

# Toelichting antwoorden AS 2.2

Sam van Roosmalen

30-05-2023

<b>Value Iteration.....</b>	<b>2</b>
Policy table:.....	2
<b>TD Learning.....</b>	<b>3</b>
gamma = 1.....	3
Policy table:.....	3
gamma = 0.5.....	4
Policy table:.....	4
gamma = 0.9.....	5
Policy table:.....	5
<b>Sarsa Control.....</b>	<b>6</b>
gamma = 1.....	6
Policy table:.....	6
gamma = 0.5.....	7
Policy table:.....	7
gamma = 0.9.....	7
Policy table:.....	8
<b>Q Learning.....</b>	<b>9</b>
gamma = 1.....	9
Policy table:.....	9
gamma = 0.5.....	10
Policy table:.....	10
gamma = 0.9.....	11
Policy table:.....	11

# Value Iteration

28.0	-1	33.0	-1	18.0	-1	0.0	40
15.0	38.0	2.0	21.0	25.0	3.0	40.0	38.0
13.0	2.0	-1	8.0	29.0	-1	26.0	0.0
30.0	37.0	32.0	25.0	38.0	37.0	38.0	39.0
8.0	18.0	-1	16.0	29.0	-1	40.0	0.0
4.0	36.0	20.0	10.0	37.0	6.0	40.0	36.0
40.0	0.0	10	8.0	4.0	-2	38.0	4.0
0.0	0.0	0.0	23.0	36.0	8.0	1.0	35.0
0.0			31.0			35.0	

Policy table:

R = Right

L = Left

U = Up

D = Down

['R', 'R', 'R', 'D']

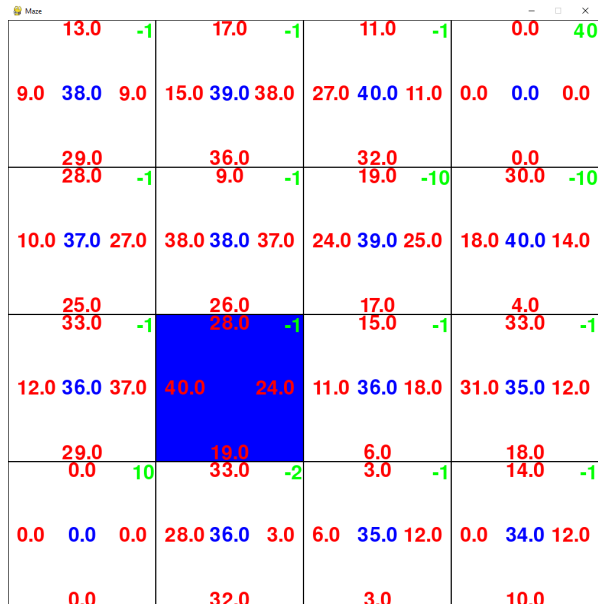
['R', 'U', 'U', 'U']

['R', 'U', 'L', 'L']

['D', 'U', 'U', 'L']

# TD Learning

gamma = 1



Policy table:

R = Right

L = Left

U = Up

D = Down

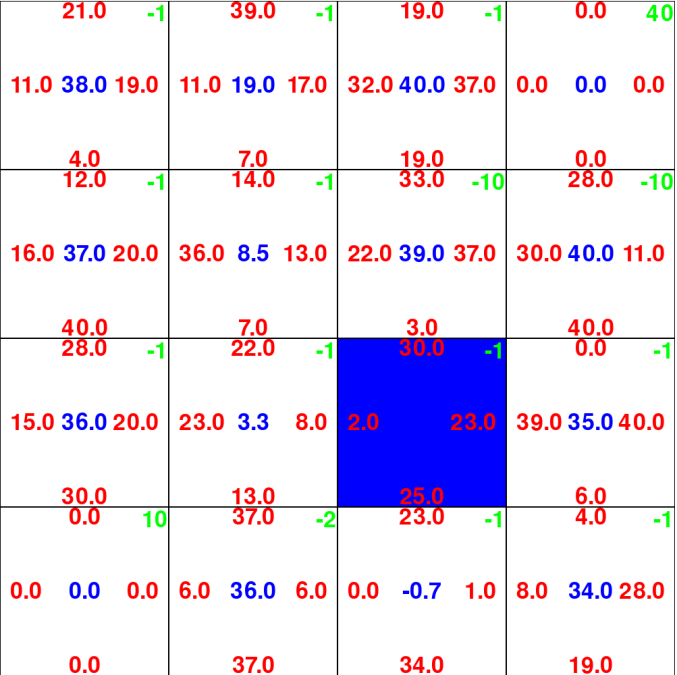
['R', 'R', 'R', 'D']

