Disclaimer

- The material provided in this document is not my original work and is a summary of some one else’s work(s).
- A simple Google search of the title of the document will direct you to the original source of the material.
- I do not guarantee the accuracy, completeness, timeliness, validity, non-omission, merchantability or fitness of the contents of this document for any particular purpose.
-

Theory and Implementation of Hidden Markov Models

Outline

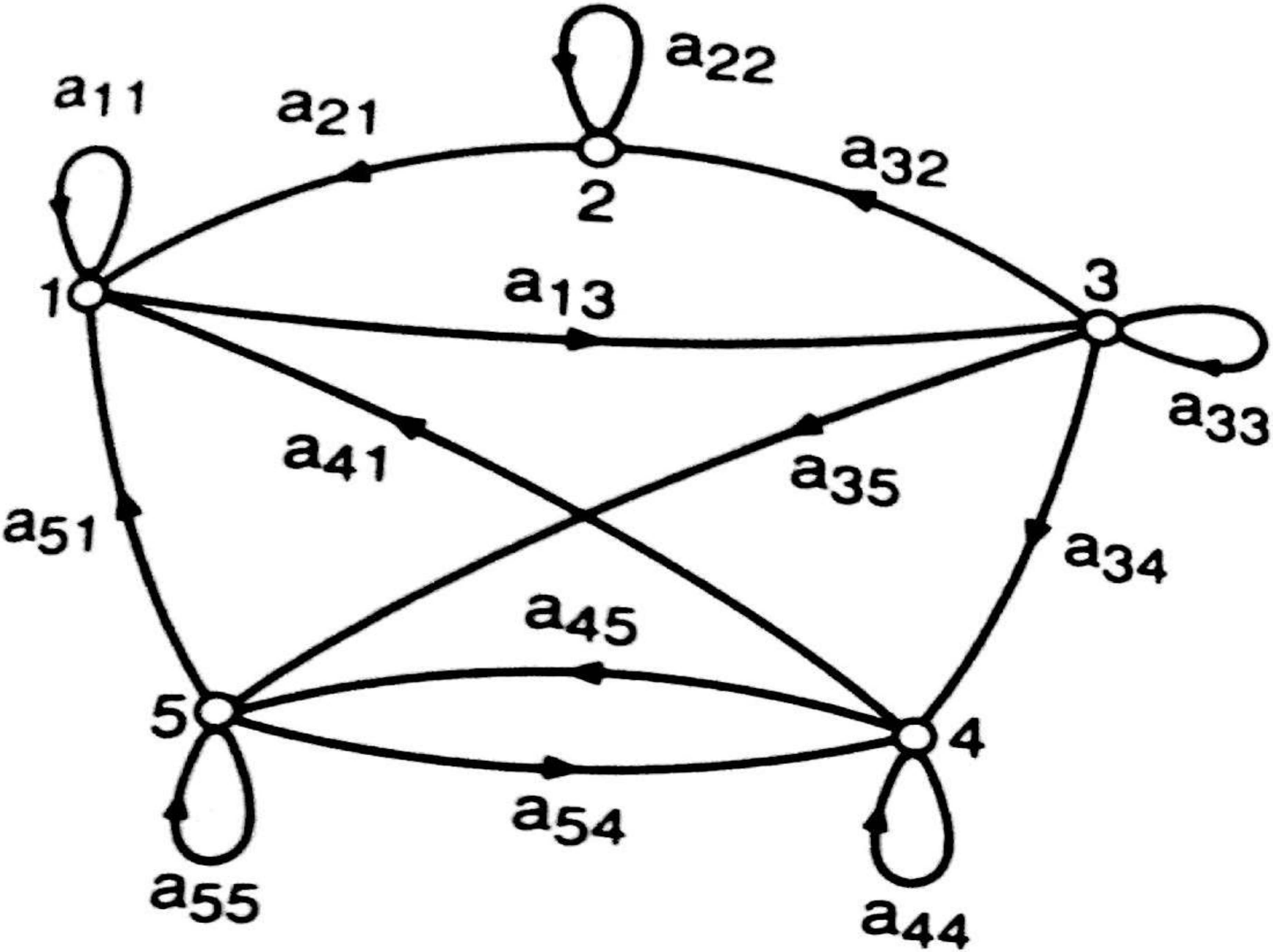
- Introduction
- Discrete Time Markov Processes
- Extensions to Hidden Markov Models
- The Three Basic Problems of HMMs
- Types of HMMs
- Continuous Observation Densities in HMMs
- Autoregressive HMMs
- Variants on HMM Structures
- Inclusion of Explicit State Duration Density in HMMs
- Optimization Criterion– ML, MMI, and MDI
- Comparisons of HMMs
- Implementation Issues for HMMs
- Improving the Effectiveness of Model Estimates
- Model Clustering and Splitting
- HMM System for Isolated Word Recognition

