Section 3: Markov Chains: Theory, Examples, and Key Concepts

Eng. Ahmed Métwalli

October 30, 2024

Introduction

Markov Chains are fundamental probabilistic models used to describe systems that transition from one state to another. They are widely used in fields such as economics, engineering, biology, and artificial intelligence. The primary characteristic of a Markov Chain is the Markov Property, which states that the probability of transitioning to a future state depends only on the current state, not on the sequence of events that preceded it.

In reinforcement learning and decision-making contexts, Markov Chains serve as the foundation for more complex models like Markov Decision Processes (MDPs) and algorithms such as Q-Learning.

1 Definition of a Markov Chain

A Markov Chain is formally defined as a tuple (S, P) where:

- S is a finite set of states, denoted as $S = \{s_1, s_2, \ldots, s_n\}$.
- P is the state transition probability matrix, where each element P_{ij} represents the probability of transitioning from state s_i to state s_j. This can be expressed as:

$$P_{ij} = P(s_{t+1} = s_j \mid s_t = s_i).$$

The Markov Property implies that the probability distribution over the next state depends only on the current state and not on the history of previous states.

2 Transition Probability Matrix

The transition probabilities for a Markov Chain are typically represented in matrix form:

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix},$$

where each row sums to 1, i.e., $\sum_j P_{ij} = 1$ for all i.

3 Example 1: Robot Navigation Model

Consider a robot that can move between three rooms: A, B, and C. The robot moves based on the following probabilities:

- If the robot is in Room A, it stays in Room A with probability 0.6, moves to Room B with probability 0.3, and moves to Room C with probability 0.1.
- If the robot is in Room B, it moves to Room A with probability 0.2, stays in Room B with probability 0.5, and moves to Room C with probability 0.3.
- If the robot is in Room C, it moves to Room A with probability 0.3 and stays in Room C with probability 0.7.

The transition probability matrix $\mathbf{P}$ is:

$$\mathbf{P} = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.3 & 0.0 & 0.7 \end{bmatrix}.$$

Steady-State Probabilities

Let the steady-state vector be:

$$\pi = \begin{bmatrix} \pi_A & \pi_B & \pi_C \end{bmatrix}.$$

We solve the system of equations:

$$\pi_A = 0.6\pi_A + 0.2\pi_B + 0.3\pi_C,$$

$$\pi_B = 0.3\pi_A + 0.5\pi_B,$$

$$\pi_C = 0.1\pi_A + 0.3\pi_B + 0.7\pi_C,$$

along with the constraint $\pi_A + \pi_B + \pi_C = 1$. Solving these, we find:

$$\pi_A = 0.4, \quad \pi_B = 0.3, \quad \pi_C = 0.3.$$

4 Example 2: Agent's Decision in a Game

In a simple game, an agent can choose between two strategies: Attack or Defend. The agent's decision depends on the opponent's last move:

- If the opponent attacked last, the agent responds by attacking with probability 0.7 and defending with probability 0.3.
- If the opponent defended last, the agent responds by attacking with probability 0.4 and defending with probability 0.6.

The states are:

$$S = \{\text{Attack}, \text{Defend}\}.$$

The transition probability matrix $\mathbf{P}$ is:

$$\mathbf{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}.$$

Steady-State Probabilities

Let the steady-state vector be:

$$\pi = \begin{bmatrix} \pi_{\text{Attack}} & \pi_{\text{Defend}} \end{bmatrix}.$$

We solve:

$$\pi_{\text{Attack}} = 0.7\pi_{\text{Attack}} + 0.4\pi_{\text{Defend}},$$

$$\pi_{\text{Defend}} = 0.3\pi_{\text{Attack}} + 0.6\pi_{\text{Defend}}.$$

With $\pi_{\text{Attack}} + \pi_{\text{Defend}} = 1$, we find:

$$\pi_{\text{Attack}} = 0.57, \quad \pi_{\text{Defend}} = 0.43.$$

5 Key Notes

Markov Chain

A Markov Chain is a mathematical model that describes a sequence of events or states where the probability of each event depends only on the current state, not on how the state was reached. This characteristic is known as the Markov Property. Markov Chains are used in various fields such as reinforcement learning, genetics, finance, and natural language processing to model systems that evolve over time with probabilistic transitions.

Markov Decision Process (MDP)

An MDP is an extension of a Markov Chain that includes actions and rewards. It models a decision-making scenario where an agent chooses actions that influence the state transitions and receives rewards based on the actions taken. MDPs provide a more comprehensive way to model situations where decisions must be made in an uncertain environment.

Q-Learning

Q-Learning is a reinforcement learning algorithm that finds the optimal policy for an agent in an MDP. It does not require prior knowledge of the transition probabilities or rewards. Q-Learning updates the value of each state-action pair using an update rule based on the immediate reward and the estimated future value.

Purpose of Relying on the Current State

The Markov Property, which relies only on the current state, simplifies modeling and analysis by ignoring the history of prior states. This assumption holds true in many real-world scenarios, allowing for efficient prediction of future states and transitions. By reducing complexity, the Markov assumption enables mathematical tools like transition matrices and steady-state analysis.

Transition Matrix

The transition matrix $\mathbf{P}$ contains the probabilities of transitioning from each state to every other state. Each entry P_{ij} in the matrix represents the probability of transitioning from state i to state j. The rows of the transition matrix sum to 1, reflecting that the system must transition to some state (including potentially remaining in the current state).

Steady-State Probability

The steady-state probability distribution π represents the long-term behavior of the Markov Chain. It indicates the proportion of time spent in each state as the system evolves over time. The steady-state distribution satisfies the condition $\pi\mathbf{P} = \pi$, meaning that the system remains in the same distribution after transitioning.

Summary

To summarize:

- Markov Chain = States + Probabilistic Transitions.
- MDP = States + Actions + Probabilistic Transitions + Rewards.
- Q-Learning = Learning optimal actions in an MDP using Q-values.