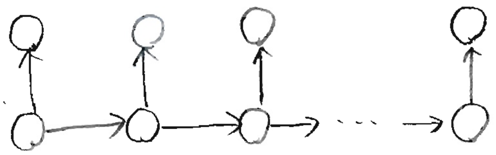
lecture 12

CSE 150 2, 22, 2010

HW 6 available on web site

$\boxed{\text{Review}}$

* Hidden Markov Models



observations $O_t \in \{1, 2, \ldots, m\}$

States $S_t \in \{1, 2, \ldots, n\}$

* Joint distribution

$$P(S, O) = P(S_1) \left[\prod_{t=2}^{T} P(S_t | S_{t-1}) \right] \left[\prod_{t=1}^{T} P(O_t | S_t) \right]$$

* Parameters

$$a_{ij} = P(S_{t+1} = j | S_t = i)$$

$$b_{ik} = P(O_t = k | S_t = i)$$

$$\pi_i = P(S_1 = i)$$

* key questions

1) How to compute $P(O_1, O_2, \ldots, O_T)$?

2) How to compute $S^* = \arg\max_S P(S|O)$?

3) How to learn $\{a_{ij}, b_{ik}, \pi_i\}$?

2) How to compute most likely state sequence?

$$S^* = \{ s_1^*, s_2^*, s_3^*, \ldots, s_T^* \}$$
$$= \underset{S}{\operatorname{argmax}}\ P(s_1, s_2, \ldots, s_T \mid o_1, o_2, \ldots, o_T)$$
$$= \underset{S}{\operatorname{argmax}} \left[\frac{P(s_1, s_2, \ldots, s_T, o_1, o_2, \ldots, o_T)}{P(o_1, o_2, \ldots, o_T)} \right]$$

Constant with respect to

$$= \underset{S}{\operatorname{argmax}}\ P(s_1, s_2, \ldots, s_T, o_1, o_2, \ldots, o_T)$$

* How to compute $S^* = \underset{S}{\operatorname{argmax}}\ P(s_1, s_2, \ldots, s_T, o_1, o_2, \ldots, o_T)$?

Define:

$$l_{it}^* = \max_{\underbrace{S}_{\{s_1, s_2, \ldots, s_{t-1}\}}} \log P(s_1, s_2, \ldots, s_{t-1}, s_t = i, o_1, o_2, \ldots, o_t)$$

= log-probability of most likely t-step state sequence that ends in state i at time t for observations $o_1, o_2, \ldots, o_t$

* Form recursion

(1) base case ($t=1$)

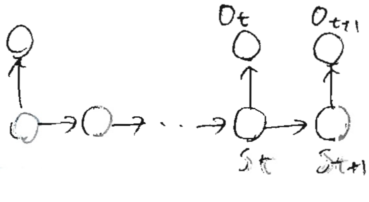
$$l_{i1}^* = \log P(s_1 = i, o_1)$$
$$= \log[P(s_1 = i)\, P(o_1 \mid s_1 = i)] \quad \text{product rule}$$
$$= \log[\pi_i\, b_i(o_1)]$$
$$= \log \pi_i + \log b_i(o_1)$$

(2) from time t to time $t+1$:

$$l_{j\,t+1}^* = \max_{s_1, s_2, \ldots, s_t} \log P(s_1, s_2, \ldots, s_t, s_{t+1} = j, o_1, o_2, \ldots, o_{t+1})$$

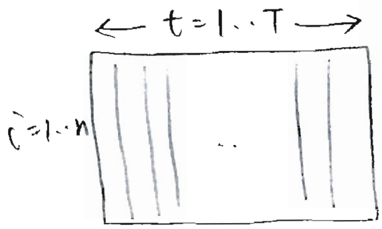
$$= \max_{s_1, s_2, \ldots, s_{t-1}} \max_{\underset{\text{representing } s_t = i}{i}} \log[P(s_1, s_2, \ldots, s_{t-1}, s_t = i, o_1, o_2, \ldots, o_t)$$

$$P(s_{t+1} = j \mid \cancel{s_1, s_2, \ldots, s_{t-1}}, s_t = i, \cancel{o_1, o_2, \ldots, o_t})$$
$$P(o_{t+1} \mid \cancel{s_1, \ldots, s_{t-1}, s_t = i}, s_{t+1} = j, \cancel{o_1, o_2, \ldots, o_t})$$



$$= \max_i \left[\max_{s_1 s_2 \ldots s_{t-1}} \log P(s_1, s_2, \ldots s_{t-1}, s_t = i, o_1, o_2, \ldots, o_t) + \log P(s_{t+1} = j \mid s_t = i) \right] + \log P(o_{t+1} \mid s_{t+1} = j)$$

$$\boxed{l^*_{j\,t+1} = \max_i \left[l^*_{it} + \log a_{ij} \right] + \log b_j(o_{t+1})}$$



filled in l^*_{it} column-by-column

* How to derive s^* from l^*?