Introduction

- In this chapter we will discuss the widely used statistical method of characterizing the spectral properties of the frames of a pattern , namely the Hidden Markov Model approach
- The underlying assumption of the HMM is that the speech signal can be well characterized as a parametric random process, and that the parameters of the stochastic process can be determined in a precise, well defined manner

Discrete Time Markov Processes

- Consider a system which may be described at any time as being in one of a set of N distinct states, indexed by 1,2…,N
- At regularly spaced discrete times, the system undergoes a change of state (possibly back to the same state) according to a set of probabilities associated with the state
- We denote the time instants associated with state changes as $t = 1, 2, \ldots$, and we denote the actual state at time t as q_t
- A full probabilistic description of the above system would, in general, require specification of the current state (at time t), as well as all the predecessor states



a_{11}
a_{21}
a_{22}
a_{32}
a_{13}
a_{33}
a_{41}
a_{35}
a_{51}
a_{34}
a_{45}
a_{54}
a_{55}
a_{44}
1
2
3
4
5

Discrete Time Markov Processes

- The above stochastic process could be called an observable Markov model since the output

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

  - State 1: Rain
  - State 2: Cloudy
  - State 3: Sunny

Discrete Time Markov Processes

Problem

What is the probability (according to the model) that the weather for eight consecutive days is "sun-sun-sun-rain-rain-sun-cloudy-sun"?

Solution

We define the observation sequence, $\mathbf{O}$, as

$$\begin{array}{lccccccccc} \mathbf{O} & = & (\text{ sunny}, & \text{sunny}, & \text{sunny}, & \text{rain}, & \text{rain}, & \text{sunny}, & \text{cloudy}, & \text{sunny }) \\ & = & (\quad 3, & 3, & 3, & 1, & 1, & 3, & 2, & 3 \quad) \\ \text{day} & & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}$$

corresponding to the postulated set of weather conditions over the eight-day period and we want to calculate $P(\mathbf{O}|\text{Model})$, the probability of the observation sequence $\mathbf{O}$, given the model of Figure 6.2. We can directly determine $P(\mathbf{O}|\text{Model})$ as:

$$\begin{aligned} P(\mathbf{O}|\text{Model}) &= P[3, 3, 3, 1, 1, 3, 2, 3|\text{Model}] \\ &= P[3]\,P[3|3]^2 P[1|3]\,P[1|1] \\ &\quad P[3|1]\,P[2|3]\,P[3|2] \\ &= \pi_3 \cdot (a_{33})^2\, a_{31}\, a_{11}\, a_{13}\, a_{32}\, a_{23} \\ &= (1.0)(0.8)^2(0.1)(0.4)(0.3)(0.1)(0.2) \\ &= 1.536 \times 10^{-4} \end{aligned}$$

Discrete Time Markov Processes

Problem

Given that the system is in a known state, what is the probability that it stays in that state for exactly

Solution

This probability can be evaluated as the probability of the observation sequence

$$\begin{array}{lcccccccc} \mathbf{O} & = (& i, & i, & i & ,\ldots, & i, & j \neq i &) \\ \text{day} & & 1 & 2 & 3 & & d & d+1 & \end{array}$$

given the model, which is

$$\begin{aligned} P(\mathbf{O}|\text{Model}, q_1 = i) &= P(\mathbf{O}, q_1 = i|\text{Model})/P(q_1 = i) \\ &= \pi_i (a_{ii})^{d-1}(1 - a_{ii})/\pi_i \\ &= (a_{ii})^{d-1}(1 - a_{ii}) \\ &= p_i(d) \end{aligned}$$

