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

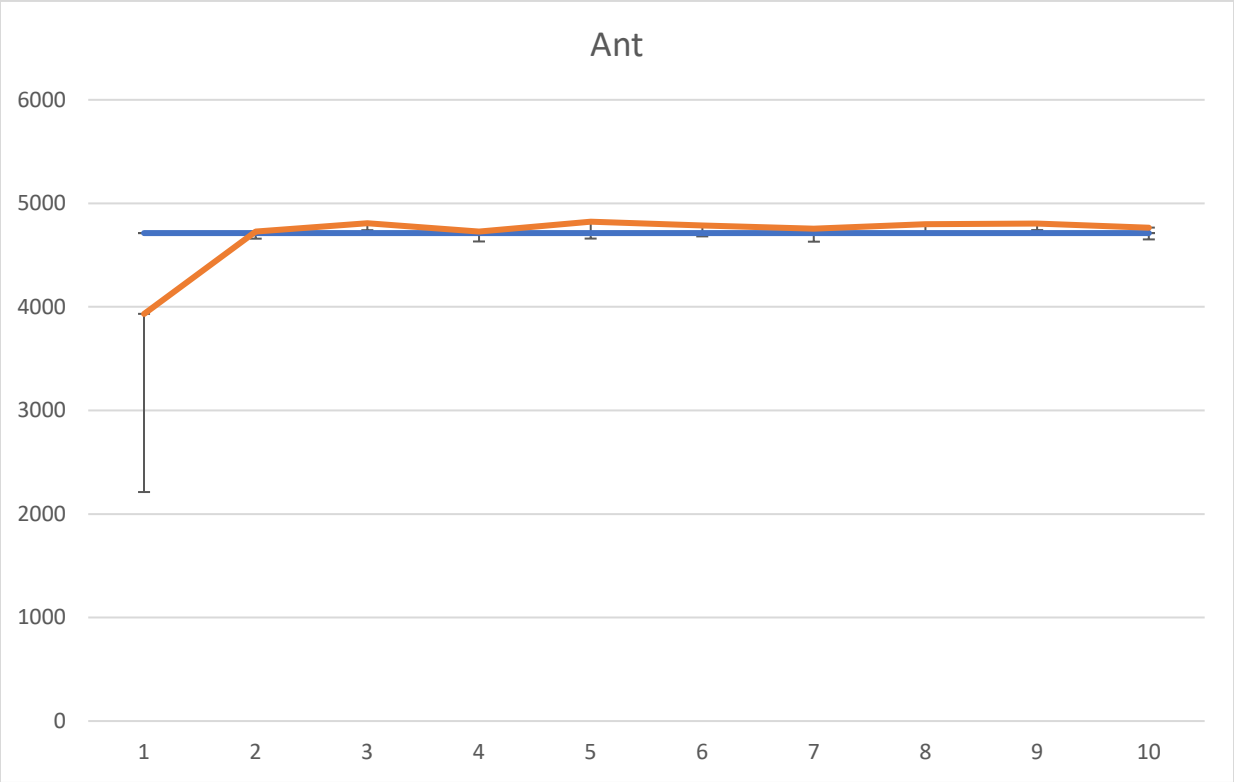
1.3:

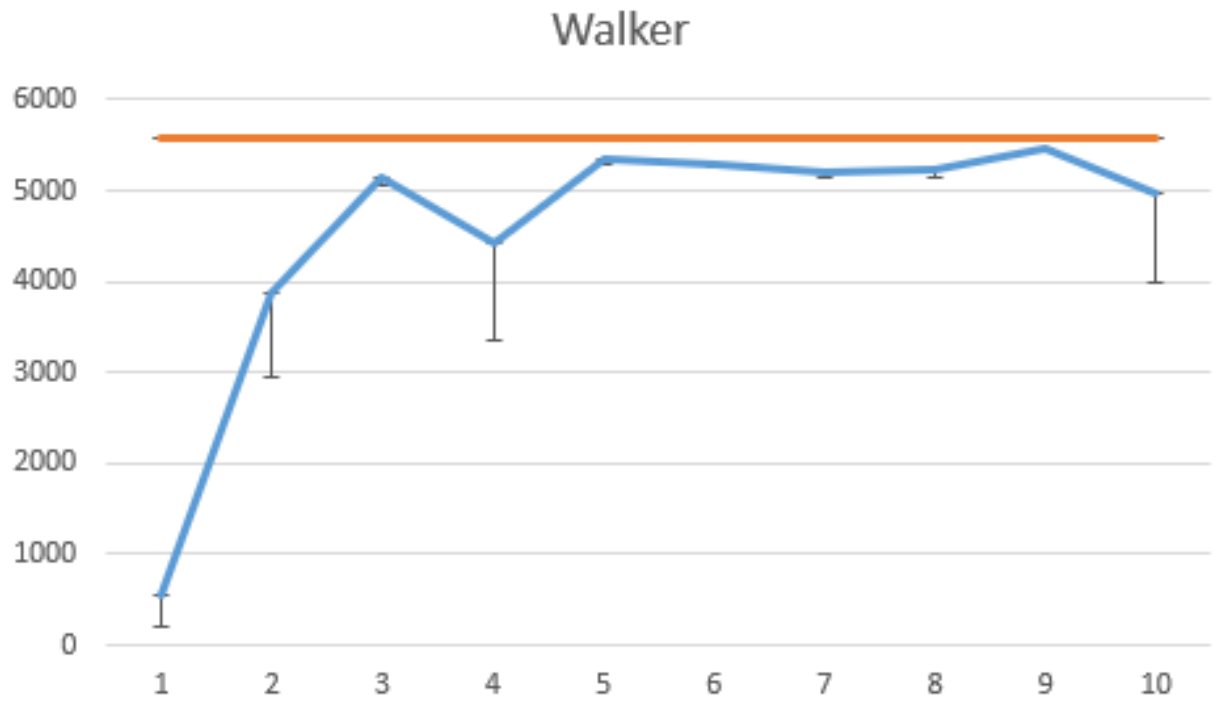
There is a linear relationship between the length of EP and the eval return until a certain threshold (ep_len == 1000).

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





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