A03 Machine Learning Concepts

Contributors: Martin Demel, Jiri Musil, Jasmine Campos, Munachimso Benson

Department of Science, Technology, Engineering & Math, Houston Community College

ITAI - 1371 Introduction to Machine Learning

Professor Vishwa Rao

September 5th, 2025.

# Introduction

# Each team member read the article on reinforcement learning, and we have summarized in our own words how an agent learns from interaction rather than labeled answers. We focused on the cycle of state, action, reward, and next state, and on how policies and value estimates emerge from experience. The two trees choices, the line world, and the grid world with a cactus, helped us understand Q values, episodes, and the balance between exploration and exploitation in the effort to maximize cumulative reward over time.

# Body

We understand that reinforcement learning is a process of learning which actions to take and when to take them to maximize cumulative reward over time. This differs from supervised learning because the agent is not given the correct action, and from unsupervised learning because the goal is not to find structure but to maximize reward. At each step, the agent observes a state, acts according to its current policy, receives a reward, and then observes the next state. Over multiple episodes, it updates a policy that maps states to actions and a value function that estimates how desirable states are when following the policy. A complete sequence from an initial state to a final state is called an episode or epoch.

The two trees scenario made these ideas concrete. The agent must choose between Tree A and Tree B without knowing which yields more bananas. Tree A produces an average of 3 bananas with a standard deviation of 1, while Tree B produces an average of 6 with a standard deviation of 2. The agent stores its experience as Q values that reflect the expected reward of each action based on past outcomes. It initially prefers Tree A because it has already seen rewards there, but by occasionally sampling Tree B, it discovers that Tree B is more fruitful and shifts its choice. This resembles the multi-armed bandit problem.

The line world updates state changes because each action moves the agent left or right. Early in training, the learned policy moves left toward a known tree, even though the right tree actually offers higher rewards. This illustrates the explore-exploit dilemma. Purely greedy behavior always chooses the option that seems best at the moment and can overlook better opportunities. The epsilon-greedy policy balances exploration with a probability epsilon and exploits otherwise. After fifteen episodes using this approach, the agent still tends to go left from all states. After 2500 episodes, the updated Q values and policy favor moving right. In both the line world and the grid world, Q values represent expected discounted cumulative rewards, so rewards received later are valued less than immediate rewards.

A screenshot of a game

AI-generated content may be incorrect.Looking at Figure 1, the grid, the agent can move up, down, left, and stepping on the cactus gives a negative reward and ends the episode. The task involves learning a path to reach rewarding states while avoiding obstacles. Adjusting epsilon influences how much the agent explores different routes, and repeated episodes improve both Q values and the policy.

Figure 1: Grid world with cactus environment used to demonstrate reinforcement learning concepts. Adapted from Wilber (2025).

# Conclusion

We have learned that reinforcement learning is a process of learning by doing, guided by rewards, and improved through progress and experience. Policies and value estimates change across episodes (epochs), and Q values show which actions are beneficial in each situation. Furthermore, we have also observed how exploration, exploitation, and discounted rewards work together so the agent can choose actions that maximize total reward over time.

References

Wilber, J. (August 8th, 2025). Reinforcement learning. MLU-Explain. <https://mlu-explain.github.io/reinforcement-learning/>

GeeksforGeeks. (2025, February 24). Reinforcement learning. GeeksforGeeks. <https://www.geeksforgeeks.org/machine-learning/what-is-reinforcement-learning/>