Adaptive system

AS2.2 - Model-free prediction and control

Naam : Hussin Almoustafa

Studentnummer : 1776495

May 31, 2023

1 Monte-Carlo evaluation

```
Value based poly ;iterations=10000  discount_rate=1 exploring_starts=True
Outcome

[[38. 39. 40.  0.]
 [37. 38. 39. 40.]
 [36. 37. 36. 35.]
 [ 0. 36. 35. 34.]]

Value based poly ;iterations=10000  discount_rate=0.9   exploring_starts=True
Outcome

[[30.5  35.   40.    0.  ]
 [26.45 30.5  35.   40.  ]
 [22.81 26.45 22.81 19.52]
 [ 0.   22.81 19.52 16.57]]

Random based poly ;iterations=10000 discount_rate=1 exploring_starts=True
Outcome

[[-14.25 -10.94   1.54   0.  ]
 [-14.01 -16.   -12.08  -3.36]
 [ -6.73 -13.27 -18.17 -17.96]
 [  0.    -7.43 -16.26 -19.32]]

Random based poly ;iterations=10000 discount_rate=0.9   exploring_starts=True
Outcome

[[-5.28 -2.83  6.69  0.  ]
 [-5.25 -7.51 -4.6   2.84]
 [-1.05 -5.69 -9.27 -8.68]
 [ 0.   -1.95 -7.32 -8.55]]
```

Figure 1: Monte-Carlo

2 Temporal Difference Learning evaluation

$y = 0.9$

```
Value based poly Temporal Difference Learning
iterations=10000    discount_rate=1 alpha=0.1   exploring_starts=True
Outcome

[[37.88 38.92 39.96  0.  ]
 [36.84 37.88 38.92 39.96]
 [35.8  36.84 35.8  34.76]
 [ 0.   35.8  34.76 33.72]]
Value based poly Temporal Difference Learning
iterations=10000    discount_rate=0.9   alpha=0.1   exploring_starts=True
Outcome

[[30.38 34.92 39.96  0.  ]
 [26.3  30.38 34.92 39.96]
 [22.62 26.3  22.62 19.31]
 [ 0.   22.62 19.31 16.33]]
Random based poly Temporal Difference Learning
iterations=10000    discount_rate=1 alpha=0.1   exploring_starts=True
Outcome

[[-11.52  -6.1    3.94   0.  ]
 [-12.02 -13.03  -7.49  -1.13]
 [ -9.58 -15.41 -18.93 -19.78]
 [  0.   -13.46 -19.78 -21.46]]
Random based poly Temporal Difference Learning
iterations=10000    discount_rate=0.9   alpha=0.1   exploring_starts=True
Outcome

[[ -4.54  -1.84   4.27   0.  ]
 [ -5.25  -7.27  -6.29  -6.68]
 [ -3.66  -5.26  -9.87 -10.72]
 [  0.    -0.77  -7.51  -8.82]]
```

