Project B

学号：2017202085

姓名：陕杰才

一、实验背景：

Gym是一个研究和开发强化学习相关算法的仿真平台，无需智能体先验知识，并兼容常见的数值运算库如 TensorFlow、Theano等。OpenAI Gym由以下两部分组成：

Gym开源库：测试问题的集合。当你测试强化学习的时候，测试问题就是环境，比如机器人玩游戏，环境的集合就是游戏的画面。这些环境有一个公共的接口，允许用户设计通用的算法。

OpenAI Gym服务：提供一个站点和API（比如经典控制问题：CartPole-v0），允许用户对他们的测试结果进行比较。

简单来说OpenAI Gym提供了许多问题和环境（或游戏）的接口，而用户无需过多了解游戏的内部实现，通过简单地调用就可以用来测试和仿真。

游戏背景：环境中有一辆小车，在一维无阻力的轨道上运动，车上连接一根活动杆，杆会左右摇摆。游戏规则很简单，我们要操纵我们的小车左右移动，使它上面的木棒能够保持平衡。当小车偏离中心4.8个单位，或杆的倾斜超过15度，任务失败。

游戏奖励（reward）：

    在gym的Cart Pole环境（env）里面，左移或者右移小车的action之后，env会返回一个+1的reward。其中CartPole-v0中到达200个reward之后，游戏也会结束，而CartPole-v1中则为500。最大奖励（reward）阈值可通过前面介绍的注册表进行修改。

二、运行官方文档示例代码：

import gym

env = gym.make('CartPole-v0')

for i\_episode in range(20):

observation = env.reset()

for step in range(100):

env.render()

print(observation)

action = env.action\_space.sample()

observation, reward, done, info = env.step(action)

if done:

print("Episode finished after {} timesteps".format(step+1))

break

由于是基于随机生成的action，所以会导致小车一直在山底左右移动

import gym

env = gym.make('CartPole-v0')

observation = env.reset()

print(observation)

#> [-0.00478028 -0.02917182 0.00313288 0.03160127]

以上代码显示了初始状态下的取值，每次调用*env.reset*( )将重新产生一个初始状态。打印出的*observation*的四个元素分别表示了小车位置、小车速度、杆子夹角及角变化率。