Toelichting antwoorden AS 2.2

Sam van Roosmalen 30-05-2023

| Value Iteration | 2 |
|-----------------|----|
| Policy table: | 2 |
| TD Learning | 3 |
| gamma = 1 | 3 |
| Policy table: | 3 |
| gamma = 0.5 | 4 |
| Policy table: | 4 |
| gamma = 0.9 | 5 |
| Policy table: | 5 |
| Sarsa Control | 6 |
| gamma = 1 | 6 |
| Policy table: | 6 |
| gamma = 0.5 | 7 |
| Policy table: | 7 |
| gamma = 0.9 | 7 |
| Policy table: | 8 |
| Q Learning | 9 |
| 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 | | | | 3.0 | 40.0 | 38.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | |
| | 40.0 | | | | | | 06.0 | | | 0.0 | |
| | 13.0 2.0 | - | | 8.0 29.0 | -1 | | 26.0 30.0 | 40 | | 0.0 11.0 | 10 |
| | 2.0 | - " | | 29.0 | 1 | | 30.0 | -10 | | 11.0 | -10 |
| | | | | | | | | | | | |
| 30.0 | 37.0 | 32.0 | 25.0 | 38.0 | 37.0 | 38.0 | 39.0 | 18.0 | 11.0 | 40.0 | 28.0 |
| | | | | | | | | | | | |
| | 0.0 | | | 16.0 | | | 400 | | | 2.0 | |
| | 8.0 18.0 | -4 | | 16.0 29.0 | -4 | | 40.0 0.0 | _4 | | 3.0 27.0 | _1 |
| | 10.0 | | | _ 5.0 | | | 0.0 | | | 27.0 | - " |
| | | | | | | | | | | | |
| 4.0 | 36.0 | 20.0 | 10.0 | 37.0 | 6.0 | 40.0 | 36.0 | 24.0 | 30.0 | 35.0 | 34.0 |
| | | | | | | | | | | | |
| | 400 | | | 8 N | | | 38.0 | | | 22 N | |
| | 40.0 0.0 | 10 | | 8.0 4.0 | -2 | | 38.0 4.0 | -1 | | 22.0 21.0 | -1 |
| | | | | | | | | _ | | | - • |
| | | | | | | | | | | | |
| 0.0 | 0.0 | 0.0 | 23.0 | 36.0 | 8.0 | 1.0 | 35.0 | 31.0 | 24.0 | 34.0 | 37.0 |
| | | | | | | | | | | | |
| | 0.0 | | | 31.0 | | | 35.0 | | | 10.0 | |
| | 0.0 | | | J 1.U | | | 55.0 | | | 10.0 | |

Policy table:

R = Right

L = Left

U = Up

D = Down

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

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

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

TD Learning

gamma = 1

| Maze | | | | | | | | | - | □ × |
|------|--------------|------|--------------|------|------|--------------|------|------|--------------|------|
| | 13.0 | -1 | 17.0 | -1 | | 11.0 | -1 | | 0.0 | 40 |
| 9.0 | 38.0 | 9.0 | 15.0 39.0 | 38.0 | 27.0 | 40.0 | 11.0 | 0.0 | 0.0 | 0.0 |
| | 29.0 28.0 | -1 | 36.0 9.0 | -1 | | 32.0 19.0 | -10 | | 0.0 30.0 | -10 |
| 10.0 | 37.0 | 27.0 | 38.0 38.0 | 37.0 | 24.0 | 39.0 | 25.0 | 18.0 | 40.0 | 14.0 |
| | 25.0 33.0 | -1 | 26.0 28.0 | -1 | | 17.0 15.0 | -1 | | 4.0 33.0 | -1 |
| 12.0 | 36.0 | 37.0 | | | 11.0 | 36.0 | 18.0 | 31.0 | 35.0 | 12.0 |
| | 29.0 0.0 | | 19.0 33.0 | | | 6.0 3.0 | | | 18.0 14.0 | |
| | 0.0 | 10 | 33.0 | -2 | | 3.0 | -1 | | 14.0 | -1 |
| 0.0 | 0.0 | 0.0 | 28.0 36.0 | 3.0 | 6.0 | 35.0 | 12.0 | 0.0 | 34.0 | 12.0 |
| | 0.0 | | 32.0 | | | 3.0 | | | 10.0 | |

