

# Q-Learning Intuition

## Q-Learning Intuition

$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$

## Q-Learning Intuition

# WHERE'S THE Q?

Artificial Intelligence A-Z

© SuperDataScience

## Q-Learning Intuition



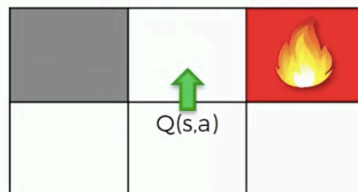
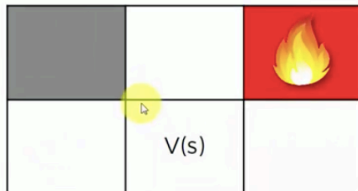
Artificial Intelligence A-Z

© SuperDataScience

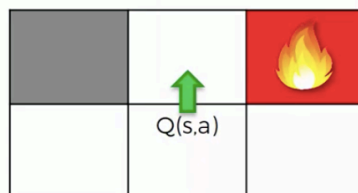
On the left is what we have done so far. This is a Markov decision process and it doesn't matter how we got here. Now it has to make an optimal decision to where it should go. The decisions are based on the current state and all the future state that come from but not from the past. there are three options and based on its experience, it calculated the values in these states for making decision.

At the right, it got modified a little bit which we're going to take the same concept but instead of looking at the values each **state** that it can end up, we are going to look at the value of each **action**. The letter Q is not exactly known, what stands for but some say it's stand for Quality.

## Q-Learning Intuition



## Q-Learning Intuition



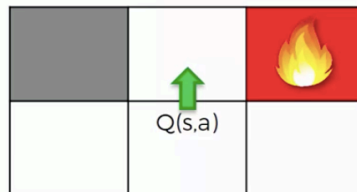
$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$

remember in that in the picture, our state is our current state not the future one. In case of  $Q(s,a)$  it means that the state in the middle bottom and the action of going one up.

## Q-Learning Intuition



$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$



$$Q(s, a) =$$



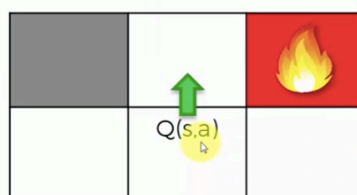
say it: maximum of all possible actions in that part.

If you notice, this is recursive, it means that the  $V$ 's are connected to other  $V$ 's

## Q-Learning Intuition



$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$



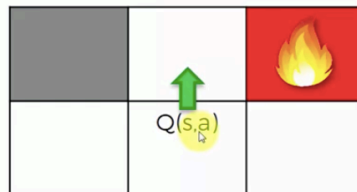
$$Q(s, a) = R(s, a) +$$



## Q-Learning Intuition

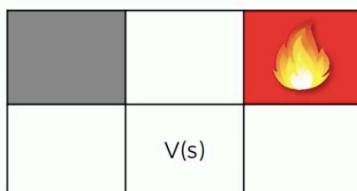


$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$

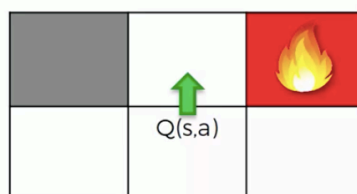


$$Q(s, a) = R(s, a) + \gamma \sum_{s'} (P(s, a, s') V(s'))$$

## Q-Learning Intuition



$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$



$$Q(s, a) = R(s, a) + \gamma \sum_{s'} (P(s, a, s') V(s'))$$



$Q(s,a)$  is exactly the same as  $V(s)$  but without considering the max.

Or in other word we always take the maximum of Q values which in here we have 4 (wall are considered to)

## Q-Learning Intuition


		
	V(s)	

$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$



		
	 Q(s,a)	

$$Q(s, a) = R(s, a) + \gamma \sum_{s'} (P(s, a, s') \overbrace{V(s')})$$

## Q-Learning Intuition

		
	V(s)	

$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$

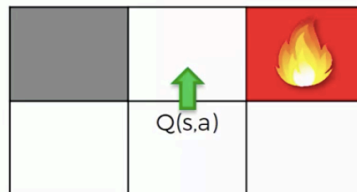
		
	 Q(s,a)	

$$Q(s, a) = R(s, a) + \gamma \sum_{s'} (P(s, a, s') \overbrace{V(s')})$$

# Q-Learning Intuition



$$V(s) = \max_a \left( R(s, a) + \gamma \sum_{s'} P(s, a, s') V(s') \right)$$



$$Q(s, a) = R(s, a) + \gamma \sum_{s'} \left( P(s, a, s') \max_{a'} Q(s', a') \right)$$

$a'$  is the action that we are going to make in the following state.  
Now we have recursive formula for our Q.

## Additional Reading

Additional Reading:

*Markov Decision Processes:  
Concepts and Algorithms*

By Martijn van Otterlo (2009)

Link:

<https://pdfs.semanticscholar.org/968b/ab782e52faf0f7957ca0f38b9e9078454afe.pdf>