['R', 'U', 'U', 'U']

['R', 'U', 'L', 'L']

['D', 'U', 'U', 'L']

gamma = 0.5



21.0	-1	39.0	-1	19.0	-1	0.0	40
11.0	38.0	19.0	11.0	19.0	17.0	32.0	40.0
4.0		7.0		19.0			0.0
12.0	-1	14.0	-1	33.0	-10		28.0
16.0	37.0	20.0	36.0	8.5	13.0	22.0	39.0
40.0		7.0		3.0			40.0
28.0	-1	22.0	-1	30.0	-1		0.0
15.0	36.0	20.0	23.0	3.3	8.0	2.0	23.0
30.0		13.0		25.0			6.0
0.0	10	37.0	-2	23.0	-1		4.0
0.0	0.0	0.0	6.0	36.0	6.0	0.0	-0.7
						1.0	8.0
0.0		37.0		34.0			19.0

Policy table:

R = Right

L = Left

U = Up

D = Down

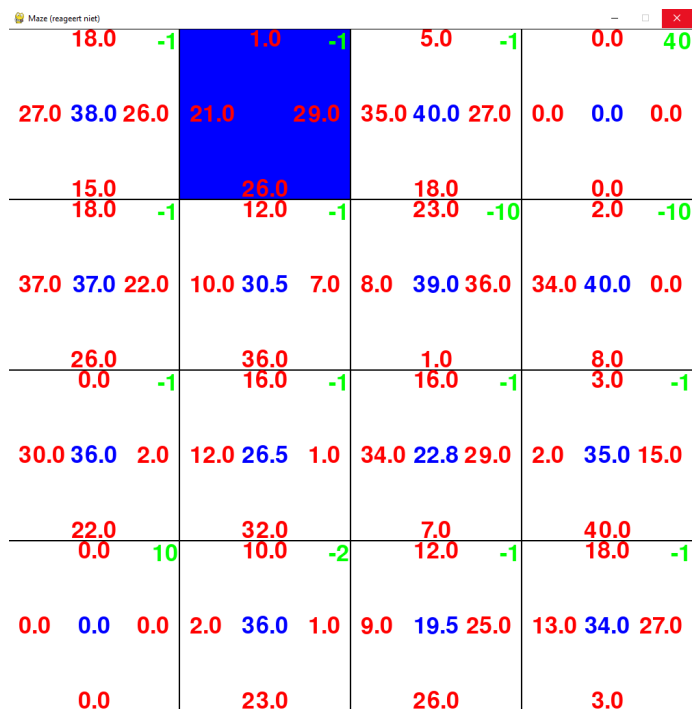
['R', 'R', 'R', 'D']

['R', 'U', 'U', 'U']

['R', 'U', 'L', 'L']

['D', 'U', 'U', 'L']

gamma = 0.9



18.0	-1	1.0	-1	5.0	-1	0.0	40
27.0	38.0	26.0	21.0	29.0	35.0	40.0	27.0
15.0		25.0		18.0		0.0	
18.0	-1	12.0	-1	23.0	-10	2.0	-10
37.0	37.0	22.0	10.0	30.5	7.0	8.0	39.0
26.0		36.0		1.0		8.0	
0.0	-1	16.0	-1	16.0	-1	3.0	-1
30.0	36.0	2.0	12.0	26.5	1.0	34.0	22.8
22.0		32.0		7.0		40.0	
0.0	10	10.0	-2	12.0	-1	18.0	-1
0.0	0.0	0.0	2.0	36.0	1.0	9.0	19.5
0.0		23.0		26.0		3.0	

