Homework #1

Mahan Fathi

February 14, 2022

1 The experiments

Kindly find the results/logs in directories with names related to the experiments, e.g. `BC-Ant-v2` or `DAgger-Walker2d-v2`. In each directory there is a `config.yaml` corresponding to the specific experiment. Moreover, `log.txt` contains all the outputs written to `stdout`.

2 Question 1.2

Table 1: Behavior Cloning **results** for `Ant-v2`.

`Ant-v2`	Behavior Cloning	Expert
`AverageReturn`	4704.03	4713.66
`StdReturn`	105.96	12.20

Table 2: Behavior Cloning **results** for `Walker2d-v2`.

`Walker2d-v2`	Behavior Cloning	Expert
`AverageReturn`	3168.86	5566.85
`StdReturn`	1872.78	9.24

Table 3: Behavior Cloning **results** for `Humanoid-v2`.

`Humanoid-v2`	Behavior Cloning	Expert
`AverageReturn`	298.61	10344.52
`StdReturn`	70.61	20.99

Table 4: Behavior Cloning **hyper-parameters**.

*	max_episode_length	train_steps_per_iter	n_layers
Ant-v2	1000	5000	3
Walker2d-v2	1000	5000	3
Humanoid-v2	1000	5000	3

Behavior Cloning fails to capture the true behavior of the expert in some environments, take `Humanoid-v2` as a case in point. It barely manages to get to 50% of average return in `Walker2d-v2`, but it has huge variance.

3 Question 1.3

I chose the amount of data provided to BC for ablations, namely `max_episode_length`, and I expect better performance with more data.

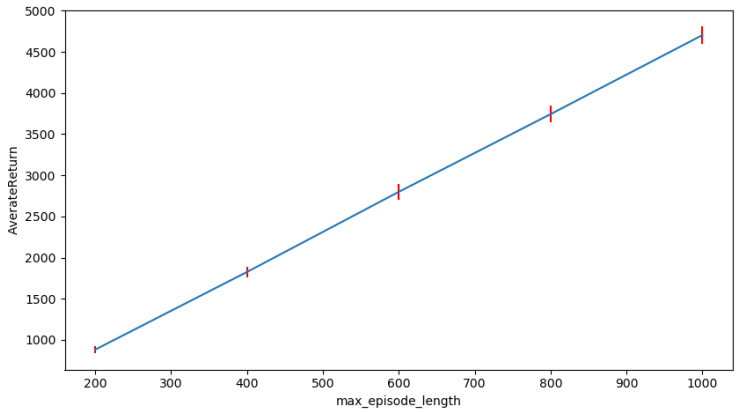

Figure 1: Ablation on the amount of data provided to BC, for `Ant-v2`.

We observe that the performance of the BC agent imporves linearly wrt the amount of data provided.

4 Question 2.2

Table 5: DAgger **best results** for `Ant-v2`.

`Ant-v2`	DAgger	Expert
`AverageReturn`	4764.02	4713.66
`StdReturn`	116.59	12.20

Table 6: DAgger **best results** for `Walker2d-v2`.

`Walker2d-v2`	DAgger	Expert
`AverageReturn`	5588.88	5566.85
`StdReturn`	29.49	9.24

Table 7: DAgger **best results** for `Humanoid-v2`.

`Humanoid-v2`	DAgger	Expert
`AverageReturn`	324.12	10344.52
`StdReturn`	69.96	20.99

- Observation: DAgger significanly helps in the case of `Walker2d-v2`, and actually cracks the task. Moreover, the performance variance is reduced significantly via DAgger.
- Obervation: DAgger fails to improve the performance in the case of `Humanoid-v2`.
- Note: The variance, i.e. error bars, seem to be large for some data points, and that is because we are using 5 trajectories during evaluation which is a small number. It was prescribed in the assignment, but I am now realizing we could benefit from a larger sample size.

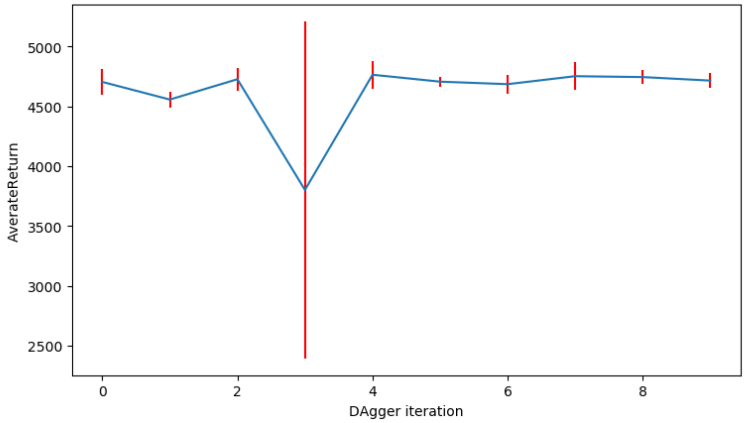

Figure 2: DAgger performance for `Ant-v2`.

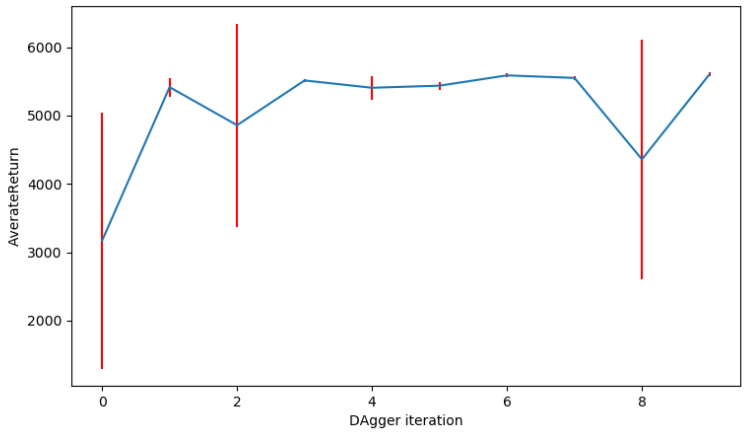

Figure 3: DAgger performance for `Walker2d-v2`.