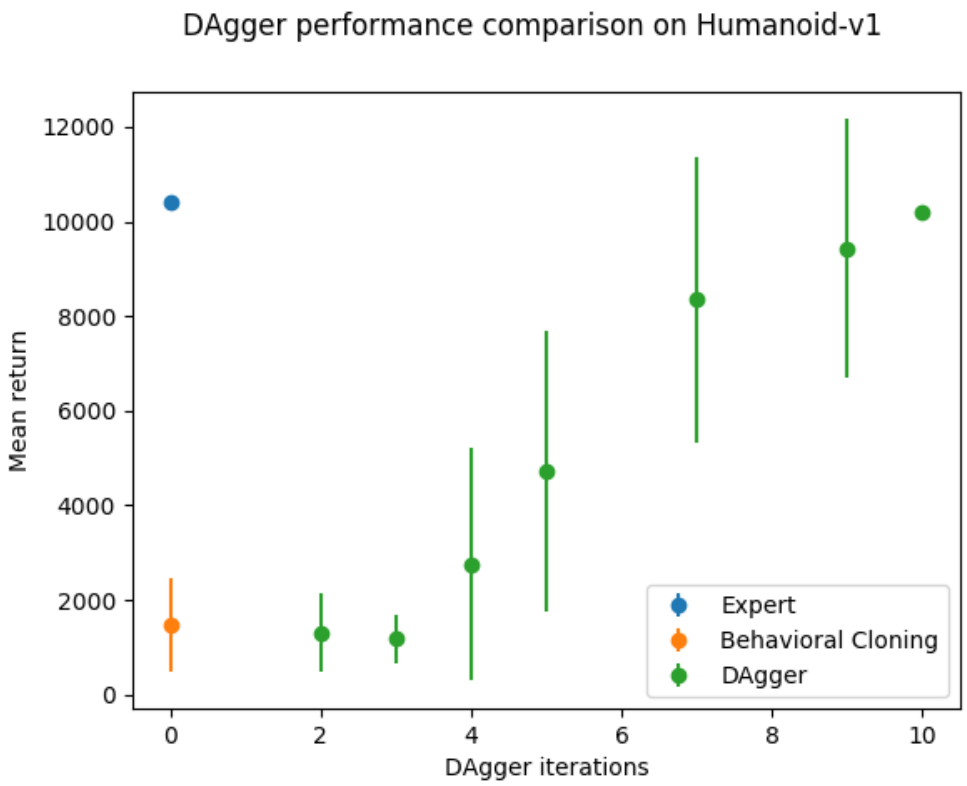

Figure 2: DAgger performance comparison (using 5 rollouts of 1000 steps at each iteration)