Shivam Bhat
PUID: 0033760929
bhat41@purdue.edu

Instructions

'q1' has question 1 code - main3.py
To run question
python main3.py --qn <Question Number>

For Q1.1
python main3.py --qn 1.1

For Q1.2
python main3.py --qn 1.2

For Q1.3
python main3.py --qn 1.3

For Q1.4
python main3.py --qn 1.4

```
~/Doc/P/purduePrivate/r/a4/s/q1 devRob *4 !6 ?24 ❯ python main3.py --qn 1.2
Obs Space:4
 Action Space:2

Iter (1/200) Mean Reward: 20.90, Loss: 162.29
```

```
~/Doc/P/purduePrivate/r/a4/s/q1 devRob *4 !6 ?24 ❯ python main3.py --qn 1.3
Obs Space:4
 Action Space:2

Iter (1/200) Mean Reward: 22.06, Loss: 0.1154807806
```

Q2

Flags - 'mode'- 0/1 and 'modelPath'
mode - 1/0 - Test/Train if you want to train the model

```
    parser.add_argument('--mode', help='1/0 - Test/Train if you want to train the model')
        parser.add_argument("--modelPath",  help='defaults  to  \'./q2Model.pkl\'', default='./ModelPolicy.pkl')
```

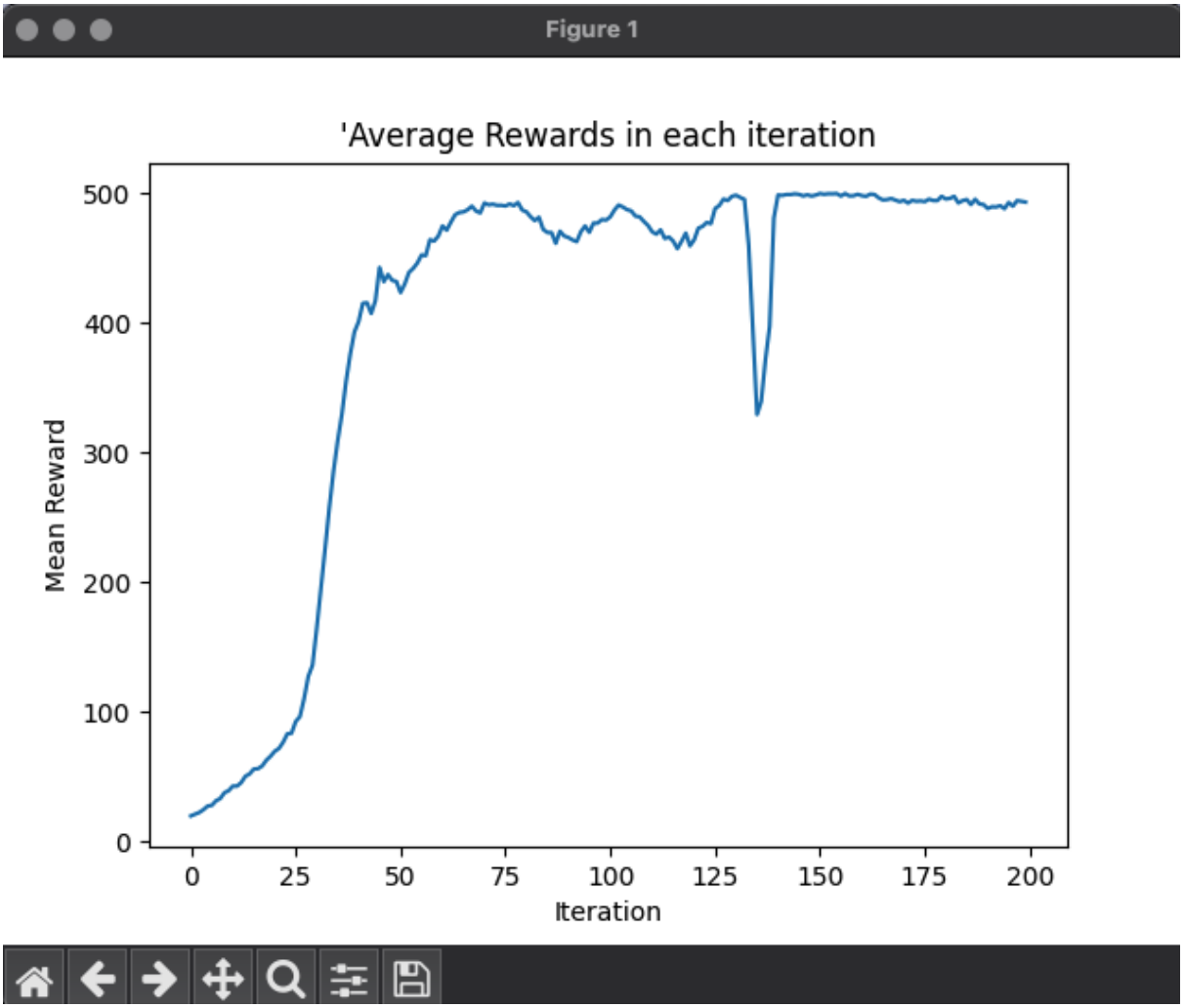
```
~/Doc/P/purduePrivate/r/a4/s/q2 devRob *4 !6 ?24 ❯ python3 main.py --mode 0       temp2l4
pybullet build time: Mar 31 2023 13:24:03
***** Training Model *****

/Users/sbhat/miniconda3/envs/temp2l4/lib/python3.10/site-packages/gym/spaces/box.py:73: UserWa
rning: WARN: Box bound precision lowered by casting to float32
  logger.warn(
Obs Space 8
 Action Space 2

argv[0]=
argv[0]=
Iter (1/200) Mean Reward: -24.77, Loss: -283.54
^Z
zsh: suspended  python3 main.py --mode 0
~/Doc/P/purduePrivate/r/a4/s/q2 devRob *4 !6 ?24 ❯                 ✘ TSTP 32s  temp2l4
```

