Hidden Markov Models

Diljeet Jagpal

-

Contents

Preface

Preface

Chapter 1

Introduction

The goal of this project is to determine if HMMs are suitable as rain generators.

The first task will be to extend on the work done by Grando. In her testing, she has come to the conclusion that HMMs are suitable as rain generators however she has used the same data for testing as she used for training. This, quite likely, has lead to bias and thus we will extend her work by conducting out-of-sample tests.

If possible, I will build the software so it is user friendly and efficient. With this, I can test data for multiple locations. This will allow me to understand if the result is truly signficant, at least more so than just one location.

Chapter 2

Preliminaries

In this section, we will briefly visit foundations on which we will build throughout this paper. For most, this will be a simple refresher.

2.1 Mathematical Foundations

We start with a few key mathematical concepts.

2.1.1 Probability Theory

To discuss any probablisitic ideas we must first understand general probability theory. This can be done through the definition of a probability space.

Definition 2.1. Probability Space
A probability space is defined by $(\Omega, \mathcal{F}, \mathbb{P})$. Ω is the non-empty set of all possible outcomes, such that all events $\omega \in \Omega$. $\mathbb{P}$ is a probability measure, a function $\mathbb{P}$(A) that maps event A to a number within [0,1] based on the liklihood of the event. $\mathcal{F}$ is a σ-algebra on Ω if

1. $\Omega \in \mathcal{F}$
2. A $\in \mathcal{F}$ implies $A^c \in \mathcal{F}$
3. if A_1, A_2, A_3,... are in $\mathcal{F}$ then so is $A_1 \cup A_2 \cup A_3$...

2.1.2 Conditional Probability

Sometimes we require the probability of an event assuming another event has occured. In such situations we require conditional probability. Given two events A and B, the probability of event A occuring conditioned on the occurance of event B can be calculated as below.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},\ \forall A \in \mathcal{F} \tag{2.1}$$

From 2.1 and the fact that for dependednt events $\mathbb{P}$(A $\cap$ B) = $\mathbb{P}$(B $\cap$ A) we can see that:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A),\ \forall A, B \in \mathcal{F} \tag{2.2}$$

Subsituting 2.2 into 2.1 we get the famous Bayes Theorem.

Theorem 2.2. *Bayes' Theorem*
For dependent events A and B with probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *, where* $\mathbb{P}$ *(B)* $\neq$ *0,*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)},\ \forall A \in \mathcal{F} \tag{2.3}$$

2.1.3 Stochastic Process

To be able to define a Markov model, of any kind, we must first define a stochastic process.

Definition 2.3. Stochastic Process
Given an ordered set T and probability space $(\Omega, \mathcal{F}, \mathbb{P})$ a stochastic process is a collection of random variables X = $\{X_t;\ t \in T\}$. Based on t $\in$ T and $\omega \in \Omega$ we get a numerical realization of the process. For simplicity, this may be viewed as a function; $X_t(\omega)$.

2.2 Applied Foundations

Chapter 3

Standard Markov and Markov Property

3.1 History

Andrei Markov discovered the Markov model while analyzing the relationship between consecutive letters from text in the Russian novel "Eugene Onegin". With a two state model (states Vowel and Consonant) he proved that the probabililty of letters being in a particular state are not independent. Given the current state he could probabilistically predict the next. This chain of states, with various probabilities to and from each state, formed the foundation of the Markov Chain.

3.2 Markov Chain

A Markov chain is a network of connceted states. At any given time the model is said to be in a particular state. At a regular discrete interval the model has the ability to change states, which state will depend on the probaility and randomness. To define a Markov chain we must first address the Markov property. This simply states that the next state depends only on the current state. A more formal definition is given below.

Definition 3.1. Markov Property
Let $\{X_t$; t $\in \mathbb{N}_0\}$ denote a stochastic proccess 2.3, where t represents discrete time. The process has the Markov property if and only if,

$$\mathbb{P}\{X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, ..., X_0 = i_0\} = \mathbb{P}\{X_{n+1} = i_{n+1} | X_n = i_n\} \tag{3.1}$$

A Markov chain is simply a model that obeys 3.1. Again a more formal definition is given below.

