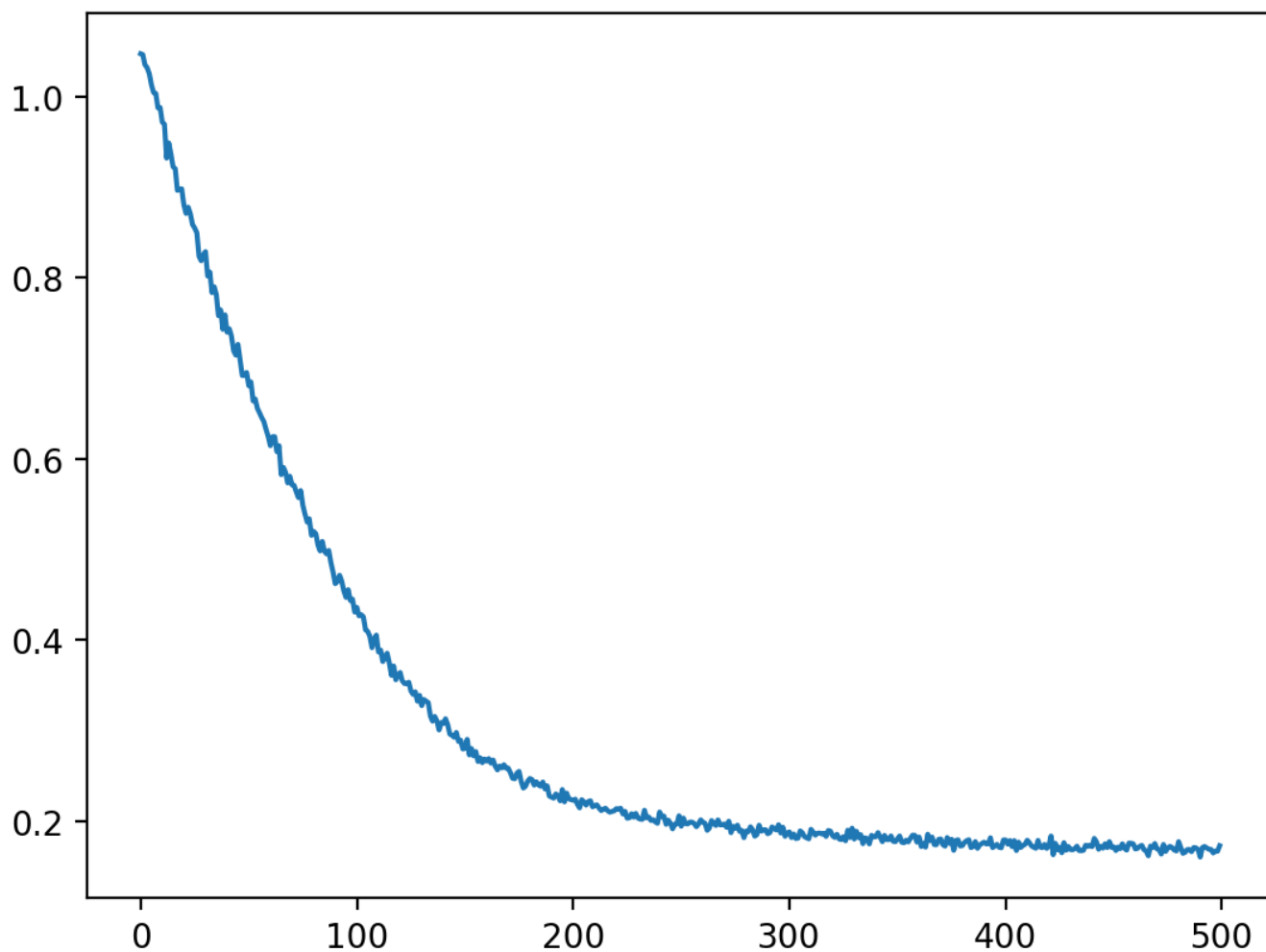


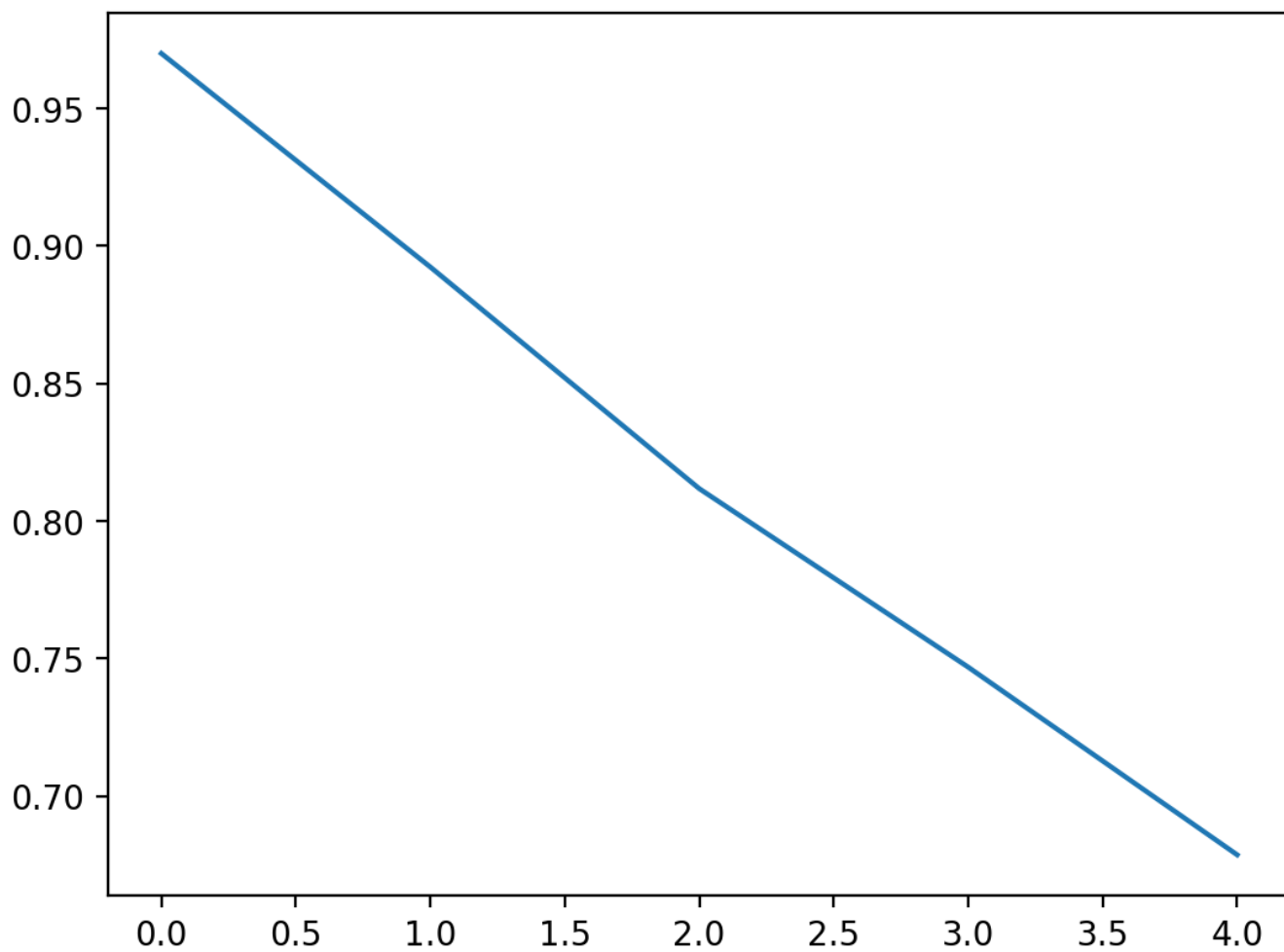
CS285 HW4 Report

Q1

