Stat 200A Homework 3

November 27, 2007

1. Work out the "Asia" example. In particular, assuming we are given all the necessary probabilities and conditional probabilities, please calculate P(lung cancer | smoking, dyspnea, positive x-ray).

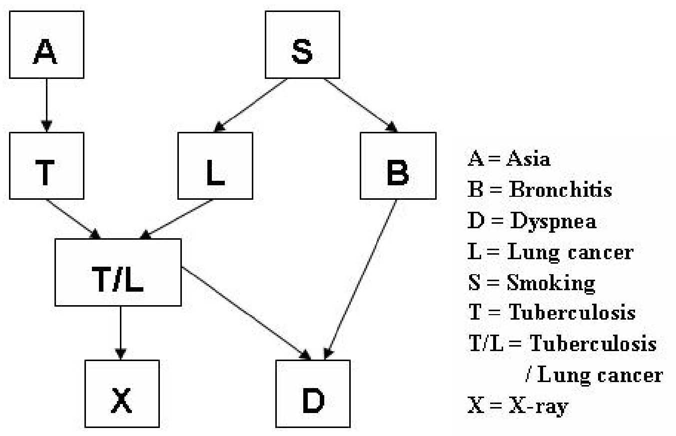

Figure 1: A schematic Bayesian network

Figure 1 shows a simple Bayesian network about diagnostics for lung related diseases, for which all the necessary marginal and conditional probabilities are known. We want to infer $P(L|S = s, X = x, D = d)$.

To begin with, we can decompose the joint distribution of all variables according to the network structure:

$$\begin{aligned} &P(A, S = s, T, L, B, T/L, X = x, D = d) \\ &= P(A)P(T|A)P(S = s)P(L|S = s)P(B|S = s) \\ &\times P(T/L|T, L)P(X = x|T/L)P(D = d|T/L, B). \end{aligned}$$

Then summing over all the unknown variables, we have

$$\begin{aligned} &P(L, S = s, X = x, D = d) \\ &= P(S = s)P(L|S = s) \left\{ \sum_{T/L} \left[\sum_{T} P(T/L|T, L) \left(\sum_{A} P(A)P(T|A) \right) \right] \right. \\ &\left. \times \left(\sum_{B} P(D = d|T/L, B)P(B|S = s) \right) P(X = x|T/L) \right\}. \end{aligned}$$

The actual calculation procedure is from the formulae within parentheses, to the formulae within brackets and then to the formulae within braces.

It follows that the probability of L given $S = s, X = x, D = d$ is

$$P(L|S = s, X = x, D = d) = \frac{P(L, S = s, X = x, D = d)}{\sum_L P(L, S = s, X = x, D = d)}.$$

2. Work out the chain example of belief propagation, assuming that we observe the head and tail of the chain. Check out Pearl's seminal paper on belief network.

$$U \to X \to Y \to Z \to W$$

Figure 2: A simple chain model for belief propagation

Figure 2 shows a simple chain model, for which all the necessary marginal and conditional probabilities are known. Given $U = u$ and $W = w$, we want to infer the conditional probabilities of X, Y, Z using belief propagation. More specifically,

$$P(X = x|U = u, W = w) = \alpha_1 \pi_1(x) \lambda_1(x),$$

$$P(Y = y|U = u, W = w) = \alpha_2 \pi_2(y) \lambda_2(y),$$

$$P(Z = z|U = u, W = w) = \alpha_3 \pi_3(z) \lambda_3(z);$$

where

$$\pi_1(x) = P(X = x|U = u),$$

$$\pi_2(y) = P(Y = y|U = u) = \sum_x P(X = x|U = u) P(Y = y|X = x) = \sum_x \pi_1(x) K_{x \to y},$$

$$\pi_3(z) = P(Z = z|U = u) = \sum_y P(Y = y|U = u) P(Z = z|Y = y) = \sum_x \pi_2(y) K_{y \to z};$$

$$\lambda_3(z) = P(W = w|Z = z),$$

$$\lambda_2(y) = P(W = w|Y = y) = \sum_z P(Z = z|Y = y) P(W = w|Z = z) = \sum_y K_{y \to z} \lambda_3(z),$$

$$\lambda_1(x) = P(W = w|X = x) = \sum_y P(Y = y|X = x) P(W = w|Y = y) = \sum_x K_{x \to y} \lambda_2(y);$$

$$\alpha_1 = [\sum_x \pi_1(x) \lambda_1(x)]^{-1}, \alpha_2 = [\sum_y \pi_2(y) \lambda_2(y)]^{-1}, \alpha_3 = [\sum_x \pi_3(z) \lambda_3(z)]^{-1}.$$

