150 HW 6

Saturday, May 26, 2018 1:42 PM

#1

Jiajin Liang

5/29/2018.

6.1 Conditional independence

Consider the hidden Markov model (HMM) shown below, with hidden states S_t and observations O_t for times $t \in \{1, 2, \ldots, T\}$. Indicate whether the following statements are true or false.

F $\quad P(S_t|S_{t-1}) = P(S_t|S_{t-1}, O_t)$

T $\quad P(S_t|S_{t-1}) = P(S_t|S_{t-1}, O_{t-1})$

F $\quad P(S_t|S_{t-1}) = P(S_t|S_{t-1}, S_{t+1})$

F $\quad P(S_t|O_{t-1}) = P(S_t|O_1, O_2, \ldots, O_{t-1})$

T $\quad P(O_t|S_{t-1}) = P(O_t|S_{t-1}, O_{t-1})$

F $\quad P(O_t|O_{t-1}) = P(O_t|O_1, O_2, \ldots, O_{t-1})$

T $\quad P(O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(O_t|O_1, \ldots, O_{t-1})$ $\quad P(O_1)P(O_2|O_1)P(O_3|O_1,O_2)$

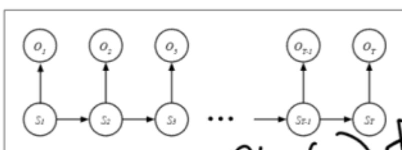
T $\quad P(S_2, S_3, \ldots, S_T|S_1) = \prod_{t=2}^{T} P(S_t|S_{t-1})$

T $\quad P(S_1, S_2, \ldots, S_{T-1}|S_T) = \prod_{t=1}^{T-1} P(S_t|S_{t+1}) = P(S_1|S_2)P(S_2|S_3) \cdots P(_{T-1}|S_T)$

F $\quad P(S_1, S_2, \ldots, S_T|O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(S_t|O_t)$

F $\quad P(S_1, S_2, \ldots, S_T, O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(S_t, O_t)$

T $\quad P(O_1, O_2, \ldots, O_T|S_1, S_2, \ldots, S_T) = \prod_{t=1}^{T} P(O_t|S_t)$



$P(O_1, S_1)$

$P(O_2, S_2) \neq P(O_2, S_2|S_1, O_1)$

$\cdots$

$P(O_T, S_T|S_1, O_1, \cdots O_{T-1}, S_{T-1})$

1

#2

6.2 More conditional independence

Indicate the **smallest** subset of evidence nodes that must be considered to compute each conditional probability shown below. The first two problems are done as examples. (You may assume everywhere that $2 < t < T-1$: i.e., do not worry about special boundary cases.)

$$P(S_t|S_1, S_2, \ldots, S_{t-1}) = P(S_t|S_{t-1})$$

$$P(O_t|S_1, S_2, \ldots, S_T) = P(O_t|S_t)$$

$$P(S_t|S_{t+1}, S_{t+2}, \ldots, S_T) = P(S_t|S_{t+1})$$

$$P(S_t|O_t, O_{t-1}, O_{t+1}) = P(S_t|O_t, O_{t-1}, O_{t+1})$$

$$P(S_t|O_t, O_{t-1}, O_{t+1}, S_{t-1}, S_{t+1}) = P(S_t|S_{t-1}, S_{t+1}, O_t)$$

$$P(S_t|S_1, S_T, O_1, O_t, O_T) = P(S_t|S_1, S_T\ O_t)$$

$$P(S_t|O_t, O_{t+1}, \ldots, O_T) = P(S_t|O_t, O_{t+1}, \ldots, O_T)$$

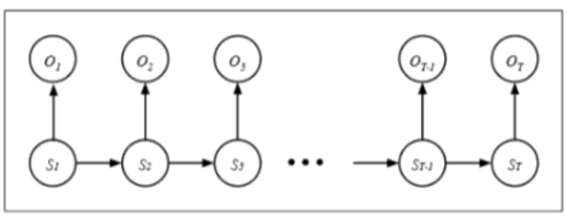
$$P(O_t|O_1, O_2, \ldots, O_{t-1}) = P(O_t|O_1, O_2, \ldots, O_{t-1})$$

$$P(O_t|O_1, O_2, \ldots, O_{t-1}, S_{t-1}) = P(O_t|S_{t-1})$$

$$P(O_t|O_1, O_2, \ldots, O_{t-1}, S_{t-2}) = P(O_t|S_{t-2}, O_{t-1})$$

$$P(O_t|S_{t-2}, S_{t-1}, S_{t+1}, S_{t+2}) = P(O_t|S_{t-1}, S_{t+1})$$

$$P(O_t|O_{t-1}, O_{t+1}, S_1, S_T) = P(O_t|O_{t-1}, O_{t+1}, S_1, S_T)$$



#3

#4

(a) $P(Y_1 | X_1) = \sum_{x_0} P(X_1, X_0 = x_0 | X_1)$ (margi.)