Policy table:

R = Right

L = Left

U = Up

D = Down

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

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

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

| Maze | | | | | | | | | | _ | □ × |
|------|--------------|------|------|--------------|------|------|--------------------|------|------|-------------|------|
| | 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 | 37.0 | 0.0 | 0.0 | 0.0 |
| | 4.0 12.0 | | | 7.0 | -1 | | 19.0 | -10 | | 0.0 28.0 | 10 |
| | 12.0 | -1 | | 14.0 | -1 | | 33.0 | -10 | | 20.0 | -10 |
| 16.0 | 37.0 | 20.0 | 36.0 | 8.5 | 13.0 | 22.0 | 39.0 | 37.0 | 30.0 | 40.0 | 11.0 |
| | 40.0 | ı | | 7.0 | | | 3.0 | | | 40.0 | |
| | 40.0 28.0 | -1 | | 7.0 22.0 | -1 | | 3.0 30.0 | -1 | | 40.0 0.0 | -1 |
| 15.0 | 36.0 | 20.0 | 23.0 | 3.3 | 8.0 | | | | 39.0 | 35.0 | 40.0 |
| | 30.0 | | | 13.0 | | | 25.0 | | | 6.0 | |
| | 30.0 0.0 | 10 | | 13.0 37.0 | -2 | | 25.0 23.0 | -1 | | 6.0 4.0 | -1 |
| 0.0 | 0.0 | 0.0 | 6.0 | 36.0 | 6.0 | 0.0 | -0.7 | 1.0 | 8.0 | 34.0 | 28.0 |
| | 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']

| | ageert niet) | | | | | | | | | _ | □ X |
|------|--------------|------|------|------------------|------|-------|--------------|------|------|--------------|------|
| | 18.0 | -1 | | 1.0 | -1 | | 5.0 | -1 | | 0.0 | 40 |
| | | | | | | | | | | | |
| 27.0 | 38.0 | 26.0 | | | 29.0 | 25.0 | 40.0 | 27.0 | 0.0 | 0.0 | 0.0 |
| 21.0 | 30.0 | 20.0 | | | 29.0 | 33.0 | 40.0 | 27.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | |
| | 15.0 18.0 | | | 26.0 12.0 | | | 18.0 23.0 | | | 0.0 2.0 | |
| | 18.0 | -1 | | 12.0 | -1 | | 23.0 | -10 | | 2.0 | -10 |
| | | | | | | | | | | | |
| 37.0 | 37.0 | 22 N | 10.0 | 30.5 | 7.0 | 8.0 | 39.0 | 36.0 | 34.0 | 40.0 | 0.0 |
| 07.0 | 07.0 | 22.0 | 10.0 | 00.5 | 7.0 | 0.0 | 03.0 | 00.0 | 04.0 | 40.0 | 0.0 |
| | | | | | | | | | | | |
| | 26.0 0.0 | | | 36.0 16.0 | | | 1.0 16.0 | | | 8.0 3.0 | |
| | 0.0 | -1 | | 16.0 | -1 | | 16.0 | -1 | | 3.0 | -1 |
| | | | | | | | | | | | |
| 30.0 | 36.0 | 2.0 | 12.0 | 26.5 | 1.0 | 3/1 (| 22.8 | 20.0 | 2.0 | 35.0 | 15.0 |
| 30.0 | 30.0 | 2.0 | 12.0 | 20.5 | 1.0 | 54.0 | 22.0 | 29.0 | 2.0 | 33.0 | 13.0 |
| | | | | | | | | | | | |
| | 22.0 0.0 | | | 32.0 10.0 | | | 7.0 12.0 | | | 40.0 18.0 | |
| | 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 | 10.5 | 25.0 | 13.0 | 34.0 | 27.0 |
| 0.0 | 0.0 | 0.0 | 2.0 | 30.0 | 1.0 | 9.0 | 19.5 | 25.0 | 13.0 | J4.0 | 21.0 |
| | | | | | | | | | | | |
| | 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']

