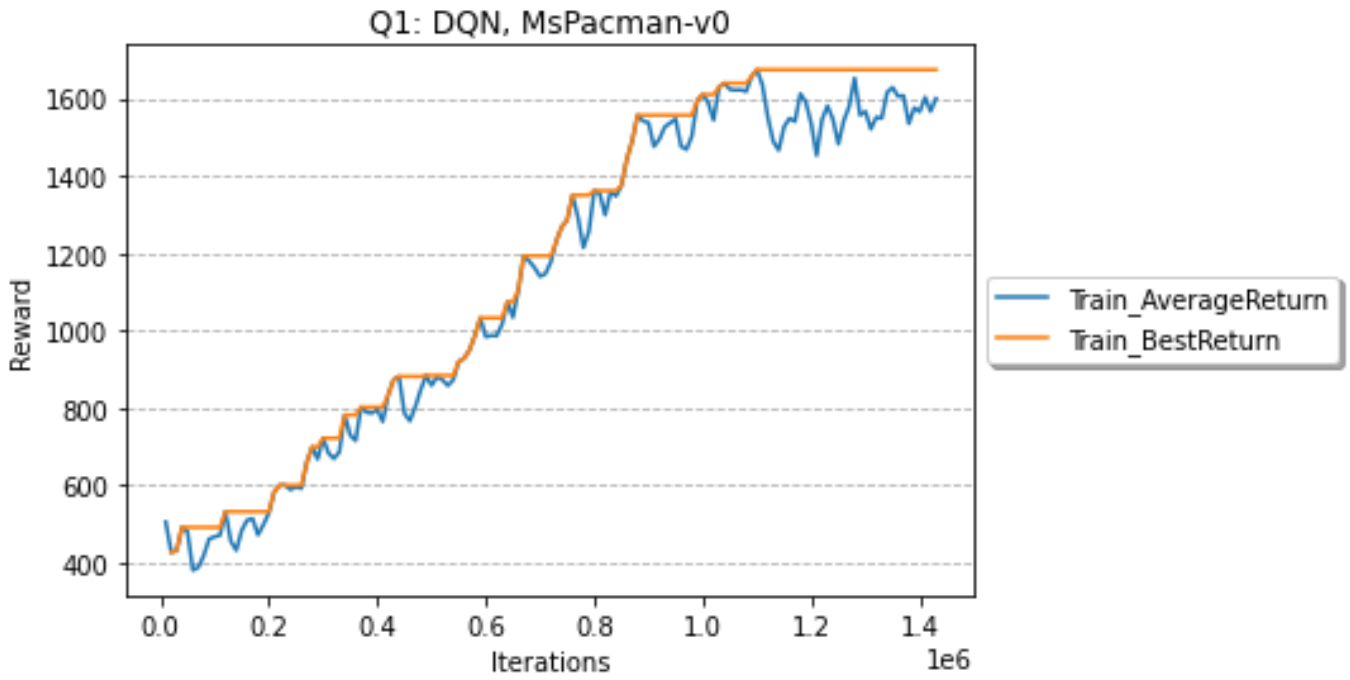
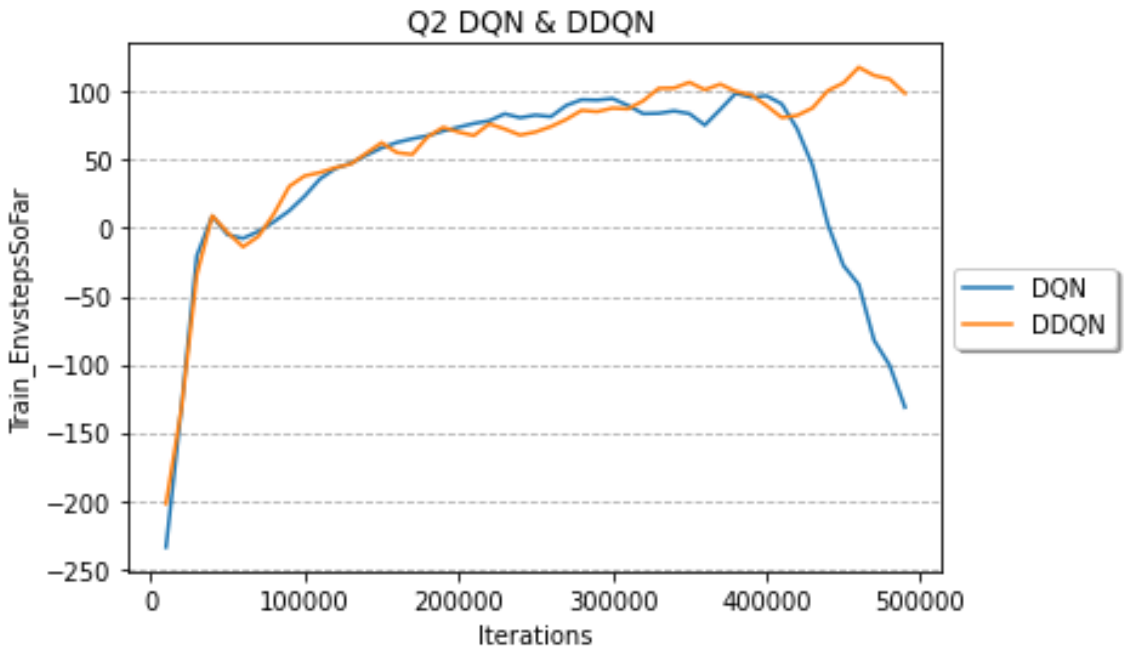
CS285 HW3 REPORT

