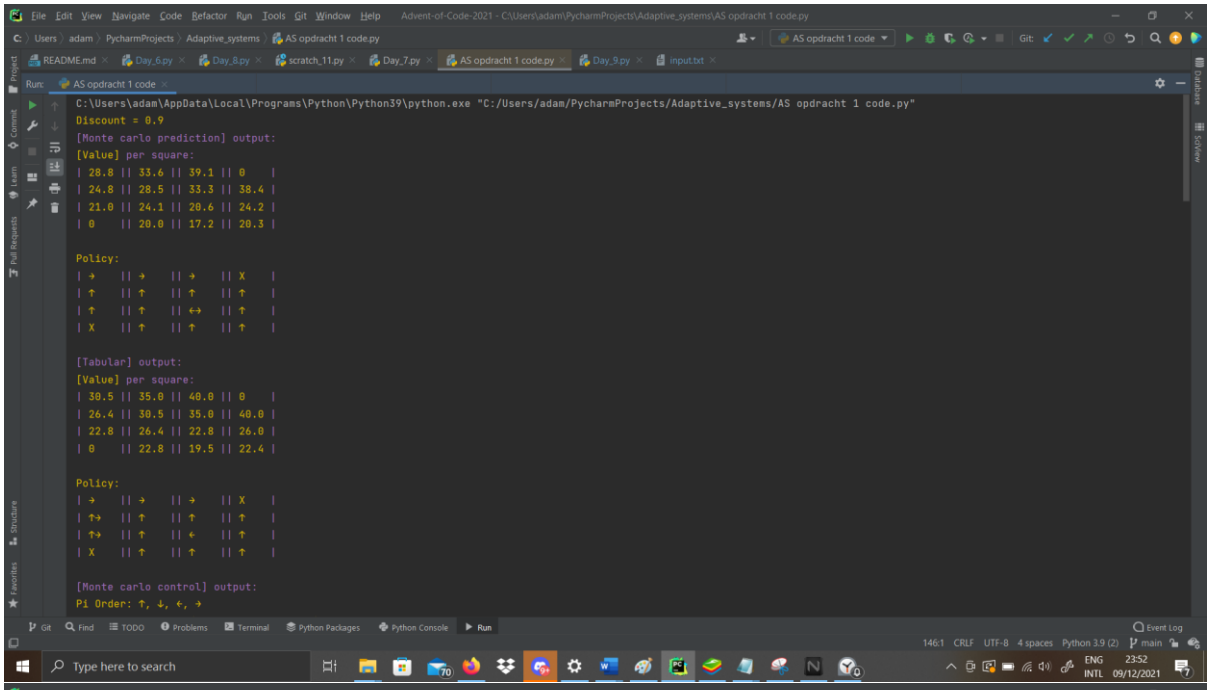
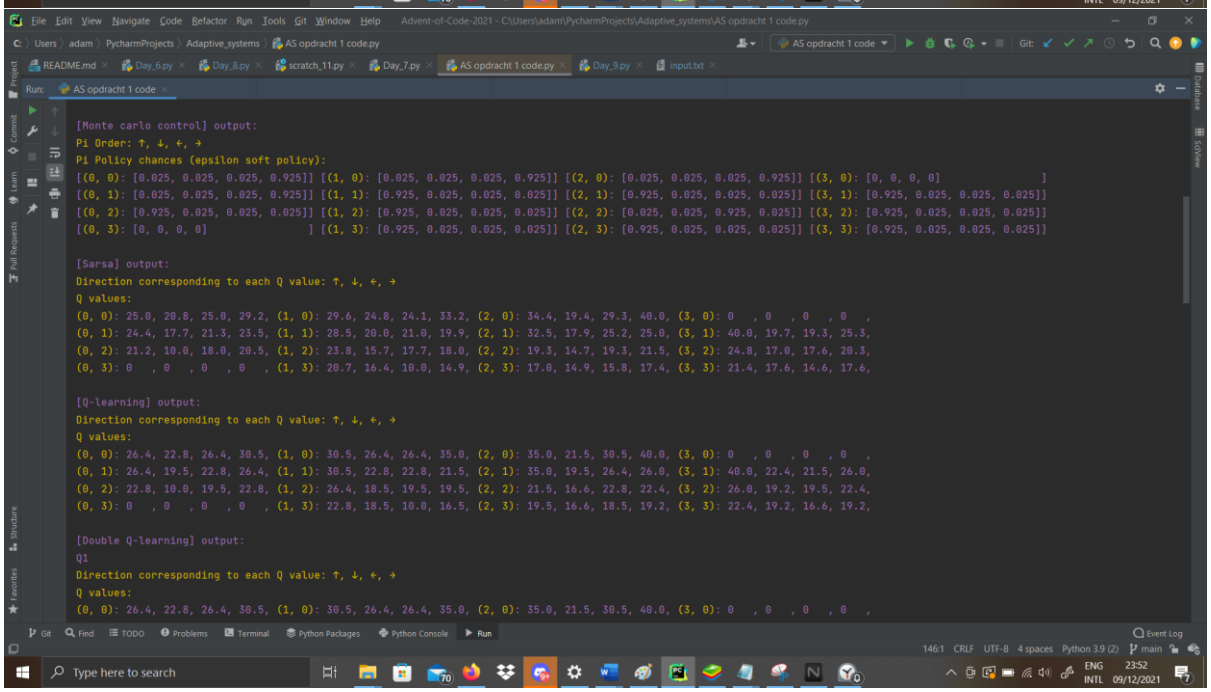
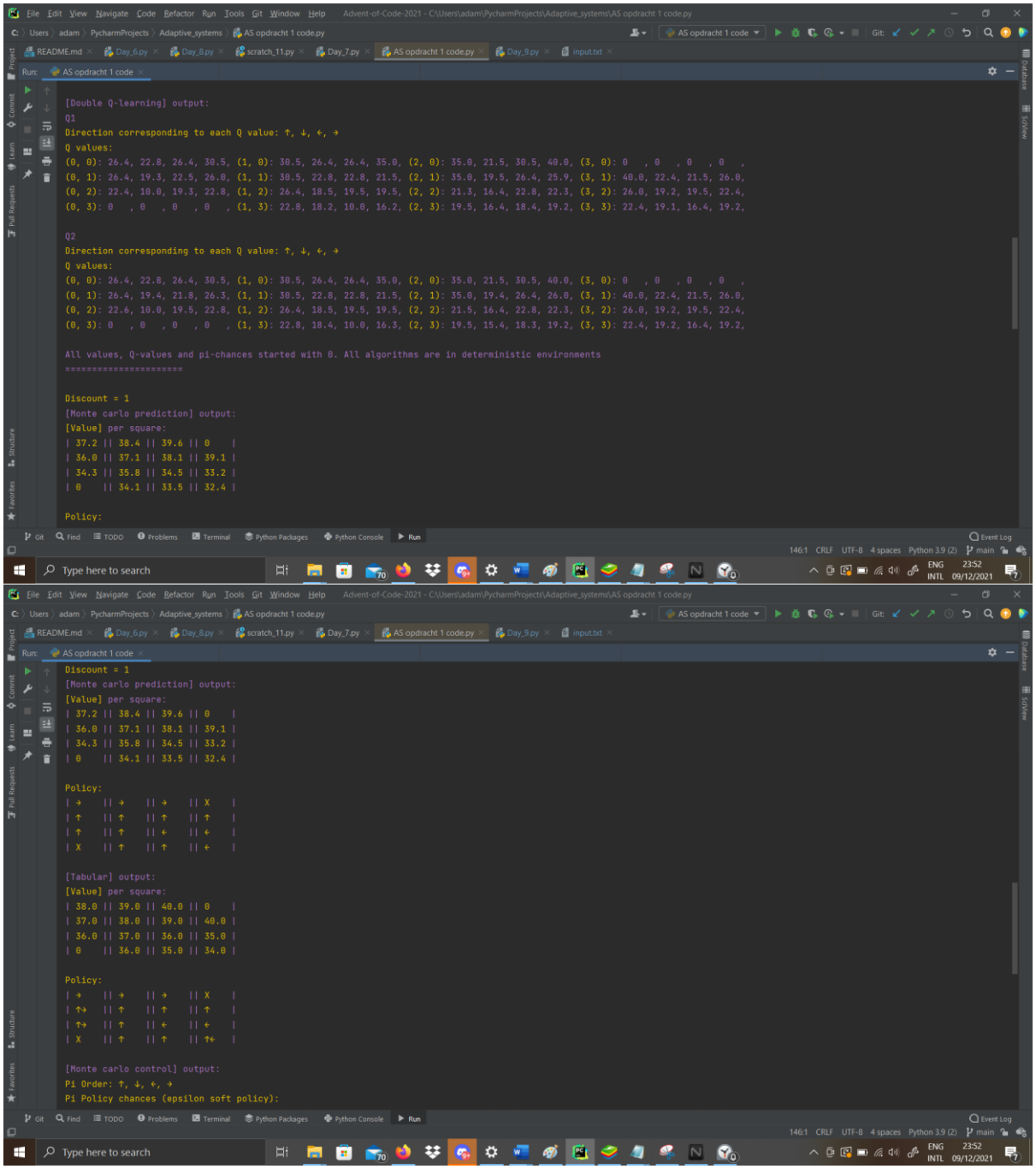
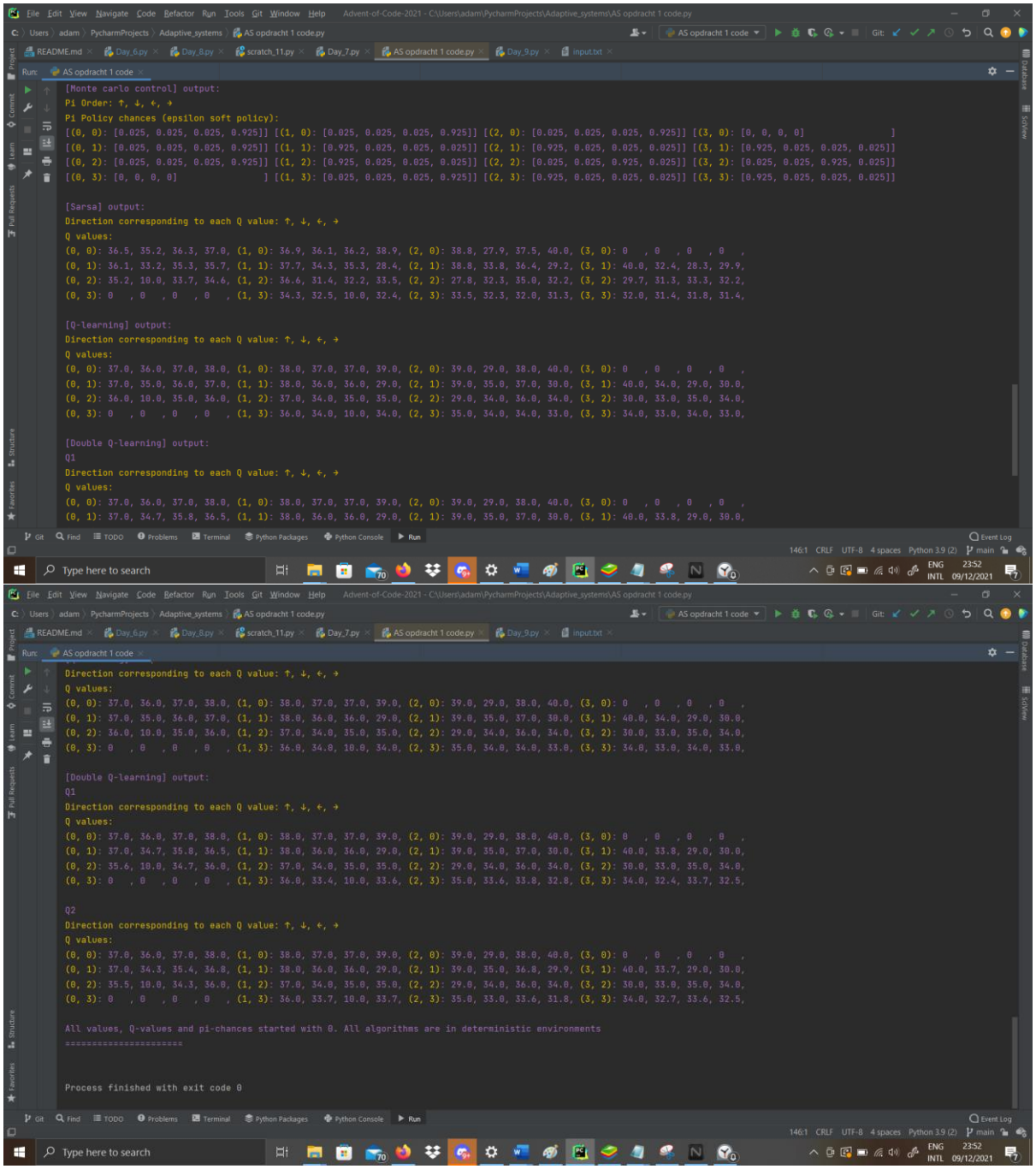

C:\Users\adam\AppData\Local\Programs\Python\Python39\python.exe "C:/Users/adam/PycharmProjects/Adaptive_systems/AS opdracht 1 code.py"
Discount = 0.9
[Monte carlo prediction] output:
[Value] per square:
28.8		33.6		39.1		0
24.8		28.5		33.3		38.4
21.0		24.1		20.6		24.2
0		20.0		17.2		20.3
Policy:
→		→		→		X
↑		↑		↑		↑
↑		↑		↔		↑
X		↑		↑		↑
[Tabular] output:
[Value] per square:
30.5		35.0		40.0		0
26.4		30.5		35.0		40.0
22.8		26.4		22.8		26.0
0		22.8		19.5		22.4
Policy:
→		→		→		X
↑→		↑		↑		↑
↑→		↑		←		↑
X		↑		↑		↑
[Monte carlo control] output:
Pi Order: ↑, ↓, ←, →


[Monte carlo control] output:
Pi Order: ↑, ↓, ←, →
Pi Policy chances (epsilon soft policy):
[(0, 0): [0.025, 0.025, 0.025, 0.925]] [(1, 0): [0.025, 0.025, 0.025, 0.925]] [(2, 0): [0.025, 0.025, 0.025, 0.925]] [(3, 0): [0, 0, 0, 0]
[(0, 1): [0.025, 0.025, 0.025, 0.925]] [(1, 1): [0.925, 0.025, 0.025, 0.025]] [(2, 1): [0.925, 0.025, 0.025, 0.025]] [(3, 1): [0.925, 0.025, 0.025, 0.025]]
