INDIAN INSTITUTE OF INFORMATION TECHNOLOGY VADODARA

GOVERNMENT ENGINEERING COLLEGE, SECTOR-28, GANDHINAGAR

GUJARAT - 382028

Artificial Intelligence
Lab Assignment 4

Markov Decision Process and Dynamic Programming

Submitted by:
Group 4
Sandhya Yadav (201652023)
Payal Mahisaniya (201652015)
Pooja Gurjar (201652016)

1 Problem Statement

Suppose that an agent is situated in the 4x3 environment as shown in Figure . Beginning in the start state, it must choose an action at each time step. The interaction with the environment terminates when the agent reaches one of the goal states, marked +1 or -1. We assume that the environment is fully observable, so that the agent always knows where it is. You may decide to take following four actions in every state: Up, Down, Left and Right. However, the environment is stochastic, that means the action that you take may not lead you to desired state. Each action achieves the intended effect with probability 0.8, but the rest of the time, the action moves the agent at right angles to the intended direction with equal probabilities. Furthermore, if the agent bumps into a wall, it stays in the same square. The immediate reward for moving to any state (s) except for the terminal states S+ is r(s)= -0.04. And the reward for moving to terminal states is +1 and -1 respectively. Find the value function corresponding to the optimal policy using value iteration.

1.1 Problem 1 -

Find the value functions corresponding optimal policy for the following:

(a) r(s)=-2
(b) r(s)=0.1
(c) r(s)=0.02
(d) r(s)=1

Observation -

```
Policy
    2
    3
    2
    1
    0
    2
    1
    3
    2
    1
    3
    1


Action Value
   -0.0360
   -0.0320
   -0.0360
   -0.0320
     -Inf
   -0.0320
   -0.0360
   -0.1320
    0.7960
   -0.1320
    0.7960
   -0.1320
|
```

(a) r(s)=-2

gbike3.m × **gbike.m** × r × **policy** × V ×

12×1 double

	1
1	2
2	3
3	2
4	1
5	0
6	2
7	2
8	3
9	2
10	3
11	3
12	4

gbike3.m × **gbike.m** × r × policy × **V** ×

12×1 double

	1
1	-1.8000
2	-1.6000
3	-1.8000
4	-1.6000
5	-Inf
6	-1.6000
7	-1.8000
8	-1.2000
9	0.6000
10	-1
11	0.6000
12	-1

(b) r(s)=0.1

gbike3.m × **gbike.m** × policy × **V** ×

12×1 double

	1
1	0.0900
2	0.0800
3	0.0900
4	0.0800
5	-Inf
6	0.0800
7	0.1000
8	-0.0200
9	0.8100
10	-0.0200
11	0.8100
12	-0.0200

12×1 double

	1
1	2
2	3
3	2
4	2
5	0
6	2
7	1
8	3
9	2
10	1
11	3
12	1

(c) r(s)=0.02

12×1 double

	1
1	0.0180
2	0.0160
3	0.0180
4	0.0160
5	-Inf
6	0.0160
7	0.0200
8	-0.0840
9	0.8020
10	-0.0840
11	0.8020
12	-0.0840

policy ×   V ×

12×1 double

	1
1	2
2	3
3	2
4	2
5	0
6	2
7	1
8	3
9	2
10	1
11	3
12	1

(d) r(s)=1

V ×   policy ×

12×1 double

	1
1	0.9000
2	0.8000
3	0.9000
4	0.8000
5	-Inf
6	0.8000
7	1
8	0.7000
9	1
10	0.7000
11	0.9000
12	0.7000

policy	1
1	2
2	3
3	2
4	2
5	0
6	2
7	3
8	3
9	4
10	1
11	3
12	1

1.2 Problem 2 -

[Gbike bicycle rental] You are managing two locations for Gbike. Each day, some number of customers arrive at each location to rent bicycles. If you have a bike available, you rent it out and earn INR 10 from Gbike. If you are out of bikes at that location, then the business is lost. Bikes become available for renting the day after they are returned. To help ensure that cars are available where they are needed, you can move them between the two locations overnight, at a cost of INR 2 per bike moved. Assumptions: Assume that the number of bikes requested and returned at each locations are Poisson random variables. Expected numbers of rental requests are 3 and 4 and returns are 3 and 2 at the first and second locations respectively. No more than 20 bikes can be parked at either of the locations. You may move maximum 5 bikes from one location to the other in one night. Consider the discount rate to be 0.9.

