Reinforcement Learning Assignment #1 Report

Hazique Quazi

mquazi@buffalo.edu

UB Person #: 50416875

Github Repo for Submissions (**@ub-rl added as collaborator**)

Abstract

This report details and demonstrates our learning of Reinforcement Learning concepts such as creating an agent and environment, running the agent for a given number of timesteps and visualizing the change of states and rewards given to the agent by the environment in response to a given action. Below I describe the execution of the agent in Deterministic and Stochastic Environment.

1. Deterministic Environment

The deterministic environment can be described as follows:

Action Set

$$A = \{Left = 0, Top = 1, Right = 2, Bottom = 3\}$$

State Set

$$S = \{0 \text{ to } 15 \text{ Discrete States}\}$$
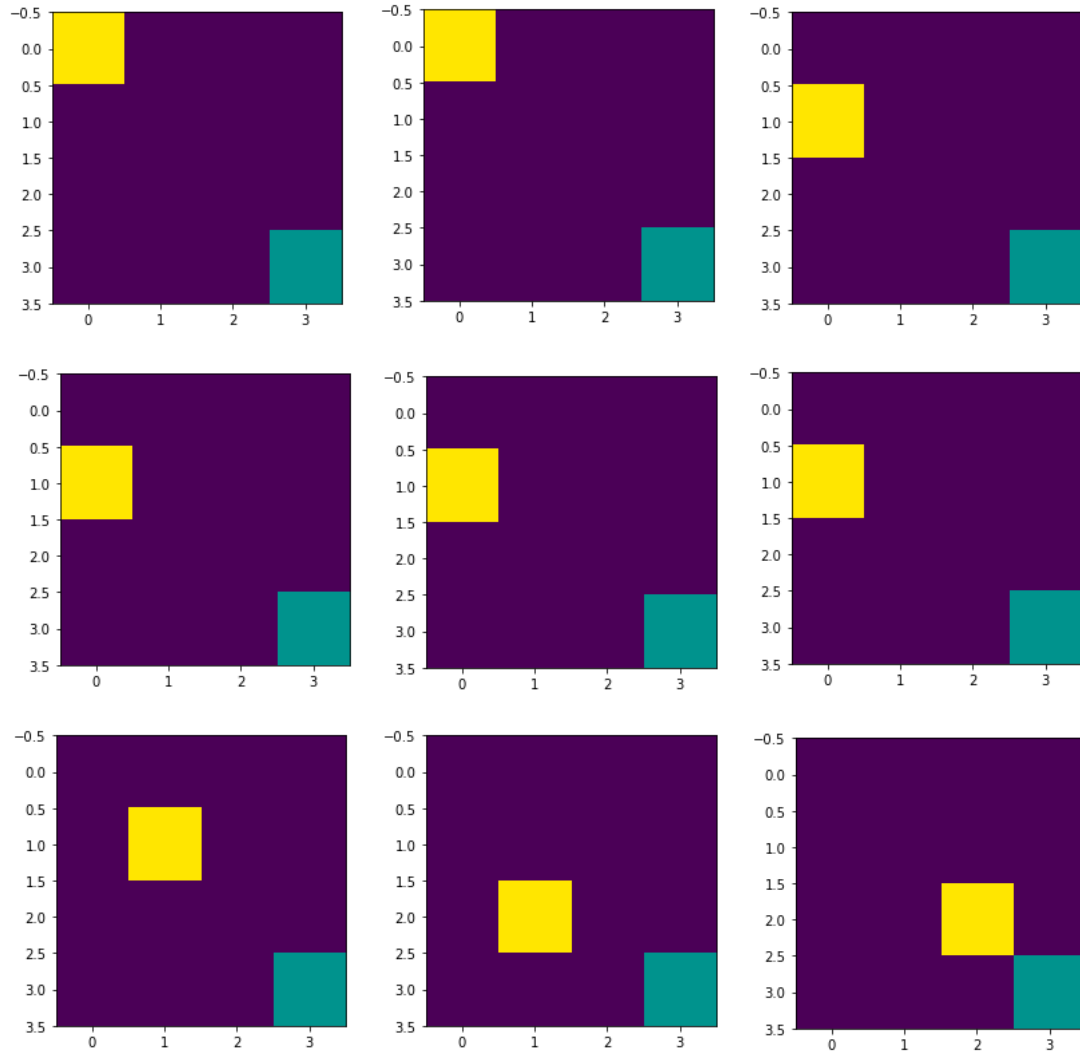
Reward Set

$$R = \{-1, 0, 1, 2\}$$

The reward received by the agent can be better represented as a 4x4 matrix

```
([[0, 1, 1, 1],
[1, -1, -1, 1],
[1, -1, -1, 1],
[1, 1, 1, 2]])
```