Figure 2: Temporal Difference Learning

3 On policy first visit Monte-carlo control

$y = 0.9$

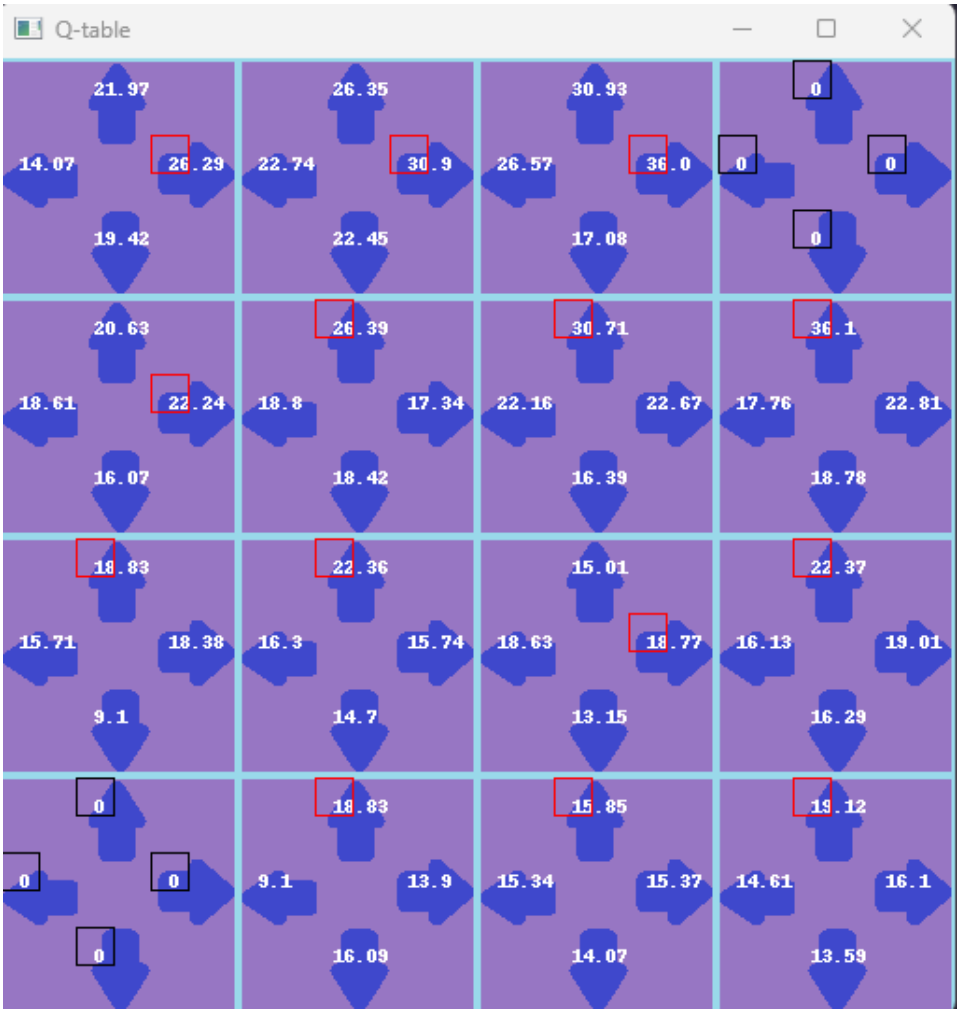

