Multi Agent Reinforcement Learning

Assignment 1 Report

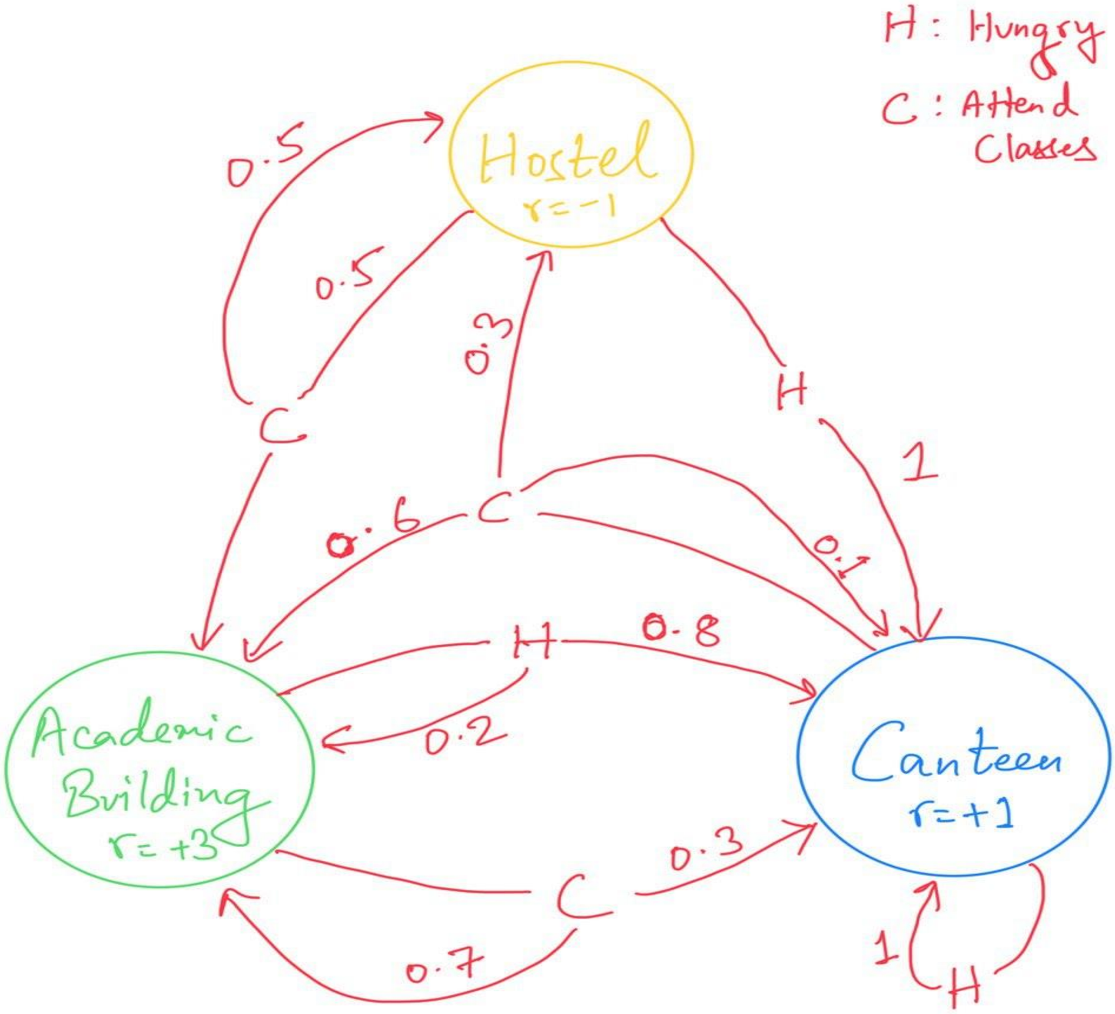
by Manish Gayen 21161

Question 1:

The MDP can be defined using the following table:

Current State	Action	Next State	Transition Probability	Reward
Hostel	Attend Classes	Canteen	0	-
Hostel	Attend Classes	Hostel	0.5	-1
Hostel	Attend Classes	Academic Building	0.5	-1
Hostel	Hungry	Canteen	1	-1
Hostel	Hungry	Hostel	0	-
Hostel	Hungry	Academic Building	0	-
Academic Building	Attend Classes	Canteen	0.3	3
Academic Building	Attend Classes	Hostel	0	-
Academic Building	Attend Classes	Academic Building	0.7	3
Academic Building	Hungry	Canteen	0.8	3
Academic Building	Hungry	Hostel	0	-
Academic Building	Hungry	Academic Building	0.2	3
Canteen	Attend Classes	Canteen	0.1	1
Canteen	Attend Classes	Hostel	0.3	1
Canteen	Attend Classes	Academic Building	0.6	1
Canteen	Hungry	Canteen	1	1
Canteen	Hungry	Hostel	0	-
Canteen	Hungry	Academic Building	0	-

