Markov Chains and Hidden Markov Models

Karthik Lakshmanan

Markov chains are basically a set of states that are connected to each other by non zero probability arrows like in a weighted directed graph. Except here, the probability distribution of the next state depends purely upon the current state, and nothing before it, i.e.

$$P(x_n|x_{n-1}, x_{n-2}, ...x_1) = P(x_n|x_{n-1})$$

This is called the Markov property. We can represent Markov chains using matrices and can find stationary states using Linear algebra, i.e. taking the given Markov chain as an adjacency matrix A ($A_{i,j}$ represents probability of going from state i to state j) and finding the required eigenvectors.

Now, if we have states that we might never reach again in a certain random walk, we call those states **transient**. If there are states where we will inevitably reach again in a certain random walk, we call them **recurrent**.

A markov chain with no transient states is called an **irreducible** Markov chain. If it only has transient states, then we can divide that Markov chain into separate irreducible Markov chains, which leads to them being called **reducible**.

Example, check Gambler's ruin. We can call those irreducible chains as communicaion classes too.

If we have a Markov chain and we need to find the probability of reaching a state j from a state i after a certain number, say n steps. To do this we need to consider all the possible paths to reach that state in those many steps and add their probabilities (Chapman Kolmogorov Theorem). This is a little tedious, but if we notice, with an adjacency matrix A, this problem can be reduced to finding the i, jth element of the A^n.

$$P_{ij}(n) = A^n_{i,j}$$

Now, if we take $\lim_{n->\infty} A^{\infty}$, each element will denote the same value as before, but after infinite steps. This is like our stationary distribution (The eigenvector thing from before) and we will see that every row vector converges to the same row vector, which means that the final probabilities of each state are independent of the starting state. This only happens when certain conditions like irreducibility and aperiodicity are satisfied.

A hidden Markov Model (HMM) is the combination of a Hidden Markov Chain (A Markov Chain with states that we cannot observe) and a set of observed variables which depend on the states attained in our Hidden Markov Chain. The matrix that contains the probabilities of observed variables is called the **emission matrix**. This can be represented by joint and conditional distributions. So, if we represent our Hidden states as X and our observed variables as Y, then we can get an estimate of the order in which transitions took place in the hidden states by

$$E(X) = argmax_{X=x_1,x_2...,x_n} P(X = x_1, x_2..., x_n|Y = y_1, y_2..., y_n)$$

As we cannot find this directly, we can use Bayes' theorem.

$$E(x) = argmax_{X=x_1,x_2...,x_n} \frac{P(Y|X)P(X)}{P(Y)}$$

We can neglect the denominator as it is independent of X. The $P(Y|X)$ term can be simplified into the product of multiple $P(y_i|x_i)$ terms as the observed variable at that time only depends upon the current state and no other state. Same for $P(X)$, which can be a product of $P(x_i|x_{i-1})$ type terms because of the Markov property. For x_0, we can use the stationary distribution. This leads us to

$$E(x) = argmax_{X=x_1,x_2...,x_n} \prod_i P(y_i|x_i) \cdot P(x_i|x_i - 1)$$

The above mentioned ideas are for an order 1 HMM. This can be extended to an order n HMM.