Figure 3: On policy first visit Monte-carlo control

$y = 1$

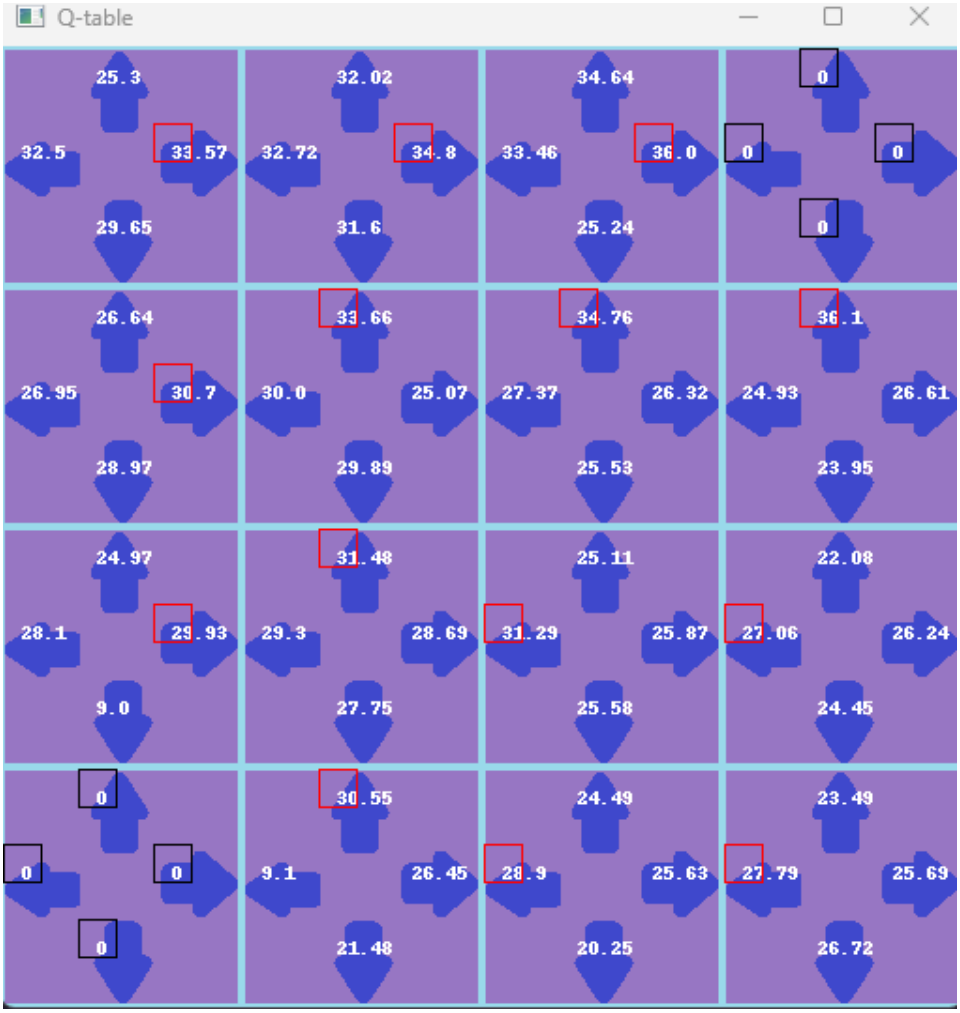

Figure 4: On policy first visit Monte-carlo control