Definition 3.2. Markov Chain
A stochastic process $\{X_t$; t $> 0\}$ is a Markov Chain if and only if it satisfies the Markov property 3.1.

To store the sequence of states a Markov chain has been through, we use the set $Q = \{q_t; t \in \mathbb{N}_0\}$, where q_t represents the state at time t. We will use this notation throughout the paper.

Example 3.3. Given a Markov Model with states S = $\{S_1, S_2, S_3\}$, if the model starts at S_2 and then goes to S_3 and then back to S_2 the state sequence Q will be $Q = \{q_1 = S_2, q_2 = S_3, q_3 = S_2\}$.

From the markov property we can see that the only thing that influences q_t is q_{t-1}. Thus we can make a prediction for q_{t+1} based on the outward transition probabililties from state q_t. If we calculate the probability of all possible states S_j given $q_t = S_i$ and find the maximum of these, we can find the most likely q_{t+1}.
i.e.
Given a Markov chain with N states including i and j and discrete time $t \in \mathbb{N}_0$:

$$\mathbb{P}(q_t = S_j | q_{t-1} = S_i)_{1 \leq i,j \leq N} \tag{3.2}$$

These probabilities can vary with time but can become quite complex. Thus, we usually assume the probabilities are constant. These special Markov models are called time-homogenous.

Definition 3.4. Time homogenous
Let $\{X_t \ ; \ \text{t} \in \mathbb{N}_0\}$ denote a stochastic proccess 2.3, where t represents discrete time, and $p(i, j)$ represent the transition probability from state i to state j.

$$\mathbb{P}\{X_n = j | X_{n-1} = i\} = p(i, j), \forall n \in \mathbb{N}_0 \tag{3.3}$$

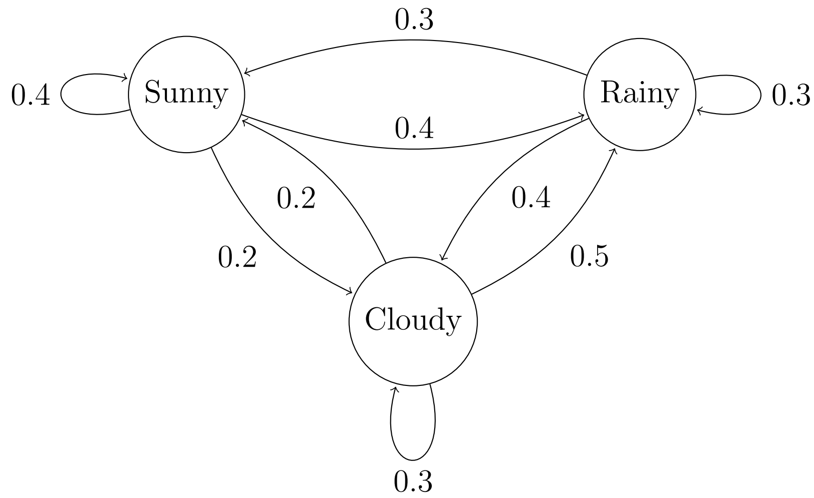
For a discrete Markov model with N states, there are N^2 possible transitions, where $p(i, j) = 0$ represents an impossible transition. We must store each of these probabilities. Given a time-homogenous Markov chain, we can create a 2-dimensional N x N matrix of transition probabilities p. Unique Markov chains have unique transition matrices. These matrices can be defined as below:

$$p = \{p(i, j) = \mathbb{P}\{X_n = j | X_{n-1} = i\}\}_{1 \leq i,j \leq N} \tag{3.4}$$

All p matrices have some special characteristics. The first, is that all values contained within p must be within $[0, 1]$. This is quite natural as all values are probabililties and thus by definition must lie within $[0, 1]$. The second is that all either the rows, columns or both form stochastic vectors. If it is the rows then the matrix is defined as a right-stochastic matrix, if it is the columns then it is called a left-stochastic matrix.

To demonstrate we will present an example where the weather is represented by the states.

Example 3.5. Let $\{X_t; t \in \mathbb{N}_0\}$ denote a Markov Chain, with state space S = {rainy, sunny, cloudy}, where t represents the number of days from start. Since any state can transition into any other state, we can say this model is ergodic. This can also be seen through the figure below as each state is connected to all others.



