Report: Assignment 1 - Defining & Solving RL Environments: Part 1.1

Assignment Overview:

The goal of the assignment is to acquire experience in defining and solving reinforcement learning environments, following OpenAI Gym standards. The assignment consists of two parts. The first focuses on defining deterministic and stochastic environments that are based on Markov decision process. In the second part we will apply two tabular methods to solve environments that were previously defined.

Problem Statement for Part 1.1

Define a deterministic environment, where $P(s', r \mid s, a) = \{0, 1\}$. Run a random agent for at least 10 timesteps to show that the environment logic is defined correctly.

Environment requirements:

- Min number of states: 12
- Min number of actions: 4
- Min number of rewards: 4

Specification of my Environment:

- No of States: 25
- Number of Actions to be taken: 4
- Number of Rewards = 4 (the yellow squares are worth +4 and the dark blue at (3,0) and (0,3) are worth -1.
- The agent start state is at the left hand top.
- The goal state is at bottom right corner.

1: Question: Describe the stochastic and deterministic environment which were defined.

The Deterministic environment:

The deterministic environment was defined in a simple way such that whatever action the agent takes, there is no influence on the transition by my environment of any sort. For example if the agent at state [0,0] decides to go down or go up, the state will change according without any uncertainty criteria (epsilon for stochastic). There are 4 rewards that are there, ie. the two negative rewards (for punishing the agent for taking a sub-optimal path and nudging it towards the optimal path) and two positive rewards (for rewarding the agent for going towards the final goal destination).

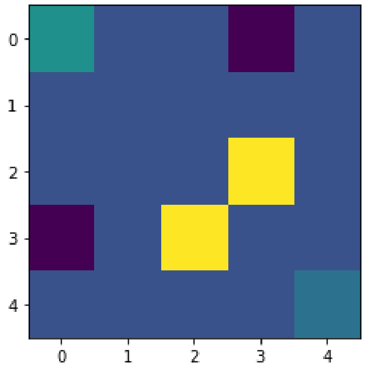
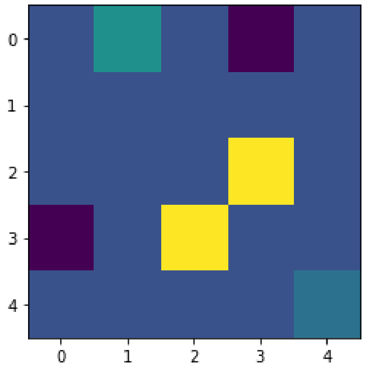
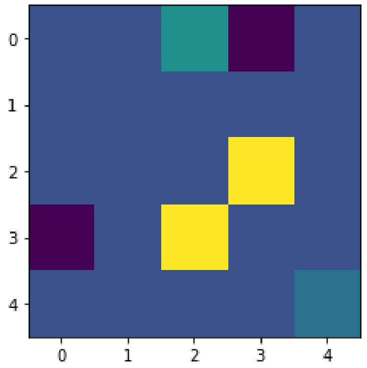
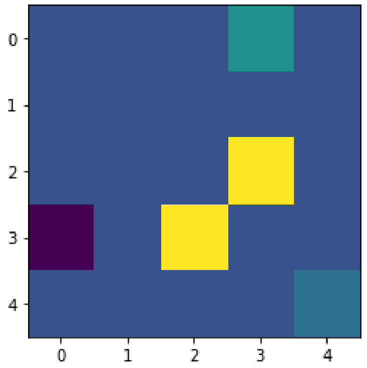
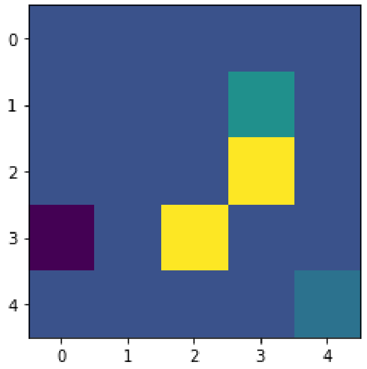
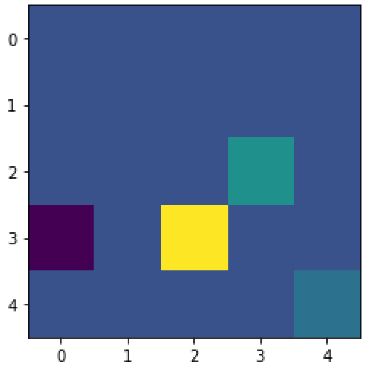
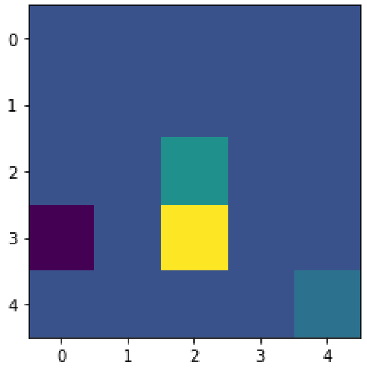
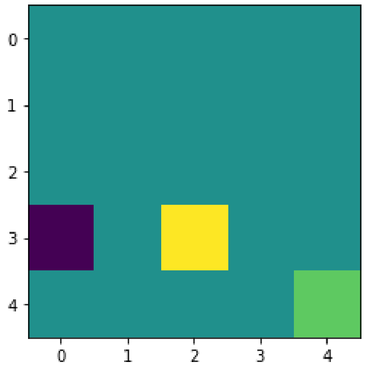
- **Set of Actions** – {0 -> up, 1 -> down, 2 -> left, 3 ->right) where (0,1,2,3) are the actions and (up, down, left, right) are the actual corresponding physical actions that the agent takes in the environment.
- **States** – {S1, S2,S3,…..S25} – These states correspond to the individual cells in the environment matrix – {[0,0],[0,1],…….[4,4]}, where [0,0] is the agent start state, [4,4] is the goal state, [0,3],[3,0],[2,3],[3,2] are the reward states.
- **Rewards** – {-1,-1,3,3} – There are four rewards in the environment, the two negative rewards for nudging the agent towards the goal state and the two positive rewards for the encouragement for taking the right steps. Once an reward has been given out, it gets deleted from the system and isn’t available for the rest of the episode.
- **The Main Objective:** The main objective of the agent is to reach the “goal state” ([4,4] in my environment) while collecting the maximum sum of rewards in least amount of time-steps.

