# Markov Decision Processes

## Markov Property and Markov Chains

Sarwan Ali

Department of Computer Science
Georgia State University

Understanding Markov Decision Processes

# Today's Learning Journey

1. Introduction to MDPs
2. The Markov Property
3. Markov Chains
4. Classification of States
5. From Markov Chains to MDPs
6. Next Steps

# What are Markov Decision Processes?

**Definition**

A **Markov Decision Process (MDP)** is a mathematical framework for modeling decision-making in situations where outcomes are partly random and partly under the control of a decision maker.

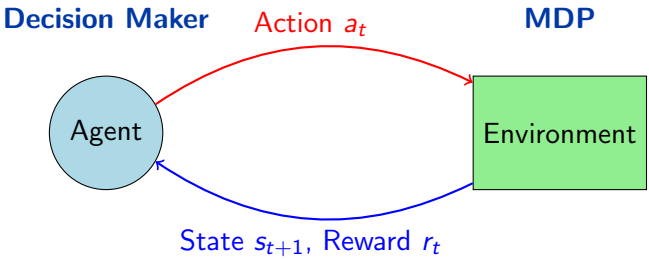
**Key Components:**

- States
- Actions
- Transitions
- Rewards
- Policy

**Applications:**

- Robotics
- Game Playing
- Finance
- Healthcare
- Resource Management

# Why MDPs Matter in Reinforcement Learning




## Core Insight

MDPs provide the mathematical foundation for formulating reinforcement learning problems where an agent learns optimal behavior through trial and error interactions with an environment.

# The Markov Property: Core Concept

## Markov Property (Memoryless Property)

The future is independent of the past given the present state.

**Mathematical Definition:**

$$P(S_{t+1} = s' | S_t = s, S_{t-1} = s_{t-1}, \ldots, S_0 = s_0) = P(S_{t+1} = s' | S_t = s) \quad (1)$$

**Markov Process:**

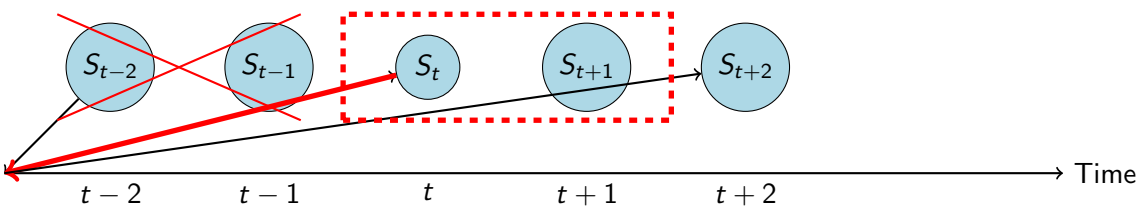
Tomorrow's weather depends only on today's weather, not on the weather from last week.

**Non-Markov Process:**

Stock prices that depend on trends from multiple previous days.

# Visualizing the Markov Property

Past is irrelevant **Only Current State Matters!**




## Key Insight

The Markov property allows us to make predictions about the future using only the current state, dramatically simplifying the computational complexity of decision-making problems.

# Examples: Markov vs Non-Markov

## Markov Examples

- **Weather:** Today's weather determines tomorrow's
- **Chess:** Current board position contains all relevant information
- **Inventory:** Current stock level determines future decisions
- **Traffic Light:** Current light color determines next state

## Non-Markov Examples

- **Stock Market:** Historical trends matter
- **Disease Diagnosis:** Patient history is crucial
- **Language:** Previous words affect meaning
- **Poker:** Memory of played cards matters

## Making Non-Markov Problems Markov

We can often make non-Markov problems Markov by **expanding the state space** to include relevant history.

# Markov Chains: Definition and Properties

## Definition

A **Markov Chain** is a sequence of random variables $S_0, S_1, S_2, \ldots$ where each variable satisfies the Markov property.

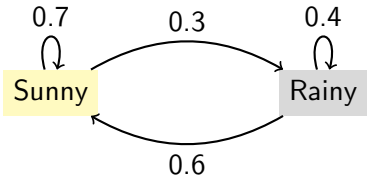
**Key Components:**

- **State Space:** $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ (finite set of states)
- **Transition Probabilities:** $P_{ij} = P(S_{t+1} = j | S_t = i)$
- **Transition Matrix:** $\mathbf{P} = [P_{ij}]_{n \times n}$
- **Initial Distribution:** $\boldsymbol{\pi}_0 = [\pi_0(s_1), \pi_0(s_2), \ldots, \pi_0(s_n)]$

## Properties of Transition Matrix

- Each row sums to 1: $\sum_{j=1}^{n} P_{ij} = 1$
- All entries are non-negative: $P_{ij} \geq 0$

# Simple Weather Example




**Transition Matrix:**

$$\mathbf{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} \tag{2}$$

**Interpretation:**

- If sunny today: 70% chance sunny tomorrow
- If rainy today: 60% chance sunny tomorrow

**Question**

If it's sunny today, what's the probability it will be sunny in 2 days?

