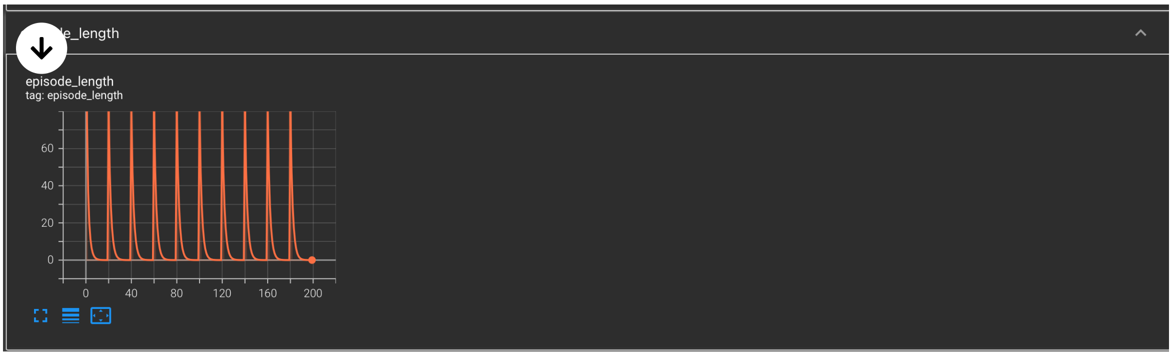
tags: `2024 年 下學期讀書計畫` `Reinforcement Learning`

