1,

110652019 林楷傑

(i)

$$L_{\pi_{\theta_1}}(\pi_{\theta_1}) = \eta(\pi_{\theta_1}) + \sum_{s\in S} d_\mu^{\pi_{\theta_1}}(s) \sum_{a\in A} \pi_{\theta_1}(a|s) A^{\pi_{\theta_1}}(s,a) = \eta(\pi_{\theta_1}) + \sum_{s\in S} d_\mu^{\pi_{\theta_1}}(s) \sum_{a\in A} \pi_{\theta_1}(a|s) \left(Q^{\pi_{\theta_1}}(s,a) - V^{\pi_{\theta_1}}(s) \right)$$

$$= \eta(\pi_{\theta_1}) + \sum_{s\in S} d_\mu^{\pi_{\theta_1}}(s) \left(V^{\pi_{\theta_1}}(s) - V^{\pi_{\theta_1}}(s) \right) = \eta(\pi_{\theta_1})$$

(ii)

$$\nabla_\theta L_{\pi_{\theta_1}}(\pi_\theta)\big|_{\theta=\theta_1} = \nabla_\theta \eta(\pi_\theta) + \sum_{s\in S} d_\mu^{\pi_{\theta_1}}(s) \nabla_\theta \sum_{a\in A} A^{\pi_{\theta_1}}(s,a) \pi_\theta(a|s) \Big|_{\theta=\theta_1}$$

$$= \nabla_{\theta_1} \eta(\pi_{\theta_1}) + \sum_{s\in S} d_\mu^{\pi_{\theta_1}}(s) \nabla_{\theta_1} \left(V^{\pi_{\theta_1}}(s) - V^{\pi_{\theta_1}}(s) \right) = \nabla_{\theta_1} \eta(\pi_{\theta_1})$$

2,

(a)

$$\mathcal{L}(\theta,\lambda) := -\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T(\theta-\theta_k) + \lambda\left(\tfrac{1}{2}(\theta-\theta_k)^T H(\theta-\theta_k) - \delta\right)$$

$$\nabla_\theta \mathcal{L}(\theta,\lambda) = -\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right) + \lambda H(\theta-\theta_k) = 0 \Rightarrow \theta = \theta_k + \frac{1}{\lambda} H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)$$

Let $V = \nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}$

$$D(\lambda) = \min_{\theta\in\mathbb{R}^d} \mathcal{L}(\theta,\lambda) = -V^T\left(\frac{1}{\lambda}H^{-1}V\right) + \frac{\lambda}{2}\left(\frac{1}{\lambda}H^{-1}V\right)^T H \left(\frac{1}{\lambda}H^{-1}V\right) - \lambda\delta$$

$$= -V^T\frac{1}{\lambda}H^{-1}V + \frac{1}{2\lambda}V^T H^{-1} V - \lambda\delta = -\frac{1}{2\lambda} V^T H^{-1} V - \lambda\delta$$

$$\frac{\partial D(\lambda)}{\partial\lambda} = \frac{1}{2\lambda^2} V^T H^{-1} V - \delta = 0 \Rightarrow \lambda^* = \left(\frac{1}{2\delta} V^T H^{-1} V\right)^{\frac{1}{2}}$$

(b) $$\mathcal{L}(\theta,\lambda^*) = -V^T(\theta-\theta_k) + \left(\frac{1}{2\delta} V^T H^{-1} V\right)^{\frac{1}{2}} \left(\frac{1}{2}(\theta-\theta_k^T) H(\theta-\theta_k) - \delta\right)$$

$$\nabla_\theta \mathcal{L}(\theta,\lambda^*) = -V + \left(\frac{1}{2\delta} V^T H^{-1} V\right)^{\frac{1}{2}} H(\theta-\theta_k) = 0$$

$$\Rightarrow \theta^* = \theta_k + \left(\frac{1}{2\delta} V^T H^{-1} V\right)^{-\frac{1}{2}} H^{-1} V$$

$$= \theta_k + \left(\frac{1}{2\delta}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)\right)^{-\frac{1}{2}} H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)$$

$$\Rightarrow \alpha = \left(\frac{1}{2\delta}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)\right)^{-\frac{1}{2}}$$ ※