Discrete Time Markov Processes

- The quantity $p_i(d)$ is the probability distribution function of duration d in state I
- Based on $p_i(d)$, we can readily calculate the expected number of observations (duration) in a state, conditioned on starting in that state as

$$\begin{aligned}\bar{d}_i &= \sum_{d=1}^{\infty} d p_i(d) \\ &= \sum_{d=1}^{\infty} d(a_{ii})^{d-1}(1-a_{ii}) = \frac{1}{1-a_{ii}}\end{aligned}$$

Discrete Time Markov Processes

Problem

Derive the expression for the mean of $p_i(d)$, i.e. Eq. (6.6b).

Solution

$$\begin{aligned}
\bar{d}_i &= \sum_{d=1}^{\infty} d p_i(d) \\
&= \sum_{d=1}^{\infty} d(a_{ii})^{d-1}(1-a_{ii}) \\
&= (1-a_{ii})\frac{\partial}{\partial a_{ii}}\left[\sum_{d=1}^{\infty} a_{ii}^d\right] \\
&= (1-a_{ii})\frac{\partial}{\partial a_{ii}}\left(\frac{a_{ii}}{1-a_{ii}}\right) \\
&= \frac{1}{1-a_{ii}}.
\end{aligned}$$

Extensions to Hidden Markov Models

- We have considered Markov models in which each state corresponded to an observable (physical) event
- We extend the concept of Markov models to include the case where the observation is a probabilistic function of the state
- The resulting model is a doubly embedded stochastic process with an underlying stochastic process that is not observable, but can only be observed through another set of stochastic processes that produce the sequence of observations

Basic Ideas in Probability

Exercise 6.1

Given a single fair coin, i.e., $P(\text{Heads}) = P(\text{Tails}) = 0.5$, which you toss once and observe Tails,

1. What is the probability that the next 10 tosses will provide the sequence $(HHTHTTHTTH)$?
2. What is the probability that the next 10 tosses will produce the sequence $(HHHHHHHHHH)$?
3. What is the probability that 5 of the next 10 tosses will be tails? What is the expected number of tails over the next 10 tosses?

Solution 6.1

1. For a fair coin, with independent coin tosses, the probability of any specific observation sequence of length 10 (10 tosses) is $(1/2)^{10}$ since there are 2^{10} such sequences and all are equally probable. Thus:

$$P(HHTHTTHTTH) = \left(\frac{1}{2}\right)^{10}.$$

2.

$$P(HHHHHHHHHH) = \left(\frac{1}{2}\right)^{10}.$$

Thus a specified run of length 10 is as likely as a specified run of interlaced H and T.

3. The probability of 5 tails in the next 10 tosses is just the number of observation sequences with 5 tails and 5 heads (in any order) and this is

$$P(5H, 5T) = \binom{10}{5}\left(\frac{1}{2}\right)^{10} = \frac{252}{1024} \cong 0.25$$

since there are $\binom{10}{5}$ ways of getting $5H$ and $5T$ in 10 tosses, and each sequence has probability of $\left(\frac{1}{2}\right)^{10}$. The expected number of tails in 10 tosses is

$$E(T \text{ in } 10 \text{ tosses}) = \sum_{d=0}^{10} d \binom{10}{d}\left(\frac{1}{2}\right)^{10} = 5.$$

Thus, on average, there will be $5H$ and $5T$ in 10 tosses, but the probability of exactly $5H$ and $5T$ is only 0.25.

Extensions to Hidden Markov Models

- We have considered Markov models in which each state corresponded to an observable (physical) event
- We extend the concept of Markov models to include the case where the observation is a probabilistic function of the state
- The resulting model is a doubly embedded stochastic process with an underlying stochastic process that is not observable, but can only be observed through another set of stochastic processes that produce the sequence of observations