CS285 - HW1

Ho Nam Nguyen

September 11, 2022

1 Behavior Cloning

Ant - expert return ≈ 4700

eval_batch_size	Eval_AverageReturn	Eval_StdReturn
5000	4664	113
10000	4665	84

Walker2d - expert return ≈ 5500

eval_batch_size	Eval_AverageReturn	Eval_StdReturn
5000	1078	1043
10000	1461	1491

Table 1: Both tasks share the same hyperparameters: network size = [64,64], data points = 100 (train_batch_size) x 1000 (gradient_steps) = 1e5, 1 training iteration for BC. Ant (table above) achieves nearly 100% expert behavior whereas Walker2d (table below) achieves around 20%.

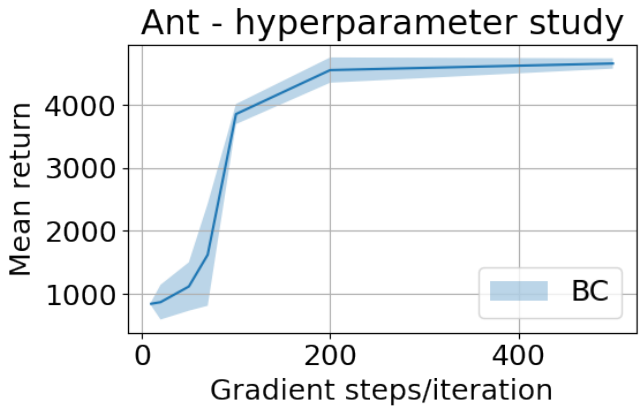

Figure 1: BC agent's performance vs. number of gradient steps per iteration. This supervised learning performance is bad when the training hasn't converged due to smaller number of gradient steps within the single iteration of BC.

2 DAgger

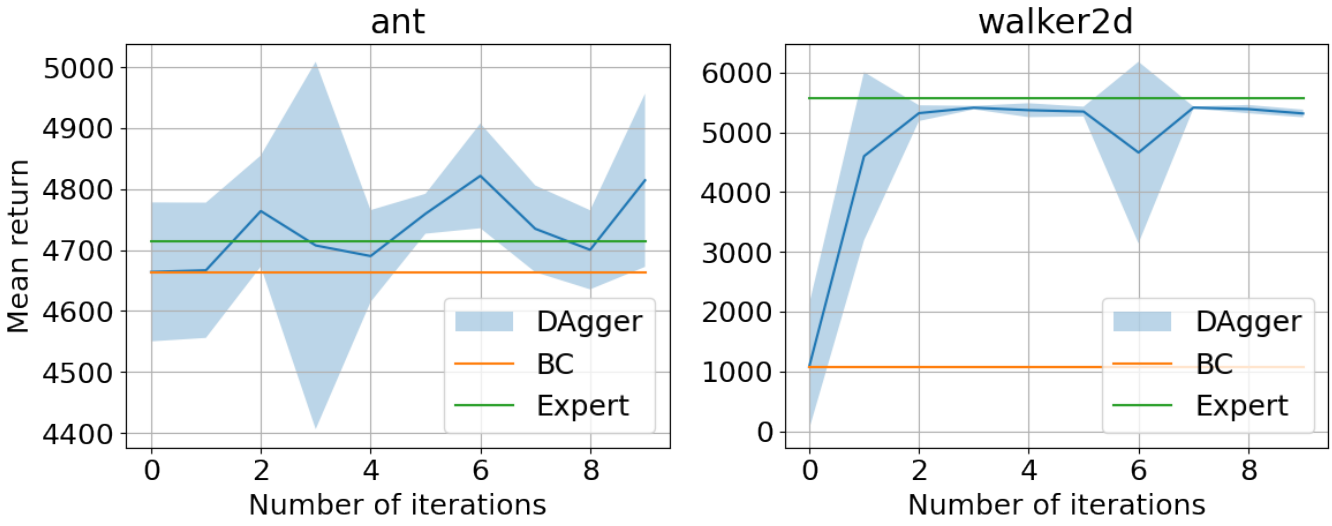

Figure 2: Policy performance vs. DAgger iterations for Ant (left) and Walker2d (right). Both tasks share the same hyperparameters: network size = [64,64], data points = 100 (train_batch_size) x 1000 (gradient_steps) = 1e5, 10 training iterations for DAgger. For the Ant environment, BC is already very good so DAgger provides a marginal improvement. However for Walker2d, DAgger manages to bring the policy performance to nearly 100% expert behavior.