CS285 - HW4

Ho Nam Nguyen

November 1, 2022

1 Problem 1

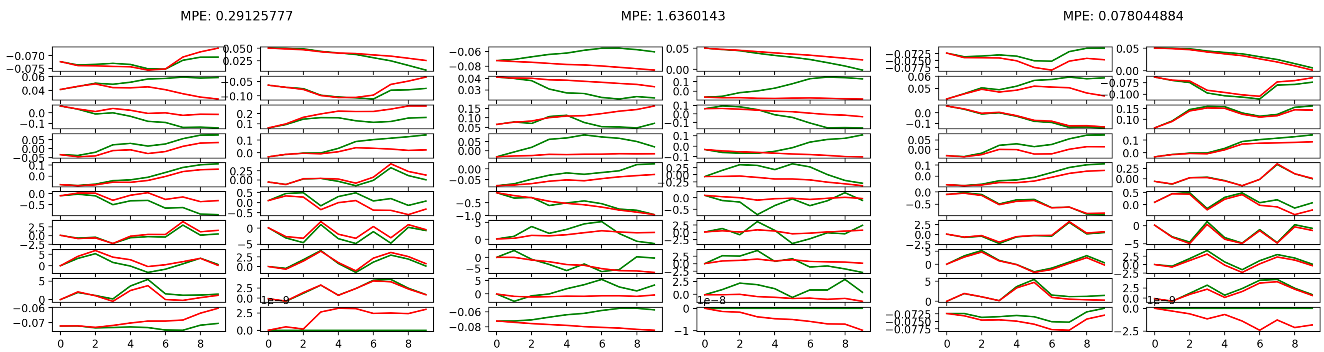

Figure 1: Left to right: n500_arch1x32, n5_arch2x250, n500_arch2x250. Best model is the last one with the most number of training steps and the most expressive architecture.

Commands

```
python cs285/scripts/run_hw4_mb.py --exp_name q1_cheetah_n500_arch1x32
--env_name cheetah-cs285-v0 --add_sl_noise --n_iter 1 --batch_size_initial 20000
--num_agent_train_steps_per_iter 500 --n_layers 1 --size 32 --scalar_log_freq -1
--video_log_freq -1 --mpc_action_sampling_strategy 'random'

python cs285/scripts/run_hw4_mb.py --exp_name q1_cheetah_n5_arch2x250
--env_name cheetah-cs285-v0 --add_sl_noise --n_iter 1 --batch_size_initial 20000
--num_agent_train_steps_per_iter 5 --n_layers 2 --size 250 --scalar_log_freq -1
--video_log_freq -1 --mpc_action_sampling_strategy 'random'

python cs285/scripts/run_hw4_mb.py --exp_name q1_cheetah_n500_arch2x250
--env_name cheetah-cs285-v0 --add_sl_noise --n_iter 1 --batch_size_initial 20000
--num_agent_train_steps_per_iter 500 --n_layers 2 --size 250 --scalar_log_freq -1
--video_log_freq -1 --mpc_action_sampling_strategy 'random'
```

