Reinforcement Learning with Q-Learning

Explore the process of learning optimal behavior with a model-free, value-based reinforcement learning algorithm.

https://github.com/ceyhunCFC/Q-Learning

CORE CONCEPTS

What is Q-Learning?

Q-Learning is a model-free algorithm that allows an agent to learn optimal behavior by interacting with its environment. It does not require prior knowledge of the environment's dynamics; the agent learns directly from experience by observing state transitions and rewards.

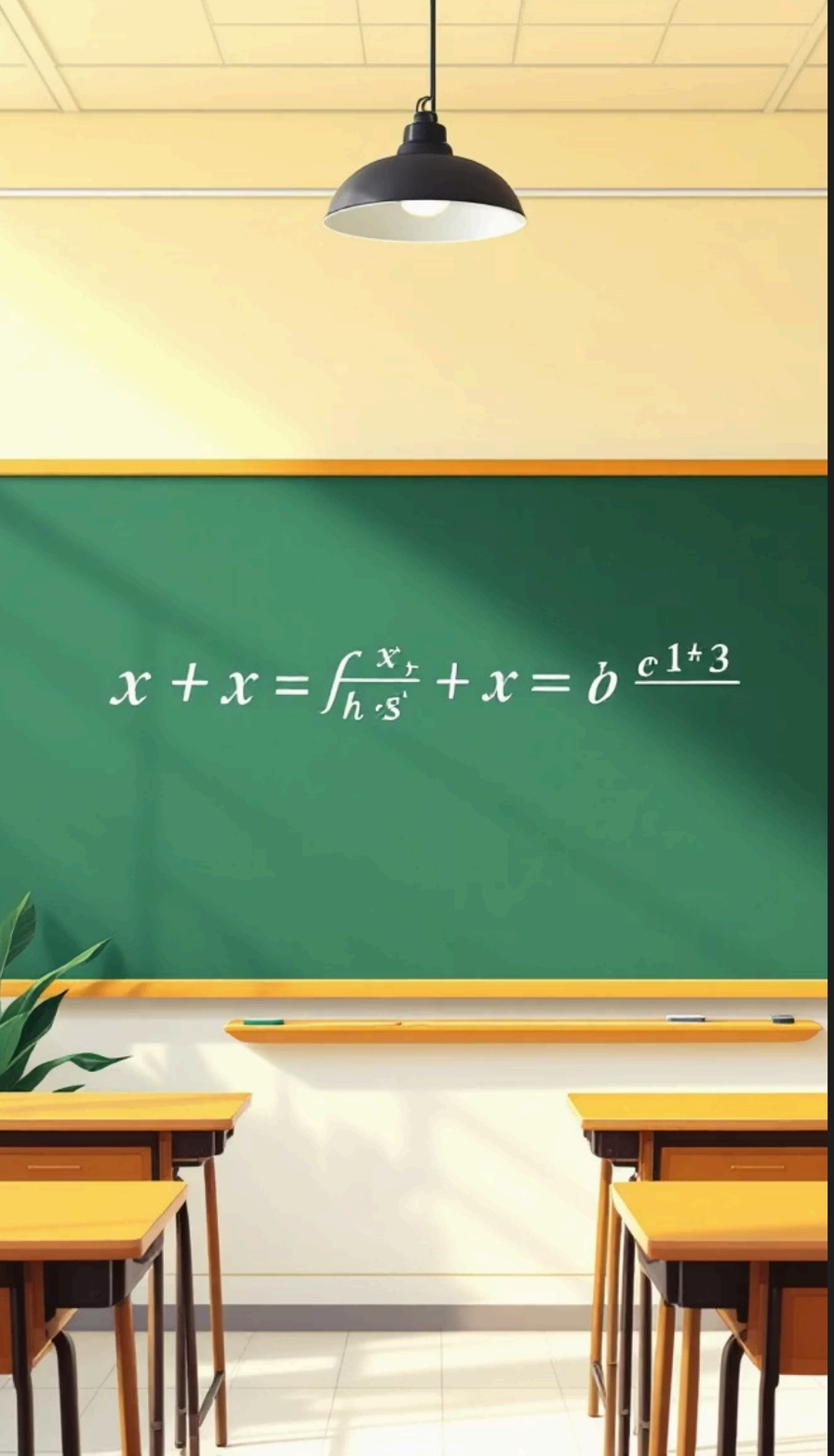
The algorithm's goal is to learn an optimal action selection policy by estimating Q-values for each state-action pair. A Q-value represents the expected cumulative reward that can be obtained by performing a specific action in a given state and then following the optimal policy.