The MDP diagram for the problem is as follows:



The results obtained are as follows:

Value Iteration Results:

Values: [12.98306307 12.98306307 12.98306307 13.39809229 13.3145874 13.3145874]

Policy: ['Eat_Food', 'Eat_Food', 'Eat_Food', 'Attend_Class', 'Attend_Class', 'Attend_Class']

Policy Iteration Results:

Values: [12.98304403 12.98304403 12.98304403 13.39807276 13.31457007 13.31457007]

Policy: ['Eat_Food', 'Eat_Food', 'Eat_Food', 'Attend_Class', 'Attend_Class', 'Attend_Class']

Discussion:
From the obtained results, the following are observed:

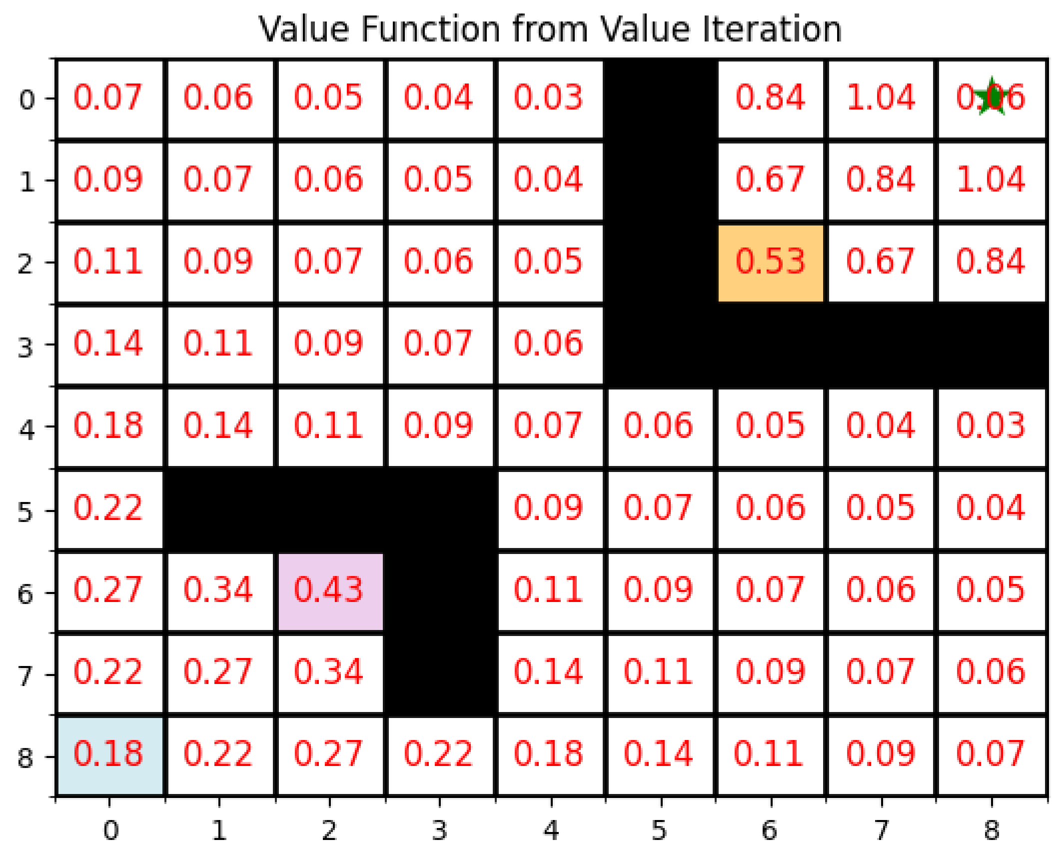
Consistency of Values: Both Value Iteration and Policy Iteration produce very similar values for each state, with only differ by an ignorable margin. This consistency indicates that both algorithms have converged to similar solutions.