Question 1

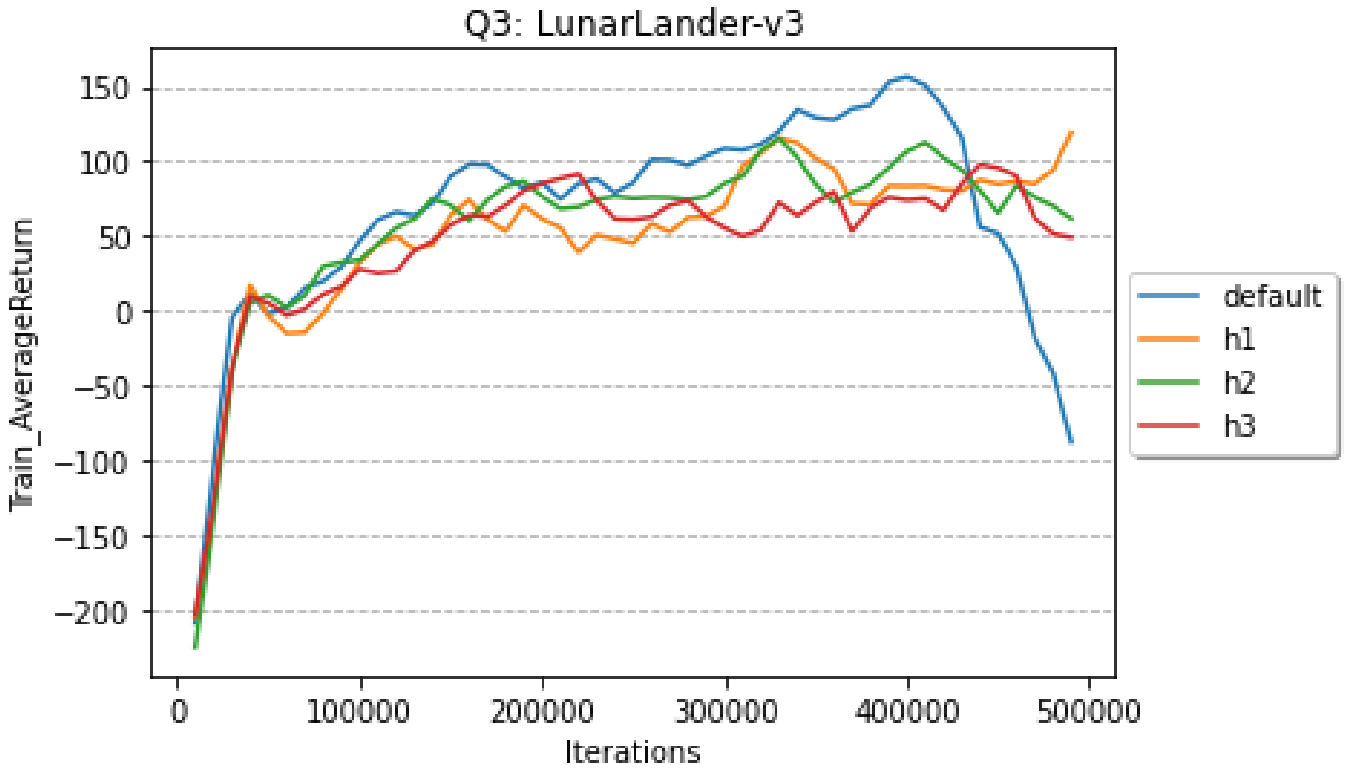


Question 2



We can see that DDQN prevents the decrease of reward after convergence.

Question 3



The parameter I chose is ϵ-greedy.

The default configuration is

```
PiecewiseSchedule(
    [(0, 1.0),(1e6, 0.1),(num_timesteps / 8, 0.01),],