Model-Free

Requires no environmental knowledge

Value-Based

Learns through Q-values



LEARNING MECHANISM

Bellman Optimality Principle

The learning process in Q-Learning is guided by the Bellman optimality principle. After an agent performs an action, it updates the Q-value using the observed reward and the maximum expected future reward from the next state.

1

Action Selection

Action is determined in the current state

2

Reward Observation

Feedback is received from the environment

3

Q-Value Update

Bellman equation is applied

4

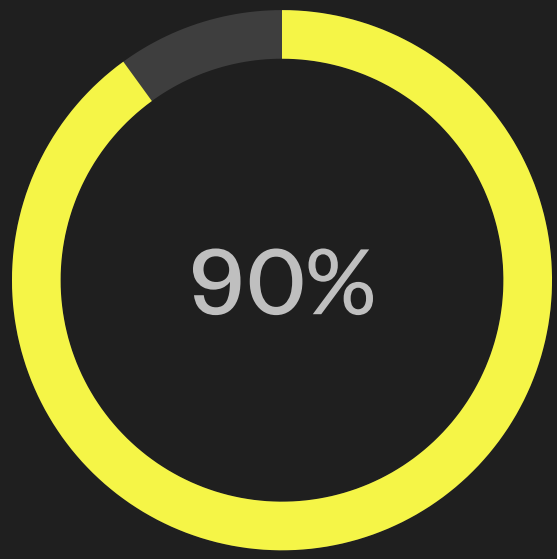
Policy Improvement

Convergence to optimal decisions

Exploration vs. Exploitation Balance

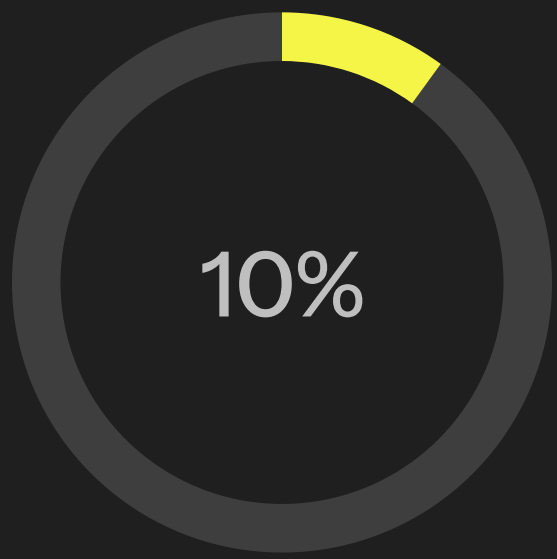
To ensure sufficient exploration of the environment, the epsilon-greedy strategy is used. In the early training stages, the agent prioritizes exploration by choosing random actions. As learning progresses, the exploration rate is reduced, and the agent increasingly uses the learned policy.

This balance allows the agent to both discover new strategies and apply the best behaviors it has learned.



Early Exploration

Random action rate



Later Exploration

Optimal policy usage

APPLICATION ENVIRONMENT

5×5 Gridworld Environment

In this project, Q-Learning was implemented in a simple 5x5 Gridworld environment. The agent starts from a fixed initial position and aims to reach a predefined goal state.

Starting Point

The agent starts from a fixed position

Action Space

Movements: Up, down, left, right

Goal State

Ends with a positive reward

Reward Structure

Efficient navigation is encouraged by providing a small negative reward for each step. Reaching the target, however, yields a large positive reward. This reward structure motivates the agent to learn the shortest path to the target location.

