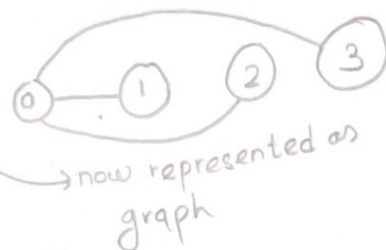


2	3	100
0	1	

State space



now represented as graph

When we call it adjacent when rows represent on set of values and column represent another one.

2 Immediate

0	0	0	100
0	1	2	3

2 Adjacent Matrix

	left	right	up	down
0	-1	1	2	-1
1	0	-1	3	-1
2	-1	3	-1	0
3	2	-1	-1	1

q Table

	left	right	up	down
0	-1	0	0	-1
1	0	-1	0	-1
2	-1	0	-1	0
3	0	-1	-1	0

Initial values 0

0	-1	0	0	-1
---	----	---	---	----

Selected state: 0

① select random action for 0:
say up

② next state from current now that we picked up action.
 $qMatrix[0][up] = 2$

③ immediate reward for this
 $= q_immediate = 0$

④ delayed reward = $\max(qTable[nextSt]) * decay$
 $= 0 * 0.9$
 $= 0$

⑤ $q_total\ reward = immediate\ reward + delayed\ reward$
 $0 + 0.$

⑥ $qTable[my\ state][my\ Action]$
 $= total\ reward$

Selected State: 2

① random action for 2:
say right

② next state from current,
 $qMatrix[2][right] = 3$

③ Immediate reward, $q_immediate = 100$
for 3.

④ delayed reward = $\max(qTable[next\ st]) * decay$
 $= 0 * 0.9 = 0$

⑤ total reward = immediate reward + delayed reward
 $= 100 + 0$

⑥ $qTable[my\ state][my\ action] = total\ reward$
 $qTable[2][right] = 100$

2	-1	100	-1	0
---	----	-----	----	---

Selected State: 3 falling if equal to goal state nothing happens, down't

① random action for 3: say down

② next state from current: $qMatrix[3][down] = 1$

③ Immediate reward, $q_immediate = 0$

④ delayed reward = $\max(qTable[next\ st]) * decay$
 $= 0 * 0.9$

⑤ total reward = $0 + 0 = 0$

⑥ $qTable[my\ state][my\ action] = total\ reward$
 $qTable[3][down] = 0$

3	0	-1	-1	0
---	---	----	----	---

No Action here

Selected State: 1

① random action for 1:
say up.

② next state from current:

$$qMatrix[1][up] = 3$$

③ Immediate reward, q -immediate = 100
for 3

④ Delayed reward = $\max(qTable[next\ st]) * decay$
= $0 * 0.9 = 0$

⑤ total reward = $100 + 0 = 100$

⑥ $qTable[my\ state][my\ Action] = 100$

$$qTable[1][up] = 100.$$

1	0	1	100	-1
---	---	---	-----	----

Selected state: 1

① random action for 1
say left

② next state from current:
 $qMatrix[1][left] = 0$

③ Immediate reward for 0, q -immediate = 0.

④ Delayed reward = $\max(qTable[next\ st]) * decay$
= $0 * decay$

⑤ total reward = 0

... 1 - 100

	90			
8	1	1	1	1
7	7	7	7	7
0		1		

next state
immediate
reward
worked

	100	
2	3	↑
0	1	

→ 100

from 1 to 0,
I got total
reward 0.

Selected State: 0

① random action for 0,
say right

② next state from current,
 $q_{Matrix}[0][right] = 1$

③ Immediate reward for 1, $q_{-immediate} = 0$

④ Delay reward = $\max(q_{Table}(next\ st)) * decay$
 $= 100 * 0.9$

⑤ total reward = ③ + ④ = 0 + 90 = 90

0	-1	90	0	-1
---	----	----	---	----

		100
90	↓	↑
	81	

Selected state: ~~0~~ 1

① random action for ~~0~~ 1,
say left

② next state from current,
 $q_{Matrix}[1][left] = 0$

③ Immediate reward for 0, $q_{-immediate} = 0$

④ Delay reward = $\max(q_{Table}(next\ st)) * decay$
 $= 90 * 0.9 = 81$

1	81	-1	100	-1
---	----	----	-----	----

Exploration