OUTPUTS

Q1.1:



Q1.2:

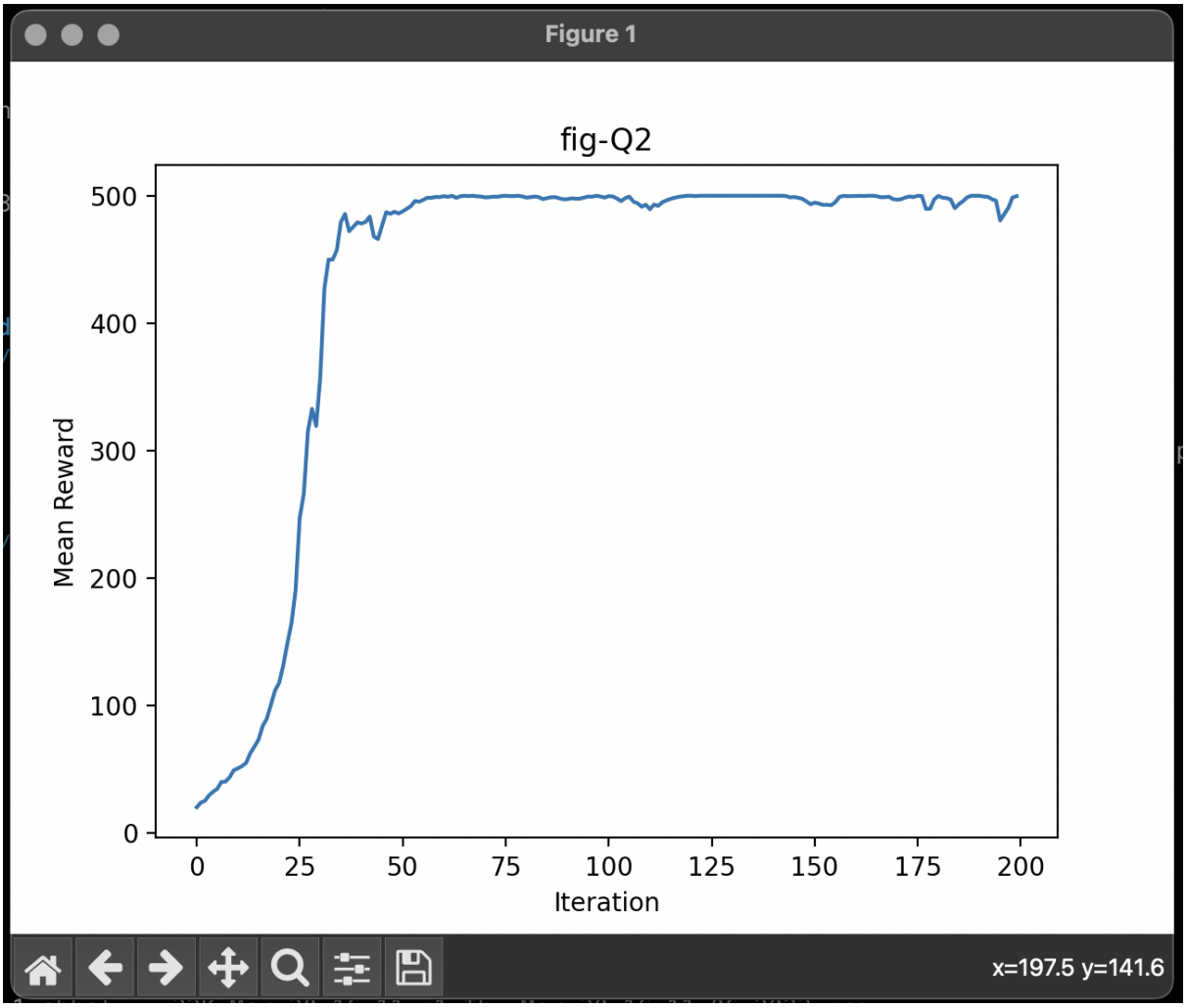
