Sample Based Learning Methods: Learning Objectives

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

Module 01: Monte Carlo Methods for Prediction & Control

Lesson 1: Introduction to Monte Carlo Methods
Understand how Monte-Carlo methods can be used to estimate value functions from sampled interaction
Identify problems that can be solved using Monte-Carlo methods
Use Monte Carlo prediction to estimate the value function for a given policy.

Lesson 2: Monte Carlo for Control
Estimate action-value functions using Monte Carlo
Understand the importance of maintaining exploration in Monte Carlo algorithms
Understand how to use monte carlo methods to implement a GPI algorithm.
Apply Monte Carlo with exploring starts to solve an MDP

Lesson 3: Exploration Methods for Monte Carlo
Understand why Exploring Starts can be problematic in real problems
Describe an alternative exploration method for Monte Carlo control

Lesson 4: Off-policy learning for prediction
Understand how off-policy learning can help deal with the exploration problem
Produce examples of target policies and examples of behavior policies.
Understand importance sampling
Use importance sampling to estimate the expected value of a target distribution using samples from a different distribution.
Understand how to use importance sampling to correct returns
Understand how to modify the monte carlo prediction algorithm for off-policy learning.

Module 2: Temporal Difference Learning Methods for Prediction

Lesson 1: Introduction to Temporal Difference Learning
Define temporal-difference learning
Define the temporal-difference error
Understand the TD(0) algorithm

Lesson 2: Advantages of TD

Understand the benefits of learning online with TD
Identify key advantages of TD methods over Dynamic Programming and Monte Carlo methods
Identify the empirical benefits of TD learning

Module 3: Temporal Difference Learning Methods for Control

Lesson 1: TD for Control

Explain how generalized policy iteration can be used with TD to find improved policies
Describe the Sarsa Control algorithm
Understand how the Sarsa control algorithm operates in an example MDP
Analyze the performance of a learning algorithm

Lesson 2: Off-policy TD Control: Q-learning

Describe the Q-learning algorithm
Explain the relationship between q-learning and the Bellman optimality equations.
Apply q-learning to an MDP to find the optimal policy
Understand how Q-learning performs in an example MDP
Understand the differences between Q-learning and Sarsa
Understand how Q-learning can be off-policy without using importance sampling
Describe how the on-policy nature of SARSA and the off-policy nature of Q-learning affect their relative performance

Lesson 3: Expected Sarsa

Describe the Expected Sarsa algorithm
Describe Expected Sarsa's behaviour in an example MDP
Understand how Expected Sarsa compares to Sarsa control
Understand how Expected Sarsa can do off-policy learning without using importance sampling
Explain how Expected Sarsa generalizes Q-learning

Module 4: Planning, Learning & Acting

Lesson 1: What is a model?

Describe what a model is and how they can be used
Classify models as distribution models or sample models
Identify when to use a distribution model or sample model
Describe the advantages and disadvantages of sample models and distribution models
Explain why sample models can be represented more compactly than distribution models

Lesson 2: Planning
Explain how planning is used to improve policies
Describe random-sample one-step tabular Q-planning

Lesson 3: Dyna as a formalism for planning
Recognize that direct RL updates use experience from the environment to improve a policy or value function
Recognize that planning updates use experience from a model to improve a policy or value function
Describe how both direct RL and planning updates can be combined through the Dyna architecture
Describe the Tabular Dyna-Q algorithm
Identify the direct-RL and planning updates in Tabular Dyna-Q
Identify the model learning and search control components of Tabular Dyna-Q
Describe how learning from both direct and simulated experience impacts performance
Describe how simulated experience can be useful when the model is accurate

Lesson 4: Dealing with inaccurate models
Identify ways in which models can be inaccurate
Explain the effects of planning with an inaccurate model
Describe how Dyna can plan successfully with a partially inaccurate model
Explain how model inaccuracies produce another exploration-exploitation trade-off
Describe how Dyna-Q+ proposes a way to address this trade-off

Lesson 5: Course wrap-up