    outside_value=0.01
)
```

h1 is

```
PiecewiseSchedule(
    [(0, 0.8),(1e6, 0.1),(num_timesteps / 8, 0.01),],
    outside_value=0.01
)
```

h2 is

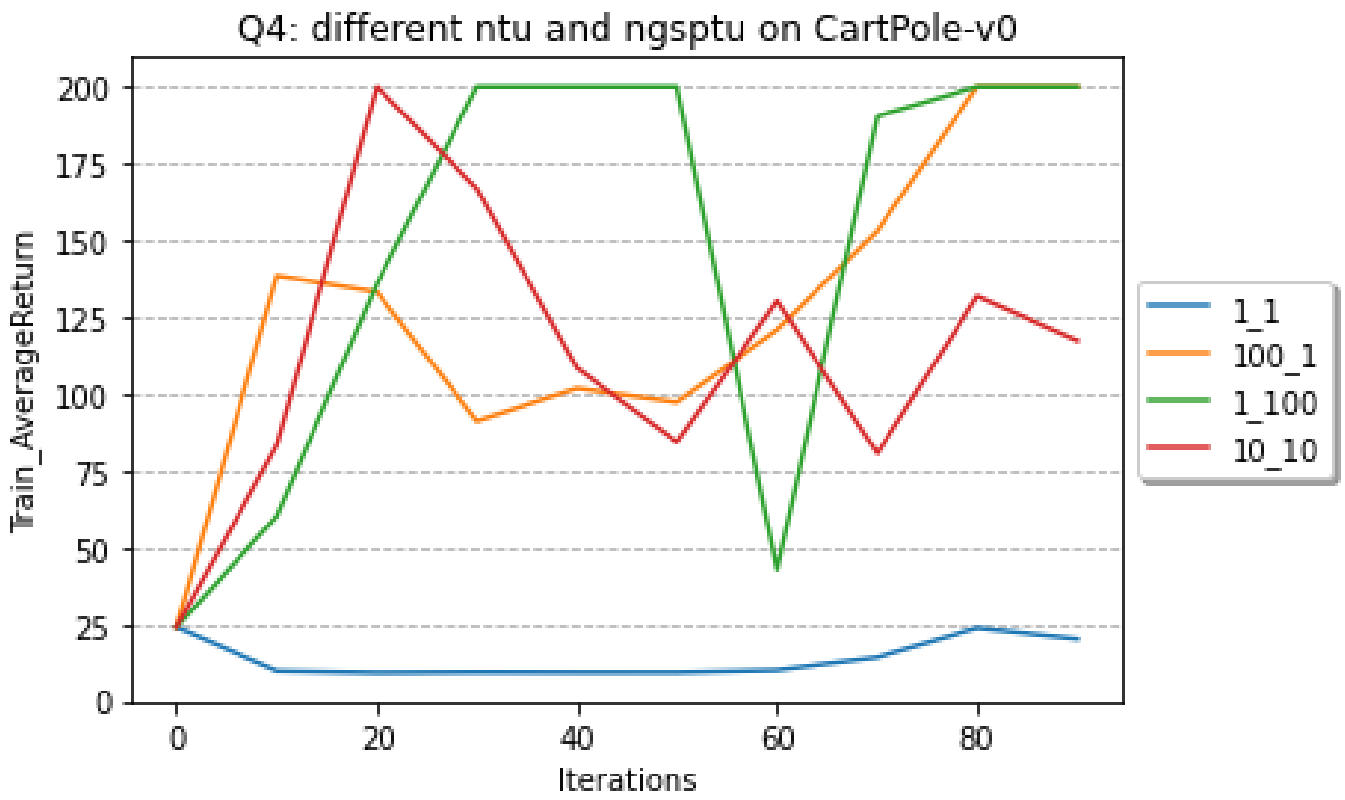
```
PiecewiseSchedule(
    [(0, 0.8),(1e6, 0.09),(num_timesteps / 8, 0.008),],
    outside_value=0.005
)
```

