COMP 480/580 — Probabilistic Algorithms and Data Structures

October 31, 2024

Lecture 19

Lecturer: Anshumali Shrivastava

Scribes: Aditi Buch(ab156), Andrew Holzbach(ath10), Meaghan Ramlakhan(mcr8), Pinar Targil(pt38)

1 Markov Chains

A Markov chain is a stochastic process describing a series of transitions between states belonging to a space Ω where the probability of transitioning to any future state depends only on the current state, not on any prior states (memoryless property). A Markov chain can be represented as a sequence of random variables X_1, X_2 ... X_i... such that the probability of moving to state X_{i+1} depends only on X_i:

$$P(S_{i+1} = s \mid S_i = s_t, S_{t-1} = s_{t-1}, \ldots, S_0 = s_0) = P(S_{t+1} = s \mid S_t = s_t)$$

For each time i, π_i is defined as the distribution of random variable X_i

1.0.1 Defining the State Space Ω

The manner in which the state space Ω is defined is crucial. A sequence of stochastic events is not considered a Markov process if more than just the current state is needed to determine the next state. For example, consider a process where we model the weather as Ω = sunny, cloudy, rainy. Suppose that today's weather depends not only on the current weather but also on the weather from the previous day. For instance:

- If it was rainy yesterday and today is cloudy, there might be a higher chance of rain tomorrow.
- However, if it was sunny yesterday and today is cloudy, there might be a higher chance of sun tomorrow.

In this case, the transition to the next state depends on both the current state and the prior state. This ”two-step” dependency breaks the memoryless property. To convert this into a Markov chain, we can expand the state space to include *pairs* of consecutive days instead of just the current day's weather. By doing so, each ”state” would represent two consecutive days of weather. This redefined state space allows the system to now satisfy the Markov property, as the probability of transitioning to the next ”state” (pair of weather conditions) depends only on the current two-day state, making it a valid Markov chain.

2 Transition Matrix

The transition matrix of a Markov chain, $\mathbb{P}$, is a $|\Omega| \times |\Omega|$ matrix, where each entry $\mathbb{P}_{ij}$ represents the probability of transitioning from a specific current state i to another state j.

$$\mathbb{P}_{ij} = P(X_{t+1} = j \mid X_t = i)$$

$\mathbb{P}$ is inherent to the stochastic process defined by the Markov chain at any point in time. Given the distribution of X_t, π_t, we can imagine taking one step forward in time to get the distribution at time t+1 by computing:

$$\pi_{t+1} = \pi_t * \mathbb{P}$$

Or, given the distribution at time t = 0, π_0, repeated multiplication by $\mathbb{P}$ will yield the distribution at time t:

$$\pi_t = \pi_0 * \mathbb{P}^t$$

2.0.1 Simulating a Markov Process

We can, naively, simulate a Markov process by beginning with an initial distribution π_0 and repeatedly multiplying by the transition matrix $\mathbb{P}$. However, each matrix multiplication is an $O(n^2)$ process, where n is the size of the state space, making sampling from the distribution π_t expensive. However, if we are given some way to sample from π_t, it is possible to sample from π_{t+1} without calculating it:

- First, we sample from π_t to get some X_i.
- Now, sample state j through a weighted random draw using row $\mathbb{P}_i$ of the transition matrix. This will be X_{t+1}.
- This j is a sample from π_{t+1} without having first calculated π_{t+1}

Now simulating the Markov chain transition is only an $O(n)$ process.

3 Eigenvalues and Eigenvectors in Markov Chains

Ideally, repeated multiplications by $\mathbb{P}$ will converge to some stable distribution π such that π does not have to be recomputed at every time step. Understanding the long-term behavior of Markov chains, such as the convergence and stability, requires a closer look at the properties of the transition matrix.

- An eigenvalue of a matrix M is a λ such that Mx = λx. That is, multiplying the vector x by m will scale it by the constant λ.
- An eigenvector of a matrix M is the corresponding vector x that is scaled constantly through multiplications by M.
- An n*n matrix M has n eigenvalues and eigenvectors, though not necessarily distinct.

3.1 Transition Matrix and Eigenvalues

To reiterate, for a Markov chain, the transition matrix P represents the probabilities of moving from one state to another in one step. P is a square matrix where:

- Each entry P_{ij} represents the probability of transitioning from state i to state j.
- Each row sums to 1: $\sum_j P_{ij} = 1$, meaning P is a **stochastic matrix**.

An important property of a transition matrix P for a finite, irreducible, and aperiodic Markov chain is that it has an eigenvalue of $\lambda = 1$, which we call the **dominant eigenvalue**. This eigenvalue plays a central role in determining the long-term behavior of the chain.

3.2 The Dominant Eigenvalue and Stationary Distribution

The largest eigenvalue of the transition matrix P is $\lambda = 1$, and it has a corresponding eigenvector π that represents the **stationary distribution**. The stationary distribution satisfies:

$$\pi P = \pi$$

This means that once the Markov chain reaches the stationary distribution, applying the transition matrix P leaves the distribution unchanged. Thus, π is the eigenvector associated with the eigenvalue $\lambda = 1$ and represents the steady-state or equilibrium behavior of the chain.

If we normalize π such that its elements sum to 1, each element π_i represents the long-term proportion of time that the Markov chain spends in state i.

