ASSIGNMENT II

MARKOV DECISION PROCESS

You will be given a grid world based MDP parameterized as below

Input: First line of the input consists of 2 space separated integers **N M** the dimensions of the grid world.

N lines follow, each having M real numbers specifying the rewards for getting into that state

Next line consists of 2 space separated integers **E W**, number of end states and number of walls

Next E lines follow - each having 2 integers - the coordinates of end states.

Next W lines follow - each having 2 integers - the coordinates of walls.

Next line has 2 integers specifying the coordinates of the start state.

Next line has 1 integer specifying the unit step reward.

The following input corresponds to the MDP below:

```
4 4
-x      0       0       0
0       0       0       0
x/10    0       0       0
0       -x/10   0       x
2 1
0 0
3 3
1 2
3 0
-x / 5
```

-X 🚫			
		▨	
+X / 10			
START	-X / 10		+X

Problem Specifications

- Rows are numbered 0,1,2,3 from top to bottom and columns are numbered 0,1,2,3 from left to right. Eg. Start cell is (3,0)

- The cell (3,3) is the positive(green) sink while (0,0) is the negative(red) sink

- The blue cells are blocked (assume them as walls)

- The borders of the grid are also walls

Agent can go North, South, East or West

Action from a state results in

○ Movement in intended direction with probability 0.8

○ Movement in directions perpendicular to the intended direction with 0.1 probability each (0.8 + 0.1 + 0.1 = 1). Eg. If action is North, then actual movement will be to North with 0.8 prob, to East with 0.1 prob, & to West with 0.1 prob.

●If an action results in reaching a cell with wall, agent will remain in the same cell

● No action needs to be performed at terminal states

NOTE:

You will be given numeric inputs for PART 1 of the assignment.

For Part 2 and 3 of the assignment replace X with your team number.

Problem Statement

The assignment consists of three parts:

Part A: Given any MDP of the aforementioned form - write a program that performs Value Iteration(VI) on that MDP and gives final utility of each state as output. For termination condition consider 1% change or less within tolerance. Print W for walls.

Output format: if the grid is n * m , output should consist of n lines - having m space separated values each upto 3 decimal precision.
For ex: if grid is 2 * 3

```
Output:       5.300 3.200 7.200

              1.200 5.900 6.100
```

Part B: Perform the VI algorithm on the above given specific MDP [here x will be your team number]. Now, carefully analyse how the policy changes over the iterations. Observe the iterations in which there is a change in the optimal action to be taken in any state and why the policy changed in that specific way.
You have to submit a handwritten / pdf note, highlighting the iterations in which the policy changes and then explain the reason for policy changing in that specific way.

Part C: Model the above problem(specific MDP) using LP shown below

$$\max(\mathbf{rx}) \,\middle|\, \mathbf{Ax} = \boldsymbol{\alpha},\ \mathbf{x} \geq 0,$$

Q1. Model the parameters r,A and α

Q2. Use the excel LP solver to compute the x values and the expected utilities for this MDP

Please verify that the expected utility obtained is equivalent to the one obtained using the VI algorithm. The VI value and LP value can differ at max by delta*(1.2)

Deliverables

You are supposed to submit :

1. A working code for general MDP solving as specified in part A. 20 marks
2. Handwritten / pdf as specified in part B. 20 marks
3. Excel file or .ods (LibreOffice) file, showing the LP solved. 20 marks

Upload the above files in a zipped folder with the name **Assignment2_teamNumber.zip** and submit the hard copy in class.

Deadline

● The zip folder needs to be uploaded on moodle on or before March 12, 11.55 pm.

Note: You need to load the solver add-in in Excel if not already installed. It comes loaded with LibreOffice packages by default.