课程：人工智能 **作业 6**

北航软件学院

学期：2023，春季 提交截止时间：2023年6月26日，11：59 PM

提醒注意：

- 本次作业发布于2023年5月30日，截止于2023年6月26日。
- 作业一分为三部分：问答题、实训题、以及实训题报告
  - 问答题答案可以手写并扫描，或者用latex（或word）手打，最终以QA.pdf文件命名。
  - 实训题按照项目共享链接内要求和基础代码进行作答。
  - 报告部分同样可以手写或者手打，以Report.pdf文件命名。
  - 作业提交格式：$< studentID >_< name >$_A6.zip。比如1921102_田嘉怡_A6.zip
  - 提交的zip文件要求（仅）包括：
    - 实训题文件：包括 main.py（或main.ipynb）。
    - 问答题答案：QA.pdf
    - 报告：Report.pdf。需要包含实训题2.3、2.6部分的运行截图。
- 作业压缩包需要在spoc平台上提交。
- 每迟交1天（不满1天按1天计算），本次作业扣除10%分数。
- 不按作业要求和格式提交，视情况扣分。不得抄袭。

第一部分：问答题

Q 1

将机器人寻路问题简化为图1的2*2的网格，假设有位于s_1位置的机器人拟从s_1这一初始位置向s_4这一目标位置移动。机器人每次只能向上或者向右移动一个方格，到达目标位置s_4则会获得奖励且游戏终止，机器人在移动过程中如果越出方格(s_d)则会被惩罚且被损坏、并且游戏终止。奖励值定义如下：当$S_{t+1} = s_4$时奖励值为1，当$S_{t+1} = s_d$时惩罚值为-1，其他情况下奖励值为0。若折扣因子 γ =0.99，智能体在s_1、s_2、s_3的策略都初始化为上，终止状态s_4、s_d的价值函数定义为0，试通过联立贝尔曼方程给出状态s_1、s_2、s_3的价值函数。

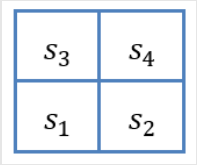

Figure 1: 2*2的机器人寻路问题

Q 2

在题1中，若每个状态的价值函数都初始化为0，智能体在s_1、s_2、s_3的策略都初始化为上，试优化智能体在状态s_3的策略。（提示：使用策略优化定理）

Q 3

在题1中，若图2表示算法的初始状态，其中a/b表示对应状态的动作-价值函数的取值，斜线左侧的a表示q_π(s,上)，斜线右侧的b表示q_π(s,右)。若α=0.5，试给出Q Learning 算法中的Q学习算法的一个片段的执行过程，并给出执行完该片段后每个状态的策略。

0.1/0	0/0
0.1/0	0.1/0

Figure 2: Q学习算法的初始状态

第二部分：实训题（共7分）

实训题要求:

- 本次作业包括1个实训题，作业要求以及基础代码以Aistudio项目的形式发布。
- 发布项目链接有效期3天，请在作业发布3天内fork这个项目，生成“我的项目”，并在自己fork的项目下进行作答，生成答案后按要求保存提交。

Q 1 机器人自走迷宫-强化学习

在本作业中，您首先需要阅读开发代码中需要用到的基础知识，包括我们代码中会用到的Maze迷宫类、QRobot类、Runner类。再之后您有一些任务需要完成，我们将根据任务完成的情况来给予您的分数。 作业分为两各部分

- 第一部分：实现2.3部分的搜索算法，您可任选深度优先搜索算法、最佳优先搜索（A*)算法实现其中一种。
- 第二部分：实现2.6部分的Deep QLearning (DQN)算法，公共部分可复用，您只需要重写Robot类中的train_update()、 test_update()方法即可，可参考QRobot类中的train_update()和test_update()接口实现的Q Learing算法。

实验介绍详情和参考基础代码请参见Aistudio中的共享项目“人工智能作业六-机器人自走迷宫”。

第三部分：实训题实验报告（共3分）

- 请按照实验报告模板完成实验报告。
- 实验报告模板是通用模板，可根据每个作业要求的差别，自由进行微调。