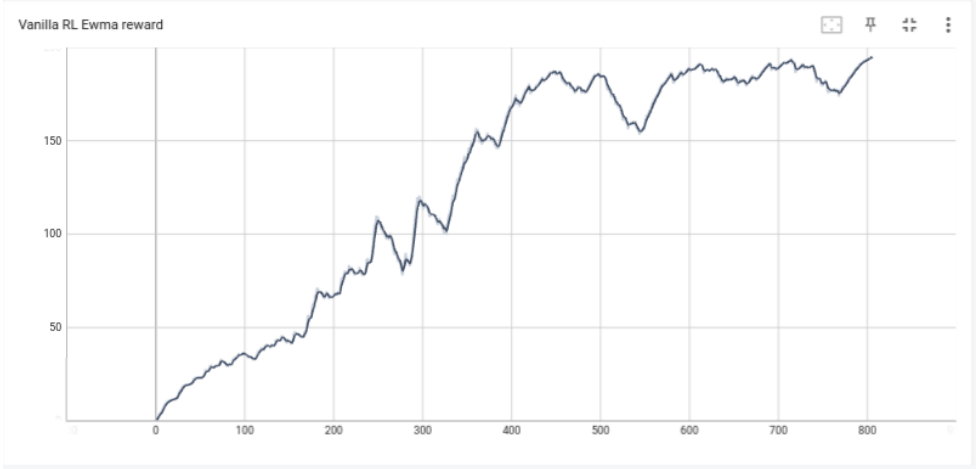
REINFORCEMENT HOMEWORK1

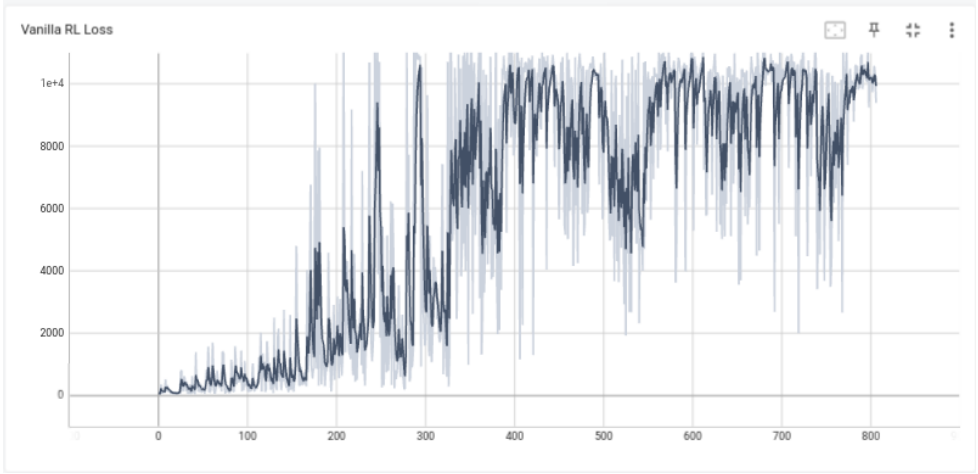
Problem 4 (Policy Gradient Algorithms With Function Approximation)

A

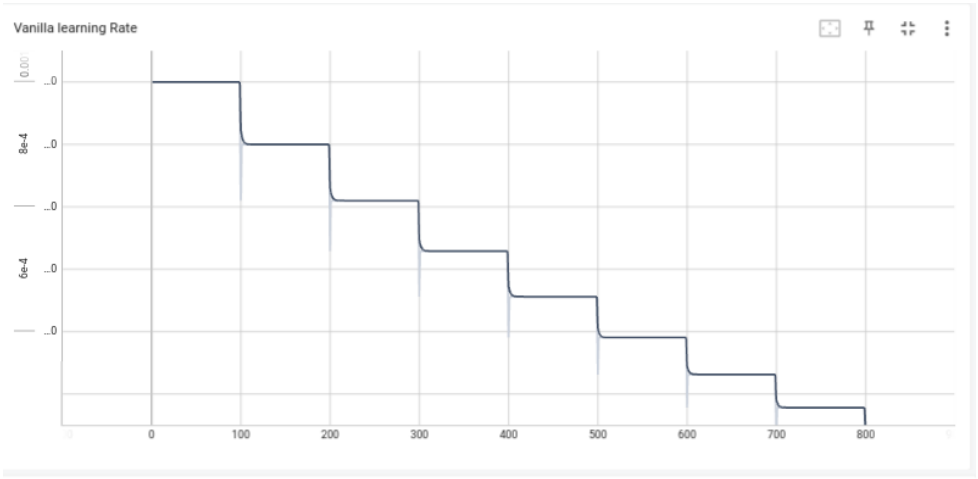
- Learning Rate (lr) = 0.001
- gamma=0.99
- scheduler optimizer = step:100, gamma:0.9
- 執行結果

