

Assignment 2: Dynamic Programming

一、 作业内容：利用强化学习动态规划算法找到最短路径（完成 dp.py 文件补全）

Part 1: 函数 `p_state_reward(state,action)`

```
177 # 定义状态动作转移，传入当前状态和执行的动作，返回当前状态下执行动作得到的转移概率、下一状态和奖励
178 def p_state_reward(state, action):
179
180
181     # 向上移动
182     if action == 0:
183         #Fill and submit this code
184
185
186     # 向下移动
187     if action == 1:
188         #Fill and submit this code
189
190
191     # 向左移动
192     if action == 2:
193         #Fill and submit this code
194
195
196
197     # 向右移动
198     if action == 3:
199         #Fill and submit this code
```

Part 2: 函数 `compute_value_fuunction(policy,gamma)`

```
201 # 策略评估：计算策略下状态的价值
202 def compute_value_function(policy, gamma):
203     #Fill and submit this code
204     return value_table
205
```

Part 3：函数 `next_best_policy(value_table,gamma)`

```
206
207 # 策略提升：更新策略
208 def next_best_policy(value_table, gamma):
209     #Fill and submit this code
210     return policy
211
```

Part 4：函数 `policy_iteration(random_policy,gamma,n)`

```
213 # 建立策略迭代函数
214 def policy_iteration(random_policy, gamma, n):
215     #Fill and submit this code
216     return new_policy
217
```

二、作业提交要求：打包成压缩包提交（实验报告+完整代码（dp.py））

1、实验报告:包括对动态规划算法的理解、动态规划设计迷宫的整体思路、实验运行截图、结果分析与遇到的问题

2、完整代码文件（dp.py）

3、请于10月8日前将压缩包提交到北航网盘class3_DP文件夹下。命名要求：学号_姓名_class3作业