ECE 566

Bayesian Inference II

Hierarchical Modeling

Suppose we did not have a chance to hear the alarm sound ourselves. Instead a friend (neighbor) called and said that they can hear our burglar alarm.

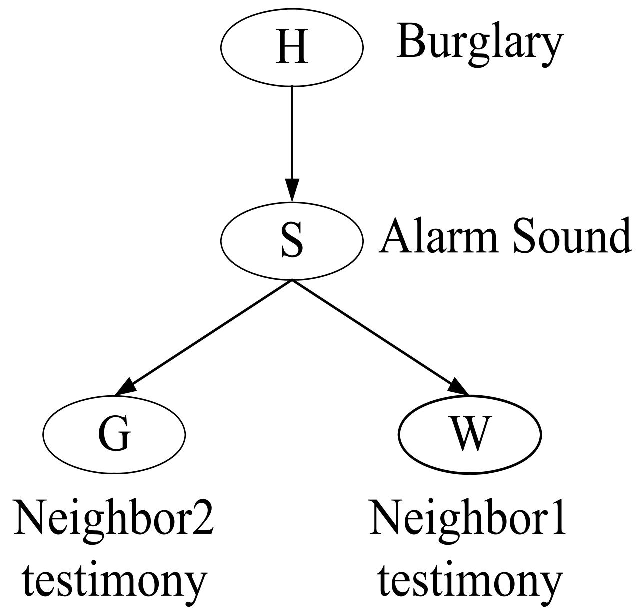
Suppose also that this friend (neighbor) sometimes plays practical jokes →

Information about alarm (evidence for Burglary) is uncertain.

We decide to call our other neighbor to see if he heard anything, but suppose he drinks →

his testimony is also uncertain.

Then we have the situation depicted on the right.



Now if we were to apply what we have learned so far, what we need to do is to compute

$$O(H \mid W) = L(W \mid H)\, O(H)$$

But we do not have L(W | H), we only have L(W | S) because the neighbor only claims that he heard the alarm. Estimating L(W | H) is even more difficult than estimating L(S | H).

We can do:

$$\begin{aligned} P(H_i \mid G, W) &= \alpha P(G, W \mid H_i) P(H_i) \\ &= \alpha P(H_i) \sum_j P(G, W \mid H_i, S_j) P(S_j \mid H_i) \end{aligned}$$

We have conditioned $(G, W \mid H_i)$ over all possible values of intermediate variable S

$S_j = S_1$ = alarm sound on

$S_j = S_2$ = alarm sound off

The testimonies (W, G) are independent of the burglary and are only dependent on the alarm sound. Using the conditional independence of G, W with respect to Hi we get:

The testimonies (W, G) are independent of the burglary and are only dependent on the alarm sound. Using the conditional independence of G, W with respect to Hi we get:

$$P(H_i \mid G, W) = \alpha P(H_i) \sum_j P(G, W \mid S_j) P(S_j \mid H_i)$$

And assuming the testimonies are independent of each other (i.e W and G are conditionally independent):

$$P(H_i \mid G, W) = \alpha P(H_i) \sum_j P(G \mid S_j) P(W \mid S_j) P(S_j \mid H_i)$$

We can interpret this computation as a three-stage process.

- Combine the likelihood vectors for G and W to get one for S.

$$P(e | S_j) = P(G | S_j)\, P(W | S_j)$$

- Propagate this result up to H (using the matrix)

$$\Lambda_i(H) = \sum_j P(e | S_j)\, P(S_j | H_i)$$

- Multiply by prior probability and compute the overall belief.

$$P(H_i | e) = \alpha P(H_i) \Lambda_i(H)$$

Example

$$\frac{P(G \mid \text{alarmsound})}{P(G \mid \neg\text{alarmsound})} = \frac{4}{1}$$

$$\frac{P(W \mid \text{alarmsound})}{P(W \mid \neg\text{alarmsound})} = \frac{9}{1}$$

$$\text{Assume } P(\text{alarm} \mid \text{burglary}) = 0.95$$

$$\text{and} \quad P(\text{alarm} \mid \neg\text{burglary}) = 0.01$$

$$\text{and} \quad P(\text{burglary}) = 10^{-4}$$

Then compute $P(H_i \mid G, W)$

Where H_1 = burglary

And H_2 = no burglary

$$M = \begin{array}{c} \\ \text{burgalry} \\ \text{noburgalry} \end{array} \begin{array}{cc} \text{alarmsound} & \text{noalarmsound} \\ \left[\begin{array}{c} 0.95 \\ 0.01 \end{array}\right. & \left.\begin{array}{c} 0.05 \\ 0.99 \end{array}\right] \end{array}$$

