COMP 480/580 — Probabilistic Algorithms and Data Structure Oct 19, 2023

Lecture 17

Lecturer: Anshumali Shrivastava *Scribe By: Benson Ning, Minh Nguyen*

1 Motivation

A Markov Chain can be considered a cornerstone of machine learning and artificial intelligence, with extremely wide-ranging applications in reinforcement learning, natural language processing, finance, weather forecasting, and speech recognition.

The phrase "The future is independent of the past given the present" also encapsulates the idea of Markov Chains: all past information has already been encapsulated in the current state, and based on the present, we can predict the future.

While this might sound somewhat extreme, it can significantly simplify the complexity of models. Therefore, Markov Chains find extensive applications in many time series models, such as Recurrent Neural Networks (RNNs), Hidden Markov Models (HMMs), and, of course, Markov Chain Monte Carlo (MCMC).

We will explore the Markov Chains, which will continue throughout subsequent lecture, underscored by a challenging problem. Our motivation emerges from the challenge of sampling from an expansive state space. Given such a state space, Ω, we wish to sample x from Ω such that the probability of getting any $x \in \Omega$ is proportional to the weight associated with x, $w(x)$. In other words,

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

We will suppose that Ω is so large that

$$\sum_{y \in \Omega} w(y)$$

is not easy to compute.

2 Markov Chains

2.1 Definitions:

- **Stochastic Process**

  In simple terms, a stochastic process is a process of predicting and dealing with certain phenomena using statistical models. For example, predicting stock prices involves using today's stock price movements to forecast the movements of stocks for tomorrow and the day after tomorrow. Weather forecasting involves using today's rainfall or lack thereof to predict whether it will rain tomorrow or the day after. These processes can all be quantitatively calculated using mathematical formulas. By calculating the probabilities of events like rain or stock price fluctuations, formulas can be used to deduce the conditions for N days into the future.

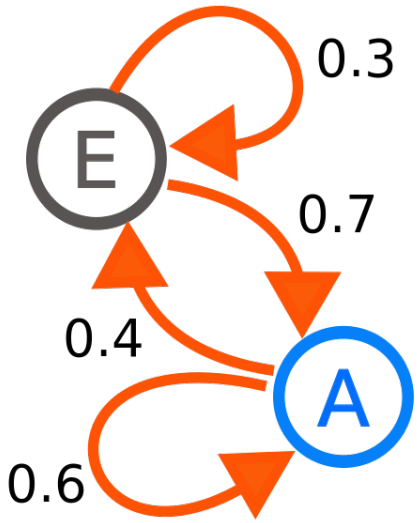

Figure 1: Markov Process

- **Markov Chain**

  Markov chain is a sequence of random variables, which take value in the in the states $X_1, X_2, ...X_t, X_{t+1}, ...$, the conditional probability of our state at time $t+1$, denoted as X_{t+1}, only depends on the state at time t. In other words,

  $$Pr(X_{t+1} = y|X_t = x, X_{t-1} = x_{t-1}, ..., X_0 = x_0) = Pr(X_{t+1} = t|X_t = x) = Pr(y|x)$$

  Since the probability of a state transition at a given moment depends solely on its preceding state, as long as we can calculate the transition probabilities between any two states in the system, the model of this Markov chain is determined. Thus, Markov chain is considered to be memory less as it only needs to consider the current state to predict the future behavior.

- **State Space**

  State space of a Markov Chain is the set of values that X_t can take, denoted as Ω. For example, $\Omega = x_1, x_2, ..., x_n$

2.2 Graph Representation

A Markov chain can represented as a directed graph $G = (V, E)$ where V is the set of vertices (or nodes) corresponding to the states in Ω, and E is the set of directed edges representing the transition probabilities between the states. For example, if there is a directed edge between node i and j, that mean there is a non-zero probability of transitioning from state i to j. In other words, this edge will have a weight of $P(j|i)$ or it can be written as $P(i, j)$. Moreover, due to the laws of probability, all the outgoing edges from any nodes must have all their weights added up to 1. The graphical representation provides a visual and intuitive way to understand the dynamics of the Markov Chain, especially the possible transitions between states and the probabilities of theses transitions. Let's represent a weather prediction of rainy vs sunny via a

weighted directed graph:

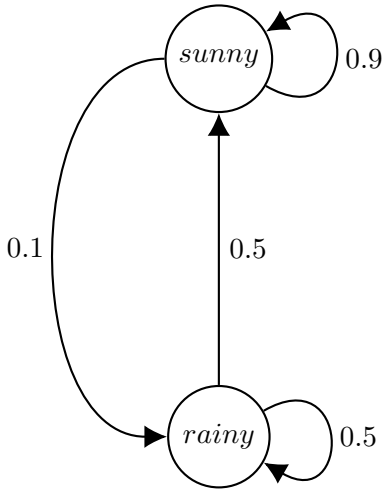

Figure 2: Rainy vs Sunny: A 2-state Markov chain

Figure 2 is a weather prediction of rainy vs sunny via a weighted directed graph. The state space of our example in Figure 2 can be written as $\Omega = sunny, rainy$. Then our conditional probability of $Pr(y|x)$ for each transition between the states can be: $Pr(sunny|sunny) = 0.9$, $Pr(rainy|sunny) = 1 - 0.9 = 0.1$, $Pr(rainy|rainy) = 0.5$, $Pr(sunny|sunny) = 1 - 0.5 = 0.5$.

This is a classic example of a simple stochastic process. In this scenario, the state of the weather on any given day (either sunny or rainy) is predicted based on the weather of the preceding day long, and not on the sequence of weather occurrences over multiple past days. It simplifies the modeling of weather dynamics by reducing the need for historical data, making predictions more efficient or memory-less.

2.3 Matrix Representation

- **Transition Matrix**

  Now consider the following matrix representation: $P = \begin{pmatrix} & A & B \\ A & p_1 & p_2 \\ B & p_3 & p_4 \end{pmatrix}$ We can easily calculate the probability of state after n times of shift of state by doing matrix multiplication. (ALso note that the summation of p should be 1)

  The probability of a state at t would be $\pi_t = \pi_0 \cdot P^t$

- **Properties**

  Assume we have the initial probability distribution $[p_0, p_1, p_2]$, and we use it in an invariant 3 by 3 transition matrix. We will realize that no matter how p_0, P_1, and P_2 are distributed, the transition matrix will converge to a stable state.

- **Sampling**

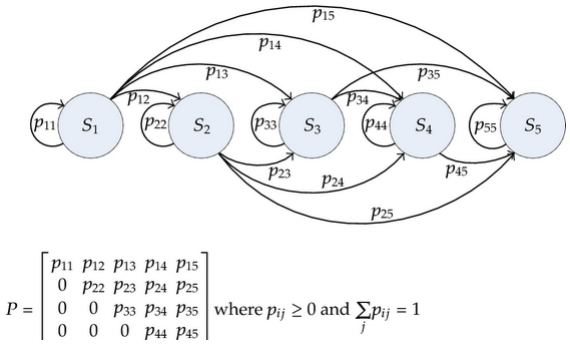

Figure 3: Transition Matrix

Given the initial distribution π_0, then the following distribution would be $\pi_1 ... \pi_i$, which converge to $\pi = \pi_n$ for large enough n. Also, for each π_i, we have $\pi_i = \pi_{i-1} P_1 = \pi_{i-2} P_2 = ... = \pi_0 P_i$

Start from π_0, sample x_0. Then, sample based on $P(x|x_0), P(x|x_1), ...$