COMP 480/580 — Probabilistic Algorithms and Data Structures

Oct 29, 2024

Lecture 18: Markov Chain

Lecturer: Anshumali Shrivastava

Scribe By: Yanxiao Cai (yc139), Kevin Ni (kn22), Sanya Garg (sg157), Caleb McKinney (cam26)

1 Motivations

Markov Chains are an essential tool in machine learning and artificial intelligence, especially for modeling sequential data with probabilistic transitions between states. They are widely applied in fields such as reinforcement learning, language modeling, and financial forecasting, where understanding dependencies over time is crucial. The core property of a Markov Chain is its "memoryless" nature, meaning that the probability of transitioning to the next state depends only on the current state, not on the full sequence of prior states. This property allows Markov Chains to simplify complex models by focusing solely on the present state.

A significant motivation for using Markov Chains is their efficiency in sampling from large or complex state spaces Ω. In many applications, we aim to sample an element $x \in \Omega$ such that the probability of selecting x is proportional to a weight $w(x)$. Mathematically, this can be expressed as

$$P(\text{sampling } x \in \Omega) = \frac{w(x)}{\sum_{y \in \Omega} w(y)}.$$

However, calculating $\sum_{y \in \Omega} w(y)$ directly is often computationally infeasible when $|\Omega|$ is large. Markov Chains provide a practical solution by enabling approximate sampling through iterative steps, avoiding the need for exhaustive computation.

A classic example of this sampling problem is the **Random Permutation Problem**. Given N items, we want to generate a random permutation where each permutation has equal probability. The state space Ω here consists of all possible orderings of the N items, so $|\Omega| = N!$, making direct sampling impractical for large N. Instead, a Markov Chain approach can be used: by repeatedly swapping pairs of items at random, the distribution over permutations gradually approaches uniformity. This method efficiently explores the space of permutations without requiring a full enumeration of all $N!$ possibilities.

Algorithm 1 Random Permutation Problem

Input: N elements to permute, K the number of iterations
Output: A random permutation of the N elements
Initialize the list with the N elements
for $k = 1$ to K **do**
    Randomly select indices $i, j \in \{1, \ldots, N\}$
    Swap the elements at positions i and j
end for
return The list as a random permutation

In this algorithm, as K increases, the list approaches a uniformly random permutation of the original elements. This random swapping process illustrates a Markov Chain approach, where each swap represents a state transition.

2 Markov Chains

2.1 Definition of Stochastic Process

A **stochastic process** is a collection of random variables $\{X_t : t \in T\}$ indexed by time T, where each random variable X_t takes values in a state space Ω, the set of all possible states the process can be in.

For each time $t \in T$, the random variable X_t represents the state of the system at that time, so $X_t : \mathcal{S} \to \Omega$, where $\mathcal{S}$ is the sample space of all possible outcomes. Each $X_t(\omega)$ denotes the state at time t corresponding to the outcome $\omega \in \mathcal{S}$. A stochastic process describes how the state X_t evolves over time in a probabilistic manner. Each possible sequence $\{X_t(\omega) : t \in T\}$ is called a realization or sample path of the process.

2.2 Definition of a Markov Chain

A **Markov Chain** is a stochastic process that satisfies the Markov property (referred to as "memorylessness") and can be used to sample from complex distributions. In simpler terms, it is a process through which future outcomes can be predicted based solely on the present state. Most importantly, such predictions are just as accurate as ones that could be made while knowing the process's entire history. In other words, conditional on the system's present state, its future, and past states are independent.

Mathematically, a Markov Chain is defined as a stochastic process $\{X_t\}_{t\geq 0}$ where each X_t represents the state at time t. Given the state space Ω, where Ω is the set of all values that X_t can take, Markov chains must satisfy the following property of "memorylessness":

$$P(X_{t+1} = x \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0) = P(X_{t+1} = x \mid X_t = x_t)$$

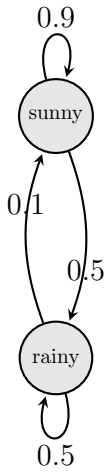
As shown above, the probability of X_{t+1} solely depends on the state of X_t, and not X_{t-1}, X_{t-2}, $X_{t-3} \ldots$. In other words, the probability of transitioning to a new state depends only on the current state and not on how the process arrived there. Markov chains can be used in a variety of applications including prediction and estimation, sampling from exponentially complex distributions, and modeling complex phenomena such as stock markets or the weather.

2.3 Graph Representation

A Markov chain can be visualized as a directed graph $G = (V, E)$, where:

1. Each node in the graph represents a state $s_i \in S$.
2. Each edge from s_i to s_j is associated with the transition probability p_{ij}.

The graph has directed edges where each edge $(s_i \to s_j)$ is labeled by p_{ij}, representing the likelihood of transitioning from state s_i to state s_j. Self-loops represent the probability of remaining in the same state.



For example, above is a graph representation for a two-state Markov chain with states "sunny" (s_1) and "rainy" (s_2) where:

1. The probability of staying "sunny" is $p_{11} = 0.9$.
2. The probability of transitioning from "sunny" to "rainy" is $p_{12} = 0.1$.
3. The probability of staying "rainy" is $p_{22} = 0.5$.
4. The probability of transitioning from "rainy" to "sunny" is $p_{21} = 0.5$.

2.4 Transition Matrix

All graphs can be represented as an adjacency matrix P, where entry $P_{i,j}$ represents the weight of the edge between vertices i and node j. Therefore, given a Markov chain, we can completely describe it with a matrix representation called the stochastic matrix. For a Markov chain with state space Ω with $|\Omega| = N$, we can construct an $N \times N$ stochastic matrix P, where entry $P_{i,j}$ gives the probability of transitioning from state i to state j.

Using this representation, we can use linear algebra to analyze the behavior of our model. One such application is computing how state probabilities will evolve over time. Let π_t be a row vector containing the probabilities of being in each state at time t (where $\pi_t[i]$ is the probability of being in state i at time t). To find the state probabilities k steps later, we use:

$$\pi_{t+k} = \pi_t \times P^k$$

where P^k represents multiplying matrix P by itself k times: $\underbrace{P \times P \times \cdots \times P}_{k \text{ times}}$.

Finally, we introduce the notion of a stationary distribution of a Markov chain, which is a probability distribution which our model will converge to as $t \to \infty$. In this matrix representation, the stationary distribution will satisfy $\pi = \pi P$. In the next lecture, we will identify key properties of the stationary distribution for a Markov chain, including methods to check for existence.