Assignment 1 - Environments and Tabular Methods

Baasit Sharief

February 21, 2022

1 Understanding the problem statement

1.1 Introduction

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.

2 Environment

The grid environment class is initialized using OpenAI gym Env class involving the basic functionality like reset, step, init and sample method which can be used for exploring. We have an action space of 4 Discrete actions, namely,

1. Up, $action = 0$
2. Down, $action = 1$
3. Right, $action = 2$
4. Left, $action = 3$

We also have an observation space of size 15, a 3x5 Grid.

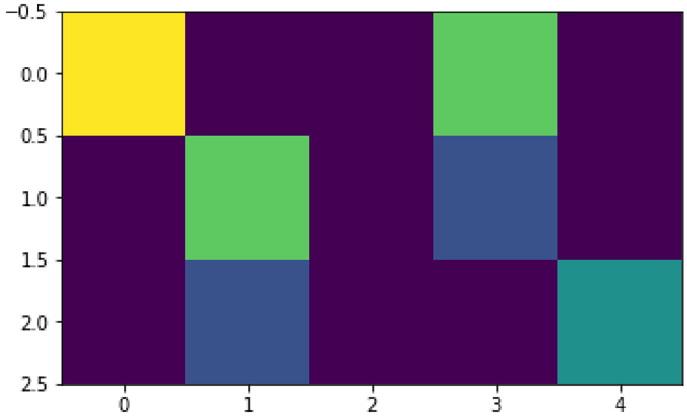

Figure 1: Grid Observation

For both environments, the initial agent position is at the top-left corner and the goal position is bottom right corner. The environment has 2 candy positions as well as 2 hole positions. The environment has a function called render which plots the grid. The goal of our agent is to reach the terminal state accumulating the maximum amount of rewards while avoiding holes as much as possible.

2.1 Deterministic Environment

The Environment has the following attributes:

1. Reward Space, $R = \{-3, 0, +5, +7, +10\}$
2. State Space, 3x5 grid space
3. Action Space, $A = \{up, down, right, left\}$

2.2 Stochastic Environment

The stochastic environment has a stochastic agent and stochastic rewards on the candy positions and hole positions. It has the following attributes:

1. Reward Space, $R = \{-3, 0, +2, +5, +6, +7, +10\}$
2. State Space, 3x5 grid space
3. Action Space, $A = \{up, down, right, left\}$

2.3 Deterministic vs Stochastic

The stochastic environment has a stochastic agent where:

1. the agent takes the action instructed with a probability of 0.9
2. the agent takes a random action from the action set

On Candy Positions:

1. Candy Position 1:
   (a) the agent receives an award of +2 with a probability of 0.5
   (b) the agent receives an award of +5 with a probability of 0.5
2. Candy Position 2:
   (a) the agent receives an award of +2 with a probability of 0.5
   (b) the agent receives an award of +7 with a probability of 0.5

On both Holes:

1. the agent receives an award of -3 with a probability of 0.9
2. the agent receives an award of +6 with a probability of 0.1

2.4 Safety of the Agent

In our use case, whenever our agent is at the edges and corners of the grid, we restrict it from not going out of it if the action results in that. In our environment we accomplish this by clipping x-coordinates and y-coordinates in the range of [0,3) and [0,4) respectively. This ensures the safety of our agent by restricting it to leave the grid environment and forcing it stay within the constraints.