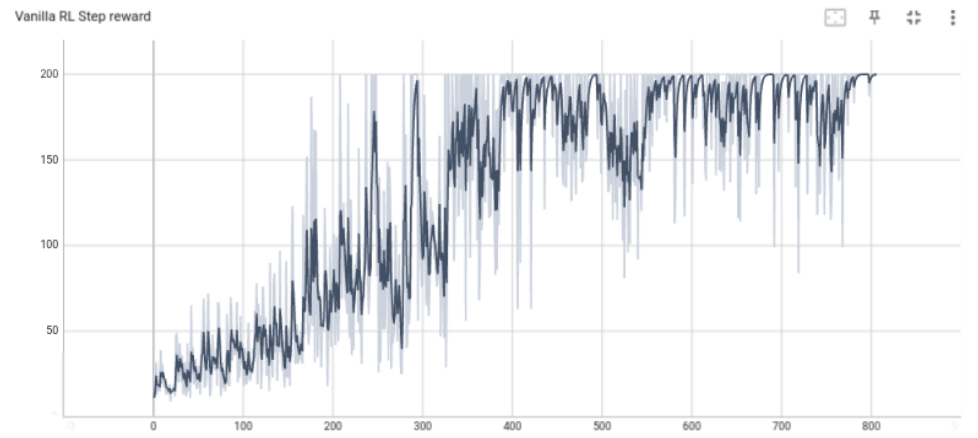
```
Episode 806     length: 199     reward: 200.0     ewma reward: 195.1805072108272
Solved! Running reward is now 195.1805072108272 and the last episode runs to 199 time steps!
/home/qmao/anaconda3/lib/python3.10/site-packages/gym/core.py:49: DeprecationWarning: WARN: You
```





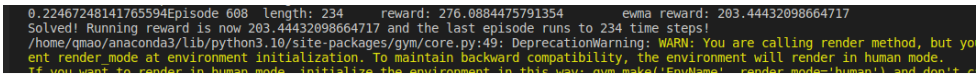
隱私權 - 條款

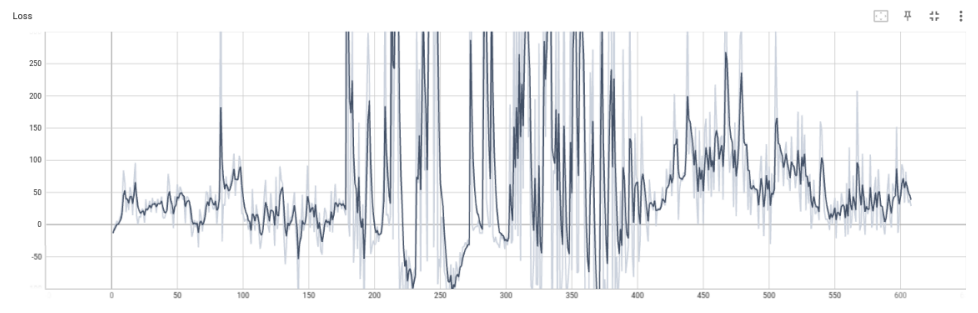


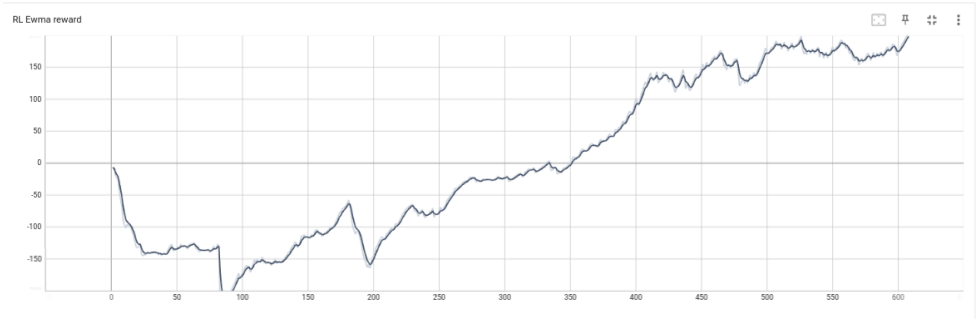


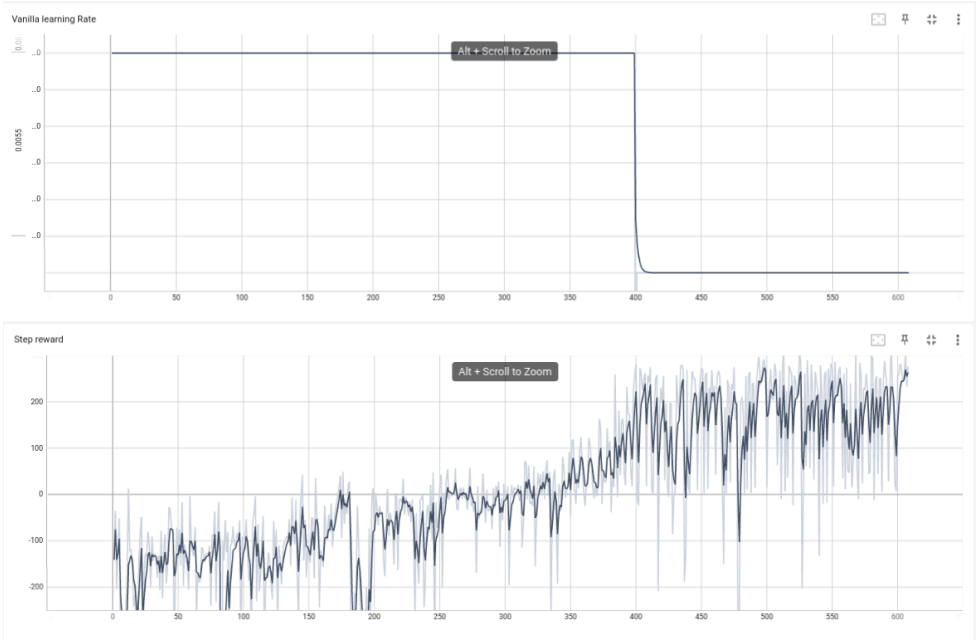
B

- Learning Rate (lr) = 0.005
- gamma=0.99
- scheduler optimizer = step:400, gamma:0.9
- huber_loss = delta =0.5
- 執行結果









C

- Learning Rate (lr) = 0.006
- gamma=0.99
- scheduler optimizer = step:400, gamma:0.9
- huber_loss = delta =0.5
- 執行 2000 epoch，可是無法收斂。