An Introduction to Reinforcement Learning: Q-Learning

Your Name

Your Institution

September 9, 2025

Outline

What is Reinforcement Learning?

- A type of machine learning where an agent learns to make decisions by taking actions in an environment to maximize cumulative reward.
- It's learning by trial and error.
- The agent is not told which actions to take, but instead must discover which actions yield the most reward by trying them.

Core Concepts

- **Agent**: The learner or decision-maker.
- **Environment**: The world the agent interacts with.
- **State (S)**: A representation of the environment's current condition.
- **Action (A)**: A move the agent can make.
- **Reward (R)**: Feedback from the environment based on the action taken. The agent's goal is to maximize the total reward.

Q-Learning: A Model-Free Approach

- Q-learning is a model-free reinforcement learning algorithm.
- It learns the value of an action in a particular state. It doesn't require a model of the environment.
- The "Q" in Q-learning stands for Quality. Quality represents how useful an action is in gaining some future reward.
- It's based on the Q-function, Q(s, a), which estimates the expected future reward for taking action 'a' in state 's'.

The Q-Table

- The Q-function is implemented as a table, called the Q-table.
- The table has states as rows and actions as columns.
- Each cell Q(s, a) in the table contains the estimated reward for taking action 'a' from state 's'.
- The agent uses this table to select the best action for a given state.

The Q-Learning Process

1. Initialize the Q-table with zeros.
2. For a number of episodes:
    1. Start in an initial state.
    2. While the episode is not finished:
        - Choose an action (using an exploration/exploitation strategy).
        - Perform the action and observe the reward and the new state.
        - Update the Q-value for the state-action pair using the Bellman equation.
        - Move to the new state.

The Dilemma

- **Exploitation**: The agent chooses the action with the highest known Q-value for the current state. This is "greedy" because it exploits known information.
- **Exploration**: The agent chooses a random action to discover new state-action pairs and potentially find better rewards.
- There is a trade-off between exploiting known good actions and exploring to find even better ones.
- A common strategy is the **Epsilon-Greedy** policy, where the agent explores with a small probability (epsilon) and exploits the rest of the time.

The Update Rule

The core of the Q-learning algorithm is the Bellman equation, which is used to update the Q-values in the table:

The Bellman Equation

$$Q(s, a) \leftarrow Q(s, a) + \alpha[R + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$

Where:

- α (alpha) is the learning rate (how much we update Q-values).
- γ (gamma) is the discount factor (importance of future rewards).

Summary and Applications

- Q-learning is a powerful, yet simple, reinforcement learning algorithm.
- It allows an agent to learn to act optimally in an environment through experience.
- It's the foundation for many more advanced deep reinforcement learning techniques.
- Applications include:
  - Game playing (e.g., simple video games)
  - Robotics for simple tasks
  - Route optimization

Thank You!