h3 is

```
PiecewiseSchedule(
    [(0, 1.0),(1e6, 0.2),(num_timesteps / 8, 0.03),],
    outside_value=0.03
)
```

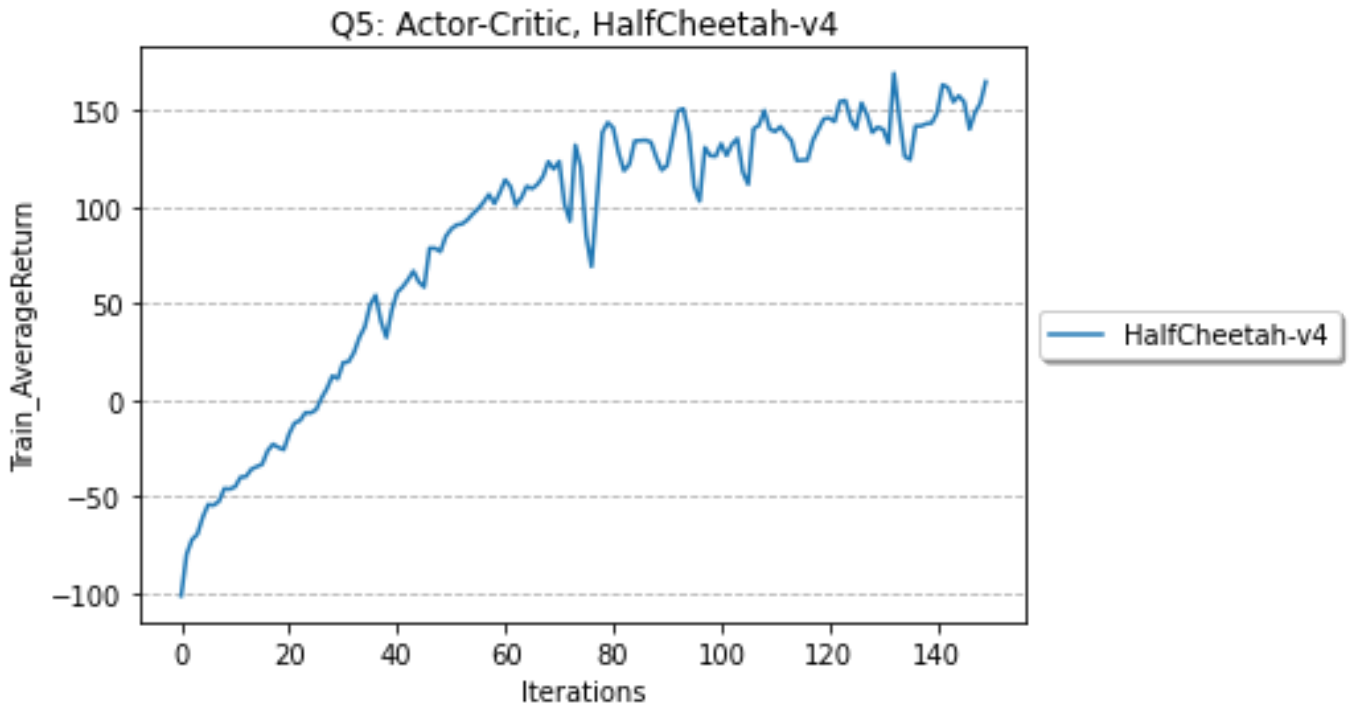
The three new configurations are not as good as the default at peak, but can alleviate the decrease after convergence.

