Multi Agent Reinforcement Learning

Assignment 1

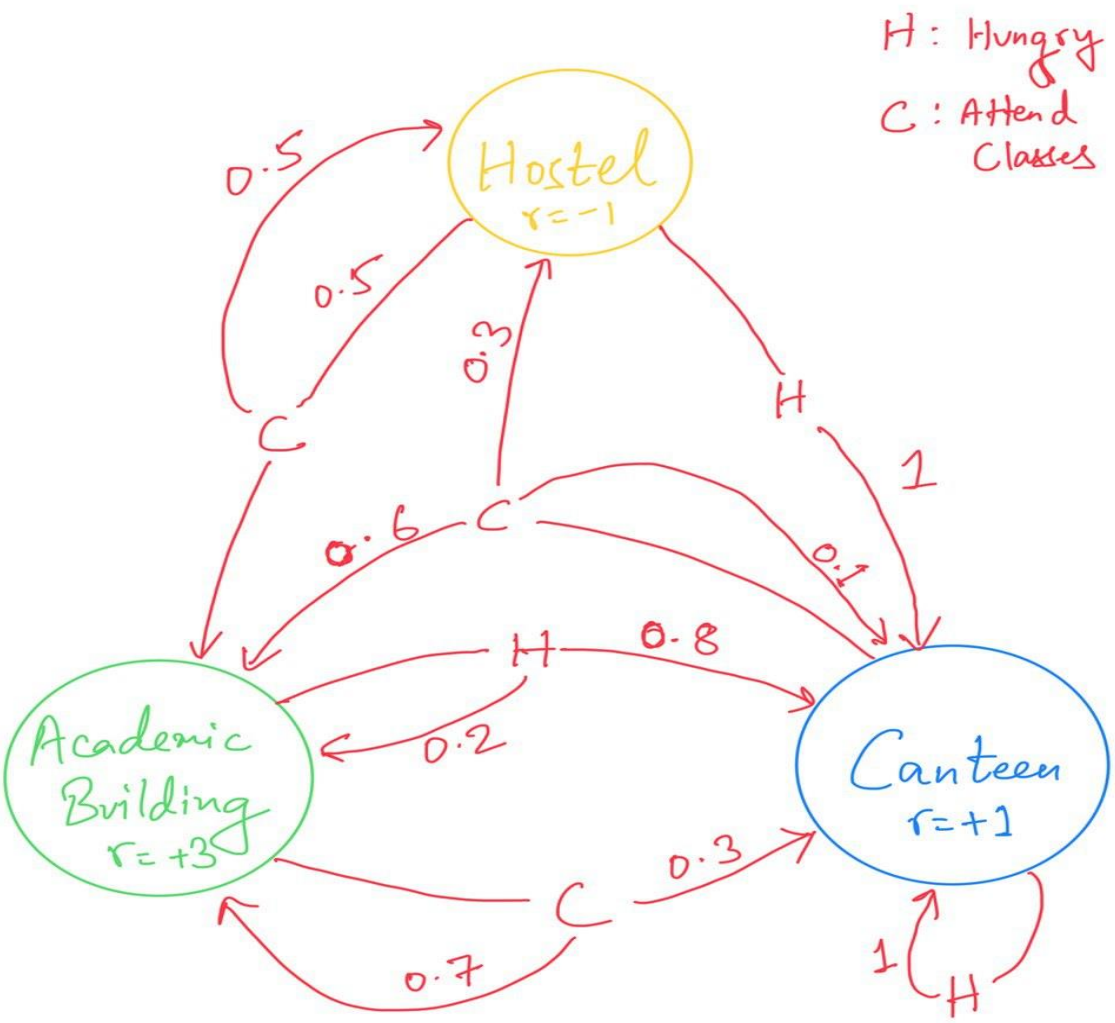
Report

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, it can be observed that both the methods produce near identical results, thereby implying that both have converged to the same optimal policy. And the optimal policy is to:
Eat_Food in the Hostel and when Hungry. Attend_Class in the Academic Building and Canteen

Question 2:

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