1 Question 2.2

task	mean-expert	mean-behavioral cloning	std-expert	std-behavioral cloning
Ant	4831.49	4612.72	121.63	93.05
Hopper	3779.38	4.38	781.79	117.69

Table 1: Results for comparison of task **Ant** and **Hopper**. With *rollouts* $= 20$ and *epochs* $= 10$. Two four-layer fully-connected network ares trained in behavioral cloning. The first 3 layers have 128 nodes and the amount of nodes in the output layer is the dimension of action.

2 Question 2.3

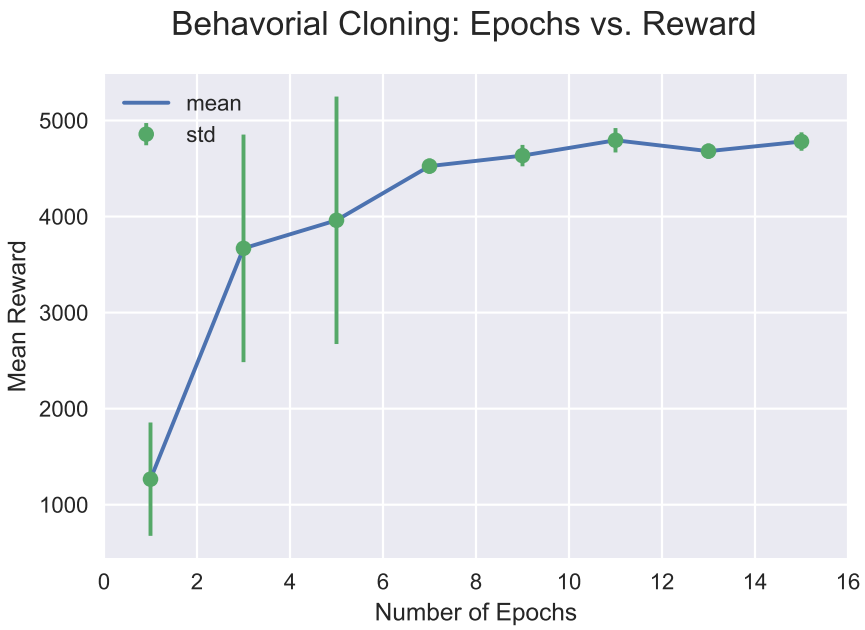

Figure 1: Figure 1 shows how the number of *epochs* affects the performance of the behavioral cloning agent. Experiments was performed on task **Ant**. As the number of *epochs* goes up, the mean *reward* is going up as well. That's because training more epochs makes the neural network to fit the dataset better, thus improves the performance of the agent.

3 Question 3.2

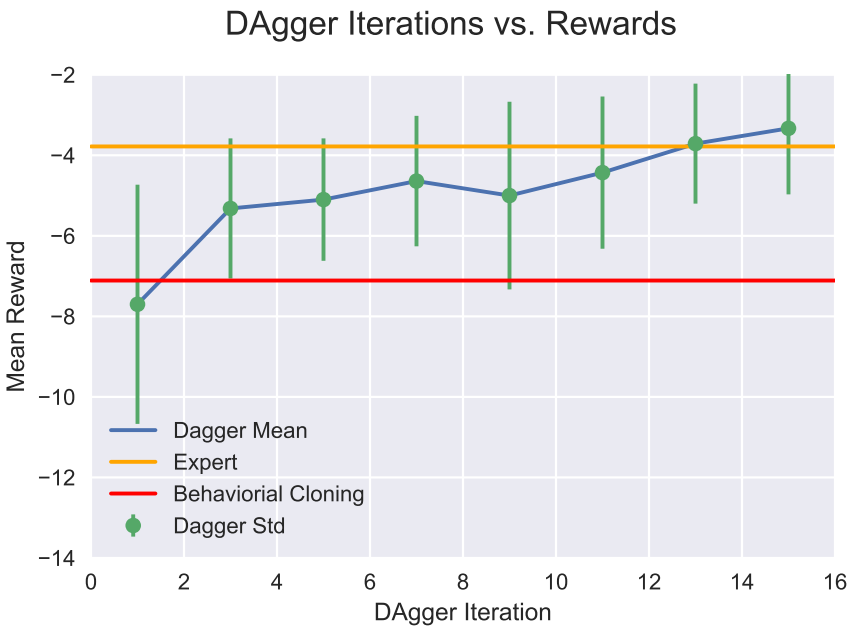

Figure 2: Figure 2 shows how the number of *iterations* affects the performance of the Dagger agent. Experiments was performed on task **Reacher**. The network structure and dataset are totally same as stated in Quesion 2.2(four-layer fully-connected network and *rollouts* $= 20$). As the number of *iterations* goes up, the mean *reward* is going up as well. That's because more *iterations* makes the Dagger agent fit the training data better and eventually outperforms the expert policy.