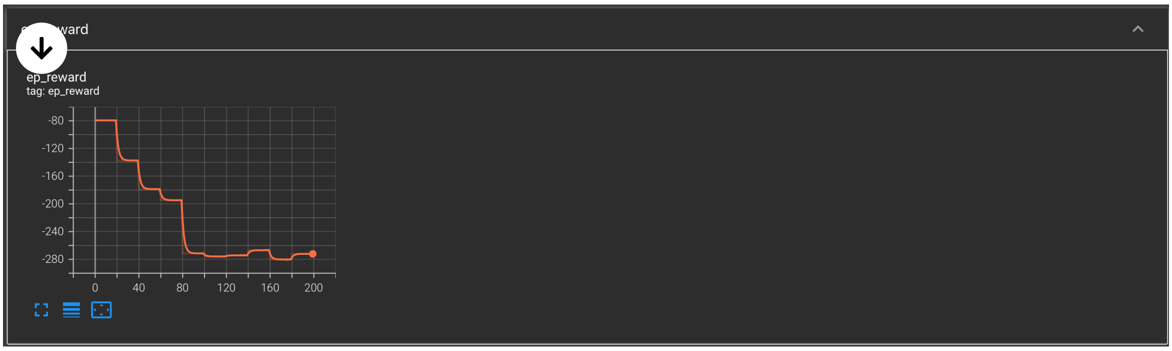
RL Homework 3: DDPG, TRPO, and PPO

a

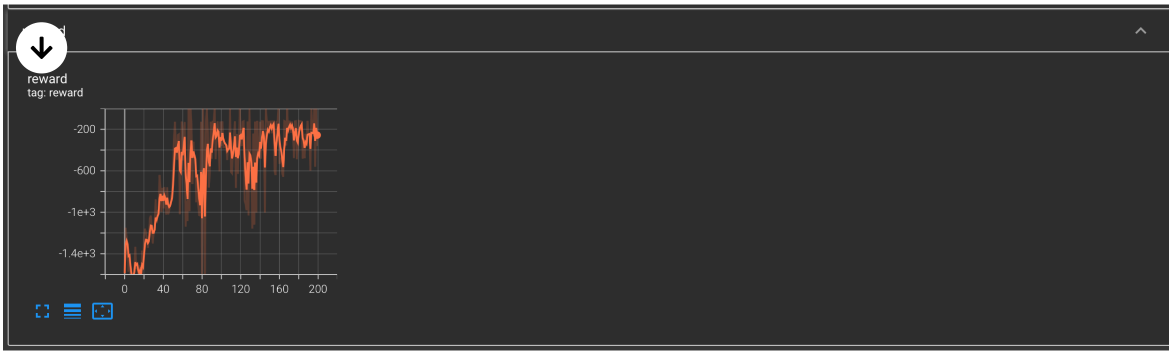
ep_len



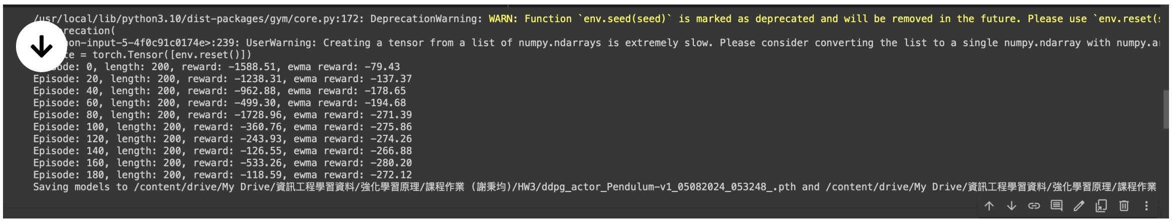
ep_reward



reward

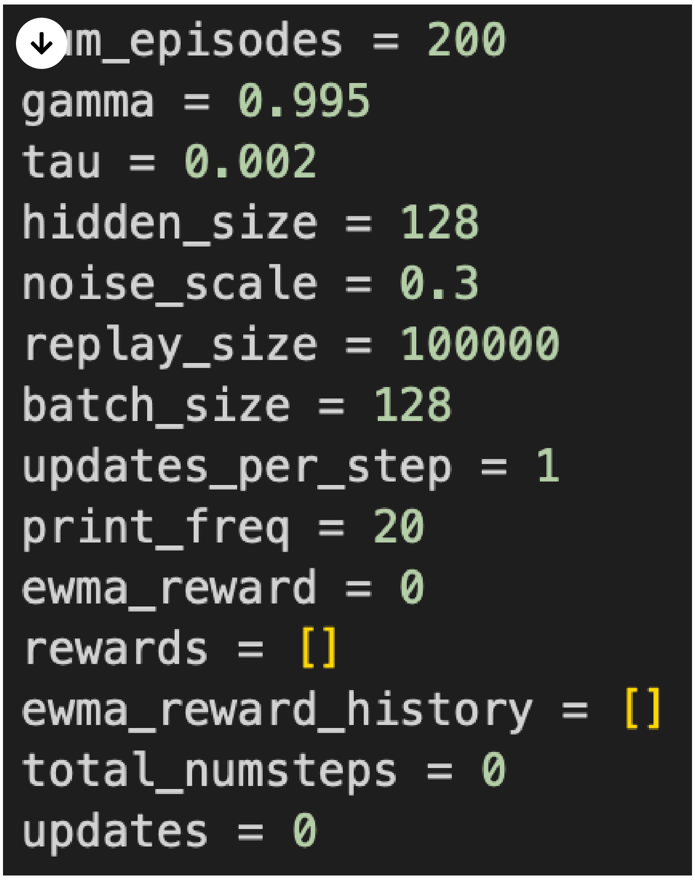


train



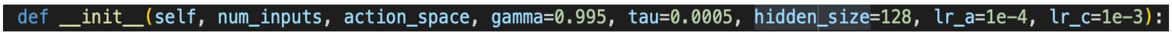
```
/usr/local/lib/python3.10/dist-packages/gym/core.py:172: DeprecationWarning: WARN: Function `env.seed(seed)` is marked as deprecated and will be removed in the future. Please use `env.reset(s
  recation(
  on-input-5-4f0c91c0174e>:239: UserWarning: Creating a tensor from a list of numpy.ndarrays is extremely slow. Please consider converting the list to a single numpy.ndarray with numpy.a
  te = torch.Tensor([env.reset()])
  sode: 0, length: 200, reward: -1588.51, ewma reward: -79.43
Episode: 20, length: 200, reward: -1238.31, ewma reward: -137.37
Episode: 40, length: 200, reward: -962.88, ewma reward: -178.65
Episode: 60, length: 200, reward: -499.30, ewma reward: -194.68
Episode: 80, length: 200, reward: -1728.96, ewma reward: -271.39
Episode: 100, length: 200, reward: -360.76, ewma reward: -275.86
Episode: 120, length: 200, reward: -243.93, ewma reward: -274.26
Episode: 140, length: 200, reward: -126.55, ewma reward: -266.88
Episode: 160, length: 200, reward: -533.26, ewma reward: -280.20
Episode: 180, length: 200, reward: -118.59, ewma reward: -272.12
Saving models to /content/drive/My Drive/資訊工程學習資料/強化學習原理/課程作業 (謝秉均)/HW3/ddpg_actor_Pendulum-v1_05082024_053248_.pth and /content/drive/My Drive/資訊工程學習資料/強化學習原理/課程作業
```