In this Markov Chain diagram, as per usual, the arrows indicate the transition between states and the values next to these correspond to the probability of this transition.

Using 3.5 we can create a matrix containing all the transition probabilities. This is called the transition matrix of the model and is usually labeled p. To build this we first create a table with our states labeled for rows and columns, where the $p(i,j)$ is the cell corresponding to row i and column j.

x	Sunny	Rainy	Cloudy
Sunny	0.4	0.4	0.2
Rainy	0.3	0.3	0.4
Cloudy	0.2	0.5	0.3

This content of this table forms the matrix p.

$$p = \begin{bmatrix} 0.4 & 0.4 & 0.2 \\ 0.3 & 0.3 & 0.4 \\ 0.2 & 0.5 & 0.3 \end{bmatrix} \tag{3.5}$$

3.3 Motivating the Hidden Markov Model

When we cannot directly calcualte the probabililties for matrix p we use data to find an approprate estimate. However, this is not always possible.

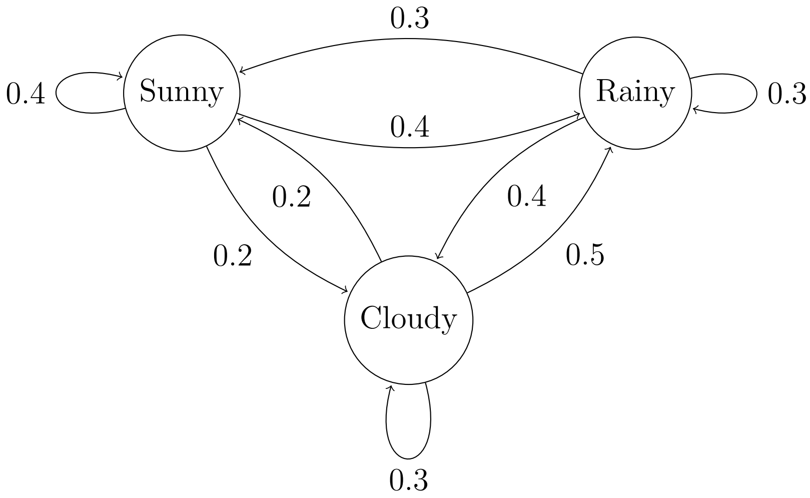
Example 3.6. Suppose Alice is hidden away from the world and has no access to information regarding the weather. She meets Bob everyday and knows how weather affects his mood. For simplicity, assume Bob only has two moods, happy and sad. Given that conditioned on the weather, his mood has the probabililties below ($\mathbb{P}(weather|mood)$) can she determine the weather?
Let (i,j) represent $p(j|i)$, for row i and column j)

x	Happy	Sad
Sunny	0.8	0.2
Rainy	0.7	0.3
Cloudy	0.6	0.4

We can let this be the Observation matrix O, since it is the condiational probabililties for the observations Alice can make.

$$O = \begin{bmatrix} 0.8 & 0.2 \\ 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} \tag{3.6}$$

We can illustrate this by adding these probabililties to 3.5.



Example 3.7.

Show how MM is a type of HMM.

The Markov model we have been discussing so far is called an observable Markov Model as we can observe its events. This is not always the case.

Chapter 4

Hidden Markov Model

4.1 Derivation

Explain all 5 inputs for HMM using example from previous chapter

Definition 4.1. Define HMM

Provide notation that will be used for the rest of the paper.

4.2 Applications of HMM

As with any mathematical model, we can use HMMs for either verification or prediction. Through the output liklihood we can find the liklihood of paticular sequences and through the given probablities we can find the most likely states in the future.

4.2.1 Predicitive Model

4.2.2 Sequence Prediction

4.2.3 Three Key Problems

4.3 Problem 1: Evaluation

4.3.1 Forward-Backward Algorithm

4.4 Problem 2: Decoding

4.4.1 Viterbi Algorithm

4.5 Problem 3: Learning

4.5.1 Expectation Maximization

4.5.2 Baum-Welch Algorithm

4.6 Modified HMM

4.6.1 GMM

Bibliography

Rabiner, L. and B. Juang (1986). "An introduction to hidden Markov models". In: *IEEE ASSP Magazine* 3.1, pp. 4–16. DOI: 10.1109/MASSP.1986.1165342.