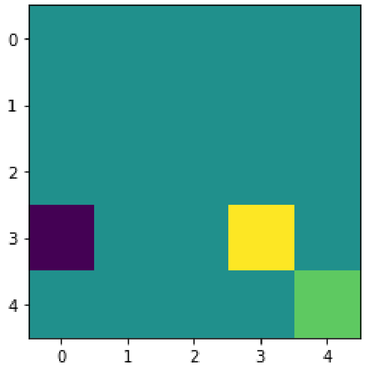
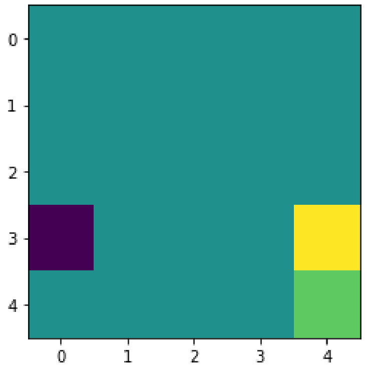
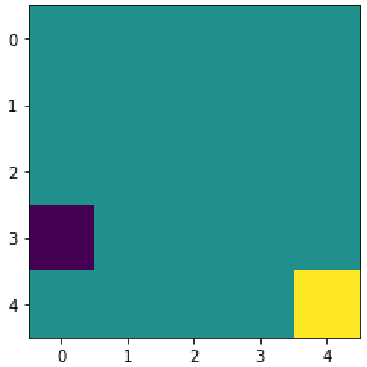
The Stochastic environment:

The Stochastic environment was designed in a way such that based on the epsilon value, the agent can take two random actions apart from the original greedy action. For example, if the agent wants to go up, there is chance that it will go down and left and so on. However the sum of all the transitions is equal to 1.