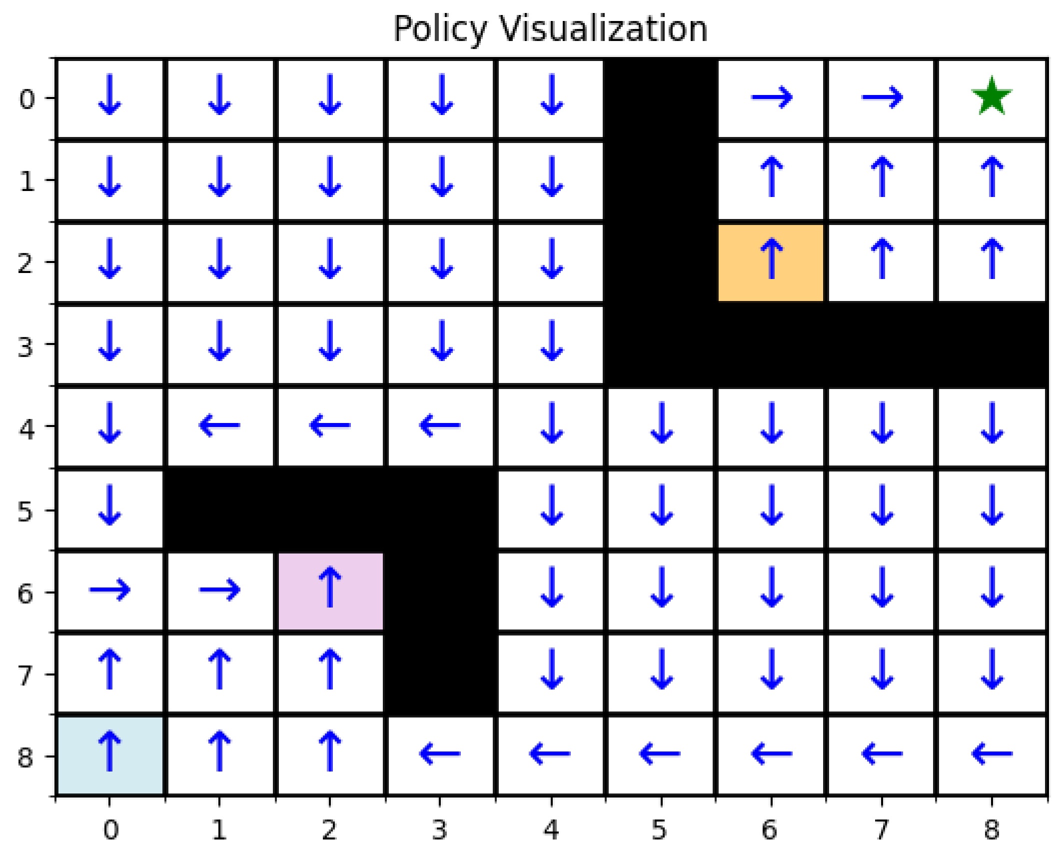
Optimal Policy: The optimal policy obtained using both methods is: ['Eat_Food', 'Eat_Food', 'Eat_Food', 'Attend_Class', 'Attend_Class', 'Attend_Class']. This indicates that for most states, the optimal action is to "Eat_Food" (in states Hostel_Attending_Classes and Hostel_Hungry), while "Attend_Class" is the optimal action for states Academic_Building_Attending_Classes, Academic_Building_Hungry, Canteen_Attending_Classes, and Canteen_Hungry.

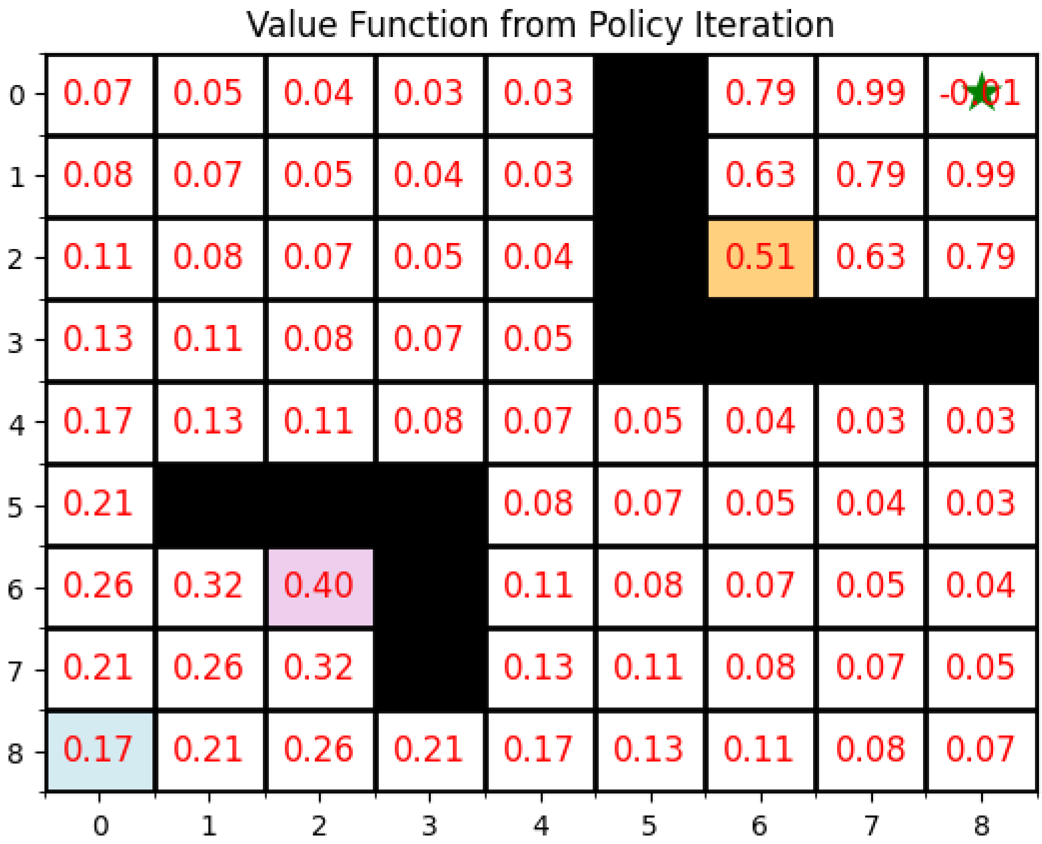
Policy Action Rationalization: Eat_Food is optimal for states related to the Hostel because eating food is essential when the student is in the hostel and hungry. It is also the best option when in the hostel and attending classes (considering the high probability of the transition). Attend_Class is optimal for the Academic Building and Canteen when attending classes as these locations have high rewards for attending classes.

