Intro To Artificial Intelligence - Exercise 5

Eran Ston (206704512) and Oded Vaalany (208230474)

July 14, 2024

1 Value Iteration

1.1

Now we want to use MDP with the following transition probabilities:

- S = {s1, s2, s3, s4, s5, s6, s7, s8, s9}
- A = {U, D, L, R}
- $P(s|s', A) = 1$ where we need to do A from s to s'
- $R(s) = -0.05$ for all $s \notin \{s_5, s_7, s_9\}$
- $R(s_5) = -10$
- $R(s_7) = 15$
- $R(s_9) = 30$
- $\gamma = 0.99$

Values of states after each iteration:

step	s1	s2	s3	s4	s5	s6	s7	s8	s9
0	0	0	0	0	0	0	0	0	0
1	-0.05	-0.05	-0.05	-0.05	-10	-0.05	15	-0.05	30
2	-0.0995	-0.0995	-0.0995	14.8	-10	29.65	15	29.65	30
3	14.602	-0.148505	29.3035	14.8	-10	29.65	15	29.65	30
4	14.602	28.960465	29.3035	14.8	-10	29.65	15	29.65	30
5	28.6208	28.96046	29.3035	14.8	-10	29.65	15	29.65	30
6	28.6208	28.96046	29.3035	28.284592	-10	29.65	15	29.65	30
7	28.6208	28.96046	29.3035	28.284592	-10	29.65	15	29.65	30

Optimal actions after each iteration:

step	s1	s2	s3	s4	s5	s6	s7	s8	s9
0	L	L	L	L	L	L	L	L	L
1	R	R	U	U	L	U	L	R	L
2	U	R	U	U	L	U	L	R	L
3	U	R	U	U	L	U	L	R	L
4	U	R	U	U	L	U	L	R	L
5	R	R	U	U	L	U	L	R	L
6	R	R	U	D	L	U	L	R	L
7	R	R	U	D	L	U	L	R	L

The optimal policy is:

←	→	←
↓	←	↑
→	→	↑

1.2

Now we want to use stochastic MDP with the following transition probabilities:

- S = {s1, s2, s3, s4, s5, s6, s7, s8, s9}
- A = {U, D, L, R}
- $P(s|s', A) = 0.9$ where we need to do A from s to s'
- $P(s'|s', A) = 0.9$ where A is not legitimate move for the state
- $P(s|s', A) = \frac{0.1}{\text{number of neighbors -1}}$ where A is not possible from s to s'(neighbors)
- $R(s) = -0.05$ for all $s \notin \{s_5, s_7, s_9\}$
- $R(s_5) = -10$
- $R(s_7) = 15$
- $R(s_9) = 30$
- $\gamma = 0.99$

Values of states after each iteration:

step	s1	s2	s3	s4	s5	s6	s7	s8	s9
0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1	-0.05000	-0.05000	-0.05000	-0.05000	-10.00000	-0.05000	15.00000	-0.05000	30.00000
2	-0.09950	-0.42785	-0.09950	12.81752	-10.00000	26.18252	15.00000	26.92750	30.00000
3	11.32806	-0.63858	23.23627	12.81507	-10.00000	26.18007	15.00000	26.92750	30.00000
4	11.30501	20.71926	23.21323	13.38074	-10.00000	27.33520	15.00000	26.92750	30.00000
5	19.73555	20.69758	26.35687	13.37960	-10.00000	27.33405	15.00000	26.92750	30.00000
6	19.71613	23.91588	26.35370	17.78188	-10.00000	27.48966	15.00000	26.92750	30.00000
7	23.01945	23.91210	26.81096	17.76457	-10.00000	27.48951	15.00000	26.92750	30.00000

Optimal actions after each iteration:

step	s1	s2	s3	s4	s5	s6	s7	s8	s9
0	L	L	L	L	L	L	L	L	L
1	L	L	L	L	L	L	L	L	L
2	R	D	L	U	L	U	L	R	L
3	U	L	U	U	L	U	L	R	L
4	U	R	U	U	L	U	L	R	L
5	R	R	U	U	L	U	L	R	L
6	R	R	U	D	L	U	L	R	L
7	R	R	U	D	L	U	L	R	L

The optimal policy is:

$\leftarrow$	$\rightarrow$	$\leftarrow$
$\downarrow$	$\leftarrow$	$\uparrow$
$\rightarrow$	$\rightarrow$	$\uparrow$

The values of the optimal policy in the stochastic MDP are lower the values of the optimal policy in the deterministic MDP. This is because the stochastic MDP has a probability of transitioning to a state that is not the desired state, which causes the values to be lower.

1.3

Given the following policy: $a_1 = \uparrow |a_2 = \rightarrow |a_3 = \uparrow |a_4 = \uparrow |a_5 = *|a_6 = \uparrow |a_7 = *|a_8 = \leftarrow |a_9 = *$