The reward 0 corresponds to the starting position of the agent whereas the 2 corresponds to the end position.

Visualization for Deterministic Environment

TimeStep	Action	Reward	Total Rewards	Done
0	0	0	0	False
1	0	0	0	False
2	2	1	1	False
3	1	1	2	False
4	1	1	3	False
5	1	1	4	False
6	3	-1	3	False
7	2	-1	2	False
8	3	-1	1	False
9	0	-1	0	True

2. Stochastic Environment

The Stochastic environment has the same Action Set, State Set & Reward Set as the Deterministic environment. It however adds stochasticity to the action the agent can take at any time step. I have defined the stochastic environment by letting an agent take an action probabilistically according to the following logic:

1. Assign random probabilities to actions (left = up = 0 to10%, right = down = 11-90%)
2. Check if direction with maximum probability has a probability value greater than 30%
   a. If False, the agent goes left instead of right OR the agent goes up instead of down.

Visualization for Stochastic Environment
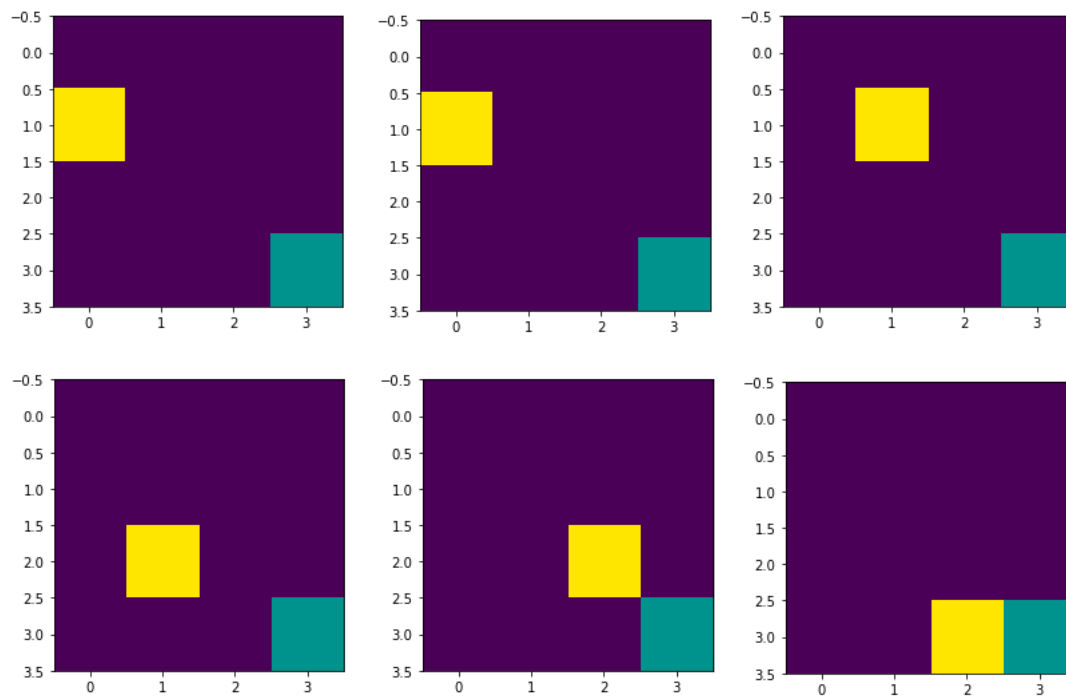
TimeStep	Action	Reward	Total Rewards	Done
0	2	1	1	False
1	1	1	2	False
2	3	-1	1	False
3	2	-1	0	False
4	3	-1	-1	False
5	2	1	0	False
6	3	2	2	True