[(0, 2): [0.925, 0.025, 0.025, 0.025]] [(1, 2): [0.925, 0.025, 0.025, 0.025]] [(2, 2): [0.025, 0.025, 0.925, 0.025]] [(3, 2): [0.925, 0.025, 0.025, 0.025]]
[(0, 3): [0, 0, 0, 0]] [(1, 3): [0.925, 0.025, 0.025, 0.025]] [(2, 3): [0.925, 0.025, 0.025, 0.025]] [(3, 3): [0.925, 0.025, 0.025, 0.025]]
[Sarsa] output:
Direction corresponding to each Q value: ↑, ↓, ←, →
Q values:
(0, 0): 25.0, 20.8, 25.0, 29.2, (1, 0): 29.6, 24.8, 24.1, 33.2, (2, 0): 34.4, 19.4, 29.3, 40.0, (3, 0): 0 , 0 , 0 , 0 ,
(0, 1): 24.4, 17.7, 21.3, 23.5, (1, 1): 28.5, 20.0, 21.0, 19.9, (2, 1): 32.5, 17.9, 25.2, 25.0, (3, 1): 40.0, 19.7, 19.3, 25.3,
(0, 2): 21.2, 10.0, 18.0, 20.5, (1, 2): 23.8, 15.7, 17.7, 18.0, (2, 2): 19.3, 14.7, 19.3, 21.5, (3, 2): 24.8, 17.0, 17.6, 20.3,
(0, 3): 0 , 0 , 0 , 0 , (1, 3): 20.7, 16.4, 10.0, 14.9, (2, 3): 17.0, 14.9, 15.8, 17.4, (3, 3): 21.4, 17.6, 14.6, 17.6,
[Q-learning] output:
Direction corresponding to each Q value: ↑, ↓, ←, →
Q values:
(0, 0): 26.4, 22.8, 26.4, 30.5, (1, 0): 30.5, 26.4, 26.4, 35.0, (2, 0): 35.0, 21.5, 30.5, 40.0, (3, 0): 0 , 0 , 0 , 0 ,