Sarsa Control

gamma = 1

| | | | | | | | | | | - | □ × |
|-----|-------------|-----|-----|------------|------|-----|--------------|------|------|--------------|------|
| | 3.6 | -1 | | 4.8 | -1 | | 11.6 | -1 | | 0.0 | 40 |
| 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 3.6 | | | 2.4 7.3 | | | 0.3 | -10 | | 0.0 | |
| | 3.0 | 7 | | 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 | |
| | 2.7 2.9 | -1 | | 2.7 2.6 | -1 | | 3.4 0.1 | -1 | | 0.5 -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 | |
| | 10.0 0.0 | 10 | | 2.8 1.8 | -2 | | -0.0 -0.0 | -1 | | -1.0 -1.1 | -1 |
| 0.0 | 0.0 | 0.0 | | | | 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']

| Maze (rea | geert 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 | | | -25 | | | -10.3 | | | 0.0 | |
| | -1.5 -1.6 | -1 | | -2.5 -1.7 | -1 | | 5.2 | -10 | | 0.0 40.0 | -10 |
| | | | | | | | | | | | |
| 4.5 | | 0.4 | 4.0 | | | 4 - | | | | | |
| -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 -1.7 | | | -1.8 | | | -1.7 | | | 0.1 | |
| | -1.7 | -1 | | -1.8 -2.5 | -1 | | -1.7 -9.1 | -1 | | 0.1 -6.6 | -1 |
| | | | | | | | | | | | |
| 1.0 | 0.0 | -1.9 | 0.7 | 0.0 | -2 B | -1.6 | 0.0 | -2.2 | -33 | 0.0 | -2.5 |
| 1.0 | 0.0 | -1.5 | 0.7 | 0.0 | -2.0 | -1.0 | 0.0 | -2.2 | -0.0 | 0.0 | -2.5 |
| | | | | | | | | | | | |
| | 10.0 0.0 | | | -2.2 -2.0 | | | -2.2 -2.5 | | | -2.2 -2.7 | |
| | 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']

| Maze | | | | | | | | | | _ | □ X |
|------|------|------|------|-----------------|-------|------|------|------|------|------|-------------|
| | -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 / | -1.5 | 0.0 | -5.6 | 0.7 | 0.0 | -5.3 | -6.3 | 0.0 | -5.4 |
| -1.5 | 0.0 | -1.4 | -1.5 | 0.0 | -5.0 | 0.7 | 0.0 | -5.5 | -0.5 | 0.0 | -5.4 |
| | | | | | | | | | | | |
| | 0.8 | _ | | -1.9 | _ | | -2.0 | -1 | | -2.9 | _ |
| | -1.0 | -1 | | -1.0 | -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 | | | -23 | | | -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 /1 | 0.0 | 0.0 | -4.1 | -4.0 | 0.0 | -4.5 |
| 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

