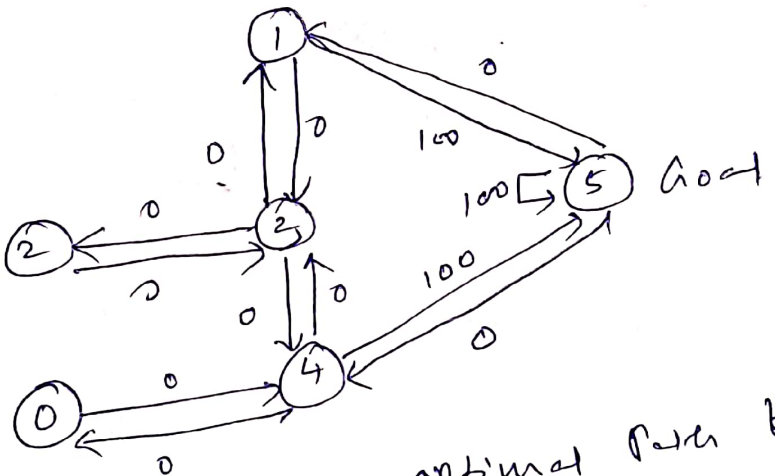
What is Q-Learning

Q-Learn is a reinforcement learning that will find the next best action, it is a model free RL reinforcement learning. The objective of this model is to find the best course of action given its current state.

Example:- Advertisement recommendation system.

In a usual ad recomendation system, the ads you get are based on your previous purchases. using Q-learning we can optimize the ad ~~recomendation~~ recomendation system to recommend products that are frequently brought together.



We need to find an optimal path from each state to goal state 5 where 5 is considered as goal state.

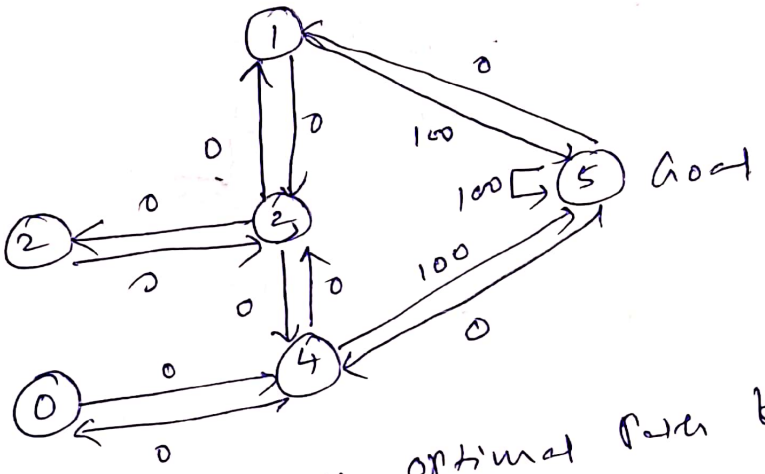
The actions which are leading to goal state are rewarded with 100. all the other actions are rewarded with 0.

What is Q-Learning

Q-Learn is a reinforcement learning that will find the next best action, it is a model free RL reinforcement learning. The objective of this model is to find the best course of action given its current state.

Example:- Advertisement recommendation system.

In a normal ad recomendation system, the ads you get are based on your previous purchases. using Q-learning we can optimize the ad ~~recommendation~~ recomendation system to recommend products that are frequently brought together.



We need to find an optimal path from each state to goal state 5 where 5 is considered as goal state.

The actions which are leading to goal state are rewarded with 100. all the other actions are rewarded with 0.

Q. Learning algorithm is applied to get optimal Q matrix.

Reward matrix:

State \ Action	0	1	2	3	4	5
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	0	-1	100
2	-1	-1	-1	0	-1	-1
3	-1	0	0	-1	0	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

$R =$

when you are in state 1 you can perform actions 3 & 5 (0) (100). -1 indicates there is no direct edge.

Learning rate = 0.8, Initial State = 1

Now initialize Q Matrix

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

$Q =$

Initial state is 1 there are two actions for this state 3 & 5 with 0 & 100 rewards. So I will select my next state as 5. immediate reward 5. from next state 5 you can perform an action 1, 4, & 5 with rewards 0, 0, & 100.

Q. Learning algorithm is applied to get optimal ②
path.

Reward matrix:

State \ Action	0	1	2	3	4	5
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	0	-1	100
2	-1	-1	-1	0	-1	-1
3	-1	0	0	-1	0	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

$R =$

when you are in state 1 you can perform actions 3 & 5 (0) (100). -1 indicates there is no direct edge.

Learning rate = 0.8,

Initial state = 1

Now initialize Q matrix

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0

$Q =$

Initial state is 1 there are two actions for this state 3 & 5 with 0 & 100 rewards. So I will select my next state as 5. immediate reward 5. from next state 5 you can perform an action 1, 4, & 5 with rewards 0, 0, & 100.