(0, 1): 26.4, 19.5, 22.8, 26.4, (1, 1): 30.5, 22.8, 22.8, 21.5, (2, 1): 35.0, 19.5, 26.4, 26.0, (3, 1): 40.0, 22.4, 21.5, 26.0,
(0, 2): 22.8, 10.0, 19.5, 22.8, (1, 2): 26.4, 18.5, 19.5, 19.5, (2, 2): 21.5, 16.6, 22.8, 22.4, (3, 2): 26.0, 19.2, 19.5, 22.4,
(0, 3): 0 , 0 , 0 , 0 , (1, 3): 22.8, 18.5, 10.0, 16.5, (2, 3): 19.5, 16.6, 18.5, 19.2, (3, 3): 22.4, 19.2, 16.6, 19.2,
[Double Q-learning] output:
Q1
Direction corresponding to each Q value: ↑, ↓, ←, →
Q values:
(0, 0): 26.4, 22.8, 26.4, 30.5, (1, 0): 30.5, 26.4, 26.4, 35.0, (2, 0): 35.0, 21.5, 30.5, 40.0, (3, 0): 0 , 0 , 0 , 0 ,


[Double Q-learning] output:
Q1
Direction corresponding to each Q value: ↑, ↓, ←, →
Q values:
(0, 0): 26.4, 22.8, 26.4, 30.5, (1, 0): 30.5, 26.4, 26.4, 35.0, (2, 0): 35.0, 21.5, 30.5, 40.0, (3, 0): 0 , 0 , 0 , 0 ,
(0, 1): 26.4, 19.3, 22.5, 26.0, (1, 1): 30.5, 22.8, 22.8, 21.5, (2, 1): 35.0, 19.5, 26.4, 25.9, (3, 1): 40.0, 22.4, 21.5, 26.0,
(0, 2): 22.4, 10.0, 19.3, 22.8, (1, 2): 26.4, 18.5, 19.5, 19.5, (2, 2): 21.3, 16.4, 22.8, 22.3, (3, 2): 26.0, 19.2, 19.5, 22.4,