Name: Kai-Jie Lin
Student ID: 110652019

Homework 2 Report

Reinforcement Learning

Name: Kai-Jie Lin
Student ID: 110652019

April 30, 2023

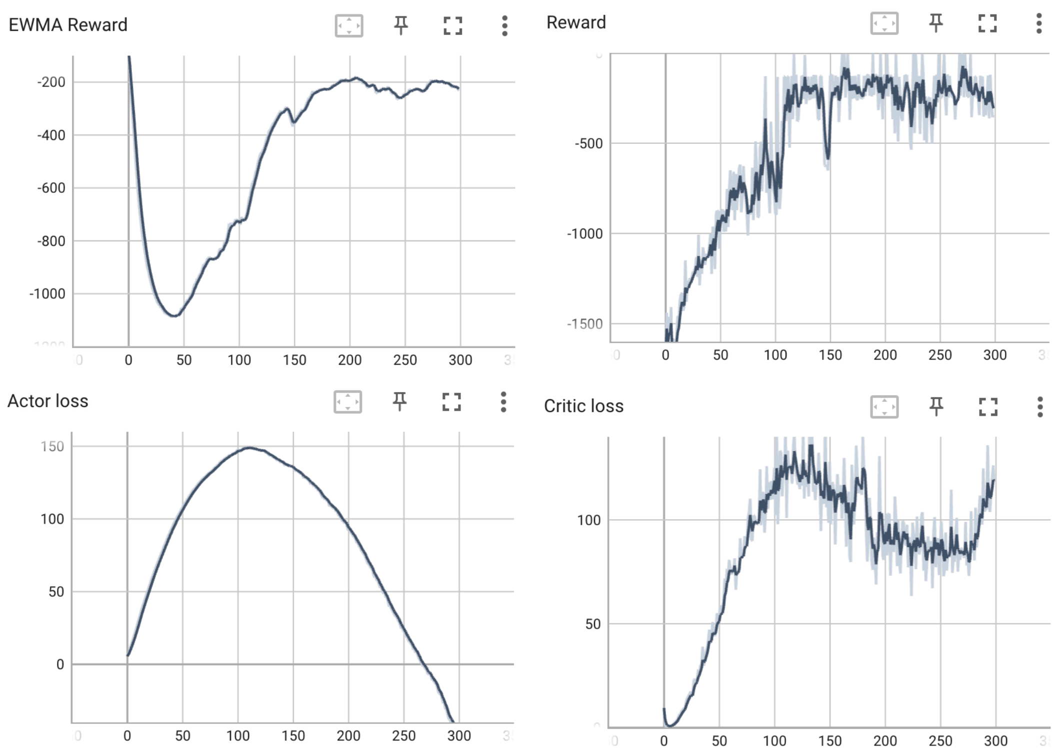
1 Experiment of DDPG

1.1 Pendulum-v1

Learning Rate(Actor)	0.0001
Learning Rate(Critic)	0.001
Batch Size	128
Hidden Size	128
Layer Number	1

Table 1: Hyperparameters

Result: Reach well policy in near 300 steps.



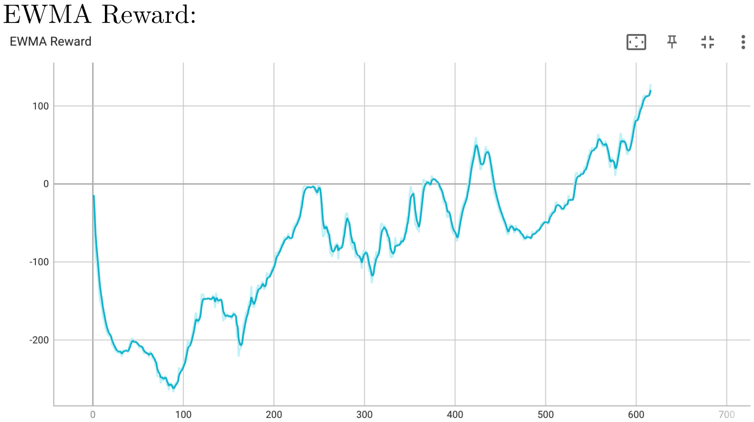
1.2 LunarLanderContinuous-v2

Use bayesian optimization to tune the hyper parameters. Reach EWMA reward=120 in 616 steps.

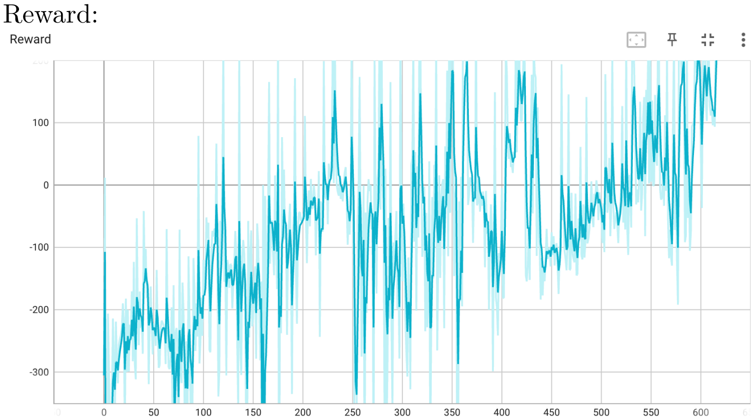
Learning Rate(Actor)	0.008
Learning Rate(Critic)	0.001
Learning Rate decay rate(Actor)	0.85
Learning Rate decay rate(Critic)	0.81
Batch Size	211
Hidden Size	128
Layer Number	2
Noise Scale	0.28

Table 2: Hyperparameters

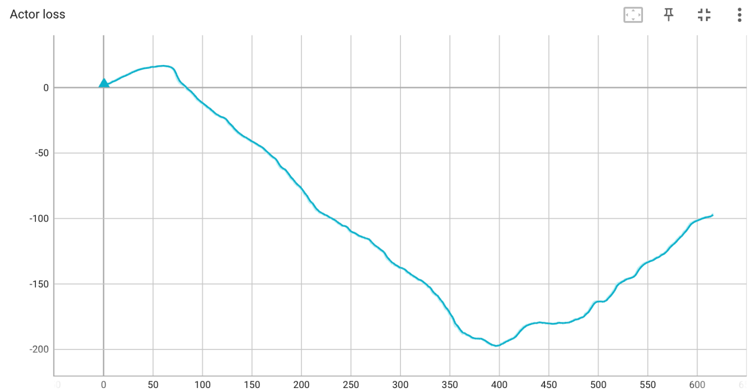
EWMA Reward:



Reward:



Loss of Actor:



Loss of Critic: