Section 2.2

Task	Comparable to expert	Mean	Standard deviation
Ant-v2	Yes	4785.48	115.59
Humanoid-v2	No	274.80	62.05

Training setups:

- Network layers: 3 fully connected layers with 1024, 512, 128 nodes in each layer.
- Expert rollouts: 20
- Training epochs: 50
- Batch size: 10
- Running rollouts: 20

Section 2.3

Task: Ant-v2

Fixed setups:

- Expert rollouts: 20
- Batch size: 10

Varying hyperparameter: Training epochs

Given enough training data, if we take more training epochs, a better approximation of the expert policy can be achieved.

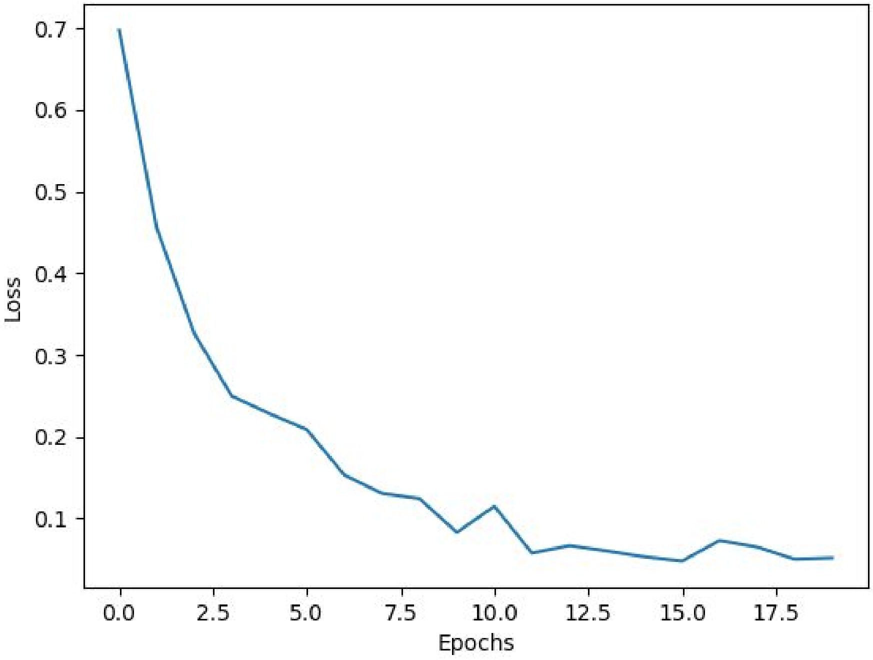

Figure 1. Training loss vs training epochs

Section 3.2

Task: Ant-v2

Fixed setups:

- DAgger iteration: 20
- Expert rollouts in each DAgger iterate: 20
- Batch size: 10

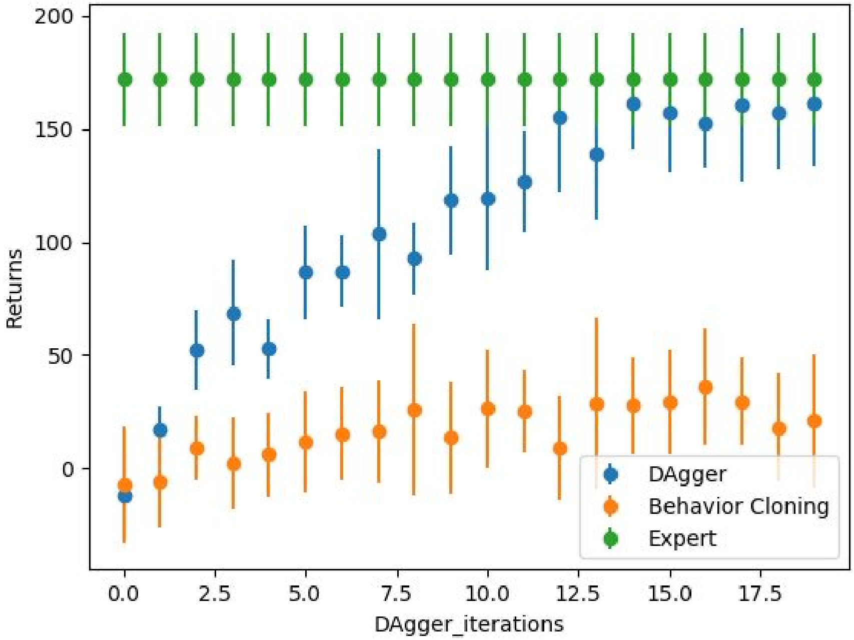

Figure 2. Performance of expert policy (green), DAgger (blue) and behavior cloning (orange)