(0, 3): 0 , 0 , 0 , 0 , (1, 3): 22.8, 18.2, 10.0, 16.2, (2, 3): 19.5, 16.4, 18.4, 19.2, (3, 3): 22.4, 19.1, 16.4, 19.2,
Q2
Direction corresponding to each Q value: ↑, ↓, ←, →
Q values:
(0, 0): 26.4, 22.8, 26.4, 30.5, (1, 0): 30.5, 26.4, 26.4, 35.0, (2, 0): 35.0, 21.5, 30.5, 40.0, (3, 0): 0 , 0 , 0 , 0 ,
(0, 1): 26.4, 19.4, 21.8, 26.3, (1, 1): 30.5, 22.8, 22.8, 21.5, (2, 1): 35.0, 19.4, 26.4, 26.0, (3, 1): 40.0, 22.4, 21.5, 26.0,
(0, 2): 22.6, 10.0, 19.5, 22.8, (1, 2): 26.4, 18.5, 19.5, 19.5, (2, 2): 21.5, 16.4, 22.8, 22.3, (3, 2): 26.0, 19.2, 19.5, 22.4,
(0, 3): 0 , 0 , 0 , 0 , (1, 3): 22.8, 18.4, 10.0, 16.3, (2, 3): 19.5, 15.4, 18.3, 19.2, (3, 3): 22.4, 19.2, 16.4, 19.2,
All values, Q-values and pi-chances started with 0. All algorithms are in deterministic environments
Discount = 1
[Monte carlo prediction] output:
[Value] per square:
37.2		38.4		39.6		0
36.0		37.1		38.1		39.1
34.3		35.8		34.5		33.2
0		34.1		33.5		32.4
Policy:
Type here to search

Discount = 1
[Monte carlo prediction] output:
[Value] per square:
37.2		38.4		39.6		0
36.0		37.1		38.1		39.1
34.3		35.8		34.5		33.2
0		34.1		33.5		32.4
Policy:
→		→		→		X
↑		↑		↑		↑
↑		↑		←		←