$= \sum_{x_0} P(X_0 = x_0 | X_1) P(Y_1 | X_0 = x_0, X_1)$ (P.R.)

$= \sum_{x_0} P(X_0 = x_0) P(Y_1 | X_0 = x_0, X_1)$ (C.I. by d-sep. #3)

(b) $P(Y_1) = \sum_{x_1} P(X_1 = x_1, Y_1)$ (margin.)

$= \sum_{x_1} \underbrace{P(X_1 = x_1)}_{\text{from CPTs}} \underbrace{P(Y_1 | X_1 = x_1)}_{\text{from (a)}}$ (P.R)

(c) $P(X_t | Y_1, Y_2, \ldots, Y_{t-1}) = P(X_t)$ since X_t is C.I. with $Y_1, Y_2, \ldots, Y_{t-1}$ when Y_t is not given by d-sep. #3 on node Y_t.

(d) $P(Y_t | X_t, Y_1, Y_2, \ldots, Y_{t-1})$

$= \sum_{x_{t-1}} P(Y_t, X_{t-1} = x_{t-1} | X_t, Y_1, Y_2, \ldots, Y_{t-1})$ (margin.)

$= \sum_{x_{t-1}} P(X_{t-1} = x_{t-1} | X_t, Y_1, Y_2, \ldots, Y_{t-1}) P(Y_t | X_{t-1} = x_{t-1}, X_t, Y_1, Y_2, \ldots, Y_{t-1})$ (P.R.)

$$= \sum_{x_{t-1}} \underbrace{P(X_{t-1}=x_{t-1} \mid Y_1, Y_2, \ldots, Y_{t-1})}_{\text{given by prev. step of recur.}} \underbrace{P(Y_t \mid X_{t-1}=x_{t-1}, X_t)}_{\text{CPT}}$$

(by d-sep. #2 & 3)

(e) $$P(Y_t \mid Y_1, Y_2, \ldots, Y_{t-1}) = \sum_{x_t} P(Y_t, X_t = x_t \mid Y_1, Y_2, \ldots, Y_{t-1})$$

(margin.)

$$= \sum_{x_t} \underbrace{P(X_t = x_t \mid Y_1, Y_2, \ldots, Y_{t-1})}_{\text{from (a)}} \underbrace{P(Y_t \mid X_t = x_t, Y_1, Y_2, \ldots, Y_{t-1})}_{\text{from (b)}}$$

(P.R.)

#5

6.5 Inference in HMMs

Consider a discrete HMM with hidden states S_t, observations O_t, transition matrix $a_{ij} = P(S_{t+1}=j|S_t=i)$ and emission matrix $b_{ik} = P(O_t=k|S_t=i)$. In class, we defined the forward-backward probabilities:

$$\begin{aligned}\alpha_{it} &= P(o_1, o_2, \ldots, o_t, S_t=i),\\ \beta_{it} &= P(o_{t+1}, o_{t+2}, \ldots, o_T|S_t=i),\end{aligned}$$

for a particular observation sequence $\{o_1, o_2, \ldots, o_T\}$ of length T. In terms of these probabilities, which you may assume to be given, as well as the transition and emission matrices of the HMM, show how to (efficiently) compute the following quantities:

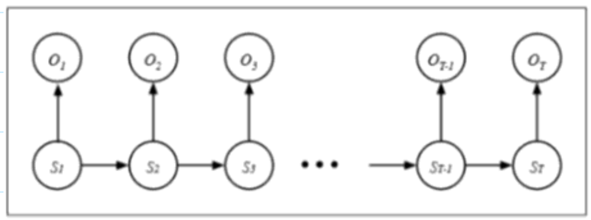
(a) $P(S_{t+1}=j|S_t=i, o_1, o_2, \ldots, o_T)$

(b) $P(S_t=i|S_{t+1}=j, o_1, o_2, \ldots, o_T)$

(c) $P(S_{t-1}=i, S_t=j, S_{t+1}=k|o_1, o_2, \ldots, o_T)$

(d) $\hat{s}_t = \mathrm{argmax}_i \left[P(S_t=i|o_{1,2}, \ldots, o_T)\right]$

In all these problems, you may assume that $t > 1$ and $t < T$; in particular, you are *not* asked to consider the boundary cases.