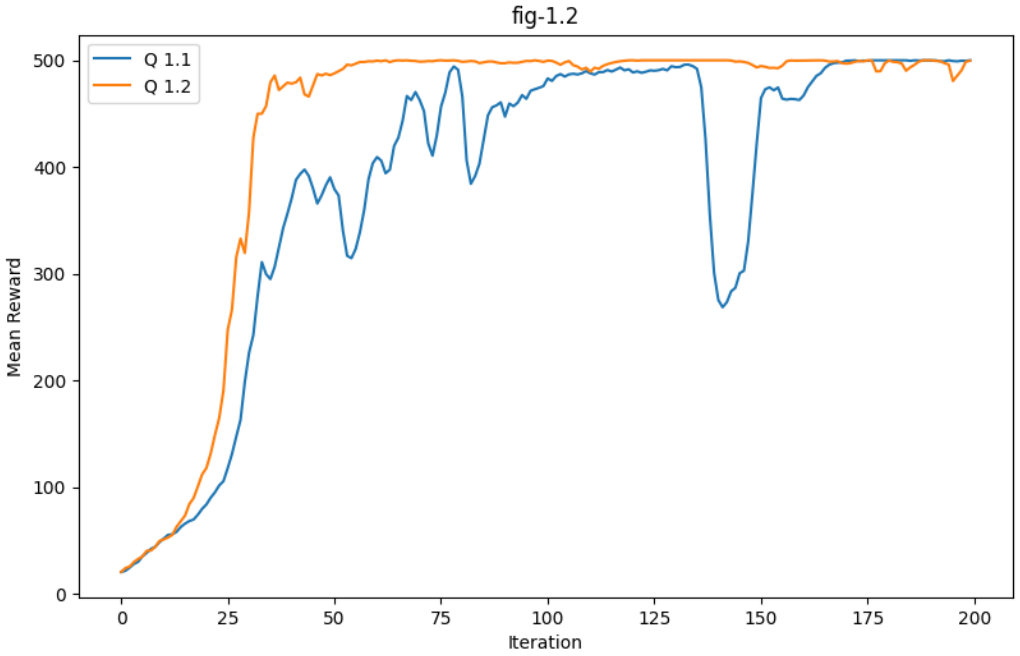
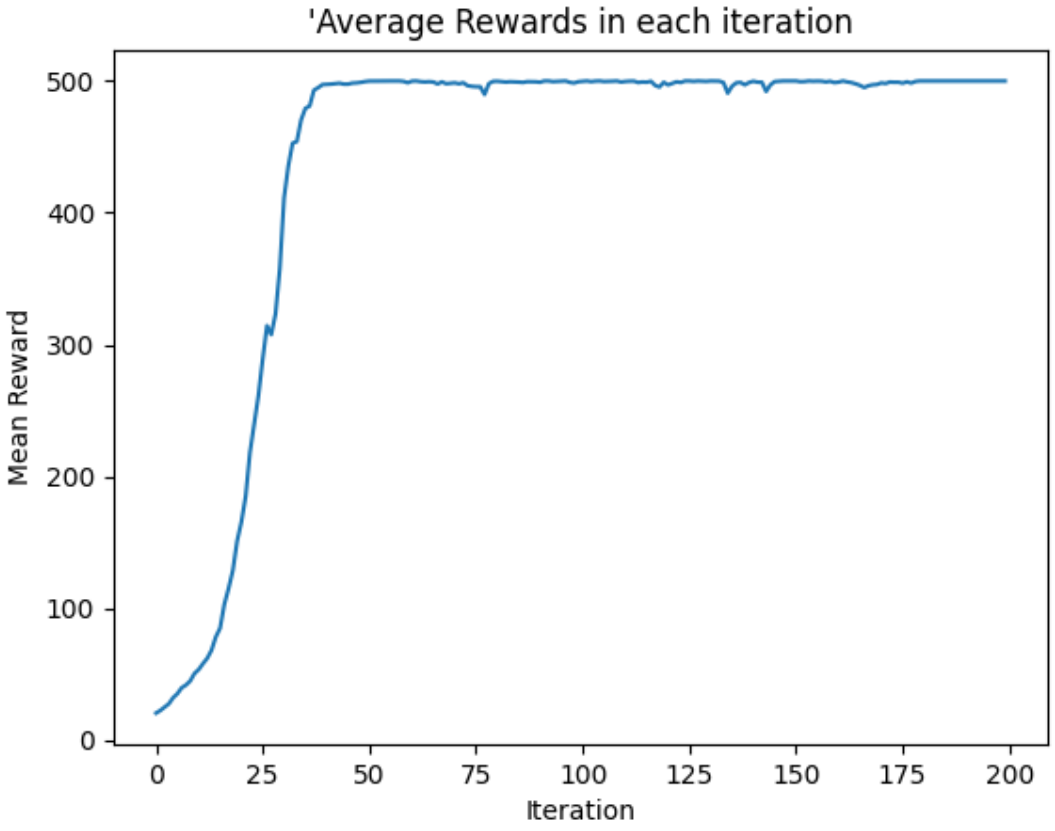