X		↑		↑		←
[Tabular] output:
[Value] per square:
38.0		39.0		40.0		0
37.0		38.0		39.0		40.0
36.0		37.0		36.0		35.0
0		36.0		35.0		34.0
Policy:
→		→		→		X
↑→		↑		↑		↑
↑→		↑		←		←
X		↑		↑		↑←
[Monte carlo control] output:
Pi Order: ↑, ↓, ←, →
Pi Policy chances (epsilon soft policy):
Type here to search


[Monte carlo control] output:
Pi Order: ↑, ↓, ←, →
Pi Policy chances (epsilon soft policy):
[(0, 0): [0.025, 0.025, 0.025, 0.925]] [(1, 0): [0.025, 0.025, 0.025, 0.925]] [(2, 0): [0.025, 0.025, 0.025, 0.925]] [(3, 0): [0, 0, 0, 0]
[(0, 1): [0.025, 0.025, 0.025, 0.925]] [(1, 1): [0.925, 0.025, 0.025, 0.025]] [(2, 1): [0.925, 0.025, 0.025, 0.025]] [(3, 1): [0.925, 0.025, 0.025, 0.025]]
[(0, 2): [0.025, 0.025, 0.025, 0.925]] [(1, 2): [0.925, 0.025, 0.025, 0.025]] [(2, 2): [0.025, 0.025, 0.925, 0.025]] [(3, 2): [0.025, 0.025, 0.925, 0.025]]
[(0, 3): [0, 0, 0, 0]
[Sarsa] output:
Direction corresponding to each Q value: ↑, ↓, ←, →
Q values:
(0, 0): 36.5, 35.2, 36.3, 37.0, (1, 0): 36.9, 36.1, 36.2, 38.9, (2, 0): 38.8, 27.9, 37.5, 40.0, (3, 0): 0 , 0 , 0 ,