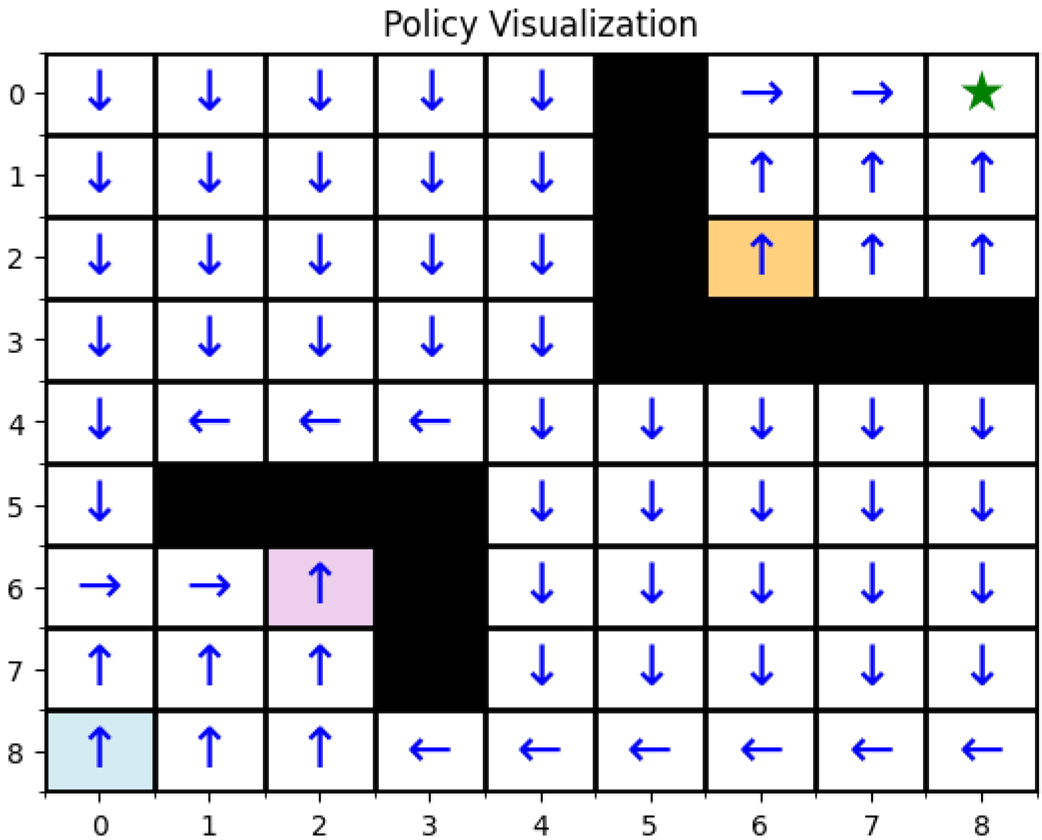
Question 2:

Question 2 has been solved in code and the resulting optimal policies are represented using quiver plots as follows:






Value Function from Policy Iteration
0.07 0.05 0.04 0.03 0.03 0.79 0.99 -0.01
0.08 0.07 0.05 0.04 0.03 0.63 0.79 0.99
0.11 0.08 0.07 0.05 0.04 0.51 0.63 0.79
0.13 0.11 0.08 0.07 0.05
0.17 0.13 0.11 0.08 0.07 0.05 0.04 0.03 0.03
0.21 0.08 0.07 0.05 0.04 0.03
0.26 0.32 0.40 0.11 0.08 0.07 0.05 0.04
0.21 0.26 0.32 0.13 0.11 0.08 0.07 0.05
0.17 0.21 0.26 0.21 0.17 0.13 0.11 0.08 0.07
0 1 2 3 4 5 6 7 8


Policy Visualization
0 1 2 3 4 5 6 7 8