2: Show your visualizations:

My Deterministic Environment Rendering

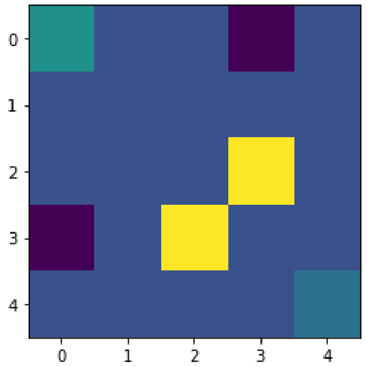
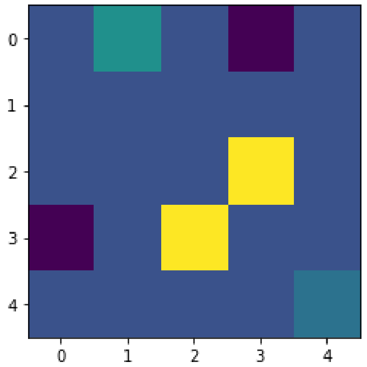
 Step 1:	 Step 2
 Step 3	 Step 4
 Step 5	 Step 6
 Step 7	 Step 8

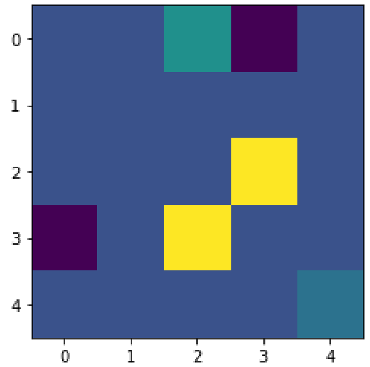
 Step 9	 Step 10
 Step 11	Final Reward : -1 + 3 + 3 = 5

3: How did you define the stochastic environment?

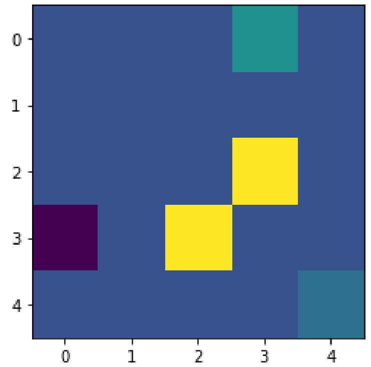
The stochastic environment was defined in a way such that, based on the epsilon value set at run-time for the agent, the environment will direct the agent based on the action it was going to take. For example, if the agent was going to take an action going up, then there is a small chance that it will go down or left. Similarly, if the agent wanted to go down, there is a small chance it can go up or right and so on. Finally the reward is calculated and episode ends.

Here is the rendering for my stochastic environment:

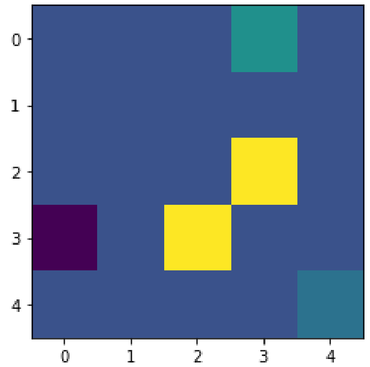
 Initial State Epsilon 0.3	 Current Agent POS [0, 0] Our Right or the Agents Left randn is 0.4019242341553725 The Agent Ended Up Going Right Final Agent POS [0, 1]



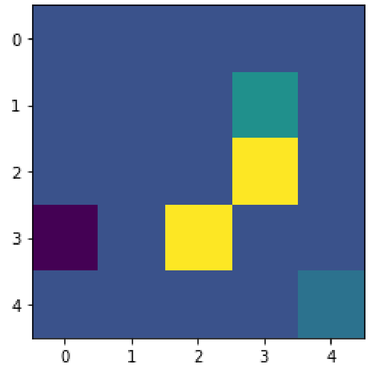
Current Agent POS [0, 1]
Our Right or the Agents Left
randn is 0.8044946340610436
The Agent Ended Up Going Right
Final Agent POS [0, 2]