4 On policy SARSA TD control

$y = 0.9$

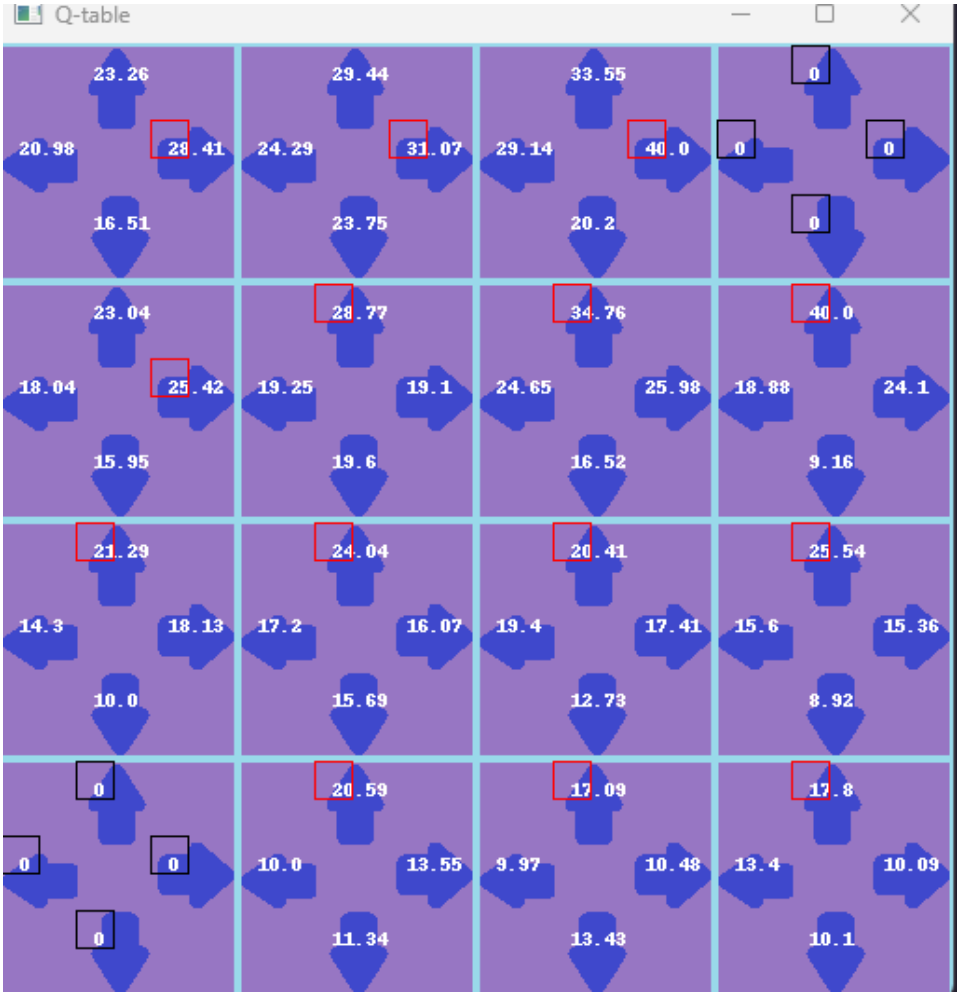

Figure 5: On policy SARSA TD control

$y = 1$