Policy table:

R = Right

L = Left

U = Up

D = Down

['R', 'R', 'R', 'D']

['R', 'U', 'U', 'U']

['R', 'U', 'L', 'L']

['D', 'U', 'U', 'L']

# Sarsa Control

gamma = 1

Max

3.6	-1	4.8	-1	11.6	-1	0.0	4.0				
3.4	0.0	4.9	3.9	0.0	14.2	5.8	0.0	40.0	0.0	0.0	0.0
3.4		2.4		0.3		0.0					
3.6	-1	7.3	-1	14.9	-10	40.0	-10				
2.7	0.0	2.4	3.1	0.0	-0.8	3.2	0.0	8.0	-1.7	0.0	5.4
2.7		2.7		3.4		0.5					
2.9	-1	2.6	-1	0.1	-1	-1.4	-1				
3.6	0.0	3.0	3.2	0.0	0.3	1.7	0.0	-0.5	-0.9	0.0	-1.0
10.0		2.8		-0.0		-1.0					
0.0	-10	1.0	-2	-0.0	-1	-1.1	-1				
0.0	0.0	0.0	10.0	0.2	1.3	0.0	-0.8	-0.6	0.0	-1.1	
0.0		2.5		-0.2		-1.2					

Policy table:

R = Right

L = Left

U = Up

D = Down

['R', 'R', 'R', '-']

['U', 'U', 'U', 'U']

['D', 'L', 'L', 'L']

['-', 'L', 'L', 'L']

gamma = 0.5

Maze (reageert niet)											
-1.8	-1	-1.3	-1	0.3	-1	0.0	40				
-1.8	0.0	-1.8	-1.8	0.0	2.9	-1.3	0.0	40.0	0.0	0.0	0.0
-1.5		-2.5		-10.3		0.0					
-1.6	-1	-1.7	-1	5.2	-10	40.0	-10				
-1.5	0.0	-2.1	-1.8	0.0	-6.4	-1.7	0.0	-0.8	-6.9	0.0	-3.0
0.2		-1.8		-1.7		0.1					
-1.7	-1	-2.5	-1	-9.1	-1	-6.6	-1				
1.0	0.0	-1.9	0.7	0.0	-2.8	-1.6	0.0	-2.2	-3.3	0.0	-2.5
10.0		-2.2		-2.2		-2.2					
0.0	10	-2.0	-2	-2.5	-1	-2.7	-1				
0.0	0.0	0.0	10.0	0.0	-2.0	-1.8	0.0	-2.2	-2.1	0.0	-2.2
0.0		-1.5		-2.2		-2.2					

Policy table:

R = Right

L = Left

U = Up

D = Down

['R', 'R', 'R', '-']

['U', 'U', 'U', 'U']

['D', 'L', 'L', 'L']

['-', 'L', 'L', 'L']

gamma = 0.9

Maze

-1.5	-1	1.7	-1	21.1	-1	0.0	40				
-1.3	0.0	-0.8	-1.6	0.0	7.3	-0.3	0.0	40.0	0.0	0.0	0.0
-1.2			-2.7			-8.9			0.0		
-1.1	-1		0.2	-1		9.6	-10		40.0	-10	
-1.3	0.0	-1.4	-1.5	0.0	-5.6	0.7	0.0	-5.3	-6.3	0.0	-5.4
0.8			-1.9			-2.0			-2.9		
-1.6	-1		-1.6	-1		-9.1	-1		-2.4	-1	
-0.5	0.0	-1.7	0.1	0.0	-2.9	-2.7	0.0	-2.9	-5.8	0.0	-5.2
10.0			-2.3			-2.8			-5.1		
0.0	10		-1.3	-2		-4.4	-1		-5.1	-1	
0.0	0.0	0.0	10.0	0.0	-3.4	0.0	0.0	-4.1	-4.0	0.0	-4.5
0.0			0.5			-3.5			-4.7		