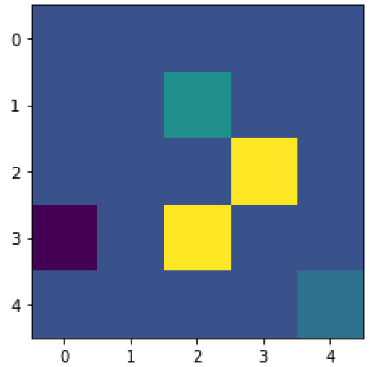
Current Agent POS [0, 2]
Our Right or the Agents Left
randn is 0.5842331718485394
The Agent Ended Up Going Right
Final Agent POS [0, 3]



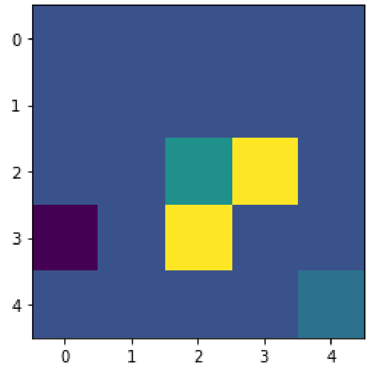
Current Agent POS [0, 3]
Down
randn is 0.27709022248561777
The Agent Ended Up Going Up
Final Agent POS [0, 3]



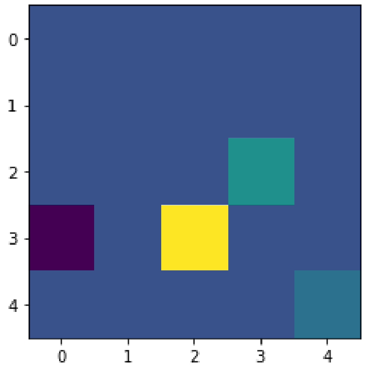
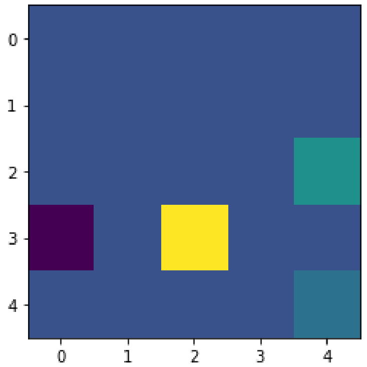
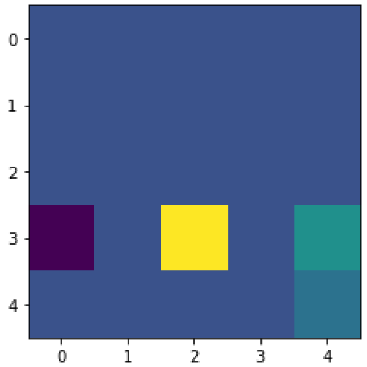
Current Agent POS [0, 3]
Down
randn is 0.49716040604898726
The Agent Ended Up Going Down
Final Agent POS [1, 3]



Current Agent POS [1, 3]
Our Left or the Agents Right
randn is 0.4103824046394997
The Agent Ended Up Going Left
Final Agent POS [1, 2]



Current Agent POS [1, 2]
Down
randn is 0.7422062057680328
The Agent Ended Up Going Down
Final Agent POS [2, 2]

 Current Agent POS [2, 2] Our Right or the Agents Left randn is 0.4940034101483233 The Agent Ended Up Going Right Final Agent POS[2,3]	 Current Agent POS [2, 3] Our Right or the Agents Left randn is 0.9714106749268795 The Agent Ended Up Going Right Final Agent POS [2, 4]
 Current Agent POS [2, 4] Down randn is 0.9981221516263096 The Agent Ended Up Going Down Final Agent POS [3, 4]	Total Reward = -1 +3 = 2

4: What is the difference between the deterministic and stochastic environments?

The main difference between deterministic and stochastic environment is that in deterministic environment, the environment doesn’t affect the agents actions apart from the limits set for it. For example if the agent chooses to go beyond the boundary of the environment or it enters a fail state (state where the episode ends), then the environment will affect the states involved. However otherwise for legal/appropriate actions (including rewards, etc), the environment will not affect the agent from taking an action. However for stochastic environments, there is a small transition probability, denoted by epsilon involved when the agent chooses to take action. For example, if the agent chooses to go “down” as in my grid-world environment, there is a small probability that it takes “left” or “up”, where “up”, “down”, “left” are the actions that an agent can take.