登陆月球

在本次作业当中，需要你采用reinforcement learning方法让月球登录器成功着陆。

你将比较几项 Deep Reinforcement Learning 方法：

Policy Gradient

Actor-Critic

作业的环境为 OpenAI 的 gym 当中的 Lunar Lander。其余请参考pynb。

请上传所有代码包括作图代码，训练好的model请上传至github-classom

其他frame请群@我

Github上请包涵

report.pdf

python files

作图代码

model param

Report

学号 姓名 专业

1. Policy Gradient 方法，请阅读及跑过范例程式，并试着改进 reward 计算的方式。请说明你如何改进 reward 的算法，而不同的算法又如何影响训练结果？试着修改与比较至少三项超参数（神经网路大小、一个 batch 中的回合数等），并说明你观察到什么。
2. Actor-Critic 方法请同学们从 REINFORCE with baseline、Q Actor-Critic、A2C 等众多方法中择一实作。请说明你的实做与前者（Policy Gradient）的差异。
3. 具体比较（数据、作图）以上几种方法有何差异，也请说明其各自的优缺点为何。