MACHINE DATA AND LEARNING

ASSIGNMENT 2: PART 1

VALUE ITERATION ALGORITHM

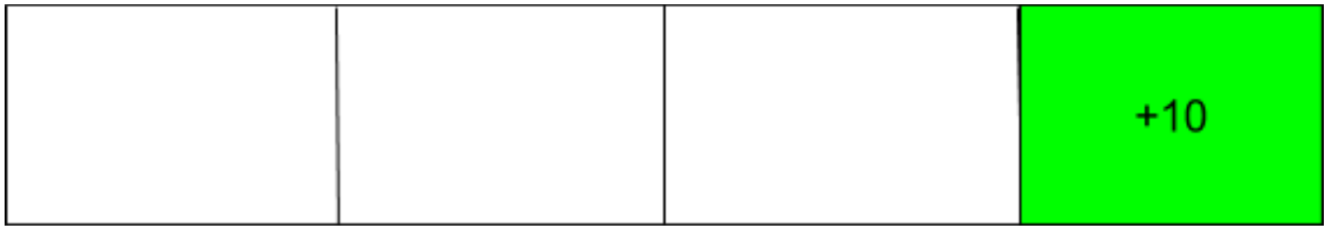
According to the value iteration algorithm , the utility $U_t(i)$ of any state i , at any given time step t is given by,

At time t = 0 , $U_t(i) = 0$ At other time , $U_t(i) = \max_a[R(i , a) + \gamma \Sigma_j U_{t-1}(j) P(j|i , a)]$

The above equation is called the Bellman Update equation. Here, we repeat this equation till the model converges. (max change in the utility of all the states is less than δ, the bellman factor)

TASK

Dry-run/apply the value iteration algorithm on the following scenario to obtain the optimal policy and the state reward values corresponding to it:



GIVEN PARAMETERS

STATES S

The different states as per the question is as follows:

- S_0
- S_1
- S_2
- S_3

Here , S_3 is an absorbant state with an initial utility value of 10.

ACTIONS A

The different supported actions are as follows:

- move left (l)

Here the agent moves left with a probability of 0.8 and right with a probability of 0.2

- move right (r)

Here the agent moves left with a probability of 0.2 and right with a probability of 0.8

TRANSITION FUNCTION P

For action l

T0 → FROM ↓	**0**	**1**	**2**	**3**
0	0.8	0.2	0	0
1	0.8	0	0.2	0
2	0	0.8	0	0.2

For action r

T0 → FROM ↓	**0**	**1**	**2**	**3**
0	0.2	0.8	0	0
1	0.2	0	0.8	0
2	0	0.2	0	0.8

REWARD FUNCTION

R(s , a) = -1

From	**l**	**r**
0	-1	-1
1	-1	-1
2	-1	-1

DISCOUNT FACTOR

$\gamma = 0.25$

BELLMAN FACTOR

$\delta = 0.01$

UTITLITY CALCULATION OF EACH STATE AT EACH TIME PERIOD

At time 0 the current state utilities are 0 0 0 10

For state 0

To	l	r
0	0.8 * 0 = 0	0.2 * 0 = 0
1	0.2 * 0 = 0	0.8 * 0 = 0
2	0 * 0 = 0	0 * 0 = 0
3	0 * 10 = 0	0 * 10 = 0
Net Value	0 * 0.25 = 0	0 * 0.25 = 0

For state 1

To	l	r
0	0.8 * 0 = 0	0.2 * 0 = 0
1	0 * 0 = 0	0 * 0 = 0
2	0.2 * 0 = 0	0.8 * 0 = 0
3	0 * 10 = 0	0 * 10 = 0
Net Value	0 * 0.25 = 0	0 * 0.25 = 0

For state 2

To	l	r
0	0 * 0 = 0	0 * 0 = 0
1	0.8 * 0 = 0	0.2 * 0 = 0
2	0 * 0 = 0	0 * 0 = 0
3	0.2 * 10 = 2	0.8 * 10 = 8
Net Value	2* 0.25 = 0.5	8* 0.25 = 2

Maximum Difference: 1.0

At time 1 the current state utilities are -1,-1,1,10

For state 0

To	l	r
0	0.8 * -1= -0.8	0.2 * -1= -0.2
1	0.2 * -1= -0.2	0.8 * -1= -0.8
2	0 * 1= 0	0 * 1= 0
3	0 * 10 = 0	0 * 10 = 0

To	l	r
Net Value	-1* 0.25 = -0.25	-1* 0.25 = -0.25

For state 1

To	l	r
0	0.8 * -1= -0.8	0.2 * -1= -0.2
1	0 * -1= 0	0 * -1= 0
2	0.2 * 1= 0.2	0.8 * 1= 0.8
3	0 * 10 = 0	0 * 10 = 0
Net Value	-0.6 * 0.25 = -0.15	0.6 * 0.25 = 0.15

For state 2

To	l	r
0	0 * -1= 0	0 * -1= 0
1	0.8 * -1= -0.8	0.2 * -1= -0.2
2	0 * 1= 0	0 * 1= 0
3	0.2 * 10 = 2	0.8 * 10 = 8
Net Value	1.2 * 0.25 = 0.3	7.8 * 0.25 = 1.95

Maximum Difference: 0.25

At time 2 the current state utilities are -1.25 -0.85 0.95 10

For state 0

To	l	r
0	0.8 * -1.25 = -1	0.2 * -1.25 = -0.25
1	0.2 * -0.85 = -0.17	0.8 * -0.85 = -0.68
2	0 * 0.95 = 0	0 * 0.95 = 0
3	0 * 10 = 0	0 * 10 = 0
Net Value	-1.17 * 0.25 = -0.292	-0.93 * 0.25 = -0.233