$$P(e \mid S_i) = P(W, G \mid S_i) = \Lambda_i(S)$$

$$= \begin{bmatrix} 4 \\ 1 \end{bmatrix} \begin{bmatrix} 9 \\ 1 \end{bmatrix} = \begin{bmatrix} 36 \\ 1 \end{bmatrix}$$

$$\Lambda_i(H) = \sum_j P(e \mid S_i) P(S_j \mid H_i)$$

$$= \begin{bmatrix} 0.95 & 0.05 \\ 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} 36 \\ 1 \end{bmatrix} = \begin{bmatrix} 34.25 \\ 0.35 \end{bmatrix}$$

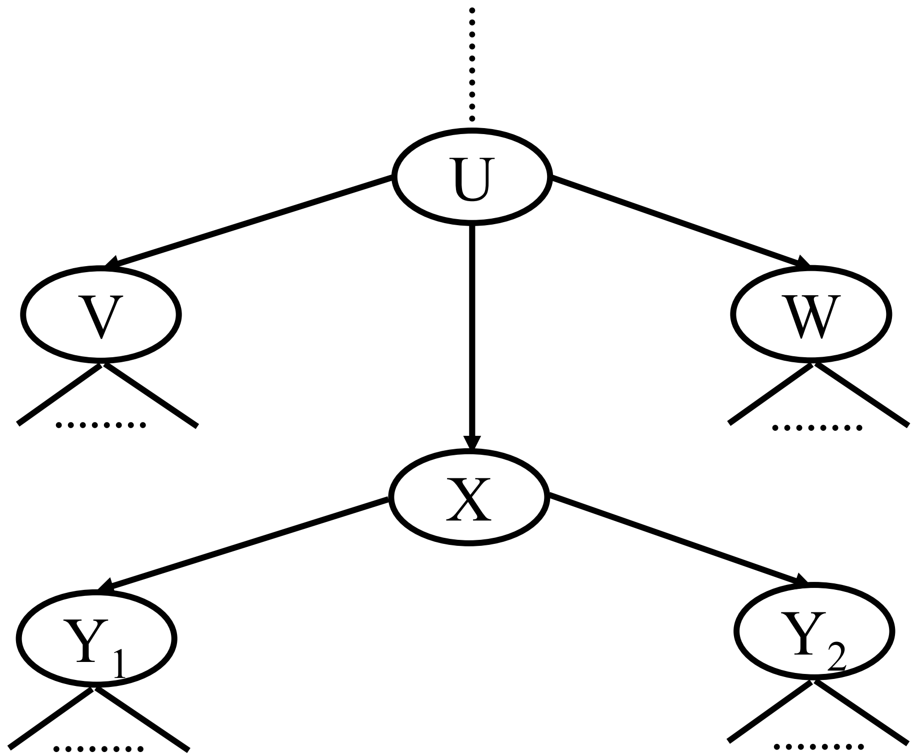
$$P(H_i \mid G, W) = \alpha P(H_i) \Lambda_i(H)$$

$$= \alpha \begin{bmatrix} 10^{-4} & 1 - 10^{-4} \end{bmatrix} \begin{bmatrix} 34.25 & 0.35 \end{bmatrix}$$

$$= \begin{bmatrix} 0.00253 & 0.99747 \end{bmatrix}$$

Belief Updating in a Hierarchy

Let X,U,V,W,Y1,Y2, stand for different hypotheses, each with possible values $X = x_1, x_2, \ldots, x_m$ and $Y_1 = y_{11}, y_{12}, \ldots, y_{1n}$.



