

Random Walk

AAIS in PKU 陈伟杰 1901111420

November 28, 2019

1 Problem Setting

已知一维 Markov 链, $\mathcal{S} = [0, 1, 2, 3, 4, 5, 6]$, $\mathcal{A} = \{-1, 1\}$, 回报函数 $R = \mathbb{I}\{s = 6\}$ 。当 $s = 0$ 或 $s = 6$ 时, 随机行走终止。

分别运用 Temporal differential 方法和 Monte Carlo 方法进行值函数迭代:

- TD: $V(s_t) \leftarrow V(s_t) + \alpha(R_{t+1} + \gamma V(s_{t+1}) - V(s_t))$
- MC: $V(s_t) \leftarrow V(s_t) + \alpha(R_\tau - V(s_t))$, 若 s_t 属于轨迹 τ

2 Experiment Result

已知该 Markov 链的精确值函数为 $V^* = [1/6, 2/6, 3/6, 4/6, 5/6]$ 。

实验中设置值函数 V 初值均为 0.5, $\gamma = 1$, 在不同的 α 下进行 $N = 1000$ 次实验求平均得到下列均方差与迭代次数的关系图

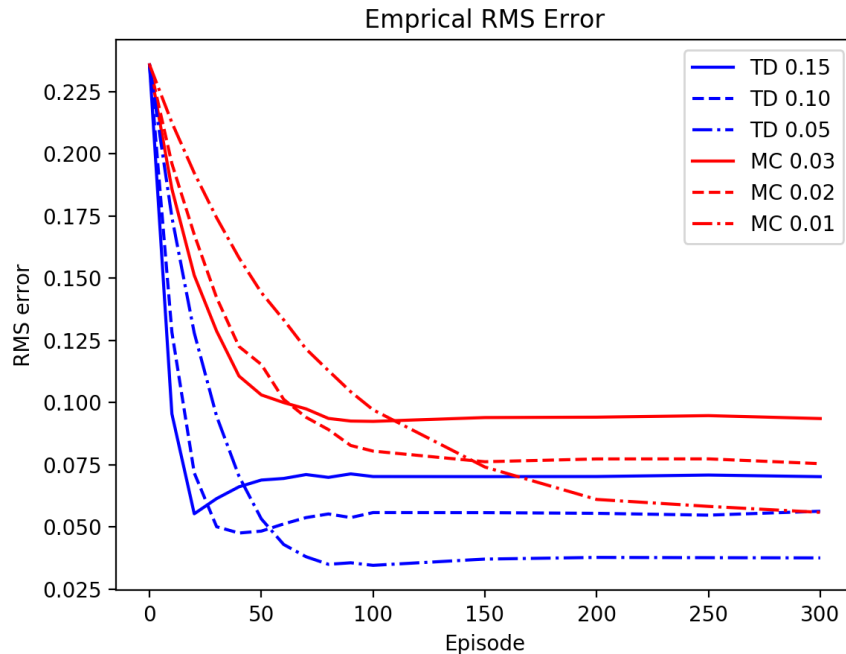


Figure 1: 均方差与迭代次数的关系图

可以明显看出 TD 方法优于 MC 方法, 并且 α 越小收敛越慢, 但能够收敛到更优的值函数。

3 README

代码包括 CarRental.py 和 test.py 两个程序，需要 python3 环境。直接运行 test.py 可以得到上述结果

- 电脑：MacBook Pro(15-inch, 2019)
- 处理器：2.6 GHz 6-Core Intel Core i7
- 内存：16 GB 2400 MHz DDR4
- 操作系统：macOS Catalina version 10.15.1
- 语言：python3.6