Figure 1
fig-Q2
Mean Reward
Iteration
0
100
200
300
400
500
0
25
50
75
100
125
150
175
200
x=197.5 y=141.6



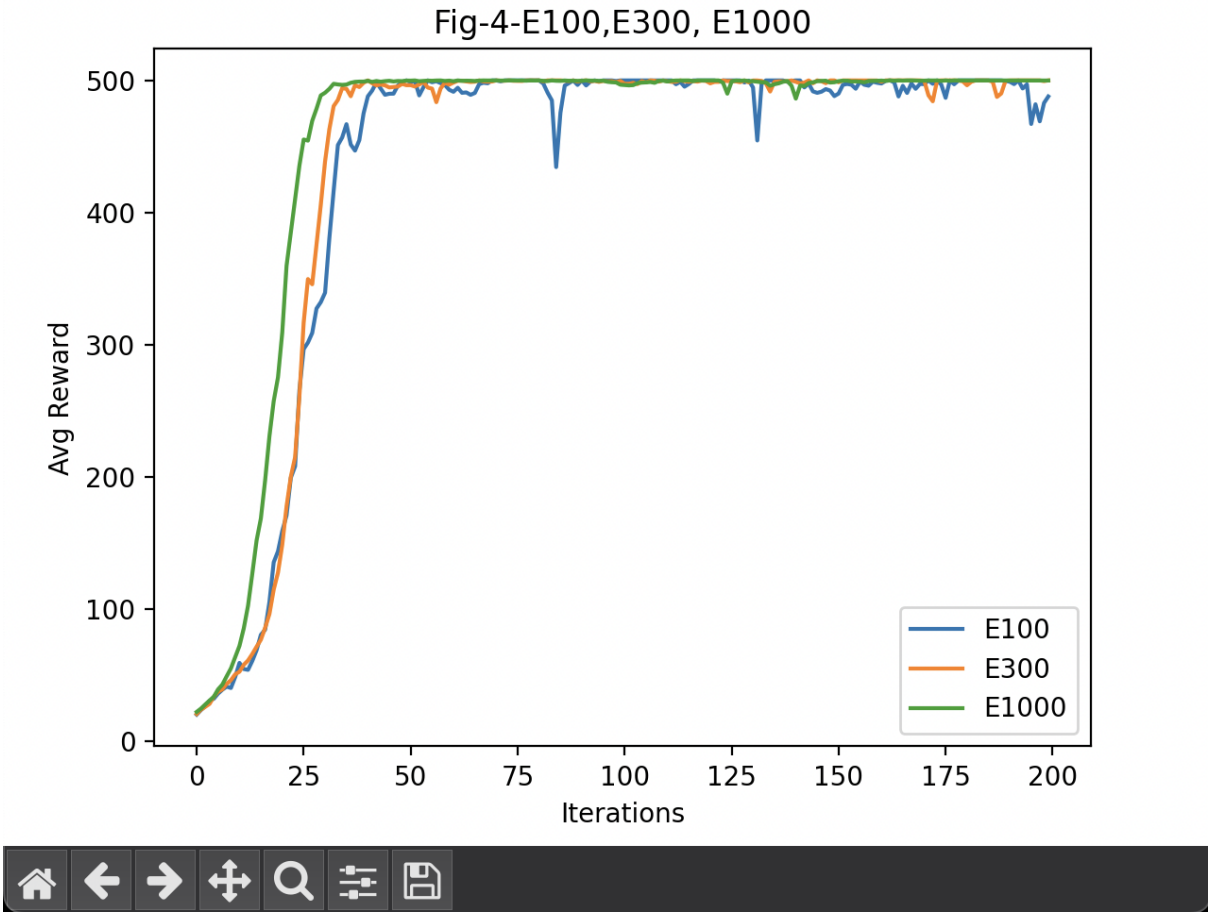
Q 1.3
EP 500



The Results here show a steeper and smoother graph than that of Q1.1 and Q1.2