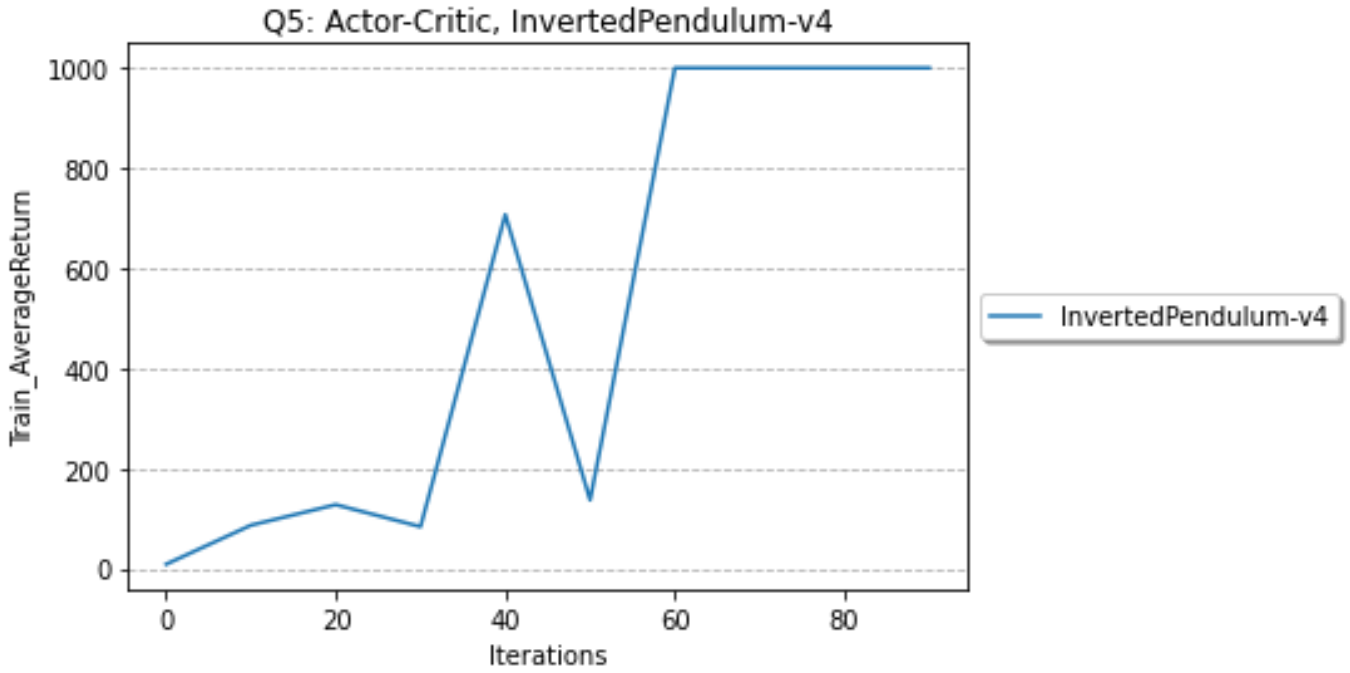
Question 4

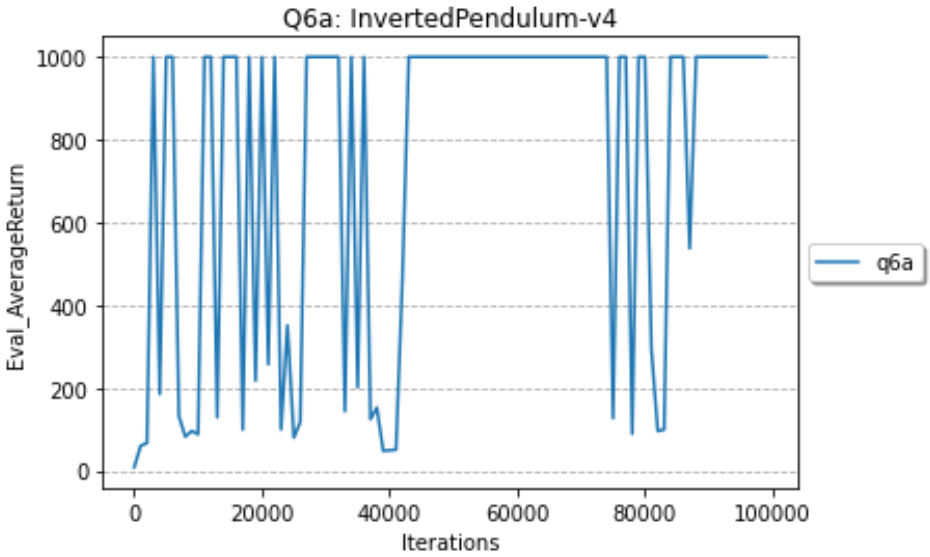


The `-ntu 1 -ngsptu 100` configuration has the best result because it holds at 200 for the longest time.

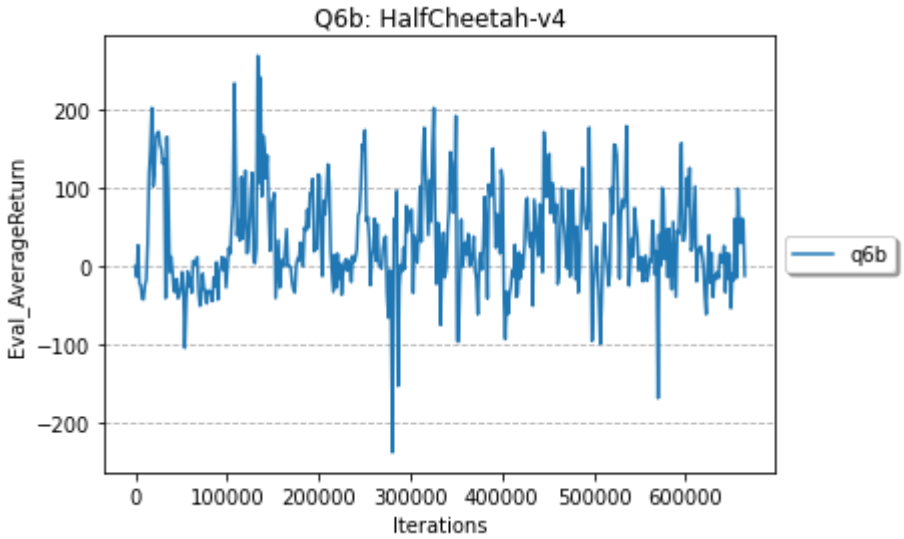
Question 5


Q5: Actor-Critic, HalfCheetah-v4
Train_AverageReturn
150
100
50
0
−50
−100
0
20
40
60
80
100
120
140
Iterations
HalfCheetah-v4


Q5: Actor-Critic, InvertedPendulum-v4
Train_AverageReturn
1000
800
600
400
200
0
0
20
40
60
80
Iterations
InvertedPendulum-v4



Question 6



My Q6b HalfCheetah does not get the expected performance.