3.3 Convergence to the Stationary Distribution

For an irreducible and aperiodic Markov chain, the transition matrix P has additional eigenvalues $\lambda_2, \lambda_3, \ldots, \lambda_n$ that satisfy:

$$|\lambda_i| < 1 \quad \text{for } i = 2, 3, \ldots, n$$

As the Markov chain evolves over time (with matrix powers P^t), the terms associated with these smaller eigenvalues decay exponentially due to $|\lambda_i| < 1$. Only the eigenvalue $\lambda = 1$ (associated with the stationary distribution) persists in the limit as $t \to \infty$, allowing the chain to converge to the stationary distribution π.

This convergence property can be expressed mathematically as:

$$\lim_{t \to \infty} P^t = \begin{bmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{bmatrix}$$

where each row converges to the stationary distribution π.

3.3.1 Convergence of Matrix Powers

Consider an $n \times n$ matrix P with an orthonormal eigenbasis $\{e_1, e_2, \ldots, e_n\}$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. We can express the power P^t acting on an initial vector $\sum_{i=1}^n a_i e_i$ as:

$$\left[\sum_{i=1}^n a_i e_i\right] P^t = \lambda_1^t \left[a_1 e_1 + \sum_{i=2}^n \left(\frac{\lambda_i}{\lambda_1}\right)^t a_i e_i\right]$$

Since $\frac{\lambda_i}{\lambda_1} \leq 1$ for $i \geq 2$, the terms $\left(\frac{\lambda_i}{\lambda_1}\right)^t$ decay exponentially as $t \to \infty$. Thus, only the term involving λ_1 remains significant in the limit, leading to convergence toward the stationary distribution.

3.4 Interpretation of Eigenvalues and Eigenvectors in Terms of System Behavior

Each eigenvalue λ_i and its associated eigenvector v_i of the transition matrix P provides insight into the behavior of the Markov chain:

- The eigenvalue $\lambda = 1$ and its eigenvector π (stationary distribution) describe the equilibrium state.
- Eigenvalues with $|\lambda_i| < 1$ represent transient behaviors that decay over time. The rate of decay for each component depends on the magnitude of λ_i.
- If the Markov chain starts from an initial distribution that is not the stationary distribution, the contributions from these transient modes (associated with $\lambda_i < 1$) will diminish as $t \to \infty$.

4 Irreducibility and Aperiodicity of a Markov Chain

If a Markov chain is irreducible and aperiodic, this means that it has a unique stationary distribution. A unique stationary distribution in a Markov chain is a probability distribution over the states that remains constant over time once it is reached. This means that if the chain starts in this distribution, it will stay in the same distribution after each transition. It must satisfy the following condition: $\pi = \pi * P$ (where π is described as the eigenvector of P for $\lambda = 1$).

4.1 Irreducible

A Markov chain is said to be irreducible if every state in the chain is reachable from every other state within a finite number of steps. Through this, it is ensured that the chain has no isolated states, and the graph visualization would be able to connect all nodes.

$$\forall x, y \in \Omega \quad \exists t \mid P^t(x, y) > 0$$

4.2 Aperiodic

A Markov chain is said to be aperiodic if there is no cyclical pattern in its state transitions. Basically, this prevents the chain from being trapped in a predictable cycle. If the chain was oscillating between states in a predictable pattern, then it would prevent convergence to that stationary distribution. Another way to describe aperiodic is to say that the greatest common divisor of the lengths of all possible return paths to a state should be 1.

$$\forall x, y : \text{GCD}\left\{t \mid P^t(x, y) = 0\right\} = 1$$

When a Markov chain is both irreducible and aperiodic, it has a unique stationary distribution that the chain will converge to over time. To recap, a stationary distribution for our

Markov chain basically meant that no matter where you start, as time goes on, the probability of being in each state settles into a predictable pattern. This helps us understand the steady-state behavior of the Markov chain.

5 Summary Notes

In this lecture, we explored the fundamental properties and behaviors of Markov chains, with a focus on the concepts of irreducibility, aperiodicity, and stationary distribution. Here's a quick recap:

- **Markov Chains**: A Markov chain is a stochastic process with the memoryless property, meaning the probability of transitioning to the next state depends only on the current state, not on the history.
- **Transition Matrix**: The transition matrix P describes the probability of moving from one state to another in a single step. This matrix is central to understanding the chain's dynamics.
- **Stationary Distribution**: A stationary distribution π is a probability distribution over states that remains constant over time. It represents the long-term probabilities of the chain being in each state.
- **Irreducibility**: A Markov chain is irreducible if every state is reachable from every other state. This property ensures that no state is isolated and that the system can freely move between all states.
- **Aperiodicity**: A Markov chain is aperiodic if it does not follow a predictable, cyclic pattern in its transitions. This property prevents the chain from oscillating between states in a fixed cycle, allowing it to converge smoothly to the stationary distribution.
- **Convergence**: When a Markov chain is both irreducible and aperiodic, it is guaranteed to converge to a unique stationary distribution. This means that, over time, the chain's behavior stabilizes, and the long-term probabilities of being in each state become predictable and independent of the starting state.

In summary, irreducibility and aperiodicity are critical properties that ensure the Markov chain has a unique stationary distribution. This steady-state behavior is essential for analyzing complex systems where long-term patterns and stability are important, such as in modeling random processes, ranking algorithms, and other various applications.