

initialize Q-table

State (pos)	action "up"	action "down"	action "left"	action "right"	action "stay"
(0, 0)	-	0	-	0	0
(0, 1)	-	0	0	0	0
(0, 2)	-	0	0	-1	0
(1, 0)	0	0	-1	0	0
(1, 1)	0	0	0	0	0
(1, 2)	0	0	0	-1	0
(2, 0)	0	-1	-1	0	0
(2, 1)	0	-1	0	0	0
(2, 2)	0	-1	0	-1	0

0,0	0,1	0,2
1,0	1,1	1,2
2,0	2,1	2,2

COSTS:

any move: $c=1$

go into grey: $c=10$

stay: $c=2$

stay in yellow: $c=0$

we define $\delta=1$

first iteration ($i=0$)

$\pi_0(0,0) = \text{action "up"}$

$$((0,0), \text{"up"}) = -$$

$$\begin{aligned} ((0,0), \text{"down"}) &= c((0,0), \text{"down"}) + \min_a [Q_0(S_{t+1}, a)] \\ &= 1 + \min_a [Q_0(1,0), a] \\ &= 1 + 0 = 1 \end{aligned}$$

$$((0,0), \text{"left"}) = -$$

$$\begin{aligned} ((0,0), \text{"right"}) &= c((0,0), \text{"right"}) + \min_a [Q_0(S_{t+1}, a)] \\ &= 1 + \min_a [Q_0(0,1), a] \\ &= 1 + 0 = 1 \end{aligned}$$

$$\begin{aligned} ((0,0), \text{"stay"}) &= c((0,0), \text{"stay"}) + \min_a [Q_0((0,0), a)] \\ &= 0 + 0 = 0 \end{aligned}$$

$\pi_0(0,1) = \text{action "up"}$

$$((0,1), \text{"up"}) = -$$

$$\begin{aligned} ((0,1), \text{"down"}) &= c((0,1), \text{"down"}) + \min_a [Q_0((0,1), a)] \\ &= 10 + 0 = 10 \end{aligned}$$

$$\begin{aligned} ((0,1), \text{"left"}) &= c((0,1), \text{"left"}) + \min_a [Q_0((0,0), a)] \\ &= 1 + 0 = 1 \end{aligned}$$

$$\begin{aligned} ((0,1), \text{"right"}) &= c((0,1), \text{"right"}) + \min_a [Q_0((0,2), a)] \\ &= 1 + 0 = 1 \end{aligned}$$

$$\begin{aligned} ((0,1), \text{"stay"}) &= c((0,1), \text{"stay"}) + \min_a [Q_0((0,1), a)] \\ &= 2 + 0 = 2 \end{aligned}$$

$\pi_0(0,2) = \text{action "up"}$

$$((0,2), \text{"up"}) = -$$

$$\begin{aligned} ((0,2), \text{"down"}) &= c((0,2), \text{"down"}) + \min_a [Q_0((1,2), a)] \\ &= 10 + 0 = 10 \end{aligned}$$

$$\begin{aligned} ((0,2), \text{"left"}) &= c((0,2), \text{"left"}) + \min_a [Q_0((0,1), a)] \\ &= 1 + 1 = 2 \end{aligned}$$

$$((0,2), \text{"right"}) = -$$

$$\begin{aligned} ((0,2), \text{"stay"}) &= c((0,2), \text{"stay"}) + \min_a [Q_0((0,2), a)] \\ &= 2 + 0 = 2 \end{aligned}$$