(0, 1): 36.1, 33.2, 35.3, 35.7, (1, 1): 37.7, 34.3, 35.3, 28.4, (2, 1): 38.8, 33.8, 36.4, 29.2, (3, 1): 40.0, 32.4, 28.3, 29.9,
(0, 2): 35.2, 10.0, 33.7, 34.6, (1, 2): 36.6, 31.4, 32.2, 33.5, (2, 2): 27.8, 32.3, 35.0, 32.2, (3, 2): 29.7, 31.3, 33.3, 32.2,
(0, 3): 0 , 0 , 0 , 0 , (1, 3): 34.3, 32.5, 10.0, 32.4, (2, 3): 33.5, 32.3, 32.0, 31.3, (3, 3): 32.0, 31.4, 31.8, 31.4,
[Q-learning] output:
Direction corresponding to each Q value: ↑, ↓, ←, →
Q values:
(0, 0): 37.0, 36.0, 37.0, 38.0, (1, 0): 38.0, 37.0, 37.0, 39.0, (2, 0): 39.0, 29.0, 38.0, 40.0, (3, 0): 0 , 0 , 0 , 0 ,
(0, 1): 37.0, 35.0, 36.0, 37.0, (1, 1): 38.0, 36.0, 36.0, 29.0, (2, 1): 39.0, 35.0, 37.0, 30.0, (3, 1): 40.0, 34.0, 29.0, 30.0,
(0, 2): 36.0, 10.0, 35.0, 36.0, (1, 2): 37.0, 34.0, 35.0, 35.0, (2, 2): 29.0, 34.0, 36.0, 34.0, (3, 2): 30.0, 33.0, 35.0, 34.0,
(0, 3): 0 , 0 , 0 , 0 , (1, 3): 36.0, 34.0, 10.0, 34.0, (2, 3): 35.0, 34.0, 34.0, 33.0, (3, 3): 34.0, 33.0, 34.0, 33.0,
[Double Q-learning] output:
Q1
Direction corresponding to each Q value: ↑, ↓, ←, →
Q values:
(0, 0): 37.0, 36.0, 37.0, 38.0, (1, 0): 38.0, 37.0, 37.0, 39.0, (2, 0): 39.0, 29.0, 38.0, 40.0, (3, 0): 0 , 0 , 0 , 0 ,
