Berkeley DRL HW1

Anton Makiievskyi

1. Behavioral cloning agent performance comparison

HalfCheetah-v2

	Mean return	Std
Expert policy	4134	63.51
BC policy	4105	82.67

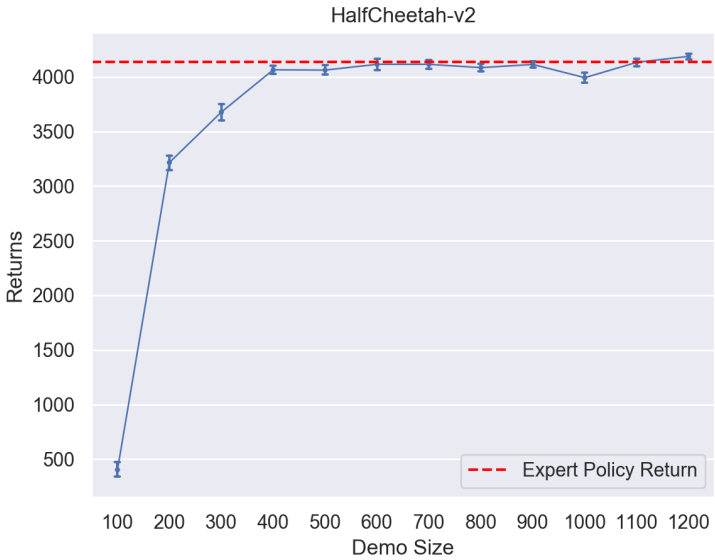
Hopper-v2

	Mean return	Std
Expert policy	3777	3.03
BC policy	857	24.64

Both tasks were trained using the same DNN architecture of 2 layer with 64 neurons in each.
Models were trained until best accuracy didn't improve for 3 consecutive epochs
Each epoch consisted of 20k observation - action pairs from the expert policy
Expert and BC policies were tested for 20 runs each
In Hopper-v2 task BC agent couldn't achieve comparable to expert performance, probably due to visiting states dissimilar from those in the training data

2. Behavioral cloning agent performance relative to number of demonstrations

The graph shows returns relative to demo size(number of experts actions). The BC agent was trained with the same architecture as in the previous example. Training set for each BC model was scaled by repeating to the contain the same number of samples for fair comparison.



3. Dagger performance

Below is a plot of dagger agent performance improvement with each iteration on a Hopper-v2 task (orange line) compared to behavioral cloning agent trained on different number of demonstrations (blue line)

Dagger Agent started with learning from one expert policy rollout of 1000 time steps. Then each iteration the learned policy was rolled out until the end of the episode, the expert policy actions on those observation added to the training set and the policy retrained.

Same NN architecture and training parameters as in the previous example was used for both agents

Dagger have quickly matched the experts performance(on 4th iteration). Interestingly the BC agent performance is consistently higher when trained in the smaller training set compared to the 20k samples training set in the first question