hyperparameters



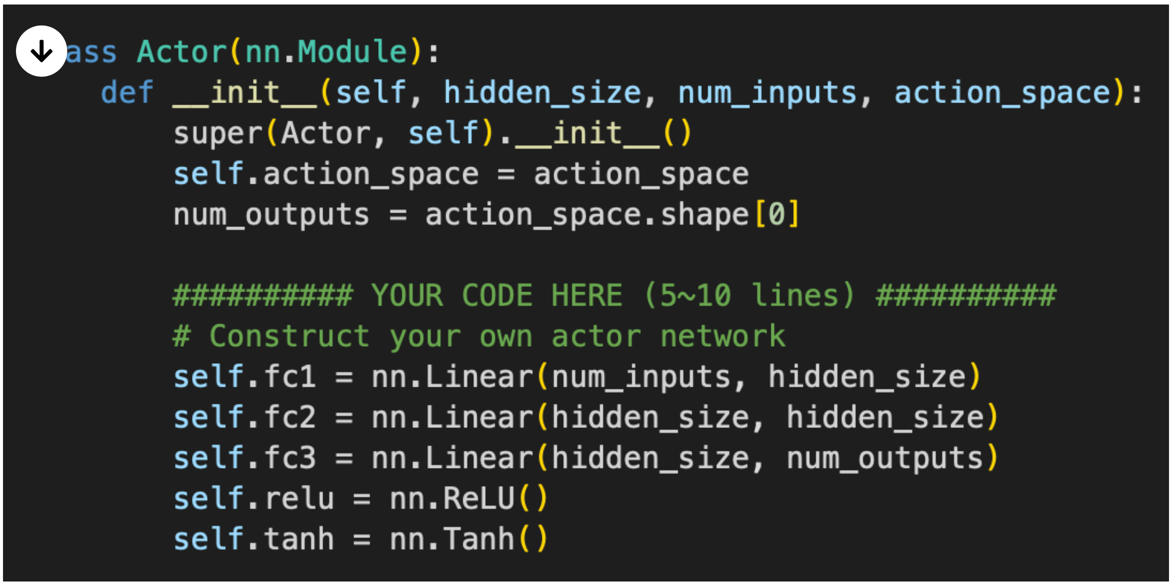
```
um_episodes = 200
gamma = 0.995
tau = 0.002
hidden_size = 128
noise_scale = 0.3
replay_size = 100000
batch_size = 128
updates_per_step = 1
print_freq = 20
ewma_reward = 0
rewards = []
ewma_reward_history = []
total_numsteps = 0
updates = 0
```