Here, all the π's are referred to as prior probability and all the λ's are referred to as likelihood.

3. Work out the forward and backward algorithms for hidden Markov model. Compare these algorithms with belief propagation by identifying π and λ messages. Check out Rabina's classical tutorial on HMM.

$$\begin{array}{ccccccccccc} X_0 & \to & X_1 & \to & X_2 & \to \cdots \to & X_t & \to \cdots \to & X_T \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ y_0 & & y_1 & & y_2 & \cdots & y_t & \cdots & y_T \end{array}$$

Figure 3: Scheme for hidden Markov model

Figure 3 shows a hidden Markov model, where $\{X_t, t = 0, \ldots, T\}$ are hidden variables and $\{y_t, t = 0, \ldots, T\}$ are corresponding observations. Assume that the hidden Markov chain has a finite state space S_X and all the observations take values on a finite set S_Y. Also assume that all transition probabilities $P(X_{t+1} = j|X_t = i) = a_{ij}$ $(i, j \in S_X, t = 0, \ldots, T-1)$ and $P(y_t|X_t = i) = b_{iy_t}$ $(i \in S_X, y_t \in S_Y, t = 0, \ldots, T)$ are known. We want to make an inference to the status of the hidden variables from the observations.

Forward Algorithm:

Let $F_t(i) = P(y_0, y_1, y_2, \ldots, y_t, X_t = i), i \in S_X, t = 0, \ldots, T$. Then we have the following induction:

$$\begin{aligned}
F_{t+1}(j) &= P(y_0, y_1, y_2, \ldots, y_t, y_{t+1}, X_{t+1} = j) \\
&= \sum_i P(y_0, y_1, y_2, \ldots, y_t, y_{t+1}, X_t = i, X_{t+1} = j) \\
&= \sum_i P(y_0, y_1, y_2, \ldots, y_t, X_t = i) P(X_{t+1} = j|X_t = i) P(y_{t+1}|X_{t+1} = j) \\
&= \sum_i F_t(i) a_{ij} b_{jy_{t+1}}
\end{aligned}$$

Using this equation, we can recursively obtain all $F_t(i)$ $(i \in S_X, t = 0, \ldots, T)$, given a prior of $P(X_0)$. Then for each t, we have

$$P(X_t = i|y_0, y_1, y_2, \ldots, y_t) = \frac{P(X_t = i, y_0, y_1, y_2, \ldots, y_t)}{P(y_0, y_1, y_2, \ldots, y_t)} = \frac{F_t(i)}{\sum_{j \in S_X} F_t(j)}.$$

Backward Algorithm:

Let $B_t(i) = P(y_{t+1}, \ldots, y_T|X_t = i), i \in S_X, t = 0, \ldots, T-1$. Then we have the following induction:

$$\begin{aligned}
B_t(i) &= P(y_{t+1}, \ldots, y_T|X_t = i) \\
&= \sum_j P(y_{t+1}, \ldots, y_T, X_{t+1} = j|X_t = i) \\
&= \sum_j P(X_{t+1} = j|X_t = i) P(y_{t+1}|X_{t+1} = j) P(y_{t+2}, \ldots, y_T|X_{t+1} = j) \\
&= \sum_j a_{ij} b_{jy_{t+1}} B_{t+1}(j)
\end{aligned}$$

Using this equation, we can recursively obtain all $B_t(i)$ $(i \in S_X, t = T-1, \ldots, 0)$ starting from a likelihood of $P(y_T|X_T)$. Then for each t, we can maximize the likelihood of $P(y_{t+1}, \ldots, y_T|X_t)$ to infer the status of each X_t.