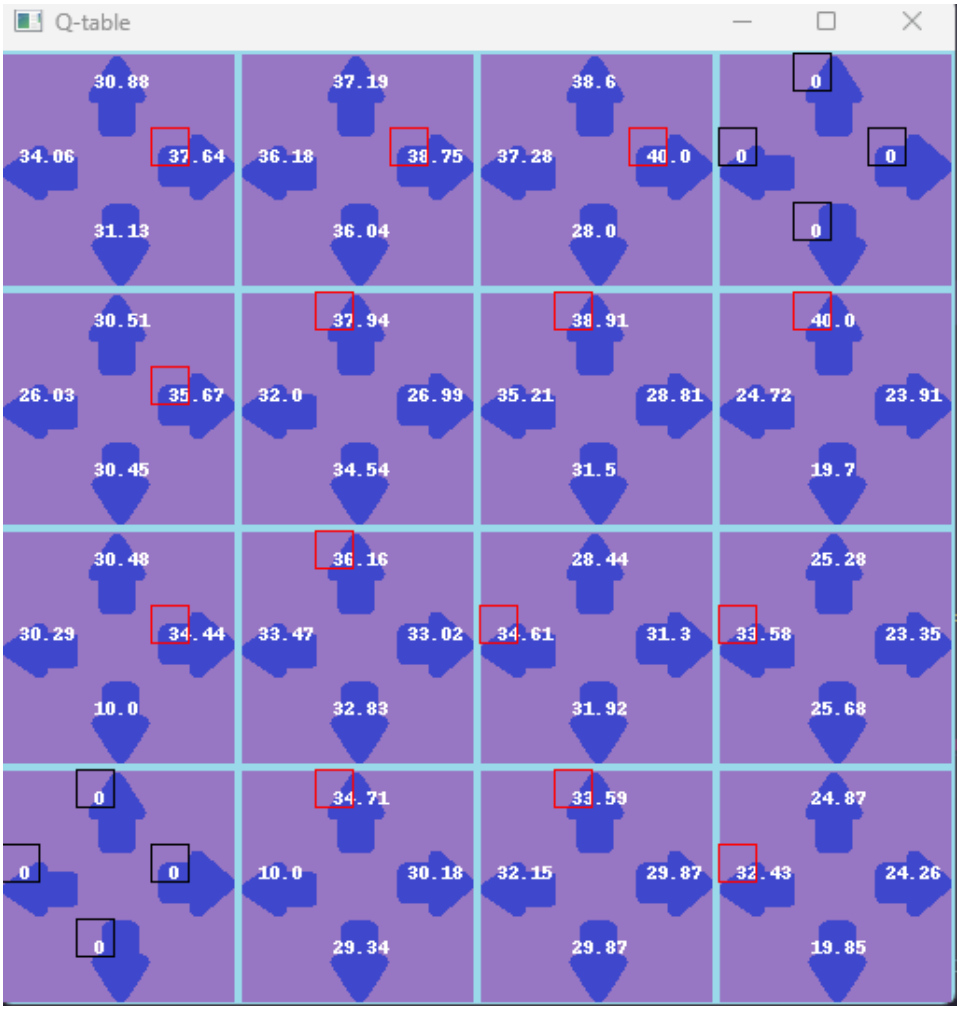

Figure 6: On policy SARSA TD control

5 Q leanring

$y = 0.9$

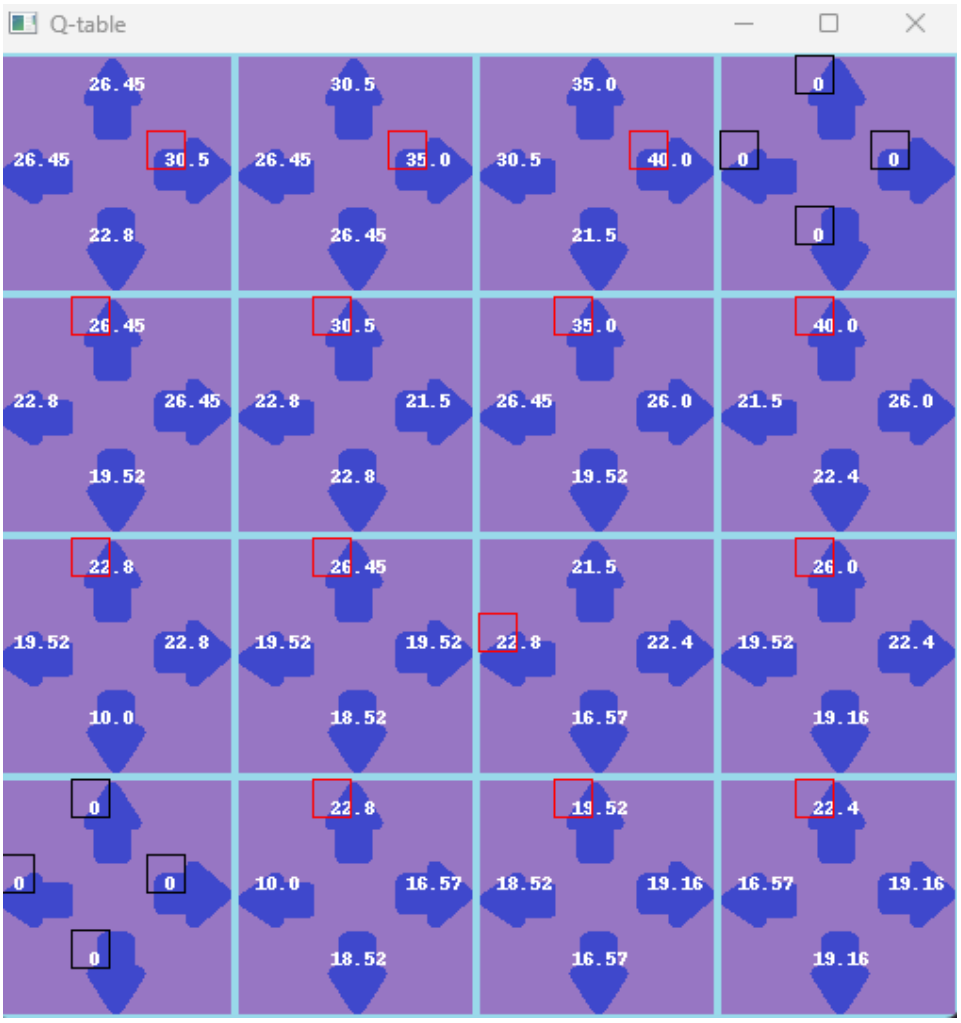

