Reinforcement learning: Q Learning

Michał Zientek

July 2025

1 Introduction

Q-learning is a reinforcement learning algorithm that trains an agent to assign values to its possible actions based on its current state, without requiring a model of the environment.

2 Reinforcement learning

Reinforcement learning involves an agent, a set of states **S**, and a set **A** of actions per state. By performing an action $a \in \mathbf{A}$, the agent transitions from state to state. Executing an action in a specific state provides the agent with a reward (a numerical score).

The goal of the agent is to maximize its total reward. It does this by adding the maximum reward attainable from future states to the reward for achieving its current state, effectively influencing the current action by the potential future reward. This potential reward is a weighted sum of expected values of the rewards of all future steps starting from the current state.

3 Bellman Equation

$$Q(s,a)_{new} = Q(s,a) + \alpha \left[R + \gamma \cdot \max_{a'} Q(s',a') - Q(s,a) \right]$$

Explanation of Symbols:

- $Q(s,a)_{new}$: new Q-value based on current Q-value and possible choices.
- $Q(s,a)$: current Q-value for state s and action a.
- α: learning rate, controls how much new information overrides the old.
- R: reward received after taking action a in state s.
- γ: discount factor, determines the importance of future rewards.
- $\max_{a'} Q(s',a')$: maximum predicted Q-value for the next state s', over all possible actions a'.

4 Q Learning Algorithm

It is an iterative learning process:

1. Initialize Q-values.
2. Choose action **a** for state **s** (best Q-value).
3. Perform action **a**, new state **s'**.
4. Measure reward **R**.
5. Update **Q** with Bellman equation $\mathbf{Q}_{new} = ...$

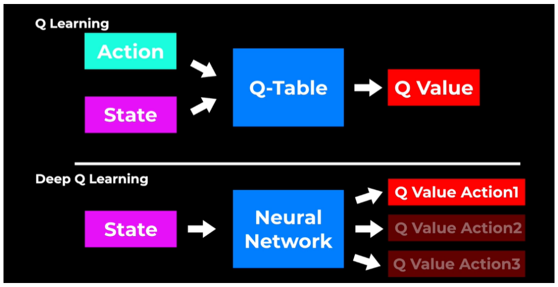
5 Exploration vs Exploitation

In the beginning choose the action randomly so that the agent can explore the environment.

The more training steps we get, the more we reduce the random exploration and use exploitation instead.

This relation is described by the Epsilon ϵ parameter.

6 Deep Q Learning



1. Initialize Q-values.
2. Choose action **a** for state **s** (best Q-value).
3. Perform action **a**, new state **s'**.
4. Measure reward **R**.
5. Update weights of Neural Network.