CS 466
Introduction to Bioinformatics
Lecture 20

Mohammed El-Kebir

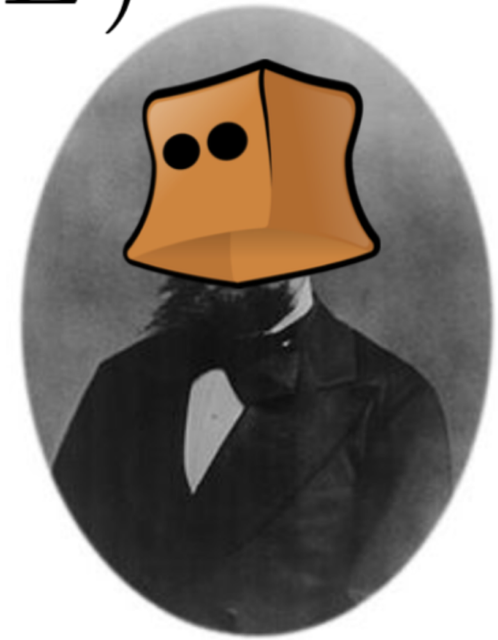
Nov 14, 2018

Course Announcements

Project Proposal due Nov 14th

Hidden Markov Model $\mathcal{M} = (Q, A, \Sigma, E)$

- Set of *hidden* states Q
  - Markov property
- Transition probabilities $A = [a_{ij}]$ on pairs of states
- Set of *emitted* symbols Σ
- Emission probabilities $E = [e_{ik}]$ on state-symbol pairs



Andrey Markov

<u>Two decisions:</u>

1. What symbol should I emit? [emission probabilities E]
2. What state should I move to next? [transition probabilities A]

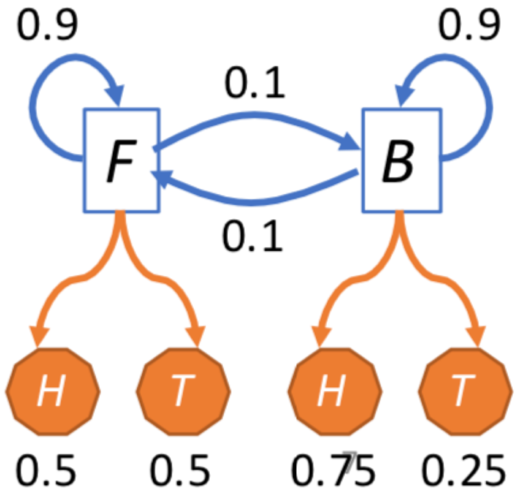
Fair Bet Casino

$$Q = \{F, B\}$$

$$A = \begin{array}{ccc} F & B & \\ 0.9 & 0.1 & F \\ 0.1 & 0.9 & B \end{array}$$

$$\Sigma = \{H, T\}$$

$$E = \begin{array}{ccc} H & T & \\ 0.5 & 0.5 & F \\ 0.75 & 0.25 & B \end{array}$$



Three Questions

Question 1:

What is the most probable path $\boldsymbol{\pi}^*$ that generated observations $\mathbf{x}$?

Question 2:

What is probability of observations $\mathbf{x}$ generated by any path $\boldsymbol{\pi}$?

Question 3:

What is the probability of observation x_i generated by state s?

Type	**Probability**	**Method**	**Solution**
Joint	$\max_{\boldsymbol{\pi}\in Q^n} \Pr(\mathbf{x}, \boldsymbol{\pi})$	Viterbi algorithm	$\Pr(\mathbf{x}, \boldsymbol{\pi}^*) = \max_{s\in Q} v[s,n]$
Marginal	$\Pr(\mathbf{x})$	Forward algorithm, or backward algorithm	$\Pr(\mathbf{x}) = \sum_{s\in Q} f[s,n]$, $\Pr(\mathbf{x}) = \sum_{s\in Q} a_{0,s} \cdot e[s, x_1] \cdot b[s,1]$
Posterior	$\Pr(\pi_i = s \mid \mathbf{x})$	Forward algorithm, and backward algorithm	$\hat{\pi}_i = \arg\max_{s\in Q} \frac{f[s,i]\cdot b[s,i]}{\sum_{t\in Q} f[t,n]}$

Table 1: Hidden Markov Models – Three different probabilities

Type	**Recurrence**
Viterbi	$v[s,i] = \begin{cases} a_{0,s} e_{s,x_1}, & \text{if } i = 1, \\ e_{s,x_i} \max_{t\in Q} v[t, i-1] \cdot a_{t,s}, & \text{if } i > 1. \end{cases}$
Forward	$f[s,i] = \begin{cases} a_{0,s} e_{s,x_1}, & \text{if } i = 1, \\ e_{s,x_i} \sum_{t\in Q} f[t, i-1] \cdot a_{t,s}, & \text{if } i > 1. \end{cases}$
Backward	$b[s,i] = \begin{cases} 1, & \text{if } i = n, \\ \sum_{t\in Q} a_{s,t} \cdot e_{t,x_{i+1}} \cdot b[t, i+1], & \text{if } 1 \leq i < n, \end{cases}$

Table 2: Hidden Markov Models – Three recurrences that each can be computed in $O(n|Q|^2)$ time and $O(n|Q|)$ space.

$\mathbf{x} =$

Viterbi	1 (T)	2 (H)	3 (T)	4 (H)	5 (H)	6 (H)	7 (T)	8 (H)	9 (H)	10 (H)	11 (T)
F											
B											

Forward	1 (T)	2 (H)	3 (T)	4 (H)	5 (H)	6 (H)	7 (T)	8 (H)	9 (H)	10 (H)	11 (T)
F											
B											

Backward	1 (T)	2 (H)	3 (T)	4 (H)	5 (H)	6 (H)	7 (T)	8 (H)	9 (H)	10 (H)	11 (T)
F											
B											

Questions:

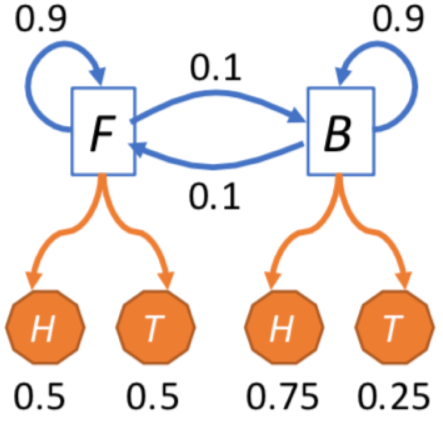
1. Compute most likely state path
2. Compute marginal probability
3. Compute posterior decoding

$Q = \{F, B\}$

$$A = \begin{array}{cc} \begin{array}{cc} F & B \end{array} & \\ \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} & \begin{array}{c} F \\ B \end{array} \end{array}$$

$\Sigma = \{H, T\}$

$$E = \begin{array}{cc} \begin{array}{cc} H & T \end{array} & \\ \begin{pmatrix} 0.5 & 0.5 \\ 0.75 & 0.25 \end{pmatrix} & \begin{array}{c} F \\ B \end{array} \end{array}$$



Type	Recurrence
Viterbi	$v[s,i] = \begin{cases} a_{0,s} e_{s,x_1}, & \text{if } i = 1, \\ e_{s,x_i} \max_{t \in Q} v[t, i-1] \cdot a_{t,s}, & \text{if } i > 1. \end{cases}$
Forward	$f[s,i] = \begin{cases} a_{0,s} e_{s,x_1}, & \text{if } i = 1, \\ e_{s,x_i} \sum_{t \in Q} f[t, i-1] \cdot a_{t,s}, & \text{if } i > 1. \end{cases}$
Backward	$b[s,i] = \begin{cases} 1, & \text{if } i = n, \\ \sum_{t \in Q} a_{s,t} \cdot e_{t,x_{i+1}} \cdot b[t, i+1], & \text{if } 1 \le i < n, \end{cases}$

Summary

- Markov property – Current state depends only on previous state
- Hidden Markov Models: states are not given only emitted symbols
- Viterbi algorithm: Find the most likely sequence of states given a set of observations
- Baum-Welch algorithm: EM-algorithm to learn ***A*** and ***E*** from training set

Reading:

- Jones and Pevzner: Chapters 11.1-11.3
- Lecture notes

Project Proposal due Nov 14th