Figure 7: Q leanring

$y = 1$

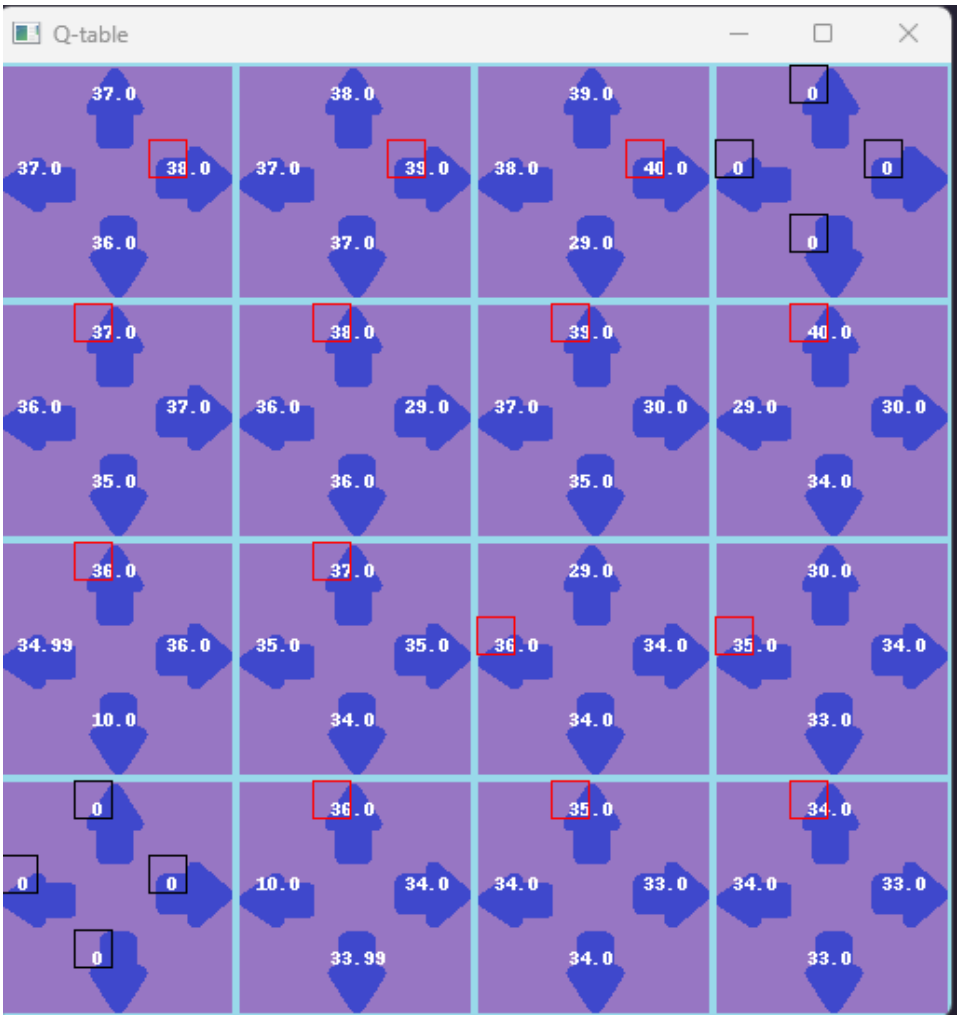

Figure 8: Q leanring

6 Double Q leanring

$y = 0.9$

