# REINFORCEMENT HOMEWORK2

## Problem 3 (Deep Deterministic Policy Gradient for Continuous Control)

### (A) Pendulum-v1

#### 1. Hyperparameter

- num_episodes = 200,
- gamma = 0.99,
- tau = 0.005,
- hidden_size = 128,
- noise_scale = 0.3,
- replay_size = 100000,
- batch_size = 128,
- updates_per_step =4,
- print_freq = 20,
- lr_a=3e-4,
- lr_c=1e-3,

#### 2. Actor NN layer

- Linear(State_input , hidden_size)
- Relu
- Linear(hidden_size , hidden_size)
- Relu
- Linear(hidden_size , action_space)
- Relu
- Tanh()

#### 3. Critic NN layer

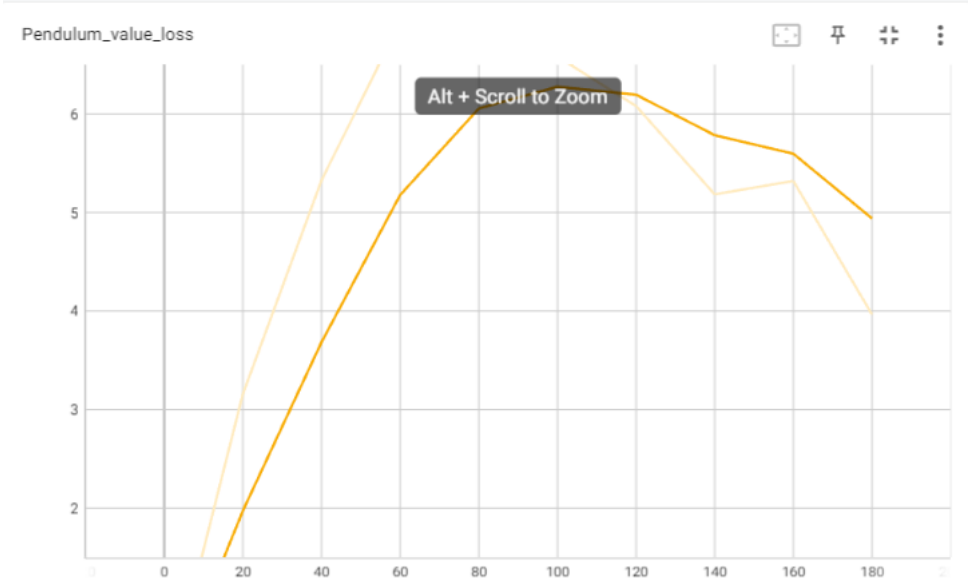
- Linear(State_input , hidden_size)
- Relu
- Linear(hidden_size + action_space , hidden_size)
- Relu

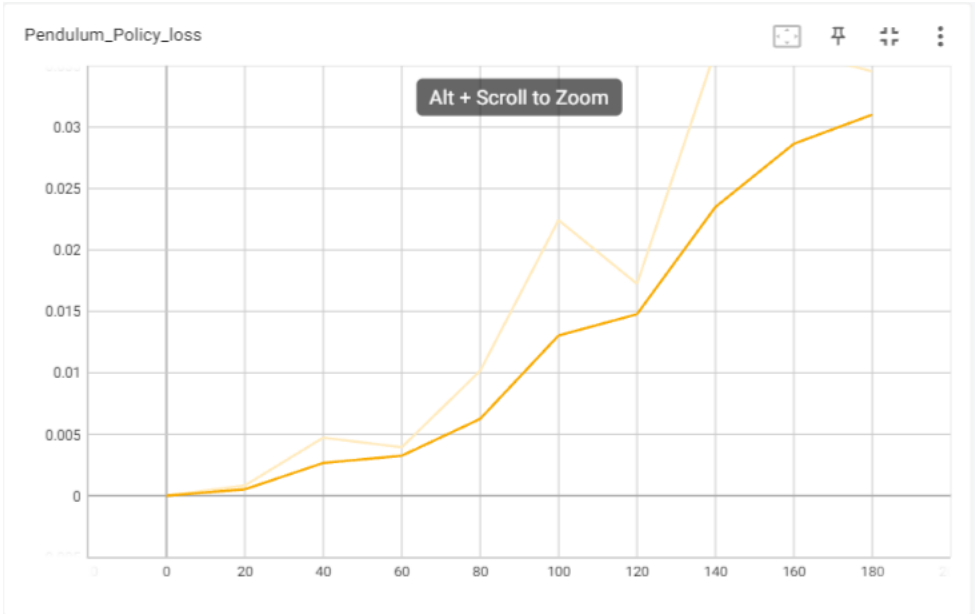
- Linear(hidden_size , 1)