learning rates

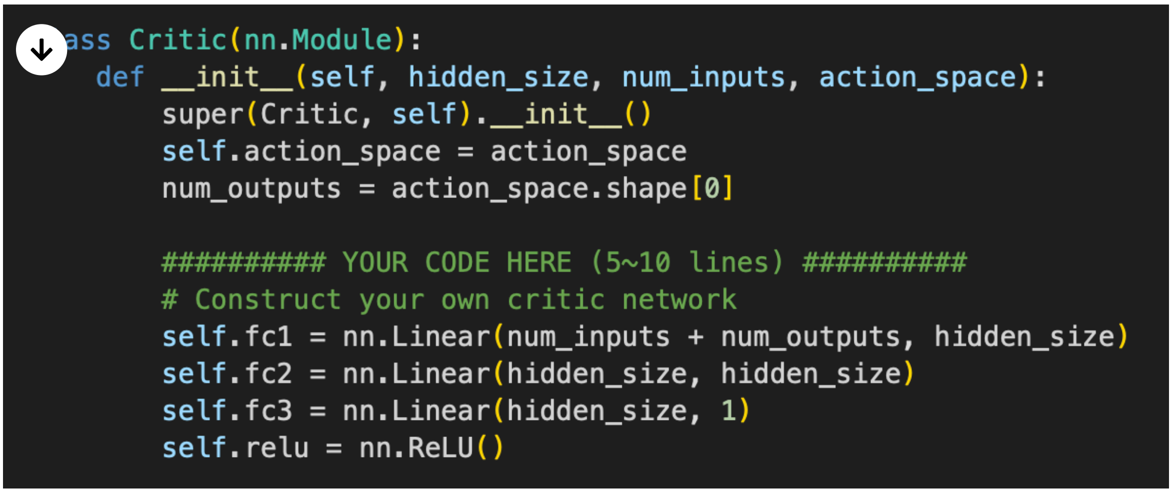


```
def __init__(self, num_inputs, action_space, gamma=0.995, tau=0.0005, hidden_size=128, lr_a=1e-4, lr_c=1e-3):
```

NN architecture



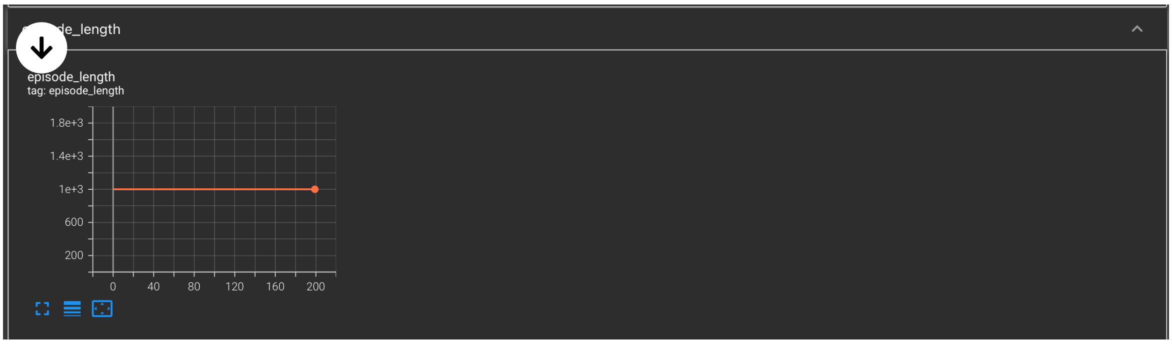
```
ass Actor(nn.Module):
    def __init__(self, hidden_size, num_inputs, action_space):
        super(Actor, self).__init__()
        self.action_space = action_space
        num_outputs = action_space.shape[0]

        ########## YOUR CODE HERE (5~10 lines) ##########
        # Construct your own actor network
        self.fc1 = nn.Linear(num_inputs, hidden_size)
        self.fc2 = nn.Linear(hidden_size, hidden_size)
        self.fc3 = nn.Linear(hidden_size, num_outputs)
        self.relu = nn.ReLU()
        self.tanh = nn.Tanh()
```



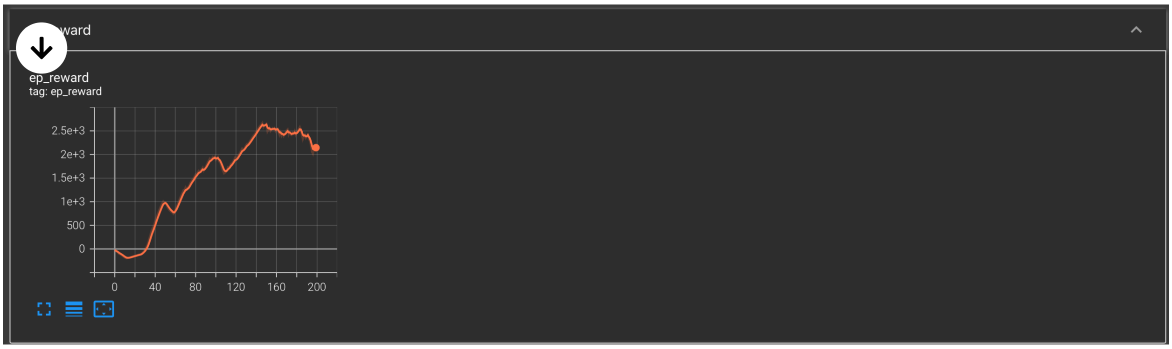
```
ass Critic(nn.Module):
    def __init__(self, hidden_size, num_inputs, action_space):
        super(Critic, self).__init__()
        self.action_space = action_space
        num_outputs = action_space.shape[0]

        ########## YOUR CODE HERE (5~10 lines) ##########
        # Construct your own critic network
        self.fc1 = nn.Linear(num_inputs + num_outputs, hidden_size)
        self.fc2 = nn.Linear(hidden_size, hidden_size)
        self.fc3 = nn.Linear(hidden_size, 1)
        self.relu = nn.ReLU()
```

