By

*Kesini Kandasamy*

Assignment 6.1: PPO Experimentation

Contents

[Scenario and Instructions 2](#_Toc208750900)

[Assignment Instructions: 2](#_Toc208750901)

[Original Source with initial output 2](#_Toc208750902)

[Assignment & Solutions 3](#_Toc208750903)

[Experiment 1: Reduce Training Epochs 3](#_Toc208750904)

[Experiment 2: Increase Hidden Layer Size 4](#_Toc208750905)

[Experiment 3: Increase Clip Ratio 5](#_Toc208750906)

|  |  |
| --- | --- |
| Assignment 6.1  PPO Experimentation |  |

# Scenario and Instructions

## **Assignment Instructions:**

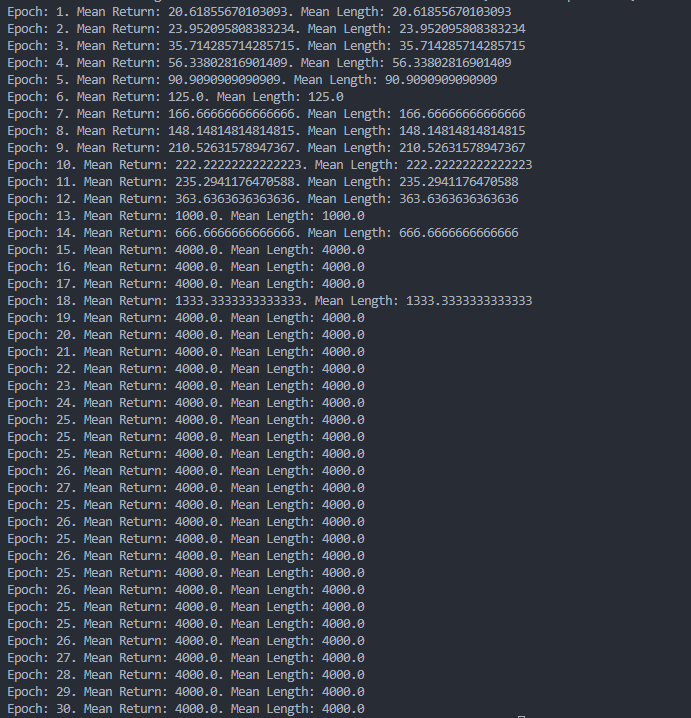
In this assignment, you will experiment with Proximal Policy Optimisation (PPO) implementation. You will work with the cartpole-v1 experiment, which is a classic Reinforcement Learning exercise wherein an AI agent learns to balance a simulated pole on a cart by moving the cart left or right. The agent is rewarded for each step it takes to keep the pole upright for as long as possible. The exercise ends when the pole falls, or the cart moves too far from the centre point.

You can find the code for this experiment from this link - https://keras.io/examples/rl/ppo\_cartpole/

You will run 3 experiments. For each experiment, modify the relevant parameters, run training, record results, and analyse your observations using the epoch number, mean return and mean episode length.

## **Original Source with initial output**

Below is the initial output from the source (<https://keras.io/examples/rl/ppo_cartpole/> ) without any modifications. This is added here for comparison later.

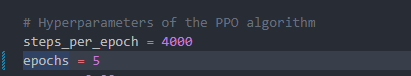


# Assignment & Solutions

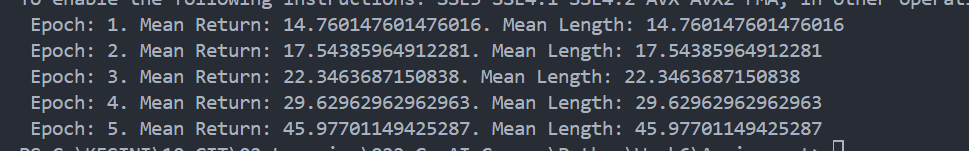
## Experiment 1: Reduce Training Epochs

What happens if we reduce training epochs from 30 to 5? Does the PPO agent learn a good policy with limited training?

* Change: epochs = 5



* Track: Mean Return, Mean Length



* Report: Is the agent improving meaningfully by epoch 5?

No. 5 total epochs seems too little for the PPO agent to learn. It indicates a reduced training so the agent has very less opportunity to update the policy.

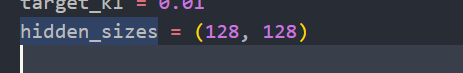
From the output, we can see that there is only a very small difference on the Mean return and Mean Length between the epochs when 5 epochs are used. The results are very far from the original result when epoch is 30.

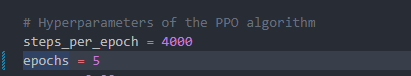
PPO agent does not learn a good policy with this much limited training. Though we see some gains, its much far from the performance when we use 30 epochs

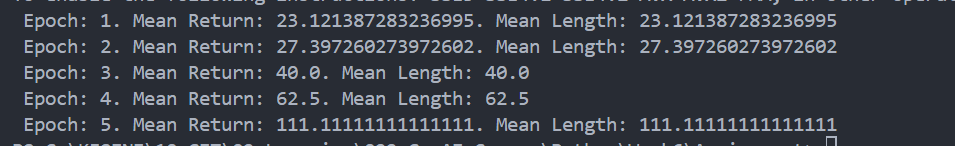
## Experiment 2: Increase Hidden Layer Size

Does increasing the hidden layer size from (64, 64) to (128, 128) speed up or stabilise learning?

* Change: hidden\_sizes = (128, 128)



* Keep: epochs = 5
  + 
* Track: Mean Return, Mean Length



* Report: Any noticeable improvement in convergence speed or return?

Increasing hidden layer size to (128, 128) did not noticeably improve convergence speed or returns within 5 epochs. Bigger networks can approximate more complex functions → in principle, they can learn better policies. However due to the limited training ( epoch 5 ) time, gains are minimal and not the same as when the epoch was at 30

## Experiment 3: Increase Clip Ratio

What happens when you increase clip\_ratio from 0.2 to 0.4? Does it improve or destabilise policy learning?

* Change: clip\_ratio = 0.4

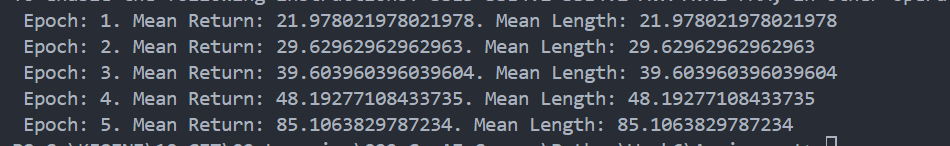


* Keep: hidden\_sizes = (64, 64), epochs = 5





* Track: Mean Return, Mean Length



* Report: Is learning faster or more unstable? Any signs of early stopping?

Setting clip\_ratio = 0.4 produced faster but early gains in Mean Return and Mean Length . It increases KL and variance across seeds and therefore destabilizes learning. Didn’t see any signs of early stopping. Tried 3 consecutive runs.