1

Lecture 10

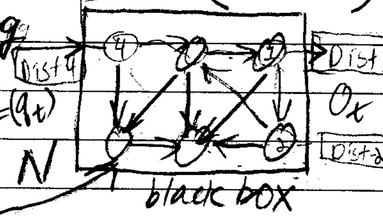
BEN KAY + JEFF LY (Notetaker + assistant)

Markov Processes; Hidden Markov Models

[See tutorial. Used in Speech recognition but also Biological sequence Analysis e.g. Durbin Et al.]

Note: A markov chain, but you don't get to see the state, but you do get to see a related output

There is an underlying discrete time finite State Markov Chain (hidden) ~~that~~ outputs Something at Each time step.



Structure: Finite state MC with N states $S_1, \ldots, S_N$

transition: The transition Matrix $a = (a_{ij})$ (like old P)

Initial dist: The Initial distribution is $\pi = (\pi_1, \ldots, \pi_N)$

Output: O_t is output of Markov Chain at time t. Values are also in finite set $\{V_1, \ldots V_m\}$. Output distribution is $b_j(k)$ = Prob. of outputting V_K when MC is in State S_j

Parameterize: A hidden Markov is characterized by a 3-tuple $\lambda = (a, b, \pi)$

Note: Matrix structure of $b_j(k)$

$$b = (b_j(k)) = \begin{array}{c c} & \xleftarrow{\quad k \quad} \\ j & \begin{bmatrix} \ddots & \\ & \ddots \end{bmatrix} \end{array}$$

4

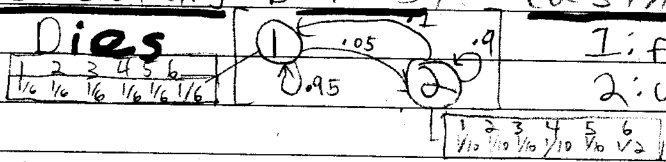
Markov Properties

1. q is Markov Chain with (a, π) as Parameters
2. $(q, O) = \{(q_t, O_t), t = 0, 1, 2, \ldots\}$ is a Markov Chain.

e.g. $P(q_{t+1}=S_j, O_{t+1}=V_k \mid q_t=S_i, O_t=V_l, q_{t-1}=S_m, O_{t-1}=V_n, \ldots, q_0, O_0)$
$= P(q_{t+1}=S_j, O_{t+1}=V_k \mid q_t=S_i) = a_{ij}\, b_j(k)$

Example (1) Occasionally Dishonest Casino

2 **Dies**



1: fair Die
2: Unfair Die

Typical Question:
Given an output sequence: e.g. 3 1 5 6 2 1 1 ...
What is the most likely path followed by the die **choice process?**

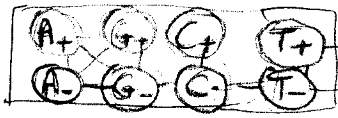
Example (2) Biological sequence analysis
DNA - A, C, G, T (single strand)
Pair - CG mutates quickly to TG in Normal Parts
In biologically important regions this mutation Process is suppressed (e.g. gene promoter region)

Locations where CG pairs are "protected" are called CpG Islands.

States of Hidden Markov Model HMM
$A+, C+, G+, T+$ - in island, $A-, C-, G-, T-$ Out of Island

Outputs are A, C, G, T (Deterministic function of state)

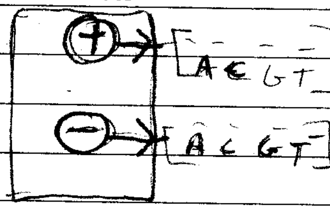
3) All connections Allowed



[T] ← Model
[T]

Simpler Model

✓ Date 5/7



But what about a much simpler model: But this is too simple because transition probabities vary based on previous letter state. **Letter sequence NOT I.I.D.** within islands.

Typical Question: Given output sequence A G C G T - - try to reconstruct the sequence of states for hidden chain.

3 Main Problems ~~Investigated with~~ For HMM

① Training or System Identification

Fix N, M, Parameter: $\lambda = (a, b, \pi)$

$a = (a_{ij})$, $a_{ij} \geq 0$, $\sum_j a_{ij} = 1$

$b = (b_j(k))$, $b_j(k) \geq 0$, $\sum_k b_j(k) = 1$

$\pi = [\pi_j]$, $\pi_j \geq 0$, $\sum \pi_j = 1$

compact state space

MLE method: Take output data $O = (O_0, O_1, \ldots O_K)$

Find λ to maximize score: $S^\lambda(O)$

$S^\lambda(O) = \log P^\lambda(O)$, choose λ to maximize likelyhood.

② Decoding: λ determined, perhaps by training

Want to identify a sequence that best generates a given output sequence

Seek $q^* = (q_0, \ldots q_K)$

$$q^* = \arg\max_q P^\lambda(q|O) = \text{values of } q \text{ that maximizes } P^\lambda(q|O)$$

$$\left\{ P^\lambda(q|O) = P^\lambda(q, O) / P^\lambda(O) \quad \text{By Bayes} \right\}$$

$$= \arg\max_q P^\lambda(q, O)$$

③ Classification or Recognition

In speech recognition, suppose we have a finite # of HMMs corresponding to words $W_1, \ldots W_L$.

For each word there is a HMM with associated parameter $\lambda_i, i=1,\ldots,L$.

output sequence: ~~Process~~ $O = (O_0, \ldots, O_K)$, want to determine

$$\lambda^* = \arg\max_{\lambda_i;\, i=1 \ldots L} P^{\lambda_i}(O).$$

(Calibrate finite # of λ, then hear a new sequence, figure out what word it is using the λ most likely to have generated it.

History:
1. Statisticians ~~discovered~~ initiated study of HMMs: Baum et al (1966 but with arcane title
2. Speech recognition (Dragon Dictate)
   Baker (CMU 1975)
   Jelinek (IBM)
3) DNA Sequencing
   Churchill et al 1989
   Borodousky et al (1980's)

5 Calculation overhead

Recall $O = (O_0, O_1 \cdots O_K)$

$\lambda = (a, b, \pi)$

Date 5/7

$$P^\lambda(O) = \sum_q P^\lambda(q, O)$$

where $q = (q_0, \cdots q_K)$

$$= \sum_q P^\lambda((q_0, O_0), (q_1, O_1), \cdots (q_K, O_K))$$

$$= \sum_q (\pi_{q_0} \cdot b_{q_0}(O_0)\, a_{q_0 q_1} b_{q_1}(O_1) \cdots a_{q_{K-1} q_K}\, b_{q_K}(O_K))$$

A lot of calculations: $2(K+1)+1$ terms in product

so, $2(K+1)$ multiplications per term

N^{K+1} terms in sum.

\# of multiplications $2(K+1)N^{K+1}$

$N^{K+1}-1$ additions

Of order KN^K opperations

Page