b

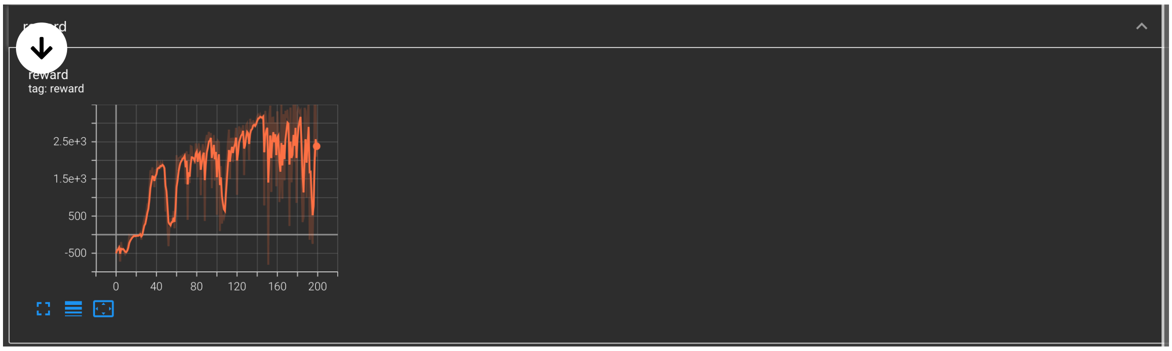
ep_len



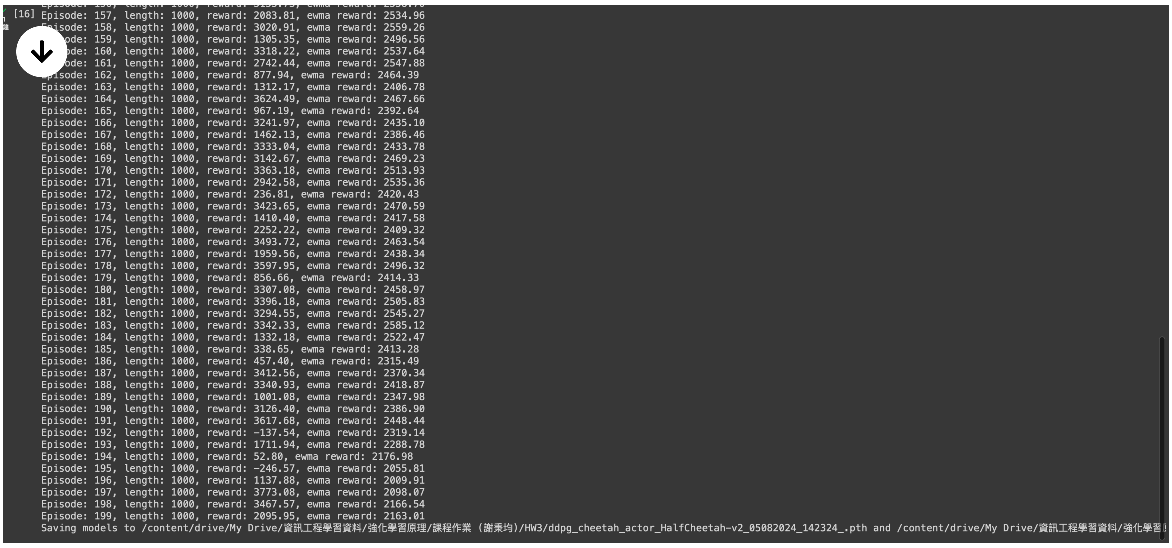
ep_reward



reward

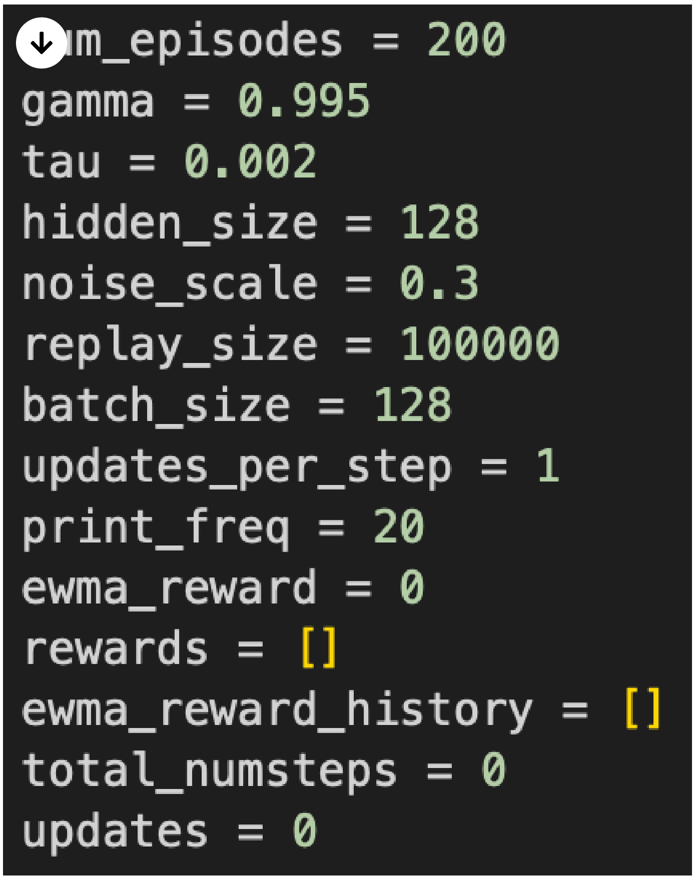


train



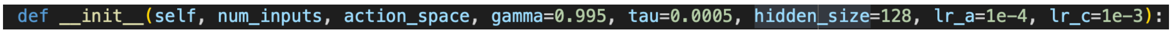
```
[16] Episode: 157, length: 1000, reward: 2083.81, ewma reward: 2534.96
     Episode: 158, length: 1000, reward: 3020.91, ewma reward: 2559.26
     Episode: 159, length: 1000, reward: 1305.35, ewma reward: 2496.56
     Episode: 160, length: 1000, reward: 3318.22, ewma reward: 2537.64
     Episode: 161, length: 1000, reward: 2742.44, ewma reward: 2547.88
     Episode: 162, length: 1000, reward: 877.94, ewma reward: 2464.39
     Episode: 163, length: 1000, reward: 1312.17, ewma reward: 2406.78
     Episode: 164, length: 1000, reward: 3624.49, ewma reward: 2467.66
     Episode: 165, length: 1000, reward: 967.19, ewma reward: 2392.64
     Episode: 166, length: 1000, reward: 3241.97, ewma reward: 2435.10
     Episode: 167, length: 1000, reward: 1462.13, ewma reward: 2386.46
     Episode: 168, length: 1000, reward: 3333.04, ewma reward: 2433.78
     Episode: 169, length: 1000, reward: 3142.67, ewma reward: 2469.23
     Episode: 170, length: 1000, reward: 3363.18, ewma reward: 2513.93
     Episode: 171, length: 1000, reward: 2942.58, ewma reward: 2535.36