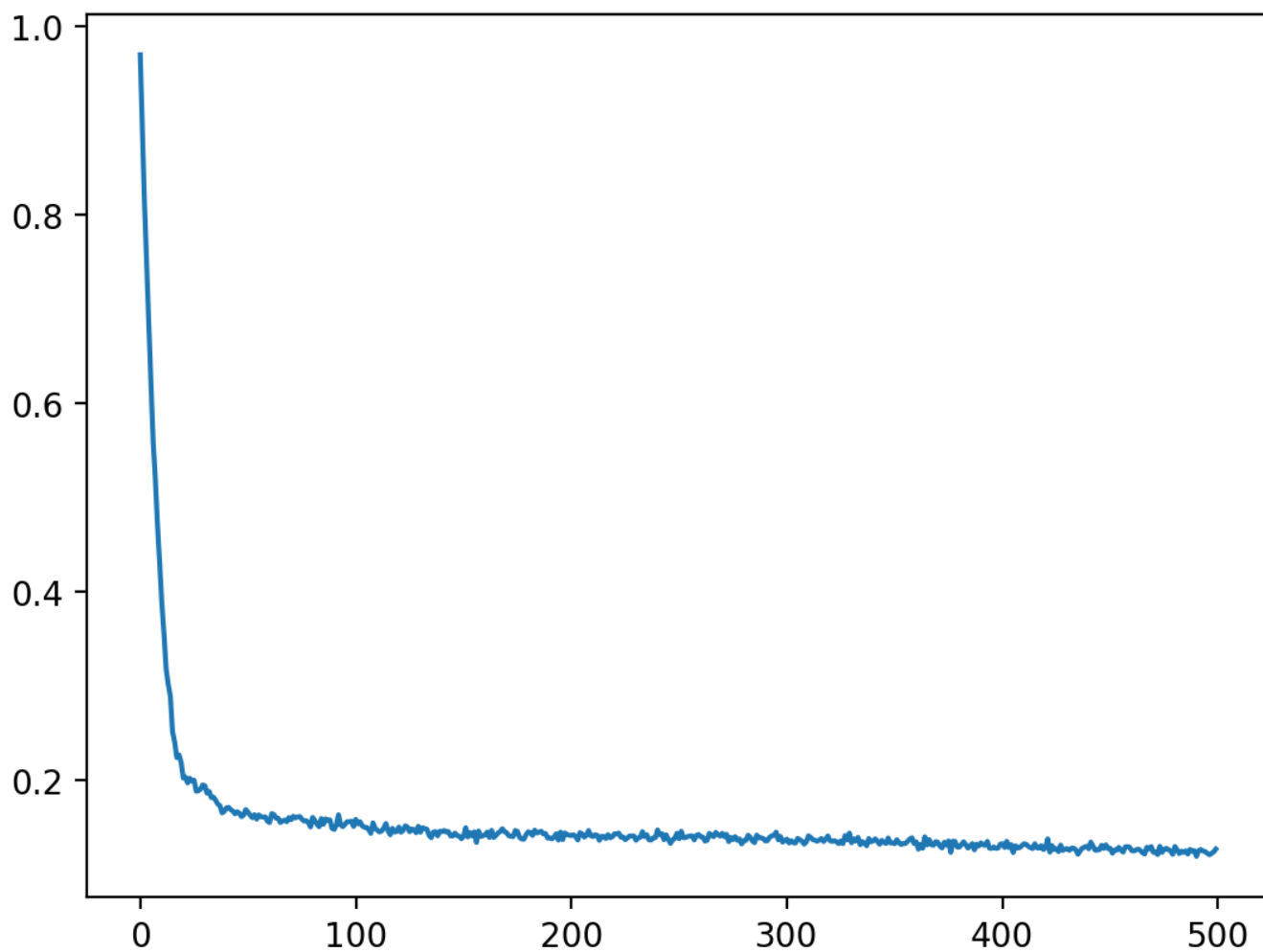
First command (q1_cheetah_n500_arch1x32):



Second command (q1_cheetah_n5_arch2x250):



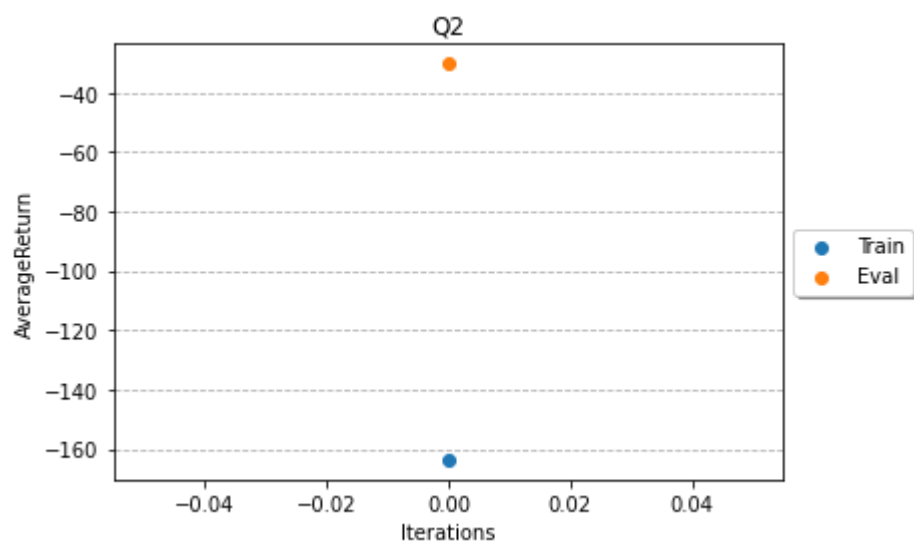
Third command (q1_cheetah_n500_arch2x250):