This is expected since we are normalising the future rewards.

Q 1.4



EP 100

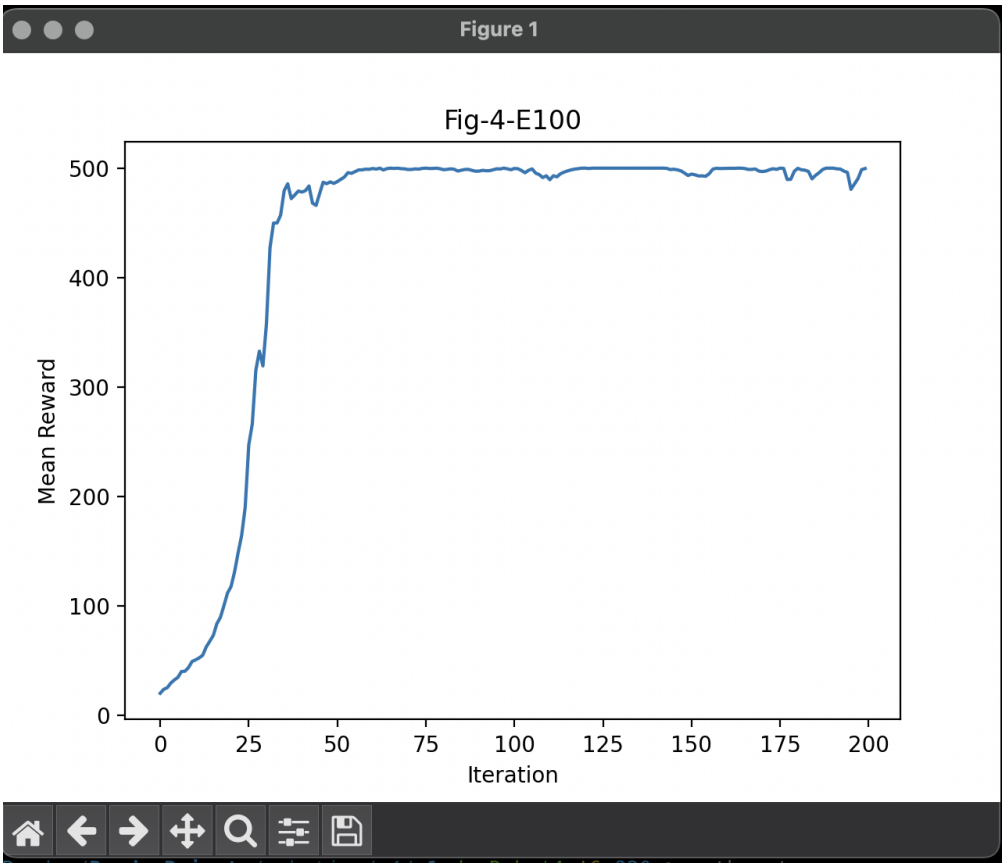


Figure 1
Fig-4-E100
Mean Reward
Iteration
0
100
200
300
400
500
0
25
50
75
100
125
150
175
200