2 Problem 2

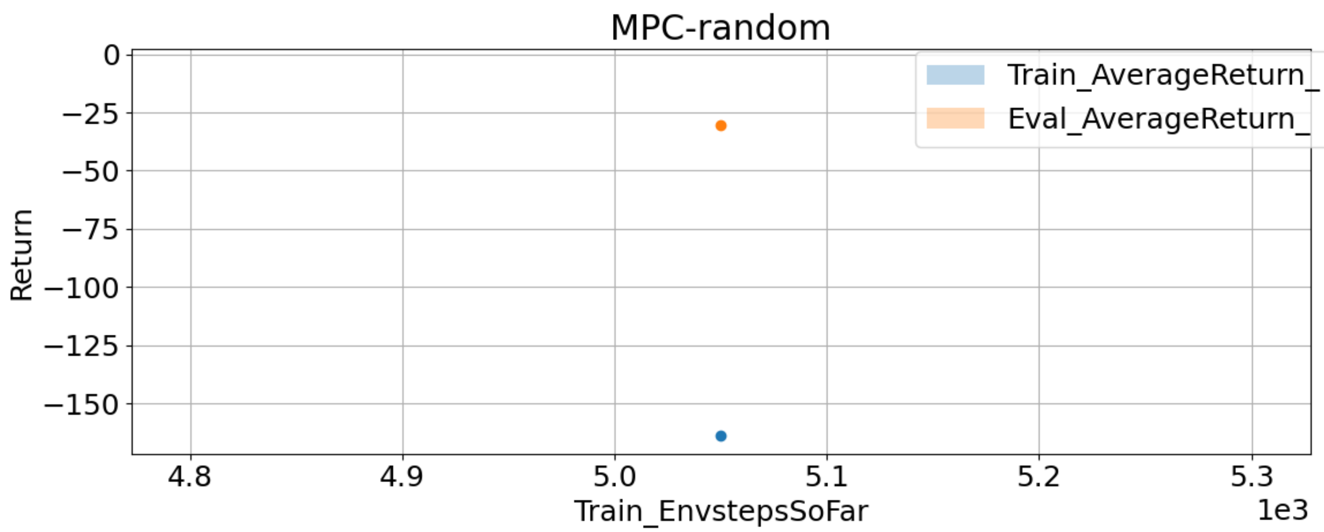

Figure 2

Commands

```
python cs285/scripts/run_hw4_mb.py --exp_name q2_obstacles_singleiteration
--env_name obstacles-cs285-v0 --add_sl_noise --num_agent_train_steps_per_iter 20
--n_iter 1 --batch_size_initial 5000 --batch_size 1000 --mpc_horizon 10
--video_log_freq -1 --mpc_action_sampling_strategy 'random'
```

3 Hyperparameters - DQN

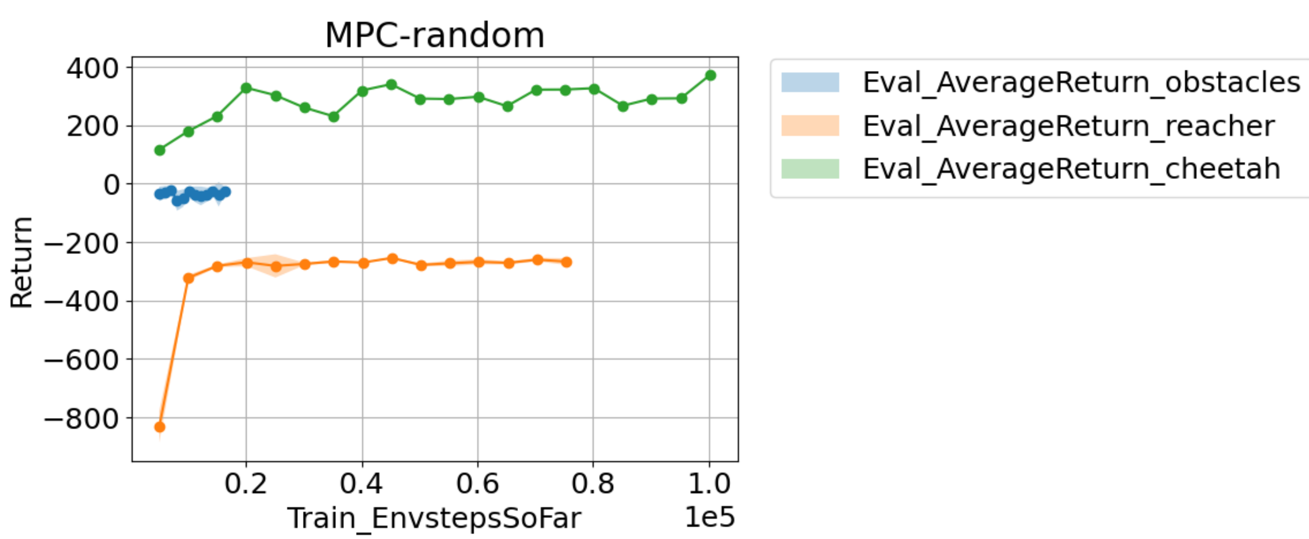

Figure 3

Commands

```
python cs285/scripts/run_hw4_mb.py --exp_name q3_obstacles --env_name obstacles-cs285-v0
--add_sl_noise --num_agent_train_steps_per_iter 20 --batch_size_initial 5000 --batch_size 1000
--mpc_horizon 10 --n_iter 12 --video_log_freq -1 --mpc_action_sampling_strategy 'random'
```

```
python cs285/scripts/run_hw4_mb.py --exp_name q3_reacher --env_name reacher-cs285-v0
--add_sl_noise --mpc_horizon 10 --num_agent_train_steps_per_iter 1000 --batch_size_initial 5000
--batch_size 5000 --n_iter 15 --video_log_freq -1 --mpc_action_sampling_strategy 'random'
```