Record most likely transitions:

$$\Phi_{t+1}(j) = \operatorname{argmax}_i \left[l^*_{it} + \log a_{ij} \right]$$

most likely state at time t given we are in state j at time $t+1$ with observations $o_1, o_2, \ldots, o_{t+1}$

* Compute s^* from back-tracking:

$$s^*_T = \operatorname{argmax}_i \left[l^*_{iT} \right]$$

$$s^*_t = \Phi_{t+1}(s^*_{t+1}) \quad \text{for } T-1, T-2, \ldots, 1$$ "Backward pass"

s^* is known as "Viterbi" state sequence or "Viterbi" path.

Viterbi algorithm is example of dynamic programming

3) Learning HMMs

Given: sequence of observations $\{o_1, o_2, \ldots, o_T\}$
(just one sequence for simplicity)

Goal: estimate (a_{ij}, b_{ik}, π_i) to maximize $P(o_1, o_2, \ldots, o_T)$

Assume: # hidden states, n, is known.

* EM algorithm in HMMs

CPTs to estimate:

$\pi_i = p(s_1 = i)$

$a_{ij} = p(s_{t+1} = j \mid s_t = i)$

$b_{ik} = p(o_t = k \mid s_t = i)$

* E-step:

$p(s_1 = i \mid o_1, o_2, \ldots, o_T)$

$p(s_t = i, s_{t+1} = j \mid o_1, o_2, \ldots, o_T)$

$p(s_t = i, o_t = k \mid o_1, o_2, \ldots, o_T) = p(s_t = i \mid o_1, o_2, \ldots, o_T)\, I(o_t, k)$ — indicator funct

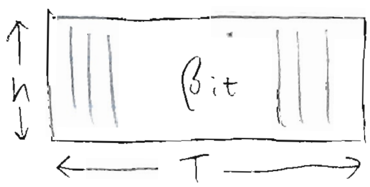
* How to compute posterior probabilities?

Analogous to $\alpha_{it} = p(o_1, o_2, \ldots, o_t, s_t = i)$,

define $\beta_{it} = p(o_{t+1}, \ldots, o_T \mid s_t = i)$ — conditioned on state $s_t = i$

* Recursion:

(1) base case: $\beta_{iT} = 1$ for all i



(2) backward step (from time t+1 to time t):

$$\beta_{it} = p(o_{t+1}, \ldots, o_T \mid s_t = i)$$

$$= \sum_j p(s_{t+1} = j, o_{t+1}, \ldots, o_T \mid s_t = i) \quad \text{marginalization}$$

$$= \sum_j p(s_{t+1} = j \mid s_t = i)\, p(o_{t+1}, \ldots, o_T \mid \cancel{s_t = i}, s_{t+1} = j) \quad \text{product rule}$$

Conditional independence

$$= \sum_j p(s_{t+1} = j \mid s_t = i)\, p(o_{t+1}, \ldots, o_T \mid s_{t+1} = j)$$

$$= \sum_j p(s_{t+1} = j \mid s_t = i)\, p(o_{t+1} \mid s_{t+1} = j)\, p(o_{t+2}, \ldots, o_T \mid s_{t+1} = j, \cancel{o_{t+1}})$$

$$\boxed{\beta_{it} = \sum_j a_{ij}\, b_j(o_{t+1})\, \beta_{jt+1}}$$

"Forward-backward" algorithm computes α_{it} and β_{it} in HMMs.

* Return to posterior probabilities in E-step:

$$P(S_{t+1}=j, S_t=i \mid O_1, O_2, \ldots, O_T) = \frac{P(S_{t+1}=j, S_t=i, O_1, O_2, \ldots, O_T)}{P(O_1, O_2, \ldots, O_T)}$$

Denominator: $P(O_1, O_2, \ldots, O_T) = \sum_{i=1}^{n} \alpha_{iT}$ (answer to question #1 in HMMs)

Numerator:

$$P(S_t=i, S_{t+1}=j, O_1, \ldots, O_T) = P(O_1, O_2, \ldots, O_t, S_t=i)\, P(S_{t+1}=j \mid S_t=i, \cancel{O_1, \ldots, O_t})$$
$$\cdot P(O_{t+1} \mid S_{t+1}=j, \cancel{S_t=i, O_1, \ldots, O_T})$$
$$\cdot P(O_{t+2}, \ldots, O_T \mid \cancel{S_t=i}, S_{t+1}=j, \cancel{O_1, \ldots, O_{t+1}})$$
$$= \alpha_{it}\, a_{ij}\, b_j(O_{t+1})\, \beta_{j\,t+1}$$

Posterior probabilities for E-step:

$$P(S_t=i, S_{t+1}=j \mid O_1, O_2, \ldots, O_T)$$
$$P(S_t=i \mid O_1, O_2, \ldots, O_T)$$
$$P(S_1=i \mid O_1, O_2, \ldots, O_T)$$

* last two are special cases of first:

$$P(S_t=i \mid O_1, O_2, \ldots, O_T) = \sum_j P(S_t=i, S_{t+1}=j \mid O_1, O_2, \ldots, O_T)$$

Set $t=1$ for last case.

Next time: M-step.