PB22111645朱恩松

1.实验流程

根据Agent_Playground.pdf中的内容，按顺序写代码即可。

2. 调试超参数的过程

未进行调参数。

3.最好的游戏结果

见result文件夹。

4.回答问题

1

对于Model-based Value Iteration，其利用Monte Carlo估计MDP的动态，使得学习过程更快。适用于状态和动作有限的 MDP，能够快速收敛到最优策略。 而 Tabular Q-learning在足够多的时间步后，Q-learning 能够收敛到最优 Q 值，从而得到最优策略。且它是 Model-free 的，不需要了解转移概率和奖励函数，因此适用于更复杂的环境。 Model-based Value Iteration 可能会因为环境过于复杂或状态空间过大、采样不足导致模型不准确因素失效。

2

Model-free, Temporal Difference, off-Policy, Value-based, on-line

5.反馈

花费时间为一天。