CS285 Assignment 1 – Imitation Learning

Alejandro Escontrela

1) Behavior Cloning

a) Performance of behavior cloning on two Mujoco environments:

Table 1: Mean return and std computed over 10000 environment steps with a max episode length of 1000 steps. The same policy architecture was used for both environments, consisting of a neural network with 2 hidden layers, each containing 64 units.

Environment Return	
Ant	Humanoid
4275.9956 ± 784.1332	4275.9956 ± 38.3869

b) Impact of train_batch_size on the mean episode return.

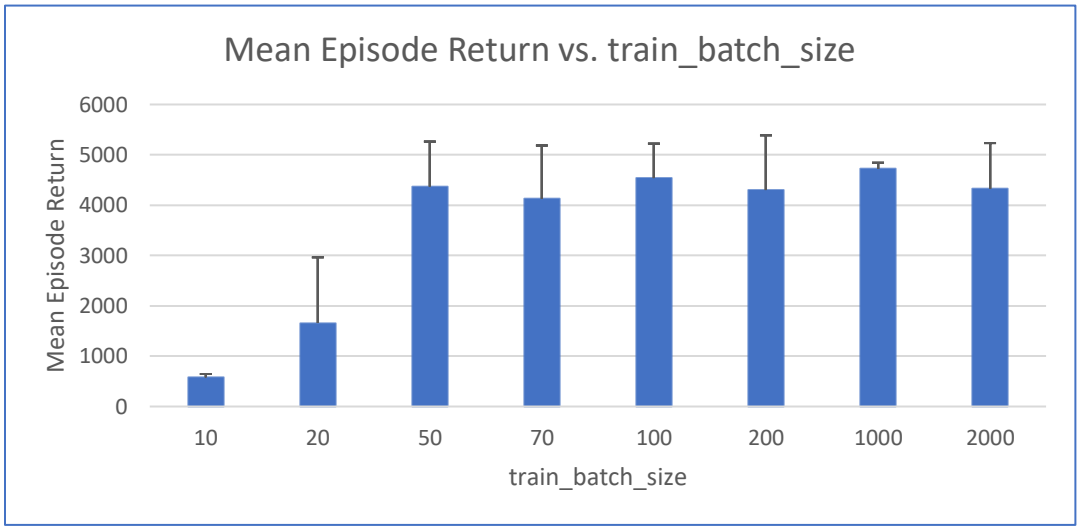

Figure 1: Effect of train_batch_size was computed by evaluating the resulting policy on 50000 ANT environment transitions with a max episode length of 1000 steps. Policy architecture consisted of 2 neural networks with 64 units for each hidden layer. It is evident that larger train batch sizes lead to more successful policies. One possible explanation is that increasing the training batch size leads to a policy gradient of lower variance.