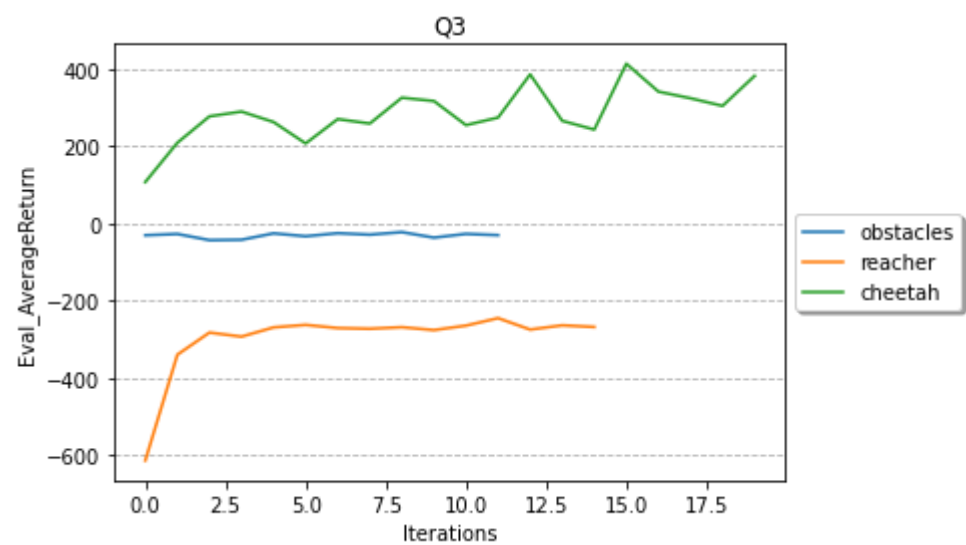
n500_arch2x250 is the best.

This network has the most parameters, and thus is the most expressive one.