Each Step

Small negative reward (-1)

Reaching Target

Large positive reward (+100)

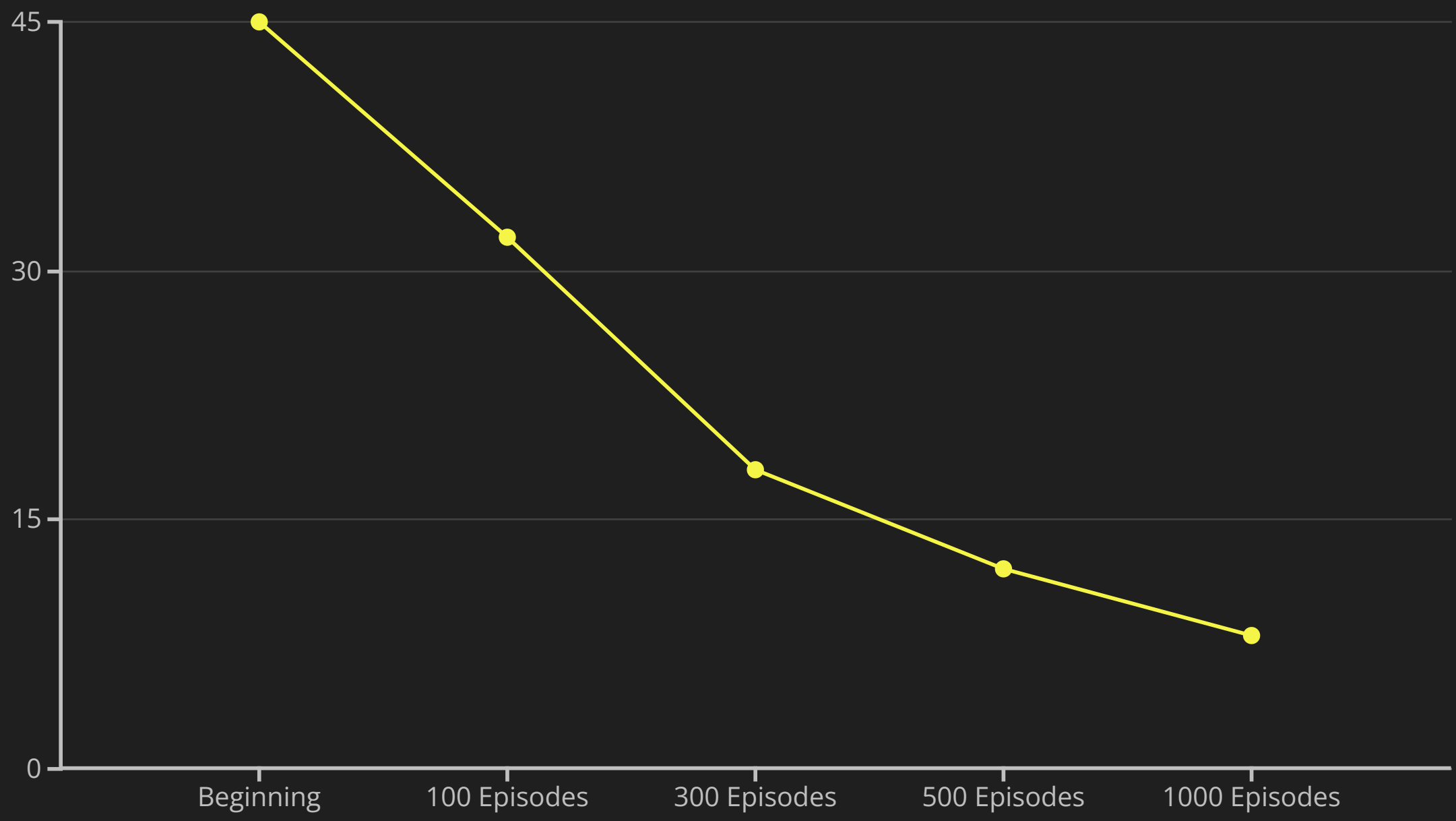
Optimal Path

Minimum number of steps

EXPERIMENTAL RESULTS

Learning Performance

Training results show a clear improvement in agent performance over time. Initially, the agent behaves randomly and requires many steps to reach the target. As learning progresses, the number of steps significantly decreases, indicating successful policy learning.