Commands

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

q1-test:
    python cs285/scripts/run_hw3_dqn.py --env_name LunarLander-v3 \
    --exp_name q1
q1:
    python cs285/scripts/run_hw3_dqn.py --env_name MsPacman-v0 \
    --exp_name q1

q2-dqn:
    python cs285/scripts/run_hw3_dqn.py --env_name LunarLander-v3 \
    --exp_name q2_dqn_1 --seed 1
    python cs285/scripts/run_hw3_dqn.py --env_name LunarLander-v3 \
    --exp_name q2_dqn_2 --seed 2
    python cs285/scripts/run_hw3_dqn.py --env_name LunarLander-v3 \
    --exp_name q2_dqn_3 --seed 3

q2-ddqn:
    python cs285/scripts/run_hw3_dqn.py --env_name LunarLander-v3 \
    --exp_name q2_doubledqn_1 --double_q --seed 1
    python cs285/scripts/run_hw3_dqn.py --env_name LunarLander-v3 \
    --exp_name q2_doubledqn_2 --double_q --seed 2
    python cs285/scripts/run_hw3_dqn.py --env_name LunarLander-v3 \
    --exp_name q2_doubledqn_3 --double_q --seed 3

q3:
    python cs285/scripts/run_hw3_dqn.py --env_name LunarLander-v3 \
    --exp_name q3_hparam1 -gt 1
    python cs285/scripts/run_hw3_dqn.py --env_name LunarLander-v3 \
    --exp_name q3_hparam2 -gt 2
    python cs285/scripts/run_hw3_dqn.py --env_name LunarLander-v3 \
    --exp_name q3_hparam3 -gt 3

q4-test:
    python cs285/scripts/run_hw3_actor_critic.py --env_name CartPole-v0 \
    -n 100 -b 1000 --exp_name q4_ac_1_1 -ntu 1 -ngsptu 1

q4:
    python cs285/scripts/run_hw3_actor_critic.py --env_name CartPole-v0 \
    -n 100 -b 1000 --exp_name q4_100_1 -ntu 100 -ngsptu 1
    python cs285/scripts/run_hw3_actor_critic.py --env_name CartPole-v0 \
    -n 100 -b 1000 --exp_name q4_1_100 -ntu 1 -ngsptu 100
    python cs285/scripts/run_hw3_actor_critic.py --env_name CartPole-v0 \
    -n 100 -b 1000 --exp_name q4_10_10 -ntu 10 -ngsptu 10

q5:
    python cs285/scripts/run_hw3_actor_critic.py \
```

```
    --env_name InvertedPendulum-v4 --ep_len 1000 \
    --discount 0.95 -n 100 -l 2 -s 64 -b 5000 -lr 0.01 \
    --exp_name q5_1_100 -ntu 1 -ngsptu 100
    python cs285/scripts/run_hw3_actor_critic.py \
    --env_name HalfCheetah-v4 --ep_len 150 \
    --discount 0.90 --scalar_log_freq 1 -n 150 -l 2\
     -s 32 -b 30000 -eb 1500 -lr 0.02 --exp_name q5_1_100\
     -ntu 1 -ngsptu 100


q6:
    python cs285/scripts/run_hw3_sac.py \
    --env_name InvertedPendulum-v4 --ep_len 1000 \
    --discount 0.99 --scalar_log_freq 1000 \
    -n 100000 -l 2 -s 256 -b 1000 -eb 2000 \
    -lr 0.0003 --init_temperature 0.1 \
    --exp_name q6a_sac_InvertedPendulum_default \
    --seed 1
    python cs285/scripts/run_hw3_sac.py \
    --env_name HalfCheetah-v4 --ep_len 150 \
    --discount 0.99 --scalar_log_freq 1500 \
    -n 2000000 -l 2 -s 256 -b 1500 -eb 1500 \
    -lr 0.0003 --init_temperature 0.1 \
--exp_name q6b_sac_HalfCheetah_default \
    --seed 1
```