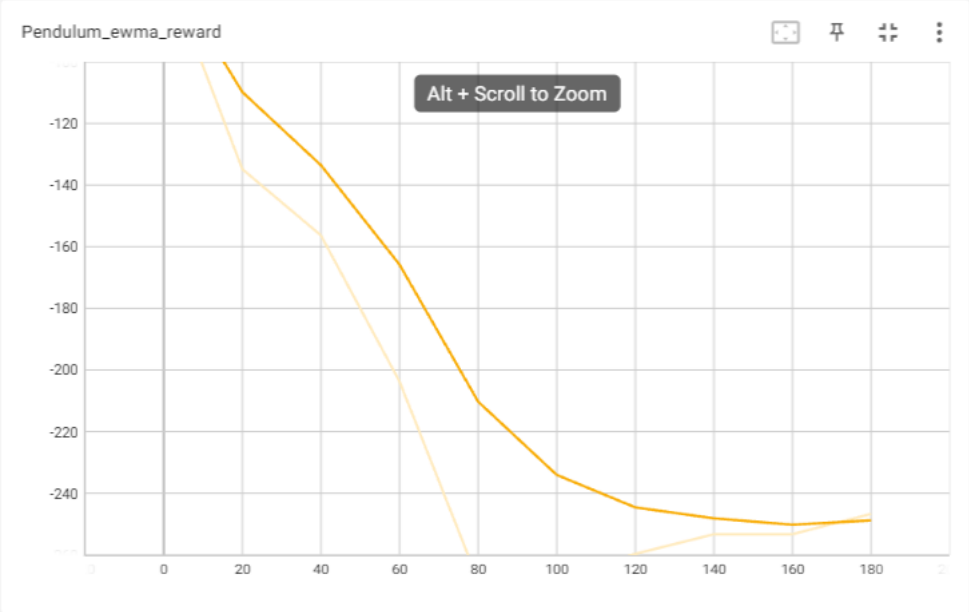
## 4. Result ( Training Success )

Reward = -117.02

```
Episode: 0, length: 200, reward: -1362.40, ewma reward: -68.12
Episode: 20, length: 200, reward: -1434.65, ewma reward: -136.45
Episode: 40, length: 200, reward: -808.52, ewma reward: -170.05
Episode: 60, length: 200, reward: -1407.06, ewma reward: -231.90
Episode: 80, length: 200, reward: -787.99, ewma reward: -259.71
Episode: 100, length: 200, reward: -250.73, ewma reward: -259.26
Episode: 120, length: 200, reward: -123.31, ewma reward: -252.46
Episode: 140, length: 200, reward: -135.12, ewma reward: -246.59
Episode: 160, length: 200, reward: -262.31, ewma reward: -247.38
Episode: 180, length: 200, reward: -117.02, ewma reward: -240.86
```






## (B) LunarLanderContinuous-v2

### 1. Hyperparameter

- num_episodes = 2000,
- gamma = 0.99,
- tau = 0.005,
- hidden_size = 128,
- noise_scale = 0.3,
- replay_size = 50000,
- batch_size = 128,
- updates_per_step =2,
- print_freq = 20,
- lr_a=3e-4,
- lr_c=3e-4,

### 2. Actor NN layer

- Linear(State_input , hidden_size)
- Relu
- Linear(hidden_size , hidden_size)
- Relu
- Linear(hidden_size , action_space)
- Relu
- Tanh()

