MIE567 Assignment 1

Part A

2. The state space is represented as s(x)(y), which refers to the x-th row and y-th column beginning with the top left. Thus, the set of feasible state spaces are:   
   {s11, s13, s14, s15, s21, s23, s24, s25, s31, s32, s33, s34, s35, s41, s45, s51, s52, s53, s54, s55}

In this problem, we begin at s51, and our goal is to reach s13.

The action space is {U, D, L, R}, which represents the action of going up, down, left, and right respectively.

1. The transition probability matrix is dependent on the action taken:

This is the set of actions that result in not moving, because they would have moved into a blocked square:



This is the set of actions that result in moving to a different (i.e. non-blocked) square:

|  |  |  |  |
| --- | --- | --- | --- |
| Origin state (s) | Action (a) | Destination state (s') | P(s'|s,a) |
| s11 | D | s21 | 1 |
| s14 | D | s24 | 1 |
| L | s13 | 1 |
| R | s15 | 1 |
| s15 | D | s25 | 1 |
| L | s14 | 1 |
| s21 | U | s11 | 1 |
| D | s31 | 1 |
| s23 | U | s13 | 1 |
| D | s33 | 1 |
| R | s24 | 1 |
| s24 | U | s14 | 1 |
| D | s34 | 1 |
| L | s23 | 1 |
| R | s25 | 1 |
| s25 | U | s15 | 1 |
| D | s35 | 1 |
| L | s24 | 1 |
| s31 | U | s21 | 1 |
| D | s41 | 1 |
| R | s32 | 1 |
| s32 | L | s31 | 1 |
| R | s33 | 1 |
| s33 | U | s23 | 1 |
| L | s32 | 1 |
| R | s34 | 1 |
| s34 | U | s24 | 1 |
| L | s33 | 1 |
| R | s35 | 1 |
| s35 | U | s25 | 1 |
| D | s45 | 1 |
| L | s34 | 1 |
| s41 | U | s31 | 1 |
| D | s51 | 1 |
| s45 | U | s35 | 1 |
| D | s55 | 1 |
| s51 | U | s41 | 1 |
| R | s52 | 1 |
| s52 | L | s51 | 1 |
| R | s53 | 1 |
| s53 | L | s52 | 1 |
| R | s54 | 1 |
| s54 | L | s53 | 1 |
| R | s55 | 1 |
| s55 | U | s45 | 1 |
| L | s54 | 1 |

1. We define the Reward function as the following:

for all other

We define a high reward (+25) for reaching the goal state of s13 because we want to be incentivized to go there.

We assign a cost of -1 to any regular feasible move so as to disincentivize taking too many moves.

1. See Gridworld.py