**Answer:** $P_{11}^2 = 0.7^2 + 0.3 \times 0.6 = 0.49 + 0.18 = 0.67$

# n-Step Transition Probabilities

## Chapman-Kolmogorov Equation

The probability of transitioning from state $i$ to state $j$ in $n$ steps:

$$P_{ij}^{(n)} = (\mathbf{P}^n)_{ij} \tag{3}$$

**For our weather example:**

$$\mathbf{P}^2 = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}^2 = \begin{bmatrix} 0.67 & 0.33 \\ 0.66 & 0.34 \end{bmatrix} \tag{4}$$

$$\mathbf{P}^3 = \begin{bmatrix} 0.667 & 0.333 \\ 0.666 & 0.334 \end{bmatrix} \tag{5}$$

## Observation

As $n \to \infty$, the transition probabilities approach a steady state, regardless of the initial state!

# Stationary Distribution

## Definition

A **stationary distribution** $\boldsymbol{\pi}$ satisfies: $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, or equivalently: $\boldsymbol{\pi}^T = \mathbf{P}^T\boldsymbol{\pi}^T$

**For our weather example:**

$$[\pi_s, \pi_r] = [\pi_s, \pi_r] \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} \tag{6}$$

$$\pi_s = 0.7\pi_s + 0.6\pi_r \tag{7}$$

$$\pi_r = 0.3\pi_s + 0.4\pi_r \tag{8}$$

With $\pi_s + \pi_r = 1$, we get: $\pi_s = \frac{2}{3}, \pi_r = \frac{1}{3}$

## Interpretation

In the long run, it's sunny $\frac{2}{3}$ of the time and rainy $\frac{1}{3}$ of the time, regardless of today's weather!

# State Classification

## Accessibility and Communication

- State $j$ is **accessible** from state $i$ if $P_{ij}^{(n)} > 0$ for some $n \geq 0$
- States $i$ and $j$ **communicate** if they are accessible from each other
- Communication is an equivalence relation, partitioning states into **communicating classes**

**Irreducible Chain:**

- All states communicate
- Only one communicating class
- Every state can reach every other state

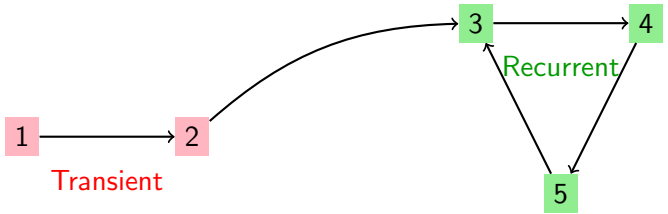
**Reducible Chain:**

- Multiple communicating classes
- Some states cannot reach others
- Chain can be "reduced"

# Transient vs Recurrent States

## Definitions

- A state is **recurrent** if, starting from that state, the probability of eventually returning to it is 1
- A state is **transient** if, starting from that state, there's a positive probability of never returning to it




## Key Property

In a finite Markov chain, not all states can be transient. There must be at least one recurrent state.

# Adding Actions and Rewards

## From Markov Chain to MDP

A Markov Chain becomes an MDP when we add:

- **Actions:** $\mathcal{A}$ - set of possible actions
- **Rewards:** $R(s, a, s')$ - reward for transitioning from $s$ to $s'$ via action $a$
- **Policy:** $\pi(a|s)$ - probability of taking action $a$ in state $s$

**State Transition becomes Action-Dependent:**

$$P(S_{t+1} = s'|S_t = s, A_t = a) = P^a_{ss'} \tag{9}$$

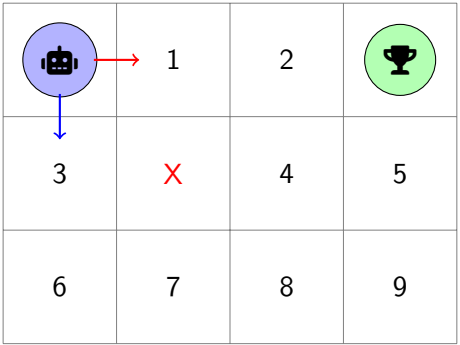
**Markov Chain:**

- Passive observation
- Fixed transition probabilities
- No control over process

**MDP:**

- Active decision making
- Action-dependent transitions
- Agent controls the process

# Example: Robot Navigation MDP




**MDP Components:**

- **States:** Grid positions (S, 1, 2, ..., 9, G)
- **Actions:** {Up, Down, Left, Right}
- **Transitions:** Move to adjacent cell (with some noise)
- **Rewards:** +10 for reaching goal, -1 for each step, -10 for obstacle

# Key Concepts Summary

## Markov Property

Future depends only on present, not on past $\Rightarrow$ Computational efficiency

## Markov Chains

Sequential process with Markov property $\Rightarrow$ Foundation for understanding state evolution

## State Classification

Transient vs Recurrent, Communicating classes $\Rightarrow$ Long-term behavior analysis

## Stationary Distribution

Long-term equilibrium probabilities $\Rightarrow$ Steady-state analysis

## Markov Chains $\rightarrow$ MDPs

Add actions and rewards $\Rightarrow$ Decision-making framework for RL

# What's Coming Next

## Finite MDPs

- States
- Actions
- Rewards
- Transition probabilities

Thank you for your attention!

Questions?