$\pi_0(1,0) = \text{action "up"}$

$$((1,0), \text{"up"}) = c((1,0), \text{"up"}) + \min_a [Q_0((0,0), a)]$$

$$1 + 0 = 1$$

$$((1,0), \text{"down"}) = c((1,0), \text{"down"}) + \min_a [Q_0((2,0), a)]$$

$$= 1 + 0 = 1$$

$$((1,0), \text{"left"}) = -$$

$$((1,0), \text{"right"}) = c((1,0), \text{"right"}) + \min_a [Q_0((1,1), a)]$$

$$= 10 + 0 = 10$$

$$((1,0), \text{"stay"}) = c((1,0), \text{"stay"}) + \min_a [Q_0((1,0), a)]$$

$$= 2 + 0 = 2$$

$\pi_0(1,1) = \text{action "up"}$

$$((1,1), \text{"up"}) = c((1,1), \text{"up"}) + \min_a [Q_0((0,1), a)]$$

$$1 + 1 = 2$$

$$((1,1), \text{"down"}) = c((1,1), \text{"down"}) + \min_a [Q_0((2,1), a)]$$

$$= 1 + 0 = 1$$

$$((1,1), \text{"left"}) = c((1,1), \text{"left"}) + \min_a [Q_0((1,0), a)]$$

$$1 + 1 = 2$$

$$((1,1), \text{"right"}) = c((1,1), \text{"right"}) + \min_a [Q_0((1,2), a)]$$

$$= 10 + 0 = 10$$

$$((1,1), \text{"stay"}) = c((1,1), \text{"stay"}) + \min_a [Q_0((1,1), a)]$$

$$= 2 + 0 = 2$$

$\pi_0(1,2) = \text{action "up"}$

$$((1,2), \text{"up"}) = c((1,2), \text{"up"}) + \min_a [Q_0((0,2), a)]$$

$$1 + 2 = 3$$

$$((1,2), \text{"down"}) = c((1,2), \text{"down"}) + \min_a [Q_0((2,2), a)]$$

$$= 1 + 0 = 1$$

$$((1,2), \text{"left"}) = c((1,2), \text{"left"}) + \min_a [Q_0((1,1), a)]$$

$$10 + 1 = 11$$

$$((1,2), \text{"right"}) = -$$

$$((1,2), \text{"stay"}) = c((1,2), \text{"stay"}) + \min_a [Q_0((1,2), a)]$$

$$= 2 + 0 = 2$$

$\pi_0(2,0) = \text{action "up"}$

$$((2,0), \text{"up"}) = c((2,0), \text{"up"}) + \min_a [Q_0((1,0), a)]$$

$$1 + 1 = 2$$

$$((2,0), \text{"down"}) = -$$

$$((2,0), \text{"left"}) = -$$

$$((2,0), \text{"right"}) = c((2,0), \text{"right"}) + \min_a [Q_0((2,1), a)]$$

$$= 1 + 0 = 1$$

$$((2,0), \text{"stay"}) = c((2,0), \text{"stay"}) + \min_a [Q_0((2,0), a)]$$

$$= 2 + 0 = 2$$

$$\pi_0(2,1) = \text{action "up"}$$

$$((2,1), \text{"up"}) = c((2,1), \text{"up"}) + \min_a [Q_0(1,1), a] \\ 10 + 1 = 11$$

$$((2,1), \text{"down"}) = -$$

$$((2,1), \text{"left"}) = c((2,1), \text{"left"}) + \min_a [Q_0((2,0), a)] \\ 1 + 1 = 2$$

$$((2,1), \text{"right"}) = c((2,1), \text{"right"}) + \min_a [Q_0((2,2), a)] \\ = 1 + 0 = 1$$

$$((2,1), \text{"stay"}) = c((2,1), \text{"stay"}) + \min_a [Q_0((2,1), a)] \\ = 2 + 0 = 2$$

$$\pi_0(2,2) = \text{action "up"}$$

$$((2,2), \text{"up"}) = c((2,2), \text{"up"}) + \min_a [Q_0(1,2), a] \\ 10 + 1 = 11$$

$$((2,2), \text{"down"}) = -$$

$$((2,2), \text{"left"}) = c((2,2), \text{"left"}) + \min_a [Q_0((2,1), a)] \\ 1 + 1 = 2$$

$$((2,2), \text{"right"}) = -$$

$$((2,2), \text{"stay"}) = c((2,2), \text{"stay"}) + \min_a [Q_0((2,2), a)] \\ = 2 + 0 = 2$$

#	first iteration i=0	state (pos)	action "up"	action "down"	action "left"	action "right"	action "stay"
	$\pi_0(0,0) = \text{"up"}$	(0,0)	-	1	-	1	0
	$\pi_0(0,1) = \text{"up"}$	(0,1)	-	10	1	1	2
	$\pi_0(0,2) = \text{"up"}$	(0,2)	-	10	2	-	2
	$\pi_0(1,0) = \text{"up"}$	(1,0)	1	1	-	10	2
	$\pi_0(1,1) = \text{"up"}$	(1,1)	2	1	2	10	2
	$\pi_0(1,2) = \text{"up"}$	(1,2)	3	1	11	-	2
	$\pi_0(2,0) = \text{"up"}$	(2,0)	2	-	-	1	2
	$\pi_0(2,1) = \text{"up"}$	(2,1)	11	-	2	1	2
	$\pi_0(2,2) = \text{"up"}$	(2,2)	11	-	2	-	2

