PSG College of Technology

Department of Applied Mathematics and Computational Sciences

Reinforcement Learning 2023-2024

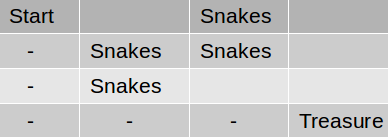
VIII Sem M.Sc Data Science/ MSc TCS

Lab 5

Imagine yourself in a treasure hunt in a maze. The game is as follows:

You start at a given position, the starting state. From any state you can go left, right, up or down or stay in the same place provided you don’t cross the premises of the maze. Each action will take you to a cell of the grid (a different state). Now, there is a treasure chest at one of the states (the goal state). Also, the maze has a pit of snakes in certain positions/states. When you place an agent in the grid (we will refer to it as our environment) it will first explore. It doesn’t know what snakes are, neither does it know what or where the treasure is. So, to give it the idea of snakes and the treasure chest we will give some rewards to it after it takes each action. For every snake pit it steps onto we will give it a reward of -10. For the treasure we will give it a reward of +10. Now we want our agent to complete the task as fast as possible (to take the shortest route). For this, we will give rest of the states a reward of -1. Then we will tell it to maximise the score.

Your goal is to travel from the starting state to the goal state by following a path that doesn’t have snakes in it



For this purpose you take a smaller maze-grid for ease.

|  |  |  |
| --- | --- | --- |
| Start  State 1 | State 2 | State 3 |
| State 4  snake | State 5  snake | State 6 |
| State 7 | State 8  snake | State 9  Treasure |

1. Apply Dynamic Programming to find an optimal policy
2. Apply Q learning to final the final Q table.
3. Apply SARSA to find the