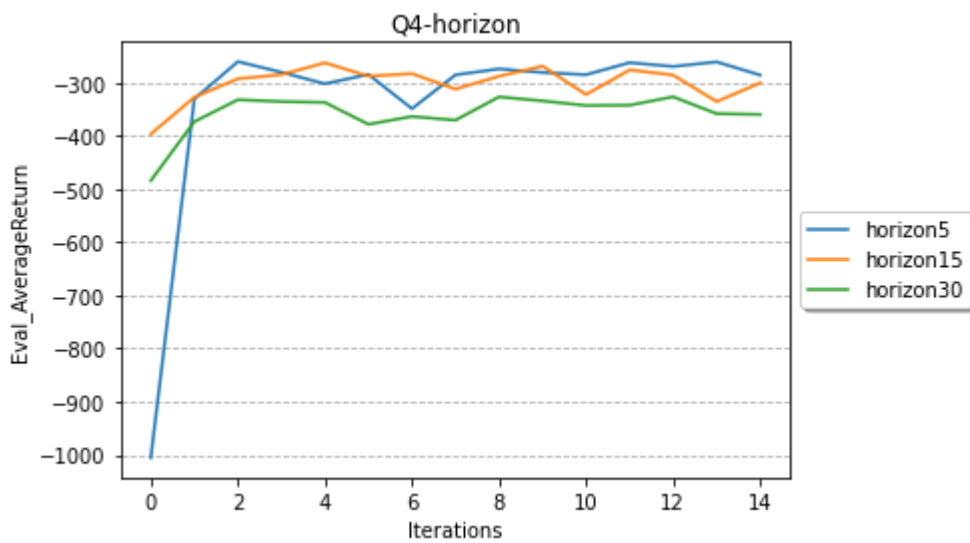
Q2



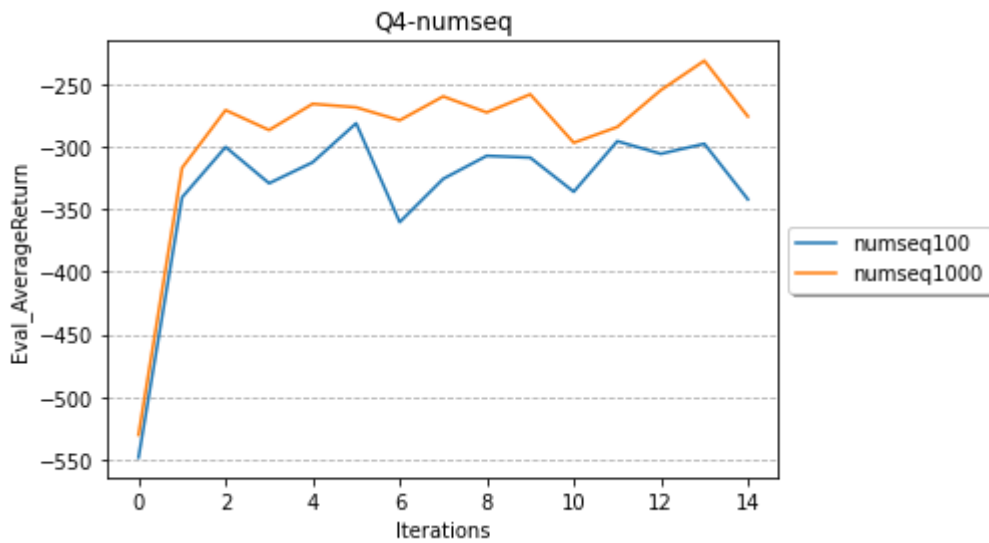
Q3



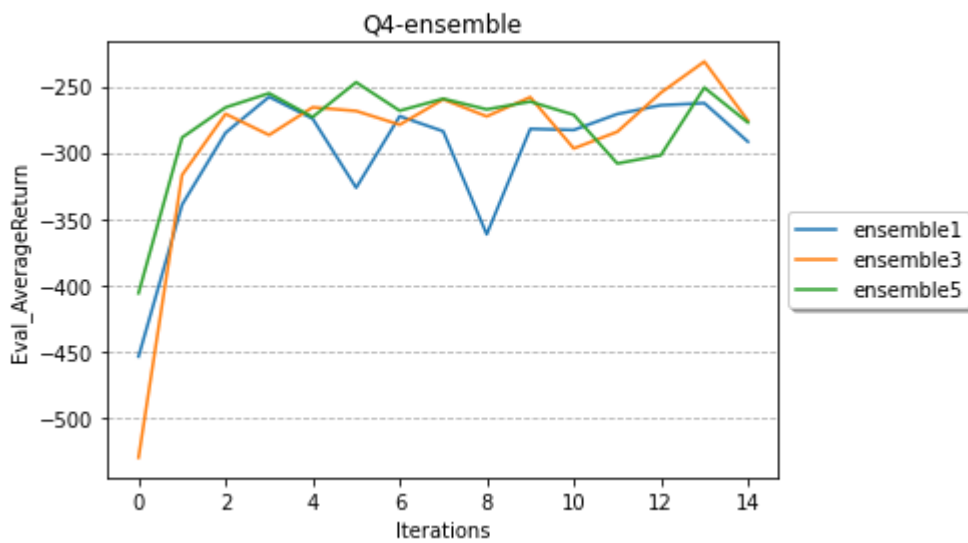
Q4



Horizon. A larger horizon makes an untrained model performe better. But a too large horizon confuses the model to select action.

