Introduction to Hidden Markov Models (HMMs)

Objectives

- To understand the basics ideas behind Hidden Markov Models
- *Notes: pp 32 - 38*

Mathematical Modelling

- Mathematical modelling for speech recognition
- Two conflicting requirements:
  - Faithful model of human speech production/perception
  - Mathematically tractable & Computationally Useful
- HMMs are the best compromise at the moment

'Divide and Conquer'

- One possible approach to ASR is sequential 'divide and conquer', e.g.
  - classify speech vectors as 'acoustic features'
  - classify sequences of acoustic features as phonemes
  - classify sequences of phonemes as words
  - classify sequences of words ...

DISASTER!!

Delayed Decision Making

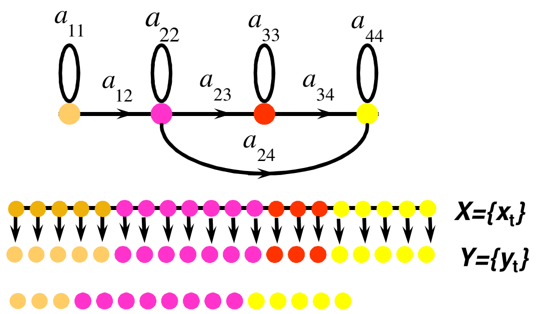
- Another name for this might be non-recoverable error propagation!
- Better to avoid all classification decisions until all sources of information are available. Then perform classification as a single, integrated process - delayed decision making
- Delayed Decision Making underlies HMM success

The 'HMM Compromise'

Assume that :

- A spoken utterance is a time-varying sequence which moves through a sequence of 'segments' – (yes)
- Underlying structure of the segments is constant w.r.t time – (no!)
- Durations of segments vary – (yes)
- All variations between different realizations of the segments are random – (no!)

Markov Model



Markov Model

Formally, a Markov model consists of:

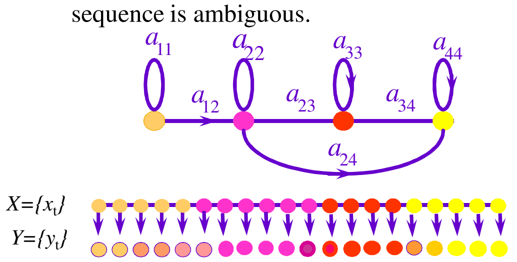
- A set of states $S = \{s_1,\ldots,s_N\}$
- A **state transition probability matrix**

  $A = [a_{ij}]_{i,j=1,\ldots N,}$

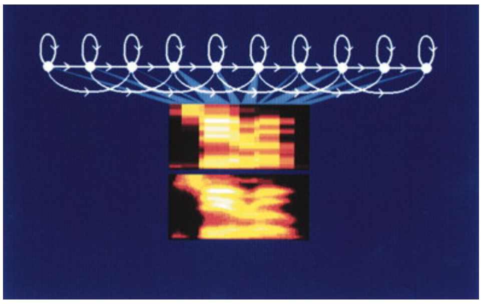
  where $a_{ij} = \text{Prob}(x_t = s_j \mid x_{t-1} = s_i)$
- The **Markov property** states that the state of the process at time $t+1$ depends on the state at time t (but is independent of the history of the process before time t)

Hidden Markov Model

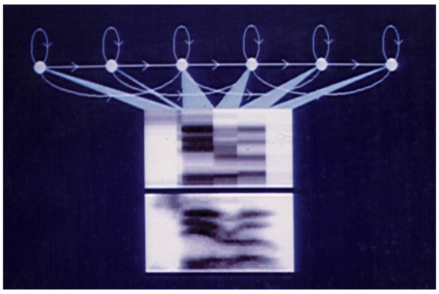
- In a **hidden** Markov model, the relationship between the observation sequence and the state sequence is ambiguous.



10 state HMM of the digit 'zero'



6 state HMM of the digit 'zero'



Multimodal Interaction Lab
b ɜː m ɪ ŋ ə m
EE4R Automatic Spoken Language Processing
UNIVERSITY OF BIRMINGHAM

Hidden Markov Models

A HMM consists of

- A set of states $S = \{s_1,...,s_N\}$
- A state transition probability matrix $A = [a_{ij}]_{i,j=1,...N}$,
  where a_{ij} =Prob(s_j at time t | s_i at time $t-1$)
- For each state s_i, a PDF b_i defined on the set of possible observations O s.t.
  $b_i(o)$ = Prob($y_t=o$ | $x_t=s_i$)
- b_i is called the **state output PDF** for state i (or the i^{th} **state output PDF**)

Multimodal Interaction Lab
b ɜː m ɪ ŋ ə m
EE4R Automatic Spoken Language Processing
UNIVERSITY OF BIRMINGHAM

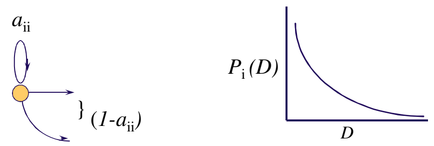
HMM Assumptions

- **Temporal Independence** - the observation y_t depends on the state x_t but is otherwise independent of the rest of the observation sequence $Y = \{y_t\}$!

  ... so, the position of the vocal tract at time *t* is independent of its position at time *t-1*!
- **Piecewise stationarity** - the underlying structure of speech is a sequence of stationary segments
- **Random variability** - variations from this underlying structure are random

HMM State Duration Model

- Constant segments correspond to the HMM states



- Probability of state duration D is given by

  $P_i\,(D) = (i - a_{ii})a_{ii}^{(D-1)}$

Summary

- Introduction to HMMs
- Markov Models
- Hidden Markov Models
- Formal definition of HMM
- HMM assumptions