| | geert niet) | | | | | | | | | - | П |
|------|--------------|------|------|---------------------|------|------|--------------|------|------|--------------|------|
| | 36.9 | -1 | | 38.0 | -1 | | 39.0 | -1 | | 0.0 | 40 |
| | | | | | | | | | | | |
| 24 5 | 0.0 | 20.0 | | | | 20.0 | 0.0 | 400 | 0.0 | 0.0 | |
| 34.5 | 0.0 | 38.0 | 37.0 | | 39.0 | 38.0 | 0.0 | 40.0 | 0.0 | 0.0 | 0.0 |
| | | | | | | | | | | | |
| | 32.0 | | | 37.0 | | | 29.0 | | | 0.0 | |
| | 32.0 34.9 | -1 | | 37.0 38.0 | -1 | | 29.0 39.0 | -10 | | 0.0 40.0 | -10 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 32.7 | 0.0 | 37.0 | 36.0 | 0.0 | 29.0 | 37.0 | 0.0 | 29.0 | 28.1 | 0.0 | 27.0 |
| | | | | | | | | | | | |
| | 3/1 8 | | | 36 N | | | 3/1 8 | | | 28.3 | |
| | 34.8 33.7 | -1 | | 36.0 37.0 | -1 | | 34.8 28.9 | -1 | | 28.3 26.1 | -1 |
| | | _ | | | _ | | | | | | |
| | | | | | | | | | | | |
| 30.8 | 0.0 | 36.0 | 35.0 | 0.0 | 35.0 | 36.0 | 0.0 | 32.0 | 31.6 | 0.0 | 29.4 |
| | | | | | | | | | | | |
| | 10.0 | | | 22.0 | | | 240 | | | 22.0 | |
| | 12.3 0.0 | 10 | | <u>33.9</u> 36.0 | -2 | | 34.0 35.0 | -1 | | 33.0 31.5 | |
| | •.• | | | | _ | | | | | •• | - |
| | | | | | | | | | | | |
| 0.0 | 0.0 | 0.0 | 10.3 | 0.0 | 34.0 | 33.5 | 0.0 | 33.0 | 34.0 | 0.0 | 32.5 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | 0.0 | | | <u> 30.6</u> | | | 34.0 | | | 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']

| Maze (re | ageert 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 | |
| | 2.9 3.3 | 7 | | 3.3 8.5 | -1 | | <u>-0.5</u> 19.0 | -10 | | <u>0.0</u> 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 1.0 | | | 1.2 | | | 4.1 | | | 4.0 10.0 | |
| | 1.0 | -1 | | 1.2 3.3 | -1 | | 4.1 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 | |
| | 10.0 0.0 | 10 | | 3.1 1.3 | -2 | | 1.0 1.6 | -1 | | 1.0 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']

| Maze (reage) | eert niet) | | | | | | | | | _ | |
|--------------|--------------|------|------|---------------------|------|------|----------------------|------|------|--------------|------|
| | 24.3 | -1 | | 30.5 | -1 | | 35.0 | -1 | | 0.0 | 40 |
| 00.0 | 0.0 | 00.5 | 00.4 | 0.0 | 05.0 | 00.5 | | 40.0 | 0.0 | | 0.0 |
| 26.0 | 0.0 | 30.5 | 26.4 | 0.0 | 35.0 | 30.5 | 0.0 | 40.0 | 0.0 | 0.0 | 0.0 |
| | 21.6 | | | 26.4 | | | 21.5 | | | 0.0 | |
| | 21.6 22.4 | -1 | | <u>26.4</u> 30.5 | -1 | | 21. <u>5</u> 35.0 | -10 | | 0.0 40.0 | -10 |
| 18.1 | 0.0 | 26.4 | 22.6 | 0.0 | 21.5 | 24.2 | 0.0 | 25.9 | 21.5 | 0.0 | 26.0 |
| | 16.6 19.8 | _ | | <u>22.6</u> 26.4 | | | 19.2 21.5 | | | 22.4 26.0 | |
| | | | | | | | | | | | |
| 16.5 | 0.0 | 22.8 | 17.7 | 0.0 | 19.2 | 22.4 | 0.0 | 22.4 | 19.2 | 0.0 | 22.4 |
| | 10.4 | | | 12.6 | | | 16.2 | | | 19.2 | |
| | 10.4 0.0 | 10 | | 12.6 16.0 | -2 | | 16.2 19.2 | -1 | | 19.2 22.4 | -1 |
| 0.0 | 0.0 | 0.0 | 101 | 0.0 | 16.0 | 10.7 | 0.0 | 10.0 | 45.0 | 0.0 | 40.7 |
| 0.0 | 0.0 | 0.0 | 10.1 | 0.0 | 16.2 | 12./ | 0.0 | 19.0 | 15.9 | 0.0 | 18.7 |
| | 0.0 | | | 11.1 | | | 16.1 | | | 18.7 | |

Policy table:

R = Right

L = Left

U = Up

D = Down

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

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

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

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