It has 3 possible actions 1, 4 & 5

Q-learning equation

$Q(state, action) = R(state, action) + Gamma * max(Q(next\ state, all\ actions))$

$Q(1,5) = R(1,5) + 0.8 \max(Q(5, (1,4,5)))$

$= 100 + 0.8 * 0 = 100.$

	0	1	2	3	4	5
0	0	0	0	0	80	0
1	0	0	0	64	0	100
2	0	0	0	64	0	0
3	0	80	51	0	80	0
4	64	0	0	64	0	100
5	0	80	0	0	80	100

for the next episode I will consider 3 as initial state.

3 it can go to 1, 2, 4.

So will select the next state as 1

$Q(3,1) = R(3,1) + 0.8 Q(\max(1, (3,5)))$

$= 0 + 0.8 \times 100 = 80$ ✓

Next I will select 0 as initial state

0, 4.

$Q(0,4) = R(0,4) + (0.8) \times \max(4, (0,3,5))$

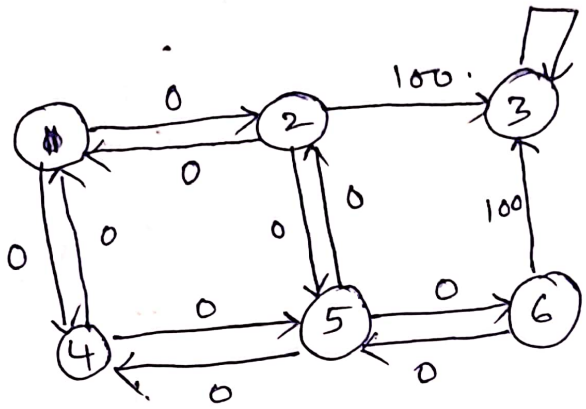
$= 0 + 0.8 \times \max = 0 + 0.8 \times 100$

$= 80$

Next I will select 1 as my initial state

$Q(1,3) = R(1,3) + 0.8 * (3, \max(1,5,4))$

Q-Learning Example. Example - 2



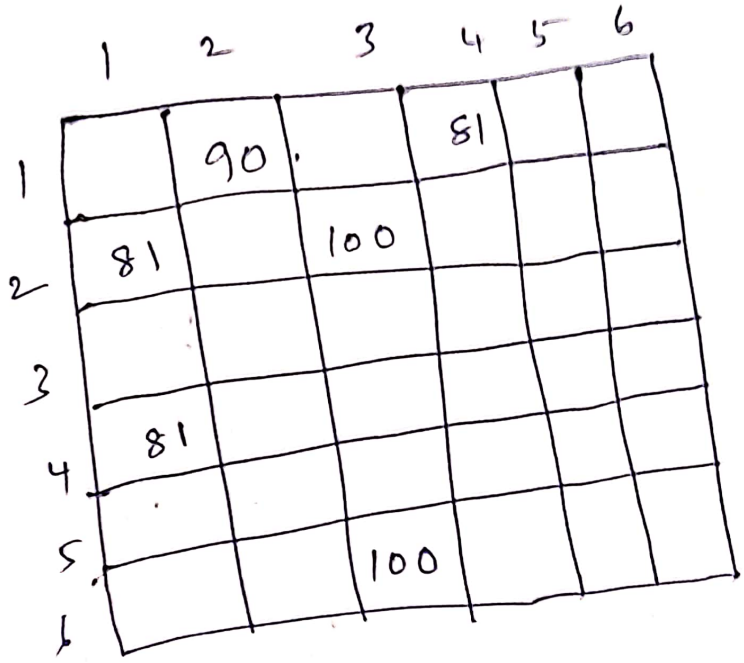
R

	1	2	3	4	5	6
1	-1	0	-1	0	-1	-1
2	0	-1	+100	-1	0	-1
3	-1	-1	100	-1	-1	-1
4	0	-1	-1	-1	0	-1
5	-1	0	-1	0	-1	0
6	-1	-1	+100	-1	0	-1

Q =

	1	2	3	4	5	6
1	0	0	0	0 81	0	0
2	0 81	0	0 100	0	72 0	0
3	0	0	0 100	0	0	0
4	0 81	0	0	0 72	0	0 90
5	0	0 90	0	0 100	0	0 90
6	0	0	0	0	0	0

Q(State, action) = R(State, action) + γ * max(Q(NextState, action))



(1,2)

$$Q(1,2) = R(1,2) + 0.9 * (\max[(2,1)(2,3)(2,5)]$$

$$= 0 + 0.9 * (\max[0, 100, 0]$$

$$= 0.9 \times 100 = 90.$$

(2,1)

$$Q(2,1) = R(2,1) + 0.9 (\max[(1,2)(1,4)]$$

$$= 0 + 0.9 (\max(90, 0)]$$

$$= 81.$$

(4,1)

$$Q(4,1) = R(4,1) + 0.9(\max(1,2)(1,4)]$$

$$= 0 + 0.9 \times (\max(81, 90)$$

$$= \underline{81} \checkmark.$$