E300

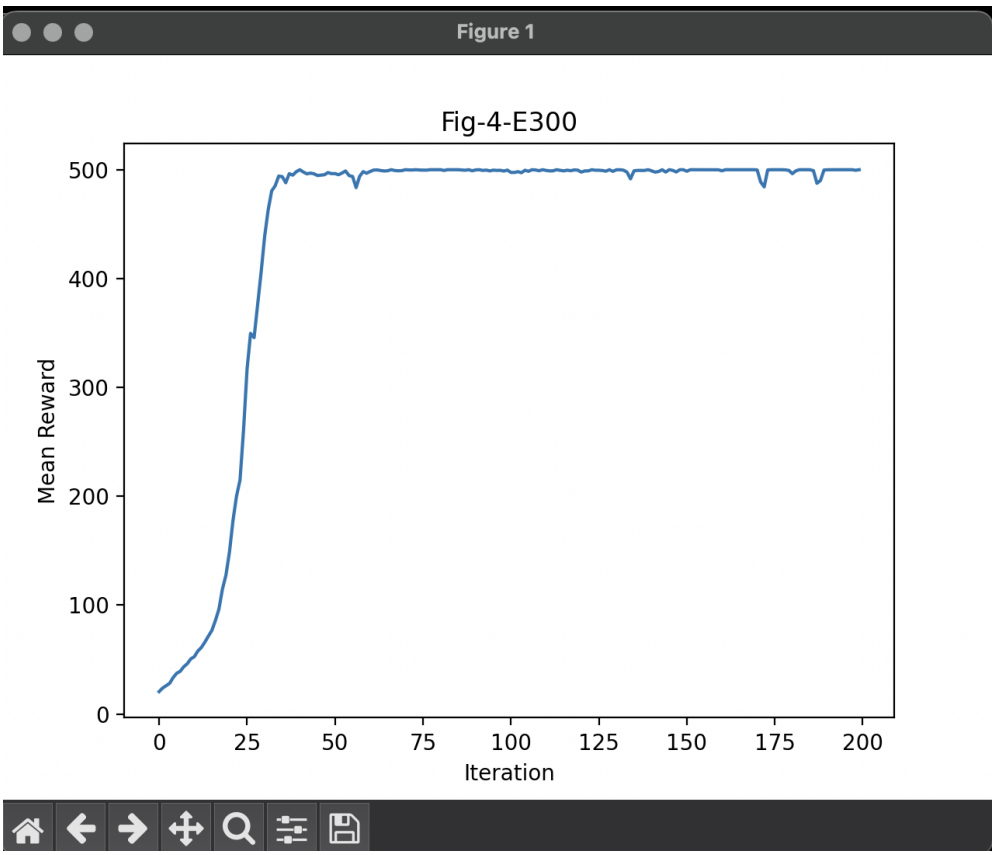