4 Problem 4

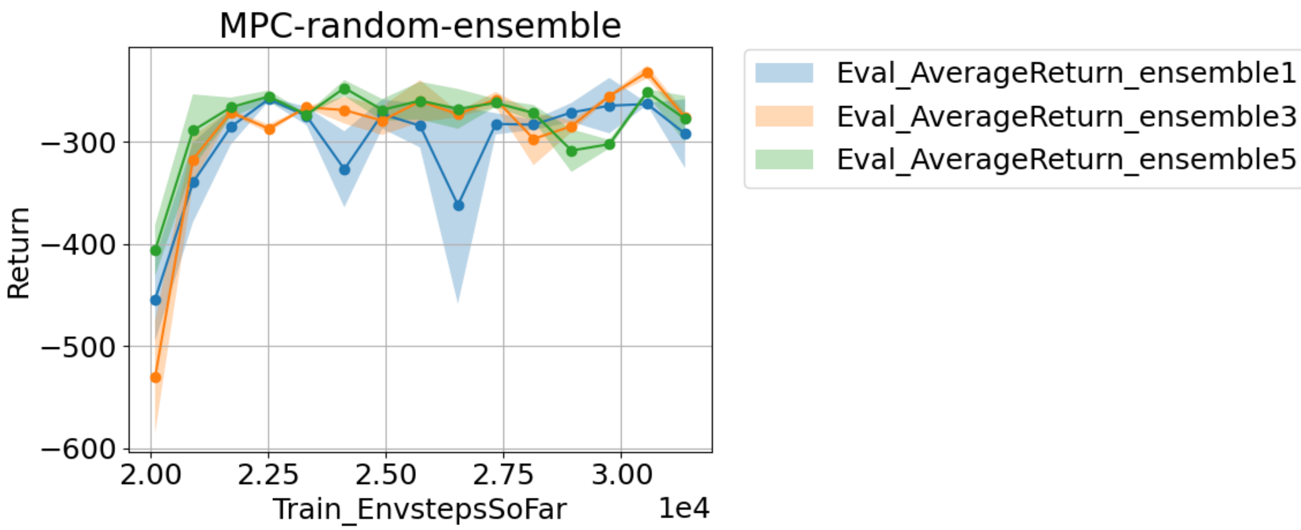

Figure 4: No significant difference

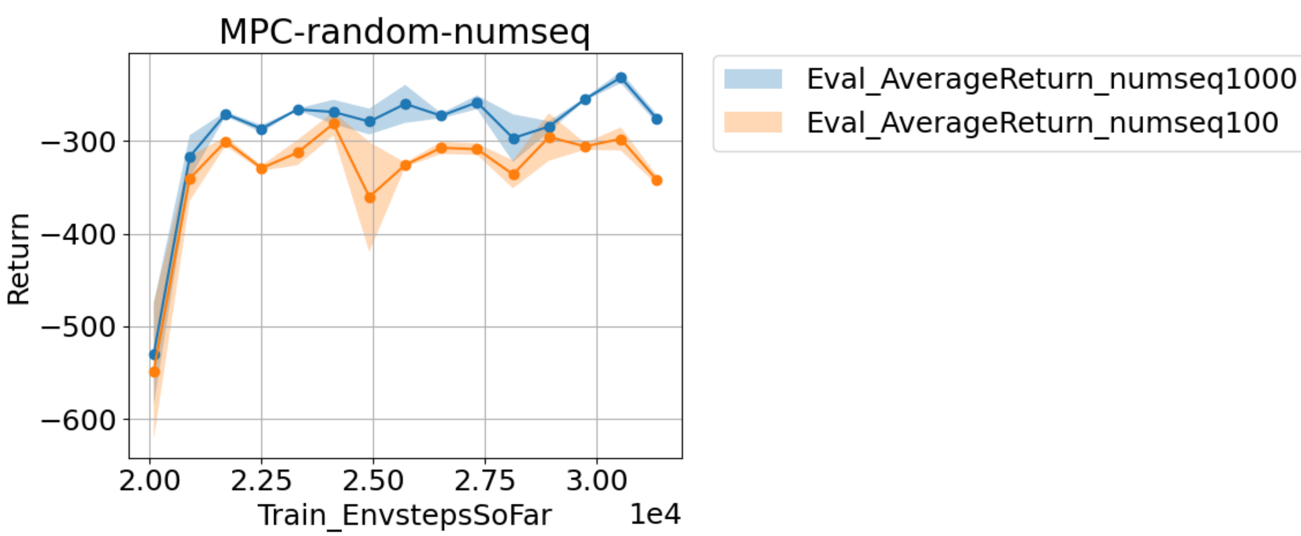

Figure 5: Larger number of candidate action sequences is better

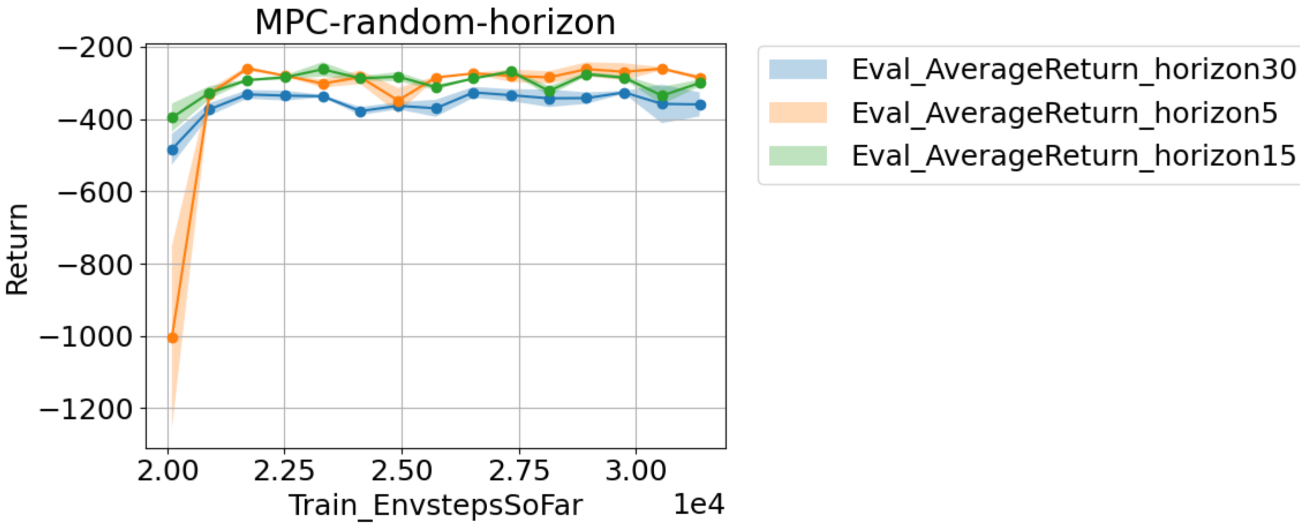

Figure 6: No significant difference

Commands

```
python cs285/scripts/run_hw4_mb.py --exp_name q4_reacher_horizon5
--env_name reacher-cs285-v0 --add_sl_noise --mpc_horizon 5
--mpc_action_sampling_strategy 'random' --num_agent_train_steps_per_iter 1000
--batch_size 800 --n_iter 15 --video_log_freq -1 --mpc_action_sampling_strategy 'random'

python cs285/scripts/run_hw4_mb.py --exp_name q4_reacher_horizon15
--env_name reacher-cs285-v0 --add_sl_noise --mpc_horizon 15
--num_agent_train_steps_per_iter 1000 --batch_size 800 --n_iter 15
--video_log_freq -1 --mpc_action_sampling_strategy 'random'

python cs285/scripts/run_hw4_mb.py --exp_name q4_reacher_horizon30
--env_name reacher-cs285-v0 --add_sl_noise --mpc_horizon 30
--num_agent_train_steps_per_iter 1000 --batch_size 800 --n_iter 15
--video_log_freq -1 --mpc_action_sampling_strategy 'random'

python cs285/scripts/run_hw4_mb.py --exp_name q4_reacher_numseq100
--env_name reacher-cs285-v0 --add_sl_noise --mpc_horizon 10