     Episode: 172, length: 1000, reward: 236.81, ewma reward: 2420.43
     Episode: 173, length: 1000, reward: 3423.65, ewma reward: 2470.59
     Episode: 174, length: 1000, reward: 1410.40, ewma reward: 2417.58
     Episode: 175, length: 1000, reward: 2252.22, ewma reward: 2409.32
     Episode: 176, length: 1000, reward: 3493.72, ewma reward: 2463.54
     Episode: 177, length: 1000, reward: 1959.56, ewma reward: 2438.34
     Episode: 178, length: 1000, reward: 3597.95, ewma reward: 2496.32
     Episode: 179, length: 1000, reward: 856.66, ewma reward: 2414.33
     Episode: 180, length: 1000, reward: 3307.08, ewma reward: 2458.97
     Episode: 181, length: 1000, reward: 3396.18, ewma reward: 2505.83
     Episode: 182, length: 1000, reward: 3294.55, ewma reward: 2545.27
     Episode: 183, length: 1000, reward: 3342.33, ewma reward: 2585.12
     Episode: 184, length: 1000, reward: 1332.18, ewma reward: 2522.47
     Episode: 185, length: 1000, reward: 338.65, ewma reward: 2413.28
     Episode: 186, length: 1000, reward: 457.40, ewma reward: 2315.49
     Episode: 187, length: 1000, reward: 3412.56, ewma reward: 2370.34
     Episode: 188, length: 1000, reward: 3340.93, ewma reward: 2418.87
