Reinforcement Learning Assignment 2

Yiheng Jiang (TA), Prof. Zou

2024-03-14

1 Introduction

The goal of this assignment is to do experiments with Monte-Carlo(MC) Learning and Temporal-Difference(TD) Learning. MC and TD methods learn directly from episodes of experience without knowledge of MDP model. TD method can learn after every step, while MC method requires a full episode to update value evaluation. Your goal is to implement MC and TD methods and test them in the small gridworld.

2 Small Gridworld

0	1	2	3	4	5
6	7	8	9	10	11
12	13	14	15	16	17
18	19	20	21	22	23
24	25	26	27	28	29
30	31	32	33	34	35

Figure 1: Gridworld

As shown in Fig.1, each grid in the gridwold represents a certain state. Let s_t denotes the state at grid t. Hence the state space can be denoted as $S = \{s_t | t \in 0, .., 35\}$. S_1 and S_{35} are terminal states, where the others are non-terminal states and can move one grid to north, east, south and west. Hence the action space is $A = \{n, e, s, w\}$. Note that actions leading out of the grid leave state unchanged. Each movement get a reward of -1 until the terminal state is reached.

3 Experiment Requirments

- Programming language: python3
- You should implement both first-visit and every-visit MC method and TD(0) to evaluate an uniform random policy $\pi(n|\cdot) = \pi(e|\cdot) = \pi(s|\cdot) = \pi(w|\cdot) = 0.25$.

4 Report and Submission

- Your reports and source files (.py) should be compressed and named after "studentID+name".
- The file should be submitted on Canvas on Mar. 21, 2024.