2) Dagger

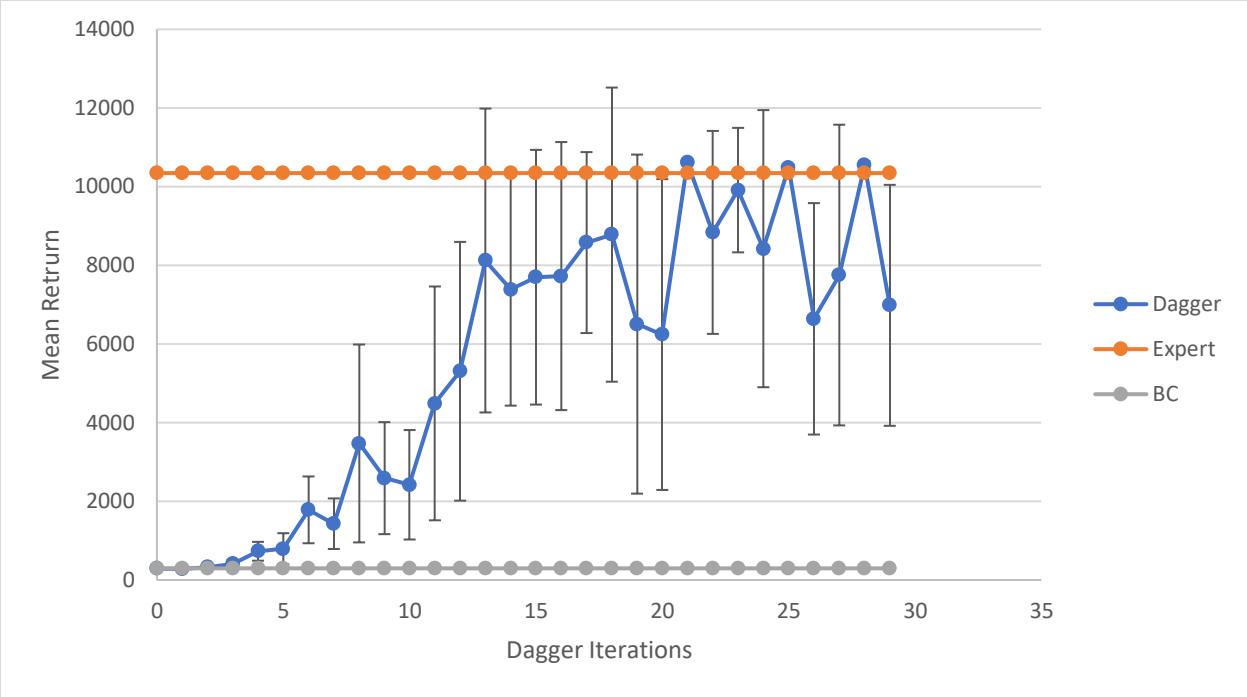

Figure 2: Performance of DAgger on Humanoid environment. Humanoid policy was left to train with 30 DAgger iterations, with a batch_size of 2500 and a train_batch_size of 500. The policy architecture consisted of 2 hidden layers with 64 neurons each.

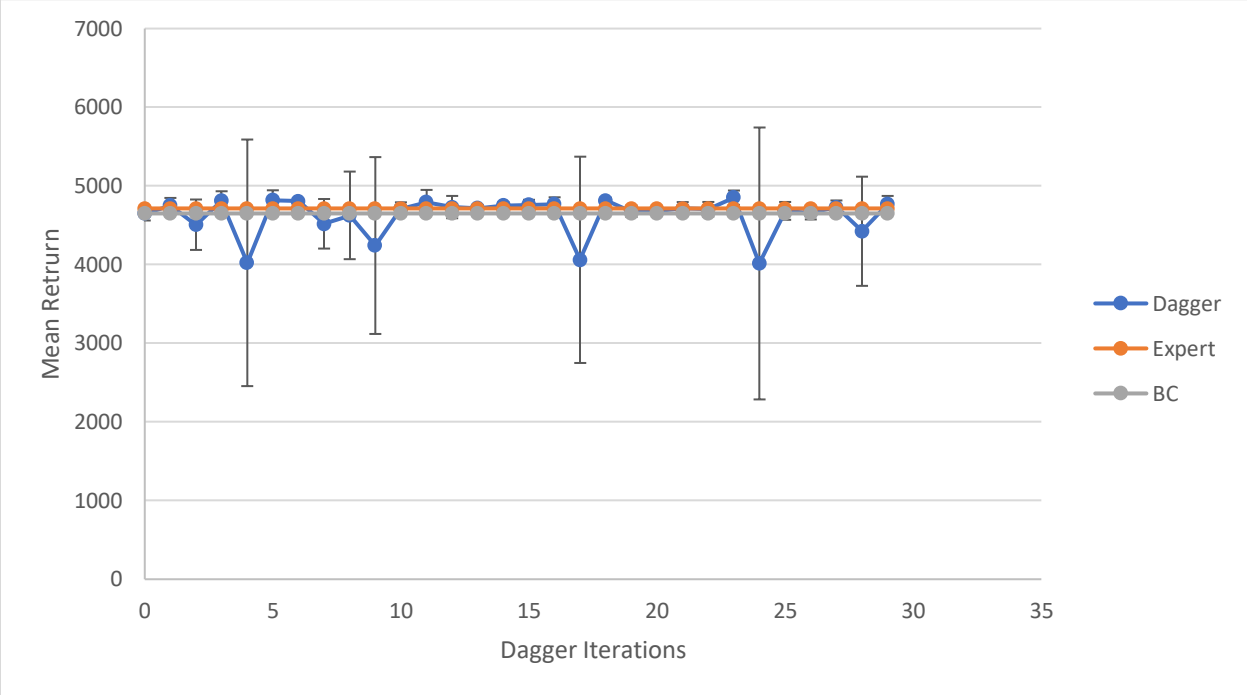

Figure 3: Performance of DAgger on Ant environment. Humanoid policy was left to train with 30 DAgger iterations, with a batch_size of 2500 and a train_batch_size of 500. The policy architecture consisted of 2 hidden layers with 64 neurons each.