HW1 Imitation Learning

Hongyi Yan

1.2:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name\Hyperparameter | Eval\_batch\_size | Train\_batch\_size | Batch\_size | N\_layers | Size | Eval\_return | Eval\_std | Train\_return |
| Ant | 5000 | 100 | 1000 | 2 | 64 | 4721 | 45 | 4713 |
| HalfCheetah | 5000 | 100 | 1000 | 2 | 64 | 3901 | 136 | 4205 |
| Hopper | 5000 | 100 | 1000 | 2 | 64 | 1108 | 85 | 3772 |
| Walker2d | 5000 | 100 | 1000 | 2 | 128 | 329 | 337 | 5566 |

1.3:

There is a linear relationship between the length of EP and the eval return until a certain threshold (ep\_len == 1000).

I think the plateau is because ep\_len had reached to self.env.spec.max\_episode\_steps.

I’m thinking of longer the step, higher the reward causing a better return on eval?

2.2

Chart, line chart

Description automatically generated

They all reached to a plateau once approach the expert\_policy.

I tried both Ant and Walker since Walker has the worst performance for my Behavior Cloning.

I was having batch\_size = 1000, eval\_batch\_size = 5000, train\_batch\_size = 100, n\_layers for MLP of 2, size of MLP of 128 and learning rate of 5e-3