

# Homework 1

Juanwu Lu (3037432593)

(M.Sc. Civil Engineering, UC Berkeley)

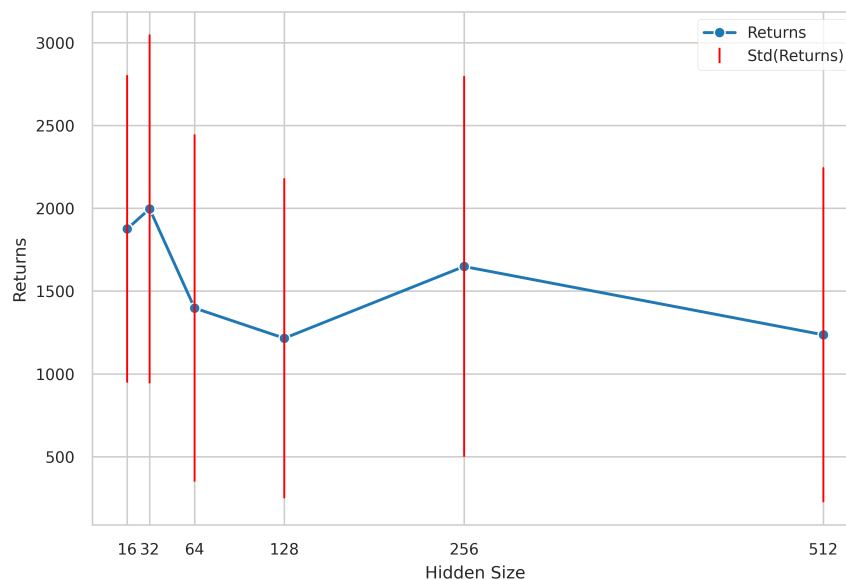
## 1 Behavior Cloning

### 1.1 Task report

**TABLE 1.** Behavior Cloning (BC) Policy’s Return over 30 rollouts. In both environments, the policy distribution mean network is an two-layer MLP with a hidden size of 64. The Ant-v4 and the HalfCheetah-v4 have action sizes of 8 and 6. The number of training iteration is one for both environments with a training batch size of 100. The `ep_len` is 1000 and the `eval_batch_size` is 30000 for both tasks. The `seed` is 42.

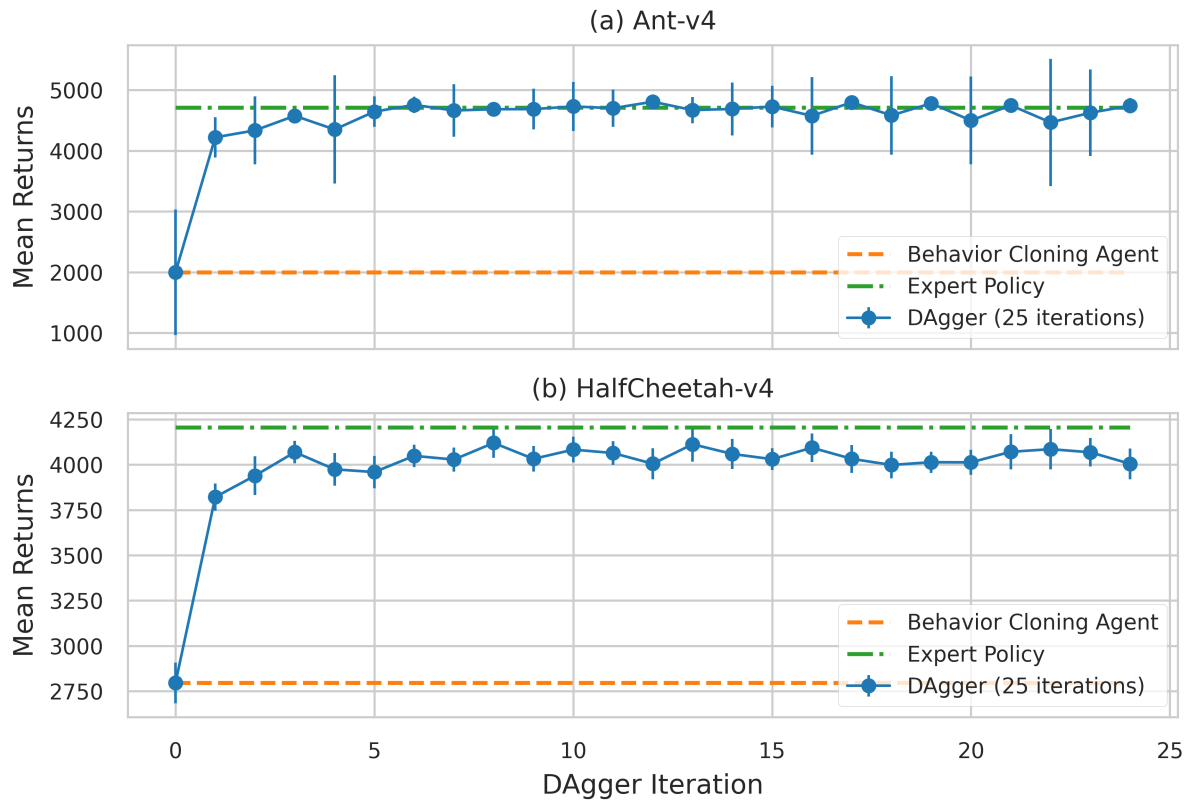
Environment	Expert Return	Mean Return	Standard Return
Ant-v4	4713.6533	1397.3680	1033.6208
HalfCheetah-v4	4205.7783	2795.4690	112.1943

### 1.2 Hyperparameter set



**Figure 1.** Evaluation Returns vs. Sizes of Hidden Layers. With different settings of the hidden layer neurons, the neural network shows different capability of representing non-linear relationship between input observation features to output action probabilities. However, the performance is not plain linear increasing with respect to increasing hidden sizes since the complexity of the network trades off marginal performance increase with fixed training steps.

## 2 DAgger



**Figure 2.** (a) The Number of DAgger Iterations vs. the Policy’s Mean Returns on the Ant-v4 Environment. Model is trained on a two-layer MLP with a hidden size of 32. The `train_batch_size`, `eval_batch_size`, and `ep_len` are 100, 30000, and 1000, respectively. The `seed` is 42. (b) The Number of DAgger Iterations vs. the Policy’s Mean Returns on the HalfCheetah-v4 Environment. Model is trained on a two-layer MLP with a hidden size of 64. The `train_batch_size`, `eval_batch_size`, and `ep_len` are 100, 30000, and 1000, respectively. The `seed` is 42.