#	second iteration i=1	state (pos)	action "up"	action "down"	action "left"	action "right"	action "stay"
	$\pi_1(0,0) = \text{"stay"}$	(0,0)	-	2	-	2	0
	$\pi_1(0,1) = \text{"left"}$	(0,1)	-	11	1	3	3
	$\pi_1(0,2) = \text{"left"}$	(0,2)	-	11	2	-	4
	$\pi_1(1,0) = \text{"up"}$	(1,0)	1	2	-	11	3
	$\pi_1(1,1) = \text{"down"}$	(1,1)	2	2	2	11	4
	$\pi_1(1,2) = \text{"down"}$	(1,2)	3	3	12	-	4
	$\pi_1(2,0) = \text{"right"}$	(2,0)	2	-	-	2	4
	$\pi_1(2,1) = \text{"right"}$	(2,1)	12	-	3	3	4
	$\pi_1(2,2) = \text{"left"}$	(2,2)	13	-	4	-	4

Third iteration $i=2$

- $\pi_2(0,0) = \text{"stay"}$
- $\pi_2(0,1) = \text{"left"}$
- $\pi_2(0,2) = \text{"left"}$
- $\pi_2(1,0) = \text{"up"}$
- $\pi_2(1,1) = \text{"up"}$
- $\pi_2(1,2) = \text{"up"}$
- $\pi_2(2,0) = \text{"up"}$
- $\pi_2(2,1) = \text{"left"}$
- $\pi_2(2,2) = \text{"left"}$

State (pos)	action "up"	action "down"	action "left"	action "right"	action "stay"
(0, 0)	-	2	-	2	0
(0, 1)	-	12	1	3	3
(0, 2)	-	13	2	-	4
(1, 0)	1	3	-	12	3
(1, 1)	2	4	2	13	4
(1, 2)	3	5	12	-	5
(2, 0)	2	-	-	4	4
(2, 1)	12	-	3	5	5
(2, 2)	13	-	4	-	6

Fourth iteration $i=3$

- $\pi_3(0,0) = \text{"stay"}$
- $\pi_3(0,1) = \text{"left"}$
- $\pi_3(0,2) = \text{"left"}$
- $\pi_3(1,0) = \text{"up"}$
- $\pi_3(1,1) = \text{"up"}$
- $\pi_3(1,2) = \text{"up"}$
- $\pi_3(2,0) = \text{"up"}$
- $\pi_3(2,1) = \text{"left"}$
- $\pi_3(2,2) = \text{"left"}$

State (pos)	action "up"	action "down"	action "left"	action "right"	action "stay"
(0, 0)	-	2	-	2	0
(0, 1)	-	12	1	3	3
(0, 2)	-	13	2	-	4
(1, 0)	1	3	-	12	3
(1, 1)	2	4	2	13	4
(1, 2)	3	5	12	-	5
(2, 0)	2	-	-	4	4
(2, 1)	12	-	3	5	5
(2, 2)	13	-	4	-	6

Fifth iteration $i=4$

- $\pi_4(0,0) = \text{"stay"}$
- $\pi_4(0,1) = \text{"left"}$
- $\pi_4(0,2) = \text{"left"}$
- $\pi_4(1,0) = \text{"up"}$
- $\pi_4(1,1) = \text{"up"}$
- $\pi_4(1,2) = \text{"up"}$
- $\pi_4(2,0) = \text{"up"}$
- $\pi_4(2,1) = \text{"left"}$
- $\pi_4(2,2) = \text{"left"}$

State (pos)	action "up"	action "down"	action "left"	action "right"	action "stay"
(0, 0)	-	2	-	2	0
(0, 1)	-	12	1	3	3
(0, 2)	-	13	2	-	4
(1, 0)	1	3	-	12	3
(1, 1)	2	4	2	13	4
(1, 2)	3	5	12	-	5
(2, 0)	2	-	-	4	4
(2, 1)	12	-	3	5	5
(2, 2)	13	-	4	-	6