# Assignment 1: Imitation Learning

**Andrew ID:** `lzaceria`
**Collaborators:** `YOUR COLLABORATORS`

## 1 Behavioral Cloning (9.75 pt)

### 1.1 Part 2 (1.5 pt)

# TODO

Table 1: Report your result in this table.

| Metric/Env | Ant-v2 | Humanoid-v2 | Walker2d-v2 | Hopper-v2 | HalfCheetah-v2 |
|---|---|---|---|---|---|
| Mean | 4713.653 | 10344.518 | 5566.846 | 3772.670 | 4205.778 |
| Std. | 12.197 | 20.981 | 9.238 | 1.948 | 83.039 |

### 1.2 Part 3 (5.25 pt)

# TODO

Table 2: BC vs Expert performance comparison.
Hyperparameters: learning_rate=4e-3, n_layers=5, n_iter=1, eval_batch_size=5000, ep_len=1000 (approximately 5 evaluation trajectories per environment).
Ant-v2 achieves 91.7% expert performance (>30% threshold), while Humanoid-v2 achieves 3.1% expert performance (<30% threshold).

| Env | Ant-v2 | | Humanoid-v2 | |
|---|---|---|---|---|
| Metric | Mean | Std. | Mean | Std. |
| Expert | 4713.653 | 12.197 | 10344.518 | 20.981 |
| BC | 4323.741 | 509.877 | 320.130 | 48.081 |

### 1.3 Part 4 (3 pt)

I tested the learning rate hyperparameter on Ant-v2, varying it from 1e-4 to 5e-2. Learning rate controls the gradient step size during training.

The results show optimal performance at lr=4e-3 (4103 return), with our Part 3 result of 4324 falling within the standard deviation range. Lower rates (1e-4) lead to undertraining, while higher rates (5e-2) cause instability. The performance drops significantly at the extremes.

See Figure 1 on the next page for a summary of these results.

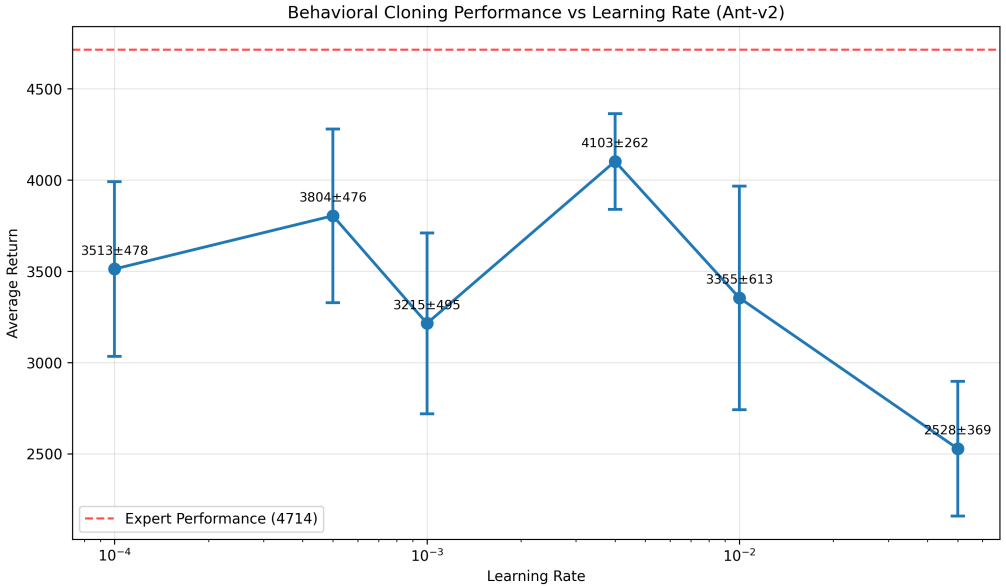


Figure 1: BC agent's performance varies with the learning rate parameter in the Ant-v2 environment. Peak performance at lr=4e-3, with worse performance at very low and high rates. Error bars show std over 5 runs.

# 2 DAgger (5.25 pt)

## 2.1 Part 2 (5.25 pt)

Report the results for Ant-v2 and HalfCheetah-v2 in Fig. 2:

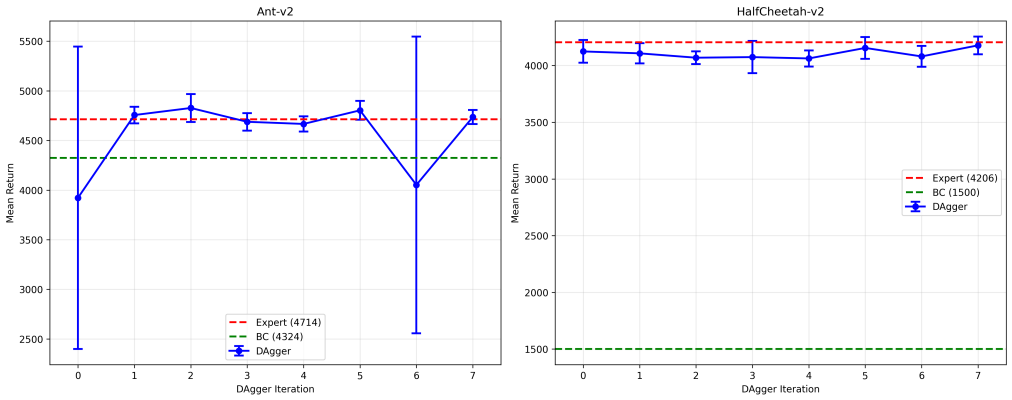


Figure 2: Learning curve, plotting the number of DAgger iterations vs. the policy's mean return, with error bars to show the standard deviation. Please show the Ant-v2 environment results on the left and the results from HalfCheetah-v2 on the right. Network architecture: 3-layer MLP. DAgger was run for 8 iterations with expert relabeling at each iteration.