--num_agent_train_steps_per_iter 1000 --batch_size 800 --n_iter 15
--mpc_num_action_sequences 100 --mpc_action_sampling_strategy 'random'

python cs285/scripts/run_hw4_mb.py --exp_name q4_reacher_numseq1000
--env_name reacher-cs285-v0 --add_sl_noise --mpc_horizon 10
--num_agent_train_steps_per_iter 1000 --batch_size 800
--n_iter 15 --video_log_freq -1 --mpc_num_action_sequences 1000
--mpc_action_sampling_strategy 'random'

python cs285/scripts/run_hw4_mb.py --exp_name q4_reacher_ensemble1
--env_name reacher-cs285-v0 --ensemble_size 1 --add_sl_noise --mpc_horizon 10
--num_agent_train_steps_per_iter 1000 --batch_size 800 --n_iter 15
--video_log_freq -1 --mpc_action_sampling_strategy 'random'

python cs285/scripts/run_hw4_mb.py --exp_name q4_reacher_ensemble3
--env_name reacher-cs285-v0 --ensemble_size 3 --add_sl_noise --mpc_horizon 10
--num_agent_train_steps_per_iter 1000 --batch_size 800 --n_iter 15
--video_log_freq -1 --mpc_action_sampling_strategy 'random'

python cs285/scripts/run_hw4_mb.py --exp_name q4_reacher_ensemble5
--env_name reacher-cs285-v0 --ensemble_size 5 --add_sl_noise --mpc_horizon 10
```

```
--num_agent_train_steps_per_iter 1000 --batch_size 800 --n_iter 15
--video_log_freq -1 --mpc_action_sampling_strategy 'random'
```