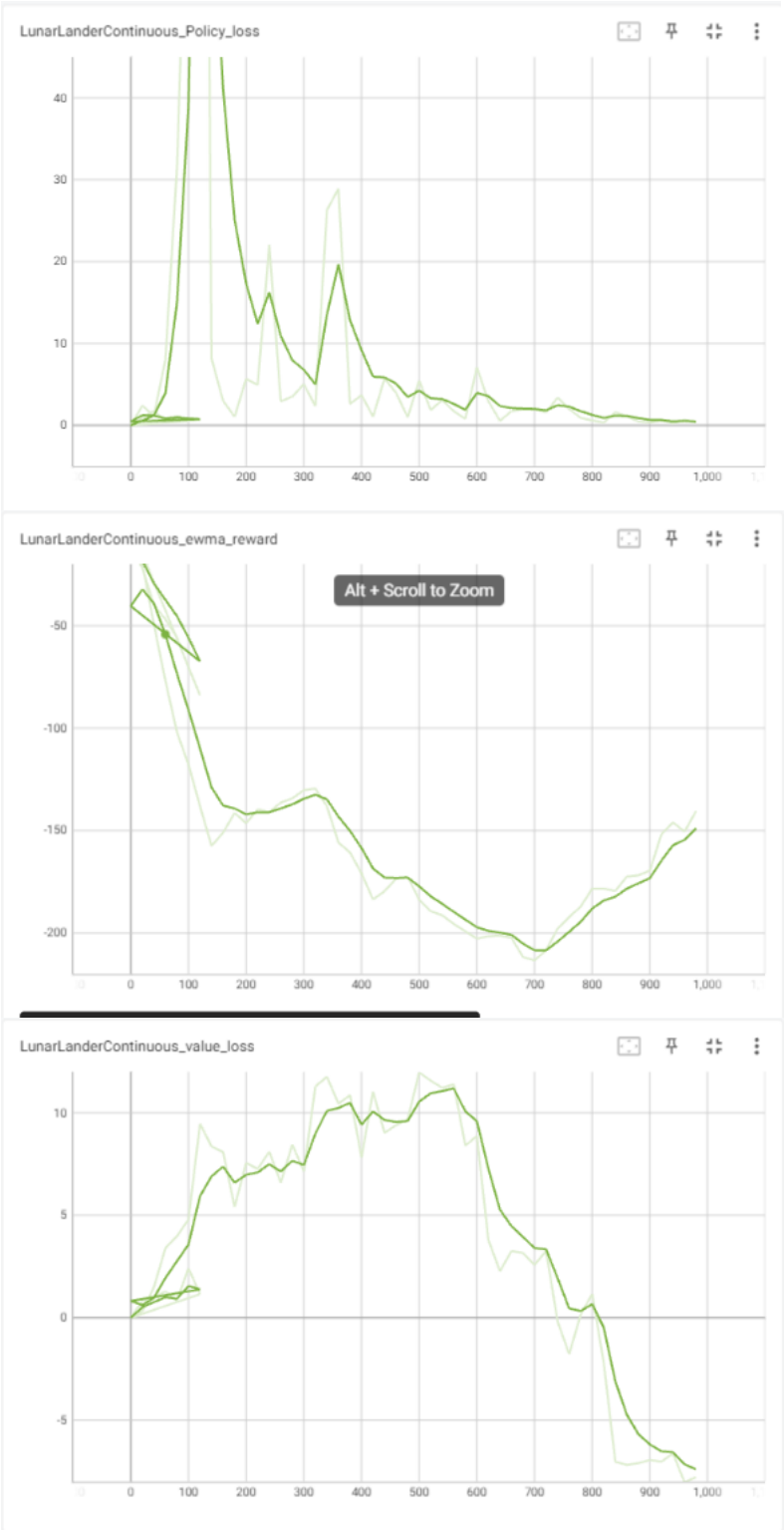
### 3. Critic NN layer

- Linear(State_input , hidden_size)
- Relu
- Linear(hidden_size + action_space , hidden_size)
- Relu
- Linear(hidden_size , 1)

## Result ( Training Fail )

Reward = -545.11

```
Episode: 0, length: 111, reward: 29.10, ewma reward: 1.45
Episode: 20, length: 110, reward: -680.77, ewma reward: -32.66
Episode: 40, length: 239, reward: -419.33, ewma reward: -51.99
Episode: 60, length: 234, reward: -124.08, ewma reward: -55.59
Episode: 80, length: 370, reward: -545.71, ewma reward: -80.10
Episode: 100, length: 453, reward: -214.07, ewma reward: -86.80
Episode: 120, length: 1000, reward: -72.62, ewma reward: -86.09
Episode: 140, length: 91, reward: -17.01, ewma reward: -82.64
Episode: 160, length: 133, reward: -192.66, ewma reward: -88.14
Episode: 180, length: 99, reward: -27.01, ewma reward: -85.08
Episode: 200, length: 78, reward: -51.02, ewma reward: -83.38
Episode: 220, length: 1000, reward: -68.73, ewma reward: -82.65
Episode: 240, length: 122, reward: -128.59, ewma reward: -84.94
Episode: 260, length: 69, reward: 1.79, ewma reward: -80.61
Episode: 280, length: 691, reward: -196.00, ewma reward: -86.38
Episode: 300, length: 126, reward: -289.58, ewma reward: -96.54
Episode: 320, length: 1000, reward: -17.00, ewma reward: -92.56
Episode: 340, length: 188, reward: 203.24, ewma reward: -77.77
Episode: 360, length: 300, reward: 218.98, ewma reward: -62.93
Episode: 380, length: 230, reward: -46.96, ewma reward: -62.13
Episode: 400, length: 197, reward: -151.26, ewma reward: -66.59
Episode: 420, length: 87, reward: -135.97, ewma reward: -70.06
Episode: 440, length: 77, reward: -123.42, ewma reward: -72.73
Episode: 460, length: 345, reward: 228.86, ewma reward: -57.65
Episode: 480, length: 140, reward: 46.45, ewma reward: -52.44
...
Episode: 1940, length: 82, reward: -767.20, ewma reward: -597.58
Episode: 1960, length: 71, reward: -642.45, ewma reward: -599.82
Episode: 1980, length: 67, reward: -545.11, ewma reward: -597.09
```


LunarLanderContinuous_Policy_loss
40
30
20
10
0
0
100
200
300
400
500
600
700
800
900
1,000
LunarLanderContinuous_ewma_reward
Alt + Scroll to Zoom
-50
-100
-150
-200
0
100
200
300
400
500
600
700
800
900
1,000
LunarLanderContinuous_value_loss
10
5
0
-5
0
100
200
300
400
500
600
700
800
900
1,000