Formulate the continuing finite MDP, where time steps are days, the state is the number of cars at each location at the end of the day, and the actions are the net number of bikes moved between the two locations overnight.
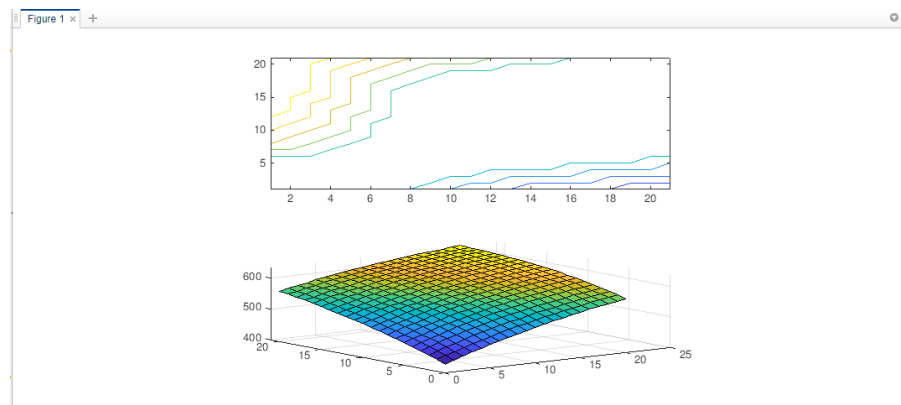
Observation -

1.3 Problem 3 -

Write a program for policy iteration and re-solve gbike bicycle rental problem with the following changes. One of your employee at the first location rides a bus home each night and lives near the second location. She is happy to shuttle one bike to the second location for free. Each additional bike still costs INR 2, as do all bikes moved in the other direction. In addition, you have limited parking space at each location. If more than 10 bikes are kept overnight at a location (after any moving of cars), then an additional cost of INR 4 must be incurred to use a second parking lot (independent of how many cars are kept there).

Observation -
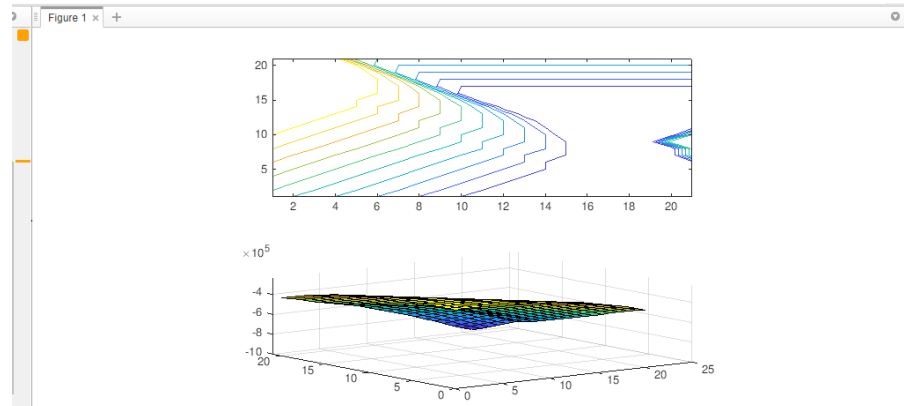
21x21 double

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0	0	-1	-1	-2	-2	-3	-3	-4	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-6
2	1	0	0	-1	-1	-2	-2	-3	-3	-4	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-6
3	1	1	0	0	-1	-1	-2	-2	-3	-3	-4	-4	-5	-5	-5	-5	-5	-5	-5	-5	-6
4	2	1	1	0	0	-1	-1	-2	-2	-3	-3	-4	-4	-5	-5	-5	-5	-5	-5	-5	-5
5	2	2	1	1	0	0	-1	-1	-2	-2	-3	-3	-4	-4	-5	-5	-5	-5	-5	-5	-5
6	3	2	2	1	1	0	0	-1	-1	-2	-2	-3	-3	-4	-5	-5	-5	-5	-5	-5	-5
7	3	3	2	2	1	1	0	0	-1	-1	-2	-2	-3	-3	-4	-4	-5	-5	-5	-5	1
8	4	3	3	2	2	1	1	0	0	-1	-1	-2	-2	-3	-3	-4	-5	-5	-5	-5	1
9	4	4	3	3	2	2	1	1	0	0	-1	-1	-2	-2	-3	-4	-5	-5	-5	2	1
10	5	4	4	3	3	2	2	1	1	0	-1	-1	-2	-3	-5	-5	-5	-5	-5	-5	1
11	5	5	4	4	3	3	2	2	1	0	0	-1	-2	-4	-5	-5	-5	-5	-5	-5	-5
12	6	5	5	4	4	3	3	2	1	1	0	-1	-3	-5	-5	-5	-5	-5	-5	-5	-5
13	6	6	5	5	4	4	3	2	2	1	0	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5
14	6	6	6	5	5	4	3	3	2	1	-1	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
15	6	6	6	6	5	4	4	3	2	0	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
16	6	6	6	6	5	5	4	3	1	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
17	6	6	6	6	6	5	4	2	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4
18	6	6	6	6	6	5	3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
19	6	6	6	6	6	4	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
20	6	6	6	6	5	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
21	6	6	6	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0