5 Problem 5 - CEM

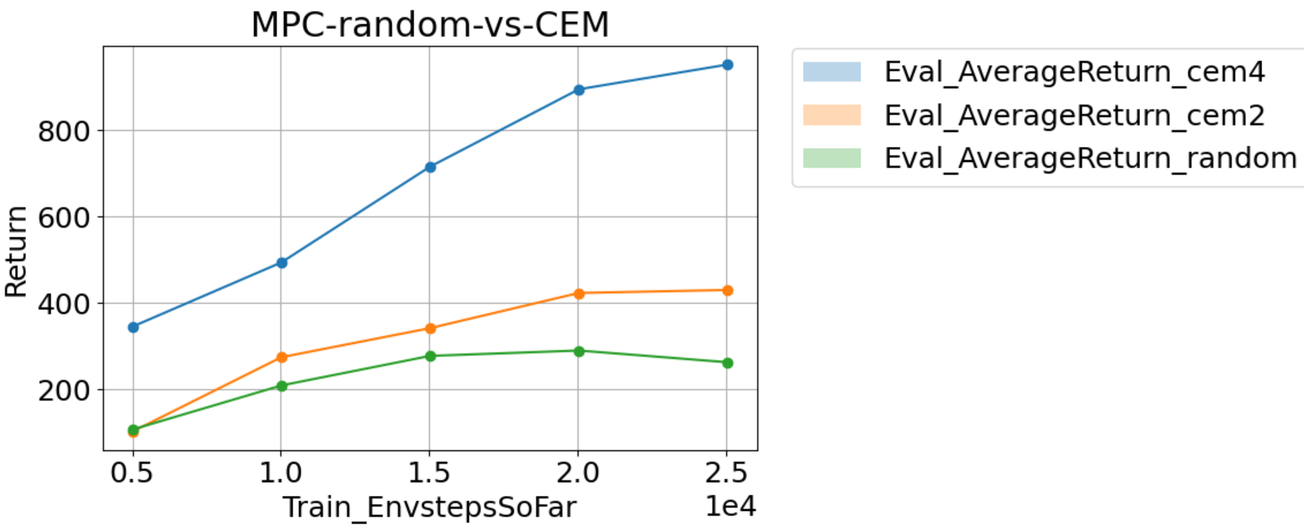

Figure 7: CEM is more effective: blue (4 iterations) is better than orange (2 iterations)

Commands

```
python cs285/scripts/run_hw4_mb.py --exp_name q5_cheetah_random
--env_name cheetah-cs285-v0 --mpc_horizon 15 --add_sl_noise
--num_agent_train_steps_per_iter 1500 --batch_size_initial 5000 --batch_size 5000
--n_iter 5 --video_log_freq -1 --mpc_action_sampling_strategy random

python cs285/scripts/run_hw4_mb.py --exp_name q5_cheetah_cem_2
--env_name 'cheetah-cs285-v0' --mpc_horizon 15 --add_sl_noise
--num_agent_train_steps_per_iter 1500 --batch_size_initial 5000 --batch_size 5000
--n_iter 5 --video_log_freq -1 --mpc_action_sampling_strategy 'cem' --cem_iterations 2

python cs285/scripts/run_hw4_mb.py --exp_name q5_cheetah_cem_4
--env_name 'cheetah-cs285-v0' --mpc_horizon 15 --add_sl_noise
--num_agent_train_steps_per_iter 1500 --batch_size_initial 5000 --batch_size 5000
--n_iter 5 --video_log_freq -1 --mpc_action_sampling_strategy 'cem' --cem_iterations 4
```