The learning curve demonstrates that Q-Learning effectively captures optimal behavior in a small, discrete environment.

MODERN DEVELOPMENTS

Limitations and Solutions of Q-Learning

While Q-Learning performs well in simple environments, it suffers from limitations such as overestimation bias and poor scalability. Modern extensions overcome these fundamental limitations.

1. **Standard Q-Learning**

   Overestimation problem and limited scalability

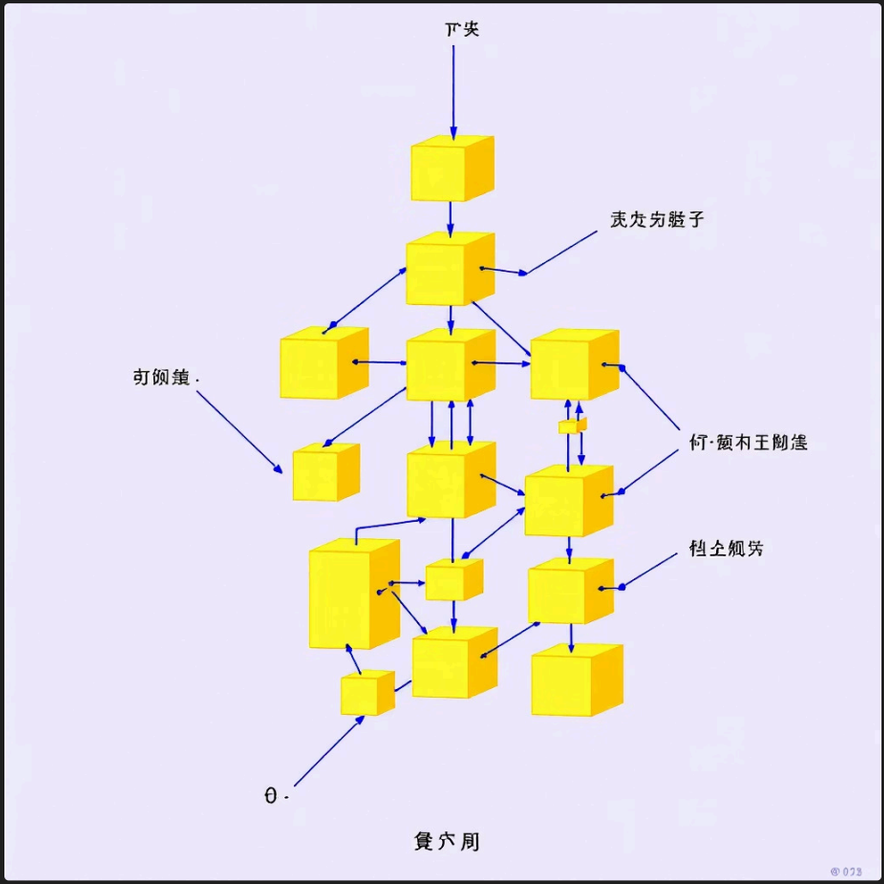
2. **Double Q-Learning**

   More stable estimates by decoupling action selection from evaluation

3. **Deep Q-Networks**

   Scalability in large state spaces with neural networks

Deep Q-Networks (DQN)



For large and high-dimensional state spaces, Deep Q-Networks replace the Q-table with a neural network that approximates the Q-function.

This approach allows Q-Learning based methods to scale to complex tasks such as video games and robotic control. DQN improves training stability using experience replay and target networks.

Video Games

Human-level performance in Atari games

Robotic Control

Complex manipulation tasks

CONCLUSION

The Power and Future of Q-Learning

This project demonstrates that Q-Learning is an effective and reliable reinforcement learning algorithm for small, discrete environments. With a simple yet accurate implementation, the agent successfully learned an optimal policy based solely on interaction with the environment.

Key Achievement

Proven effectiveness and reliability in simple environments

Modern Extensions

Enhanced capabilities with Double Q-Learning and DQN

Real-World Applications

Scalable solutions for complex problems