     Episode: 189, length: 1000, reward: 1001.08, ewma reward: 2347.98
     Episode: 190, length: 1000, reward: 3126.40, ewma reward: 2386.90
     Episode: 191, length: 1000, reward: 3617.68, ewma reward: 2448.44
     Episode: 192, length: 1000, reward: -137.54, ewma reward: 2319.14
     Episode: 193, length: 1000, reward: 1711.94, ewma reward: 2288.78
     Episode: 194, length: 1000, reward: 52.80, ewma reward: 2176.98
     Episode: 195, length: 1000, reward: -246.57, ewma reward: 2055.81
     Episode: 196, length: 1000, reward: 1137.88, ewma reward: 2009.91
     Episode: 197, length: 1000, reward: 3773.08, ewma reward: 2098.07
     Episode: 198, length: 1000, reward: 3467.57, ewma reward: 2166.54
     Episode: 199, length: 1000, reward: 2095.95, ewma reward: 2163.01
     Saving models to /content/drive/My Drive/資訊工程學習資料/強化學習原理/課程作業 (謝秉均)/HW3/ddpg_cheetah_actor_HalfCheetah-v2_05082024_142324_.pth and /content/drive/My Drive/資訊工程學習資料/強化學習
```

hyperparameters



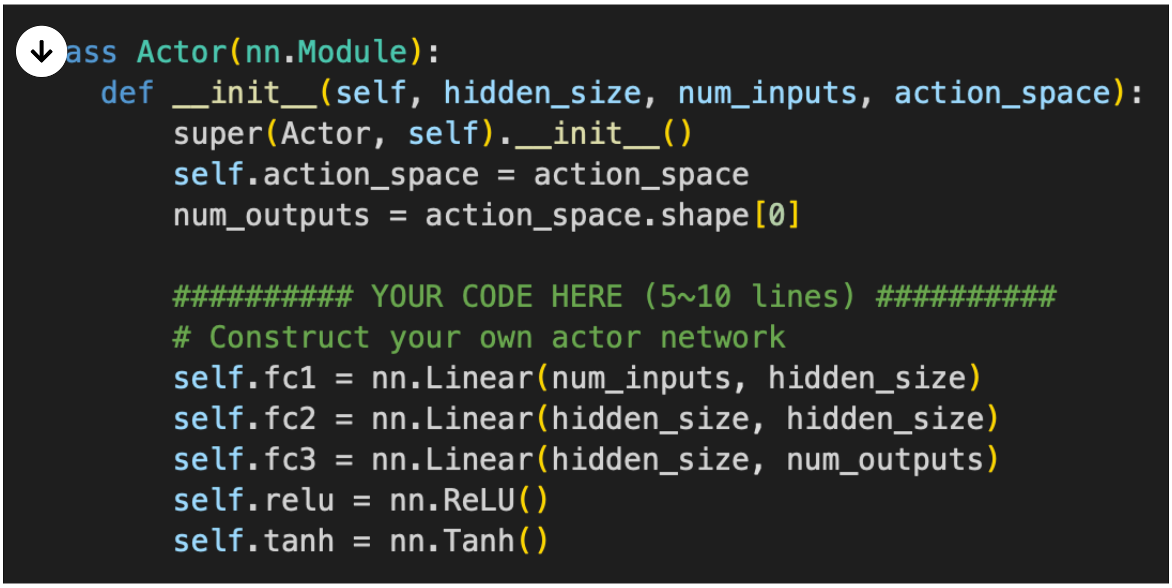
```
num_episodes = 200
gamma = 0.995
tau = 0.002
hidden_size = 128
noise_scale = 0.3
replay_size = 100000
batch_size = 128
updates_per_step = 1
print_freq = 20
ewma_reward = 0
rewards = []
ewma_reward_history = []
total_numsteps = 0
updates = 0
```