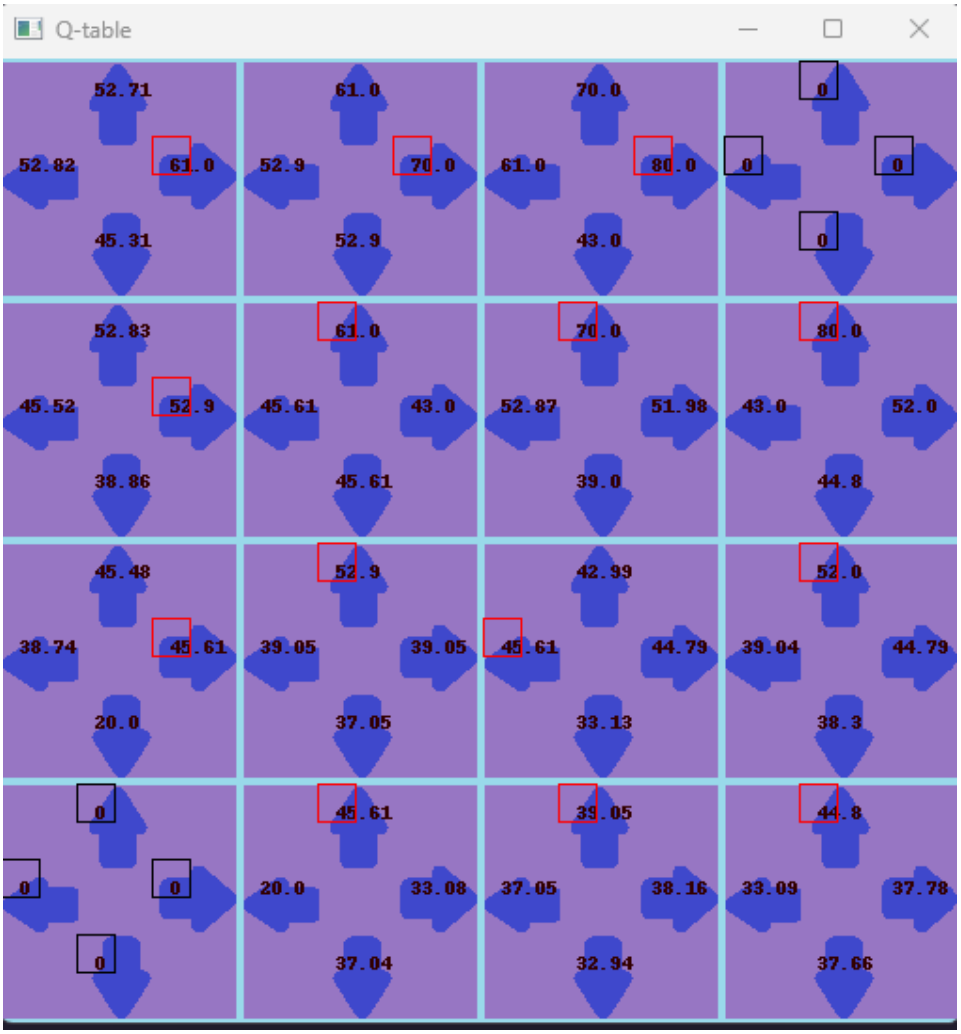

Figure 9: Double Q leanring

$y = 1$

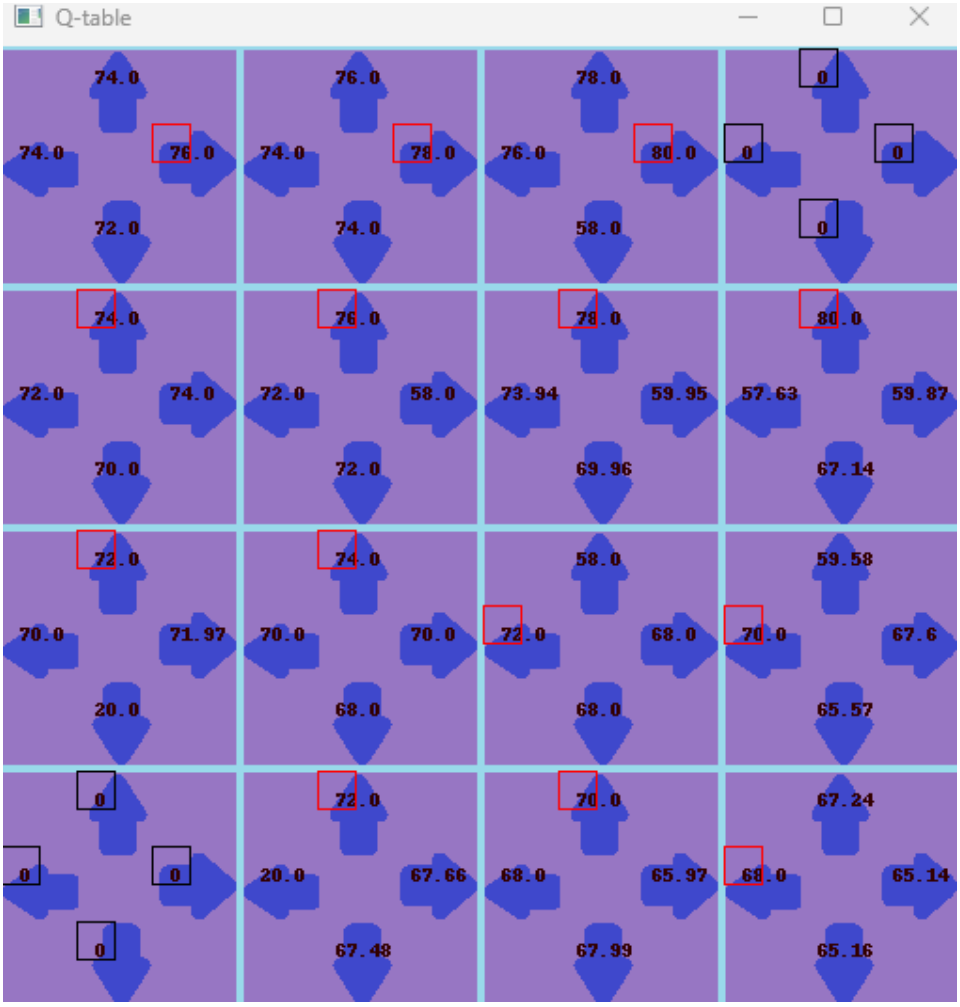

Figure 10: Double Q leanring