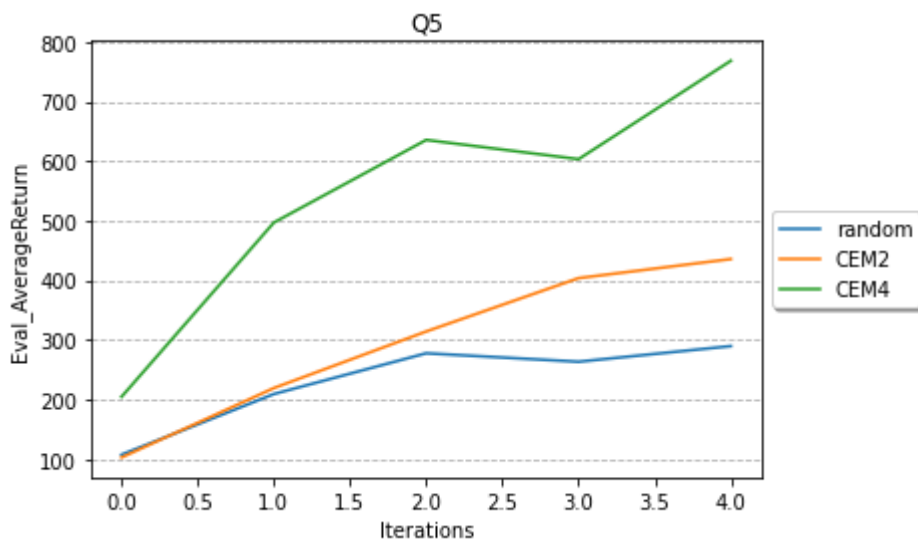


With a large numseq, the model can get a higher performance.



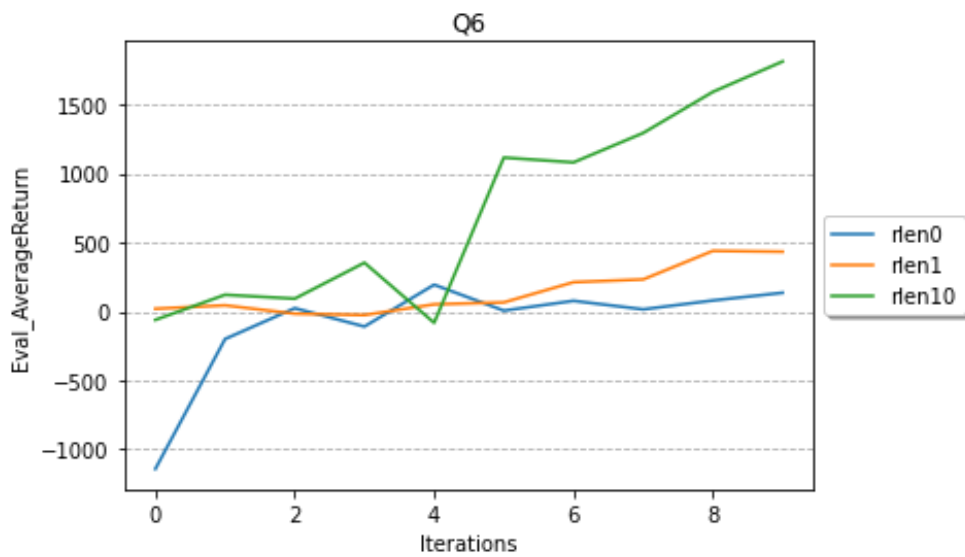
With a larger ensemble, the model performance is more stable.

Q5



A larger number of sampling iteration not only enables the model learns faster, but also improves its convergence performance.

Q6



Commands

All the commands are in a Makefile and can be used by `make q[1-6]`

submit:

```
-rm data.zip run_logs.zip
zip cs285.zip -r cs285
zip run_logs.zip -r data
```

q1:

q1 - 1

```
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
```

q1 - 2

```
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
```

q1 - 3

```
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
```

q2:

q2

```
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'
```

q3:

q3 - 1

```
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'
# q3 - 2
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'
# q3 - 3
python cs285/scripts/run_hw4_mb.py --exp_name q3_cheetah \
--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 20 --video_log_freq -1 \
--mpc_action_sampling_strategy 'random'

q4:
# q4 - 1
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'
# q4 - 2
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'
# q4 - 3
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'
# q4 - 5
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'
# q4 - 6
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'
# q4 - 7
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'
# q4 - 8
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'
# q4 - 9
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'

q5:
# q5 - 1
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'
# q5 - 2
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
# q5 - 3
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

q6:
# q6 - 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 1000 --mbpo_rollout_length 0
# q6 - 2
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
# q6 - 3
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

```