6 SAC using learned model

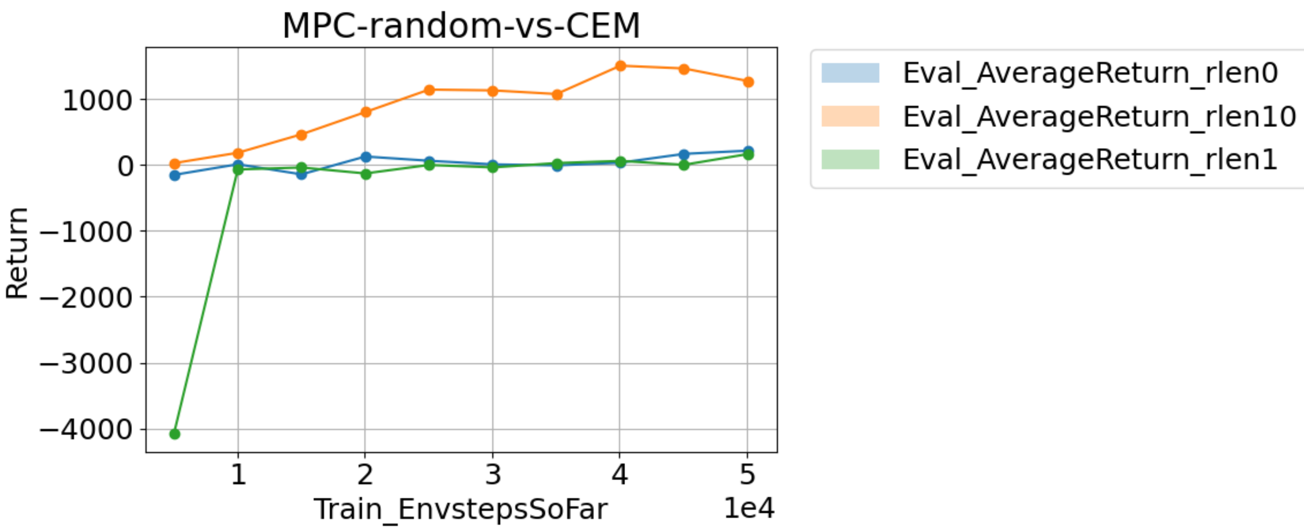

Figure 8: More rollouts from model increase the performance dramatically

Commands

```
python cs285/scripts/run_hw4_mbpo.py --exp_name q6_cheetah_rlen0
--env_name 'cheetah-cs285-v0' --add_sl_noise --num_agent_train_steps_per_iter 1500
--batch_size_initial 5000 --batch_size 5000 --n_iter 10 --video_log_freq -1 --sac_discount 0.99
--sac_n_layers 2 --sac_size 256 --sac_batch_size 1500 --sac_learning_rate 0.0003
--sac_init_temperature 0.1 --sac_n_iter 1000 --mbpo_rollout_length 0

python cs285/scripts/run_hw4_mbpo.py --exp_name q6_cheetah_rlen1
--env_name 'cheetah-cs285-v0' --add_sl_noise --num_agent_train_steps_per_iter 1500
--batch_size_initial 5000 --batch_size 5000 --n_iter 10 --video_log_freq -1 --sac_discount 0.99
--sac_n_layers 2 --sac_size 256 --sac_batch_size 1500 --sac_learning_rate 0.0003
--sac_init_temperature 0.1 --sac_n_iter 5000 --mbpo_rollout_length 1

python cs285/scripts/run_hw4_mbpo.py --exp_name q6_cheetah_rlen10
--env_name 'cheetah-cs285-v0' --add_sl_noise --num_agent_train_steps_per_iter 1500
--batch_size_initial 5000 --batch_size 5000 --n_iter 10 --video_log_freq -1 --sac_discount 0.99
--sac_n_layers 2 --sac_size 256 --sac_batch_size 1500 --sac_learning_rate 0.0003
--sac_init_temperature 0.1 --sac_n_iter 5000 --mbpo_rollout_length 10
```