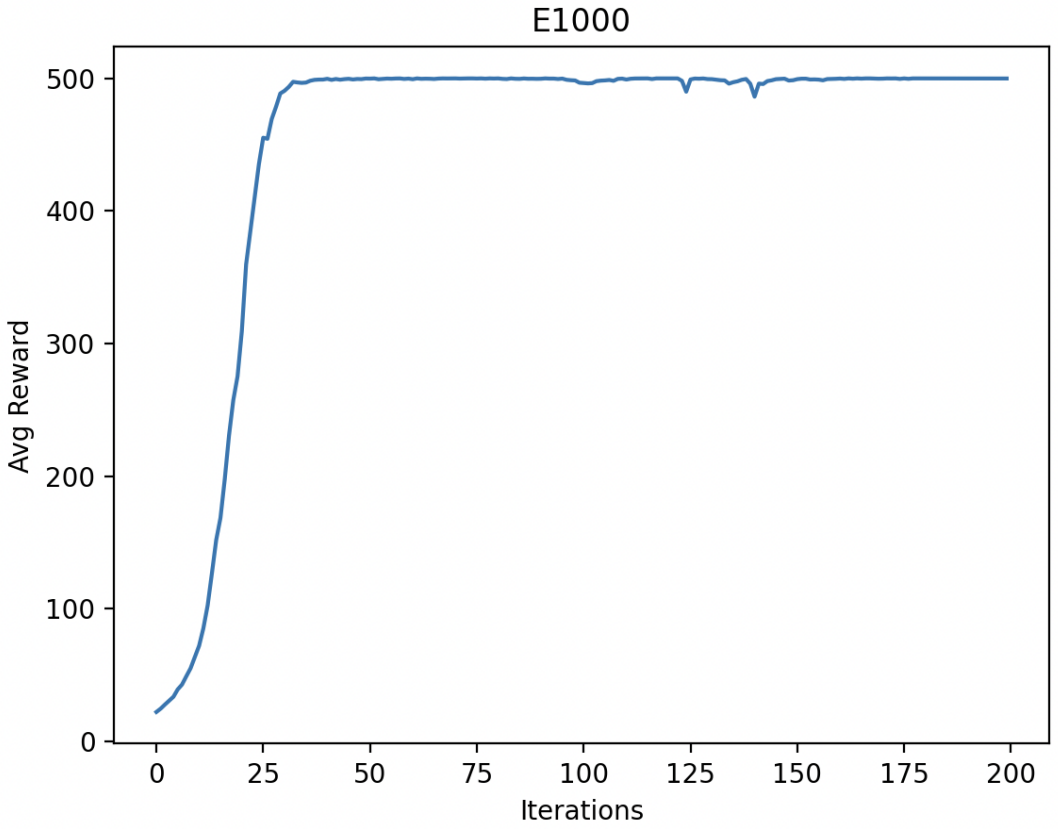
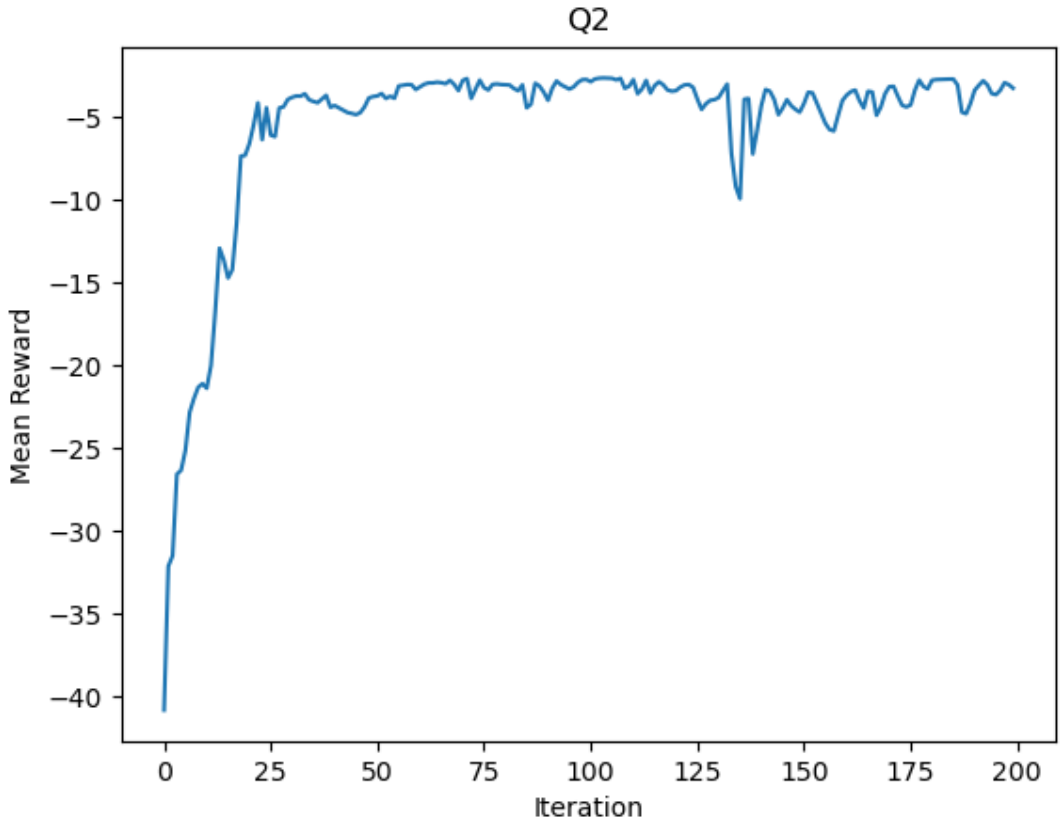