(a) $$P(S_{t+1} = j \mid S_t = i, o_1, o_2, \ldots, o_T)$$

$$= P(S_{t+1} = j \mid S_t = i, o_{t+1}, \ldots, o_T)$$

($o_1, \ldots, o_t$ are C.I. with S_{t+1} by d-sep #1)

$$= \frac{P(o_{t+1}, \ldots, o_T \mid S_t = i, S_{t+1} = j) P(S_{t+1} = j \mid S_t = i)}{P(o_{t+1}, \ldots, o_T \mid S_t = i)}$$

(Baye's Rule)

$$= \frac{P(o_{t+1} \mid S_t = i, S_{t+1} = j) P(o_{t+2}, \ldots, o_T \mid S_t = i, S_{t+1} = j) \cdot a_{ij}}{\beta_{it}}$$

(P.R.)
(substitution)

$$= P(o_{t+1} \mid S_{t+1} = j)\; P(o_{t+2}, \ldots, o_T \mid S_{t+1} = j) \cdot \frac{a_{ij}}{\beta_{it}}$$

(C.I.)

$$= \frac{b_{j(o_{t+1})} \cdot \beta_{j(t+1)} \cdot a_{ij}}{\beta_{it}}$$

(b) $P(S_t = i \mid S_{t+1} = j, o_1, o_2, \dots o_T)$

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

1st term in numerator is from (a), 2nd term is

$$P(S_t = i \mid O_1, O_2, \dots, O_T) = \frac{P(S_t = i, O_1, O_2, \dots, O_T)}{P(O_1, O_2, \dots, O_T)}$$

$$= \frac{P(S_t = i, O_1, O_2, \dots, O_t)\, P(O_{t+1}, O_{t+2}, \dots, O_T \mid S_t = i)}{\sum_{k=1}^{n} P(S_T = k, O_1, O_2, \dots, O_T)}$$

(P.R., C.I., margin.)

$$= \frac{\alpha_{it} \cdot \beta_{it}}{\sum_{k=1}^{n} \alpha_{kT}}$$

Similarly, denumerator $P(S_{t+1} = j \mid O_1, O_2, \dots, O_T)$

$$= \frac{\alpha_{j(t+1)} \cdot \beta_{j(t+1)}}{\sum_{k=1}^{n} \alpha_{kT}}$$

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

$b_{j(o_{t+1})} \cdot \cancel{\beta_{j(t+1)}} \cdot a_{ij} \quad \alpha_{it} \cdot \cancel{\beta_{it}} \qquad a_{ij} \cdot b_{j(o_{t+1})}\, \alpha_{it}$

$$= \frac{b_{j(o_{t+1})} \cdot \cancel{\beta_{j(t+1)}} \cdot a_{ij}}{\cancel{\beta_{it}}} \cdot \frac{\alpha_{it} \cdot \cancel{\beta_{it}}}{\alpha_{j(t+1)} \cdot \cancel{\beta_{j(t+1)}}} = \frac{a_{ij} \cdot b_{j(o_{t+1})} \alpha_{it}}{\alpha_{j(t+1)}}$$

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

$= P(S_t=j \mid O_1, \ldots, O_T) P(S_{t-1}=i \mid S_t=j, O_1, \ldots, O_T) P(S_{t+1}=k \mid$
$S_t=j, S_{t-1}=i, O_1, \ldots, O_T)$ (P.R., C.I.)

$$= \frac{\alpha_{jt} \cdot \beta_{jt}}{\sum_{k=1}^{n} \alpha_{kT}} \cdot \frac{a_{ij} \cdot b_{j(o_t)} \alpha_{i(t-1)}}{\alpha_{jt}} \cdot \frac{b_{k(o_{t+1})} \cdot \beta_{k(t+1)} \cdot a_{jk}}{\beta_{jt}}$$

$$= \frac{a_{ij} \cdot b_{j(o_t)} \cdot \alpha_{i(t-1)} \cdot b_{k(o_{t+1})} \cdot \beta_{k(t+1)} \cdot a_{jk}}{\sum_{k=1}^{n} \alpha_{kT}}$$

(d) $\hat{S}_t = \underset{i}{\operatorname{argmax}}\ P(S_t=i \mid O_1, O_2, \ldots, O_T)$

$$= \underset{i}{\operatorname{argmax}} \left(\frac{\alpha_{it} \cdot \beta_{it}}{\sum_{k=1}^{n} \alpha_{kT}} \right)$$ (from prev. step)

6.6

(a) No. $\{S_1^*, \ldots, S_T^*\}$ is the sequence that makes $P(S_1, S_2, \ldots, S_T \mid O_1, O_2, \ldots, O_T)$ largest.

so $P(\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_T \mid O_1, O_2, \ldots, O_T)$ can only be smaller or equal to $P(S_1^*, \ldots, S_T^* \mid O_1, \ldots, O_T)$

since $\hat{s}_1, \ldots, \hat{s}_T$ are obtained individually.

(b) Yes. As mentioned in (a), it's possible that $\hat{s}_t = s_t^*$ for all t.

(c) No. In most cases, they will not be equal.

(d) Yes. $P(\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_T) = \frac{1}{n^T}$ where n is the number of possible values s_t can take, so it always > 0.