294 HW1 Report

Naijia Fan
09/07/18

	Expert		Behavioral Cloning	
	Mean	Std	Mean	Std
Ant-v2	4800.491780799946	88.96709092942973	4378.190431864148	89.5943074599234
HalfCheetah-v2	4185.471779088302	79.65318220438478	4040.1804336249106	77.94271191799399
Hopper-v2	3778.591392827052	3.4351661338379147	1353.4823545696288	212.96869439305058
Humanoid-v2	10414.375173397892	38.05366182896654	302.1723042136717	49.441034121994306
Reacher-v2	-4.172768952457888	1.6138287052793907	-9.812060014535234	2.905922934376134
Walker2d-v2	5421.924048900647	457.58841956787876	175.20579650240822	285.33613883638907

Table 1: Question 2.2. In the same network size, amount of data, and number of training iterations, the comparison of expert policy and behavioral cloning is above. The yellow highlighted one, HalfCheetah-v2, achieves comparable performance to the expert, and red one, Hopper-v2, does not. Both rollout number is 20 and max timesteps is set by the same environment.

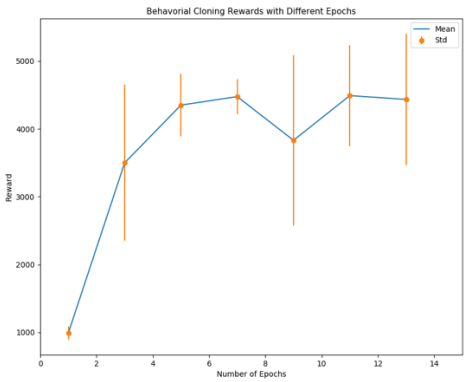

Figure 1: See source data in Table 2. In the same network size, amount of data, the comparison of rewards given different epochs number. The task chosen is Ant-v2. Both rollout number is 20 and max timesteps is set by the same environment.

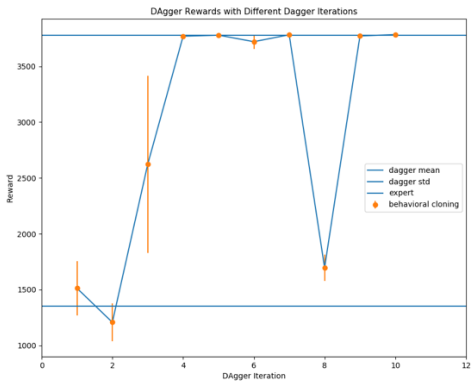

Figure 2: See source data in Table 3. In the same network size, amount of data, and number of training iterations, the comparison of rewards given different dagger iterations. The task chosen is Hopper-v2. Both rollout number is 20 and max timesteps is set by the same environment.

Appendix

Epochs	1	4	8	12	16	20	24
Mean	423.98828961402234	3789.323619971577	4350.861610917002	4476.067357747206	3832.079715068327	4491.220839019459	4434.5819579225945
Std	4.81546533697781	1091.0106601064917	464.8945713422921	259.3228763250365	1258.2538081014657	744.6349768561219	967.0909398729037

Table 2. Mean and std of rewards of behavioral cloning policy with different epochs number for cloning learning.

Iterations	1	2	3	4	5
Mean	1512.4344919914515	1207.1934207421843	2622.5493232814893	3769.397087119257	3779.1785373483167
Std	243.8018102481672	169.4559711718996	792.2196985176346	3.22897032893531	4.341987597202986
Iterations	6	7	8	9	10
Mean	3720.4268652771307	3782.5436357287967	1695.9485864344504	3772.1398833895755	3784.5361999363477
Std	62.640423688618775,	2.7732957173867856	116.32946914788599	3.582557223402759	3.4321169281471904

Table 3. Mean and std of rewards of dagger policy with increase of dagger iterations.