CS 312: Artificial Intelligence Laboratory
Task 7: Reinforcement Learning

Goal:Solve an MDP problem using policy and value iteration.

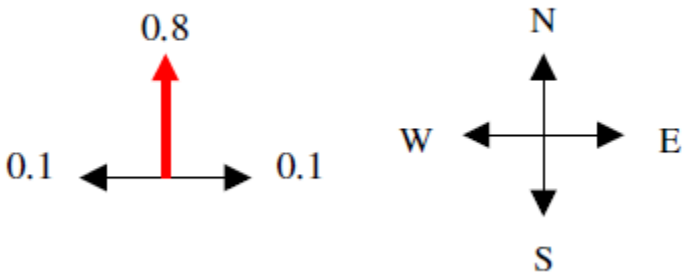
Note: This assignment is to be done individually.

MDP(Markov Decision Process): Grid World Problem:

There will be a grid, location of the player in the grid will represent a state, there will be a starting state, there will be two absorbing states having very different rewards like +1 and -20 while other states will have negative reward -1 associated with them, movement to that state will incur this negative reward. The black block is a wall where your agent won't be able to penetrate through. The transition probabilities for moving from one state to another are also given below. We need to find optimal movement direction for each state.

				End +10
				End -200
Start				

Below are the transition probabilities



1. Develop code for solving the MDP problem using policy and value iteration.
2. Write a report clearly describing the above MDP considered and your observations on running the policy and value iteration algorithms on the formulated MDP.
3. Further, one should also suggest ways to check whether the algorithm yields optimal policy for the setting considered.

Submission: ***This assignment is to be submitted individually.***
Please submit a zip file ***<Roll_number>.zip*** with the following contents

1. Program: ***<Roll_number>.<extension>*** (e.g., 1800100xx.c/cpp)
2. Report: ***<Roll_number>.<extension>*** (e.g., 1800100xx.pdf). Report should be in pdf format.
3. Readme file: readme.txt (Execution details)

Report Format :

1. [1 mark] MDP Description: Clearly describe (S, A, P, R, N)
2. [5 marks] State-transition Graph for the MDP
3. [1 mark] Optimal Policy: Suggest ways to check whether the algorithm yields optimal policy for the setting considered.
4. [5 marks] Experimental Results: Vary the gamma parameter, show the policy found in each case by both algorithms
5. [2 marks] Comparison of Policy Iteration and Value Iteration
6. [1 marks] Conclusions

Evaluation Criteria:
Correctness: 30 [15(Policy Iteration) + 15 (Value Iteration)]
Report: 15
Code Quality: 5

Deadline: *9th April 2021 11:49PM*

Late Submission Policy: *Penalty of 10% will be issued per day if the deadline is not met. If found copied, 0% score will be awarded*

For Reference :
Reinforcement-Learning:
http://www.cse.iitm.ac.in/~ravi/courses/Reinforcement%20Learning.html
See lectures 15 -25.