## Introduction

This homework is designed to follow up on the lecture about policy gradient algorithms. For this assignment, you will need to know about the basics of the policy gradient algorithm we talked about in class, specifically REINFORCE and PPO. If you have not already, we propose you brush up on the lecture notes.

**You are allowed to discuss this homework with your classmates. However, any work that you submit must be your own -- which means you cannot share your code, and any work that you submit with your write-up must be written by only you.**

## Code folder

Find the folder with the provided code in the following google drive folder: <https://drive.google.com/drive/folders/1lm9-9in2OheyPowIWnxkCykXfp-u76Go?usp=sharing> Download all the files in the same directory, and run the run.py file to run your code. You will have to complete the TODOs in ppo/ppo.py to complete this homework.

Thanks to Eric Yu for this code.

## Environment

We will reuse the environment from homework 2, so you will not need to install anything else on top of it. If you need more directions about setting up with it, see here: <https://docs.google.com/document/d/1p_mU1jZEQZk7gP_qgwVPtv6iae4bnjMa2FHkqV5t4K4/edit>

## Submission

Please submit your homework using this google form link: <https://forms.gle/FDgzwJhWSjKtysWi8>

Deadline for submission**: October 22th, 2021, 11:59 PM Eastern Time.**

## Points

* Questions 1-2 are 5 points each.
* Total: 10 points

## Questions

1. In the code folder, you will find already available code for running REINFORCE. Run this code on the following environments: Pendulum-v0, BipedalWalker-v3, and LunarLanderContinuous-v2. It is okay if REINFORCE does not perform as well in these environments. Generate the plot over training times for these 3 environments over three different seeds, and create three plots that show the average performance of REINFORCE on each environment. Why do you think REINFORCE suffers in these environments?
2. Now, complete the PPO code found in ppo/ppo.py. You will find a few different TODOs for you. Follow the original PPO pseudocode if you need to. Once again, use the previous three environments and three different seeds to plot your training rewards. **Clearly show the comparison between REINFORCE and PPO in your plots.**  
     
   Your expected mean performance should be AT LEAST:  
   Pendulum: -400  
   BipedalWalker: 125  
   LunarLanderContinuous: 100  
     
   Submit your writeup, along with your ppo.py file as your submission. If you change any of the hyperparameters, include run.py as well.