Forward and backward belief propagation:

For any t, we can use all the observations to infer X_t by

$$P(X_t = i|y_0, y_1, \ldots, y_T) = \alpha_t \pi_t(i) \lambda_t(i),$$

where

$$\begin{aligned}
\pi_t(i) &= P(X_t = i|y_0, y_1, \ldots, y_{t-1}), \\
\lambda_t(i) &= P(y_t, y_{t+1}, \ldots, y_T|X_t = i), \\
\alpha_t &= [\sum_i \pi_t(i) \lambda_t(i)]^{-1}.
\end{aligned}$$

So it is just left to derive the induction equations of $\pi_t(i)$ and $\lambda_t(i)$, which needs to modify a little the forward and backward algorithms.

For prior message propagation, we have the following induction:

$$\begin{aligned}
\pi_{t+1}(j) &= P(X_{t+1} = j | y_0, y_1, y_2, \ldots, y_t) \\
&= \sum_i P(X_t = i, X_{t+1} = j | y_0, y_1, y_2, \ldots, y_t) \\
&= \frac{\sum_i P(X_t = i | y_0, y_1, y_2, \ldots, y_{t-1}) P(y_t | X_t = i) P(X_{t+1} = j | X_t = i)}{\sum_i P(X_t = i | y_0, y_1, y_2, \ldots, y_{t-1}) P(y_t | X_t = i)} \\
&= \frac{\sum_i \pi_t(i) b_{iy_t} a_{ij}}{\sum_i \pi_t(i) b_{iy_t}}.
\end{aligned}$$

For likelihood message propagation, we have the following induction:

$$\begin{aligned}
\lambda_t(i) &= P(y_t, \ldots, y_T | X_t = i) \\
&= \sum_j P(y_t, \ldots, y_T, X_{t+1} = j | X_t = i) \\
&= \sum_j P(X_{t+1} = j | X_t = i) P(y_t | X_t = i) P(y_{t+1}, \ldots, y_T | X_{t+1} = j) \\
&= \sum_j a_{ij} b_{iy_t} \lambda_{t+1}(j).
\end{aligned}$$

4. Prove the Markov property of the Ising model.

The Ising model assume that the joint distribution of $(X_1, \ldots, X_n)$ is

$$P(X_1, \ldots, X_n) = \frac{1}{Z(\beta, h)} e^{-\beta \sum_{i=1}^{n-1} x_i x_{i+1} - h \sum_{i=1}^{n} x_i},$$

where

$$Z(\beta, h) = \sum_{x_1, \ldots, x_n} e^{-\beta \sum_{i=1}^{n-1} x_i x_{i+1} - h \sum_{i=1}^{n} x_i}.$$

It can be shown that the Ising model has the Markov property. Denote

$$f(x_{[-i,-(i-1),-(i+1)]}) = e^{-\beta \sum_{j=1}^{i-2} x_j x_{j+1} - \beta \sum_{j=i+1}^{n-1} x_j x_{j+1} - h \sum_{j=1}^{i-1} x_j - h \sum_{j=i+1}^{n-1} x_j}.$$

Then we have

$$\begin{aligned}
P(X_i | X_{[-i]}) &= \frac{P(X_i, X_{[-i]})}{P(X_{[-i]})} \\
&= \frac{Z(\beta, h)^{-1} f(x_{[-i,-(i-1),-(i+1)]}) e^{-\beta(x_{i-1} x_i + x_i x_{i+1}) - h x_i}}{Z(\beta, h)^{-1} f(x_{[-i,-(i-1),-(i+1)]}) \sum_{x_i} e^{-\beta(x_{i-1} x_i + x_i x_{i+1}) - h x_i}} \\
&= \frac{e^{-\beta(x_{i-1} x_i + x_i x_{i+1}) - h x_i}}{\sum_{x_i} e^{-\beta(x_{i-1} x_i + x_i x_{i+1}) - h x_i}} \\
&= P(X_i | X_{i-1}, X_{i+1}),
\end{aligned}$$

which complete the proof.