Policy table:

R = Right

L = Left

U = Up

D = Down

['R', 'R', 'R', '-']

['D', 'U', 'U', 'U']

['D', 'L', 'L', 'U']

['-', 'L', 'L', 'L']



# Q Learning

gamma = 1

Maze (reageert niet)

36.9	-1	38.0	-1	39.0	-1	0.0	40
34.5	0.0	38.0	37.0	39.0	38.0	0.0	40.0
32.0		37.0		29.0		0.0	
34.9	-1	38.0	-1	39.0	-10	40.0	-10
32.7	0.0	37.0	36.0	0.0	29.0	37.0	0.0
34.8		36.0		34.8		28.3	
33.7	-1	37.0	-1	28.9	-1	26.1	-1
30.8	0.0	36.0	35.0	0.0	35.0	36.0	0.0
12.3		33.9		34.0		33.0	
0.0	10	36.0	-2	35.0	-1	31.5	-1
0.0	0.0	0.0	10.3	0.0	34.0	33.5	0.0
0.0		30.6		34.0		34.0	0.0
						32.7	

Policy table:

R = Right

L = Left

U = Up

D = Down

['R', 'R', 'R', '-']

['R', 'U', 'U', 'U']

['R', 'U', 'L', 'D']

['-', 'U', 'U', 'L']

gamma = 0.5

Maze (reageert niet)											
4.5	-1	8.5	-1	19.0	-1	0.0	40				
4.1	0.0	8.5	3.3	0.0	19.0	8.5	0.0	40.0	0.0	0.0	0.0
2.9		3.3		-0.5		0.0					
3.3	-1	8.5	-1	19.0	-10	40.0	-10				
1.4	0.0	3.4	1.5	0.0	0.4	3.0	0.0	11.0	-0.5	0.0	10.0
4.0		1.2		4.1		4.0					
1.0	-1	3.3	-1	0.0	-1	10.0	-1				
4.0	0.0	1.0	4.0	0.0	1.4	1.3	0.0	4.0	1.0	0.0	4.0
10.0		3.1		1.0		1.0					
0.0	10	1.3	-2	1.6	-1	4.0	-1				
0.0	0.0	0.0	10.0	0.0	0.5	3.0	0.0	1.3	1.0	0.0	1.6
0.0		3.1		0.8		1.6					

Policy table:

R = Right

L = Left

U = Up

D = Down

['R', 'R', 'R', '-']

['D', 'U', 'U', 'U']

['D', 'L', 'R', 'U']

['-', 'L', 'L', 'U']

gamma = 0.9

Maze (reageert niet)

24.3 26.0 21.6 22.4 18.1 16.6 19.8 16.5 10.4 0.0 0.0 0.0	-1 0.0 -1 -1 0.0 -1 -1 0.0 10 0.0	30.5 26.4 26.4 30.5 22.6 22.6 26.4 17.7 12.6 16.0 10.1 11.1	-1 0.0 -1 -1 0.0 -1 -2 0.0	35.0 30.5 21.5 35.0 24.2 19.2 21.5 22.4 16.2 19.2 12.7 16.1	-1 0.0 -10 -1 0.0 -1 -1 0.0	0.0 0.0 0.0 40.0 21.5 22.4 26.0 19.2 19.2 22.4 15.9 18.7	40 0.0 -10 -1 22.4 -1
---	--	--	---	--	--	---	--------------------------------------

Policy table:

R = Right

L = Left

U = Up

D = Down

['R', 'R', 'R', '-']

['R', 'U', 'U', 'U']

['R', 'U', 'R', 'U']

['-', 'R', 'U', 'U']