CS 294-112 HW1

Yujia Luo

25296369

I used the same deep model for behavior cloning and DAgger: a simple neural net with two 64-unit fully connected hidden layers trained with batch size = 2048 and an Adam optimizer.

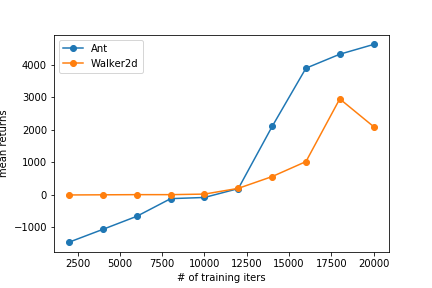
**Behavior Cloning**

2.2

|  |  |  |  |
| --- | --- | --- | --- |
| ­­­ | Mean Return | Std Return | #Rollouts |
| Ant (Expert) | 4732.42 | 503.27 | 100 |
| Ant (BC) | 4624.35 | 424.38 | 50 |
|  | | | |
| Walker2d (Expert) | 5526.56 | 75.39 | 100 |
| Walker2d (BC) | 2095.27 | 2322.00 | 50 |

2.3

The following plot shows how the two agents’ mean return changes as *the number of training iterations* increases (bc agent is evaluated every 2000 training iterations):



I chose the number of training iterations to how differently 4-leg-agent and 2-leg agent learn and how fast they converge given the same expert data

**DAgger**

3.2

The following plot shows how Walker2d-v2 agent’s performance changes as the number of training iterations increases using three different policies: expert, behavior cloning, and DAgger (#rollout = 40, and agents are evaluated every 3000 training iterations):