learning rates

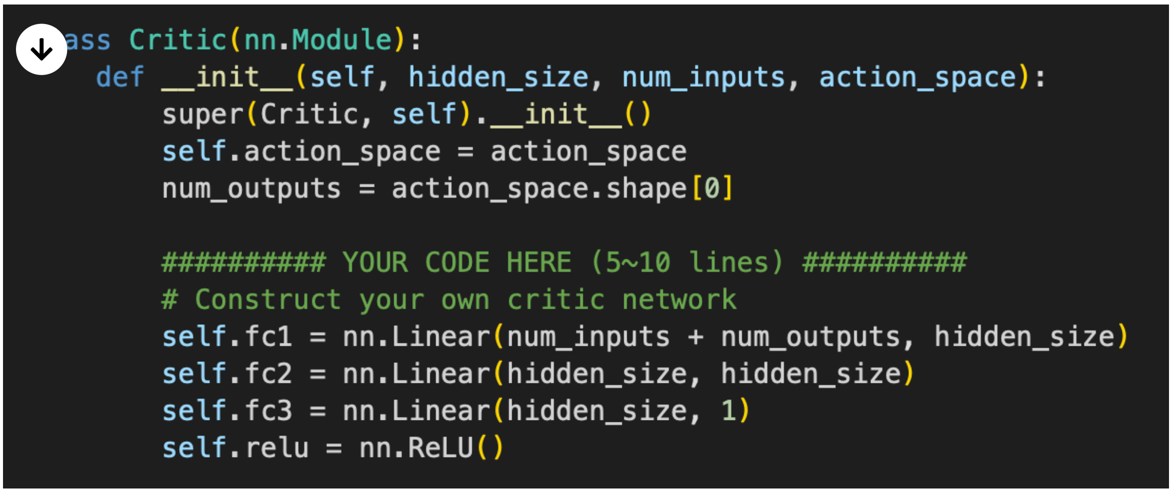


```
def __init__(self, num_inputs, action_space, gamma=0.995, tau=0.0005, hidden_size=128, lr_a=1e-4, lr_c=1e-3):
```

NN architecture



```
ass Actor(nn.Module):
    def __init__(self, hidden_size, num_inputs, action_space):
        super(Actor, self).__init__()
        self.action_space = action_space
        num_outputs = action_space.shape[0]

        ########## YOUR CODE HERE (5~10 lines) ##########
        # Construct your own actor network
        self.fc1 = nn.Linear(num_inputs, hidden_size)
        self.fc2 = nn.Linear(hidden_size, hidden_size)
        self.fc3 = nn.Linear(hidden_size, num_outputs)
        self.relu = nn.ReLU()
        self.tanh = nn.Tanh()
```



```
ass Critic(nn.Module):
    def __init__(self, hidden_size, num_inputs, action_space):
        super(Critic, self).__init__()
        self.action_space = action_space
        num_outputs = action_space.shape[0]

        ########## YOUR CODE HERE (5~10 lines) ##########
        # Construct your own critic network
        self.fc1 = nn.Linear(num_inputs + num_outputs, hidden_size)
        self.fc2 = nn.Linear(hidden_size, hidden_size)
        self.fc3 = nn.Linear(hidden_size, 1)
        self.relu = nn.ReLU()
```