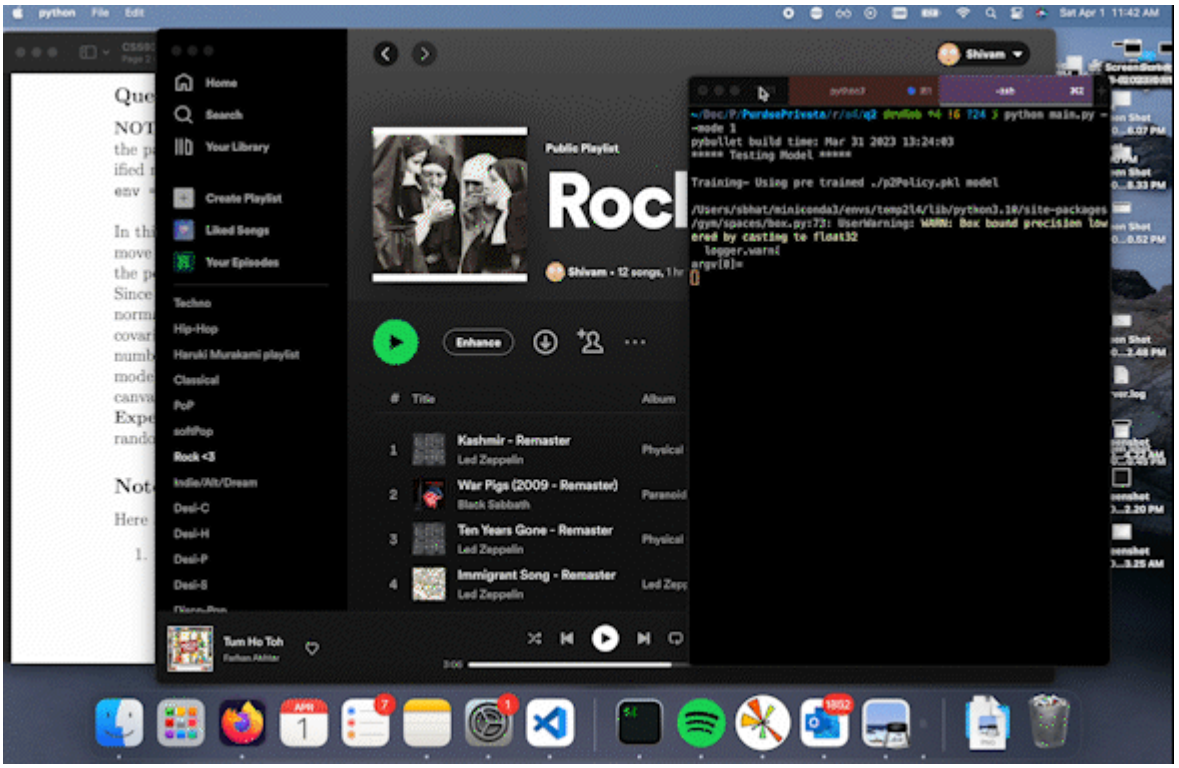
Figure 1
Fig-4-E300
Mean Reward
Iteration
0
100
200
300
400
500
0
25
50
75
100
125
150
175
200

EP 1000


E1000
Avg Reward
Iterations
0
100
200
300
400
500
0
25
50
75
100
125
150
175
200

Q2





I ran it for 200 iterations 500 episodes to get the best reward graph with loss reaching the required threshold of -3.5 as mentioned in https://piazza.com/class/lcoyhxjko165ul/post/112 .
I used my Q 1.2's policy since Q1.3 was still a work in progress and was taking time, so I went with visualising using policy two to complete Q2.

With 300 episodes the model wasn't as accurate. But did cross -3.5 but was not as smooth

If I run it with more iterations=2000 then episodes came down ~ 64. However the question did not state the iteration count so I went with 200 only.