(0, 1): 37.0, 34.7, 35.8, 36.5, (1, 1): 38.0, 36.0, 36.0, 29.0, (2, 1): 39.0, 35.0, 37.0, 30.0, (3, 1): 40.0, 33.8, 29.0, 30.0,
(0, 2): 35.6, 10.0, 34.7, 36.0, (1, 2): 37.0, 34.0, 35.0, 35.0, (2, 2): 29.0, 34.0, 36.0, 34.0, (3, 2): 30.0, 33.0, 35.0, 34.0,
(0, 3): 0 , 0 , 0 , 0 , (1, 3): 36.0, 33.4, 10.0, 33.6, (2, 3): 35.0, 33.6, 33.8, 32.8, (3, 3): 34.0, 32.4, 33.7, 32.5,
Q2
Direction corresponding to each Q value: ↑, ↓, ←, →
Q values:
(0, 0): 37.0, 36.0, 37.0, 38.0, (1, 0): 38.0, 37.0, 37.0, 39.0, (2, 0): 39.0, 29.0, 38.0, 40.0, (3, 0): 0 , 0 , 0 , 0 ,
(0, 1): 37.0, 34.3, 35.4, 36.8, (1, 1): 38.0, 36.0, 36.0, 29.0, (2, 1): 39.0, 35.0, 36.8, 29.9, (3, 1): 40.0, 33.7, 29.0, 30.0,
(0, 2): 35.5, 10.0, 34.3, 36.0, (1, 2): 37.0, 34.0, 35.0, 35.0, (2, 2): 29.0, 34.0, 36.0, 34.0, (3, 2): 30.0, 33.0, 35.0, 34.0,
(0, 3): 0 , 0 , 0 , 0 , (1, 3): 36.0, 33.7, 10.0, 33.7, (2, 3): 35.0, 33.0, 33.6, 31.8, (3, 3): 34.0, 32.7, 33.6, 32.5,
All values, Q-values and pi-chances started with 0. All algorithms are in deterministic environments
Process finished with exit code 0