Fundamentals of Reinforcement Learning: Learning Objectives

Module 00: Welcome to the Course
Understand the prerequisites, goals and roadmap for the course.

Module 01: The K-Armed Bandit Problem

Lesson 1: The K-Armed Bandit Problem
- Define reward
- Understand the temporal nature of the bandit problem
- Define k-armed bandit
- Define action-values

Lesson 2: What to Learn? Estimating Action Values
- Define action-value estimation methods
- Define exploration and exploitation
- Select actions greedily using an action-value function
- Define online learning
- Understand a simple online sample-average action-value estimation method
- Define the general online update equation
- Understand why we might use a constant stepsize in the case of non-stationarity

Lesson 3: Exploration vs. Exploitation Tradeoff
- Define epsilon-greedy
- Compare the short-term benefits of exploitation and the long-term benefits of exploration
- Understand optimistic initial values
- Describe the benefits of optimistic initial values for early exploration
- Explain the criticisms of optimistic initial values
- Describe the upper confidence bound action selection method
- Define optimism in the face of uncertainty

Module 02: Markov Decision Processes

Lesson 1: Introduction to Markov Decision Processes
- Understand Markov Decision Processes, or MDPs
- Describe how the dynamics of an MDP are defined
- Understand the graphical representation of a Markov Decision Process
- Explain how many diverse processes can be written in terms of the MDP framework

Lesson 2: Goal of Reinforcement Learning
- Describe how rewards relate to the goal of an agent

Understand episodes and identify episodic tasks

Lesson 3: Continuing Tasks

Formulate returns for continuing tasks using discounting
Describe how returns at successive time steps are related to each other
Understand when to formalize a task as episodic or continuing

Module 03: Values Functions & Bellman Equations

Lesson 1: Policies and Value Functions

Recognize that a policy is a distribution over actions for each possible state
Describe the similarities and differences between stochastic and deterministic policies
Identify the characteristics of a well-defined policy
Generate examples of valid policies for a given MDP
Describe the roles of state-value and action-value functions in reinforcement learning
Describe the relationship between value functions and policies
Create examples of valid value functions for a given MDP

Lesson 2: Bellman Equations

Derive the Bellman equation for state-value functions
Derive the Bellman equation for action-value functions
Understand how Bellman equations relate current and future values
Use the Bellman equations to compute value functions

Lesson 3: Optimality (Optimal Policies & Value Functions)

Define an optimal policy
Understand how a policy can be at least as good as every other policy in every state
Identify an optimal policy for given MDPs
Derive the Bellman optimality equation for state-value functions
Derive the Bellman optimality equation for action-value functions
Understand how the Bellman optimality equations relate to the previously introduced Bellman equations
Understand the connection between the optimal value function and optimal policies
Verify the optimal value function for given MDPs

Module 04: Dynamic Programming

Lesson 1: Policy Evaluation (Prediction)

Understand the distinction between policy evaluation and control

Explain the setting in which dynamic programming can be applied, as well as its limitations

Outline the iterative policy evaluation algorithm for estimating state values under a given policy

Apply iterative policy evaluation to compute value functions

Lesson 2: Policy Iteration (Control)

Understand the policy improvement theorem

Use a value function for a policy to produce a better policy for a given MDP

Outline the policy iteration algorithm for finding the optimal policy

Understand “the dance of policy and value”

Apply policy iteration to compute optimal policies and optimal value functions

Lesson 3: Generalized Policy Iteration

Understand the framework of generalized policy iteration

Outline value iteration, an important example of generalized policy iteration

Understand the distinction between synchronous and asynchronous dynamic programming methods

Describe brute force search as an alternative method for searching for an optimal policy

Describe Monte Carlo as an alternative method for learning a value function

Understand the advantage of Dynamic programming and “bootstrapping” over these alternative strategies for finding the optimal policy