For state 1

To	l	r

To	l	r
0	0.8 * -1.25 = -1	0.2 * -1.25 = -0.25
1	0 * -0.85 = 0	0 * -0.85 = 0
2	0.2 * 0.95 = 0.19	0.8 * 0.95 = 0.76
3	0 * 10 = 0	0 * 10 = 0
Net Value	-0.81 * 0.25 = -0.203	0.51 * 0.25 = 0.128

For state 2

To	l	r
0	0 * -1.25 = 0	0 * -1.25 = 0
1	0.8 * -0.85 = -0.68	0.2 * -0.85 = -0.17
2	0 * 0.95 = 0	0 * 0.95 = 0
3	0.2 * 10 = 2	0.8 * 10 = 8
Net Value	1.32 * 0.25 = 0.33	7.83 * 0.25 = 1.958

Maximum Difference: 0.023

At time 3 the current state utilities are -1.232 -0.873 0.958 10

For state 0

To	l	r
0	0.8 * -1.232 = -0.986	0.2 * -1.232 = -0.246
1	0.2 * -0.873 = -0.175	0.8 * -0.873 = -0.698
2	0 * 0.958 = 0	0 * 0.958 = 0
3	0 * 10 = 0	0 * 10 = 0
Net Value	-1.161 * 0.25 = -0.29	-0.945 * 0.25 = -0.236

For state 1

To	l	r
0	0.8 * -1.232 = -0.986	0.2 * -1.232 = -0.246
1	0 * -0.873 = 0	0 * -0.873 = 0
2	0.2 * 0.958 = 0.192	0.8 * 0.958 = 0.766
3	0 * 10 = 0	0 * 10 = 0

To	l	r
Net Value	-0.794 * 0.25 = -0.199	0.52 * 0.25 = 0.13

For state 2

To	l	r
0	0 * -1.232 = 0	0 * -1.232 = 0
1	0.8 * -0.873 = -0.698	0.2 * -0.873 = -0.175
2	0 * 0.958 = 0	0 * 0.958 = 0
3	0.2 * 10 = 2	0.8 * 10 = 8
Net Value	1.302 * 0.25 = 0.326	7.825 * 0.25 = 1.956

Maximum Difference: 0.004

At time 4 the current state utilities are -1.236,-0.87,0.956,10

For state 0

To	l	r
0	0.8 * -1.236 = -0.989	0.2 * -1.236 = -0.247
1	0.2 * -0.87 = -0.174	0.8 * -0.87 = -0.696
2	0 * 0.956 = 0	0 * 0.956 = 0
3	0 * 10 = 0	0 * 10 = 0
Net Value	-1.163 * 0.25 = -0.291	-0.943 * 0.25 = -0.236

For state 1

To	l	r
0	0.8 * -1.236 = -0.989	0.2 * -1.236 = -0.247
1	0 * -0.87 = 0	0 * -0.87 = 0
2	0.2 * 0.956 = 0.191	0.8 * 0.956 = 0.765
3	0 * 10 = 0	0 * 10 = 0
Net Value	-0.798 * 0.25 = -0.199	0.518 * 0.25 = 0.129

For state 2

To	l	r

To	l	r
0	0 * -1.236 = 0	0 * -1.236 = 0
1	0.8 * -0.87 = -0.696	0.2 * -0.87 = -0.174
2	0 * 0.956 = 0	0 * 0.956 = 0
3	0.2 * 10 = 2	0.8 * 10 = 8
Net Value	1.304 * 0.25 = 0.326	7.826 * 0.25 = 1.956

Maximum Difference: 0.0

At time 5 the current state utilities are -1.236 -0.871 0.956 10

For state 0 : r

To	l	r
0	0.8 * -1.236 = -0.989	0.2 * -1.236 = -0.247
1	0.2 * -0.871 = -0.174	0.8 * -0.871 = -0.696
2	0 * 0.956 = 0	0 * 0.956 = 0
3	0 * 10 = 0	0 * 10 = 0
Net Value	-1.163 * 0.25 = -0.291	-0.944 * 0.25 = -0.236

For state 1 : r

To	l	r
0	0.8 * -1.236 = -0.989	0.2 * -1.236 = -0.247
1	0 * -0.871 = 0	0 * -0.871 = 0
2	0.2 * 0.956 = 0.191	0.8 * 0.956 = 0.765
3	0 * 10 = 0	0 * 10 = 0
Net Value	-0.797 * 0.25 = -0.199	0.518 * 0.25 = 0.13

For state 2 : r

To	l	r
0	0 * -1.236 = 0	0 * -1.236 = 0
1	0.8 * -0.871 = -0.696	0.2 * -0.871 = -0.174
2	0 * 0.956 = 0	0 * 0.956 = 0
3	0.2 * 10 = 2.0	0.8 * 10 = 8.0

To	l	r
Net Value	1.304 * 0.25 = 0.326	7.826 * 0.25 = 1.956

Maximum Difference: 0.0

FINAL VALUES AT EACH TIME PERIOD

t	$U_t(S_0)$	$U_t(S_1)$	$U_t(S_2)$	$U_t(S_3)$	Difference
0	0	0	0	10	0
1	-1	-1	1	10	1
2	-1.25	-0.85	0.95	10	0.25
3	-1.232	-0.873	0.958	10	0.023
4	-1.236	-0.87	0.956	10	0.004
5	-1.236	-0.871	0.956	10	0
6	-1.236	-0.87	0.956	10	0
7	-1.236	-0.871	0.956	10	0
8	-1.236	-0.87	0.956	10	0
9	-1.236	-0.87	0.956	10	0