The conditional probabilities between any two nodes can be represented by a matrix:

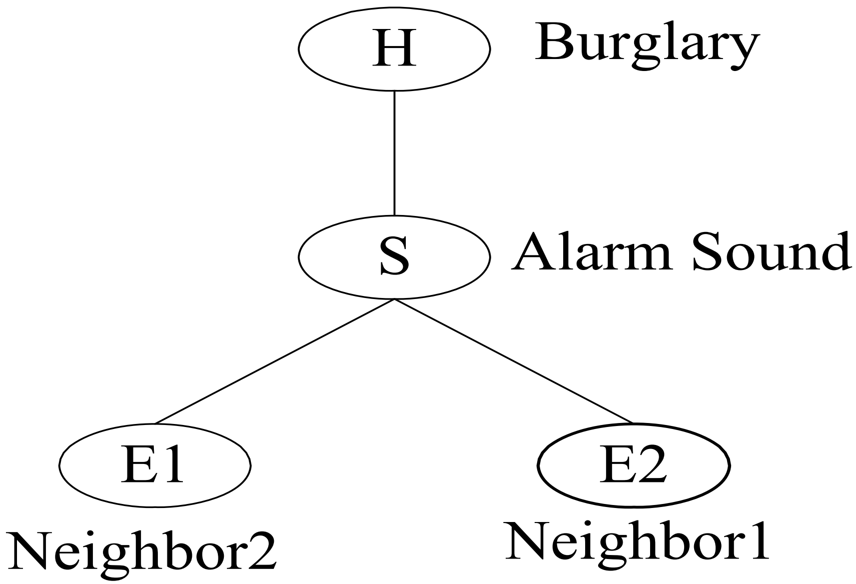
$$M_{Y|X} = \begin{bmatrix} P(Y_{11} | X_1) & P(Y_{12} | X_1) & \cdots & P(Y_{1n} | X_1) \\ P(Y_{11} | X_2) & P(Y_{12} | X_2) & \cdots & P(Y_{1n} | X_2) \\ \vdots & \vdots & & \vdots \\ P(Y_{11} | X_m) & P(Y_{12} | X_m) & \cdots & P(Y_{1n} | X_m) \end{bmatrix}_{m \times n}$$

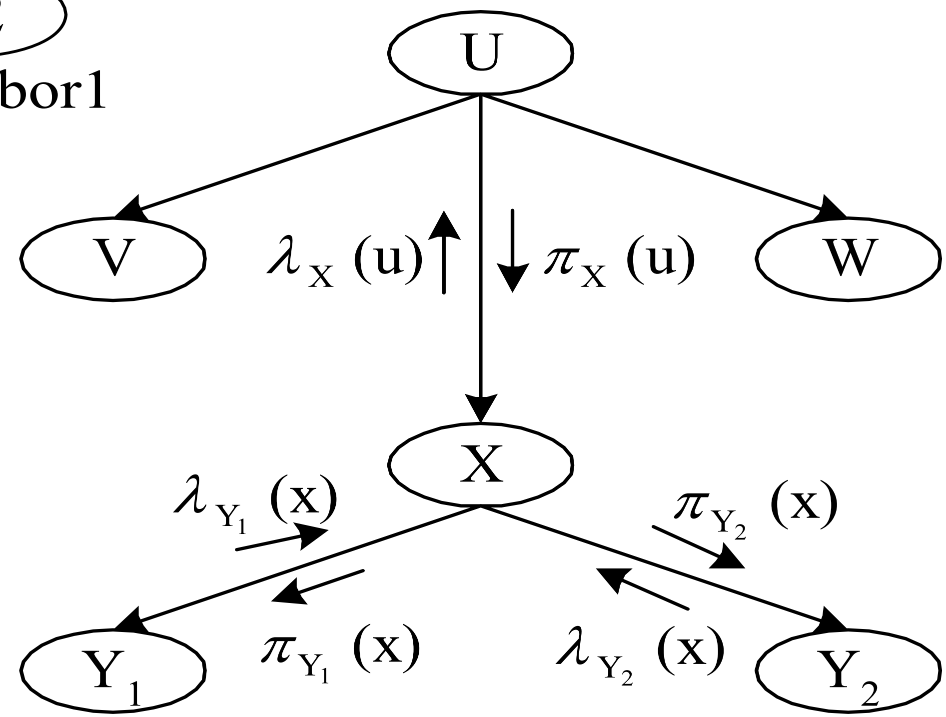
For each node (hypothesis) X in the tree, we separate the total evidence into two parts. λ(X) will be a vector which represents all support that node X receives from its descendents, e.g. Y_1, Y_2, …

$$\lambda(X) = [P(e_{Y1}, e_{Y2} \mid X = X_1)\ P(e_{Y1}, e_{Y2} \mid X = X_2) \\ \ldots P(e_{Y1}, e_{Y2} \mid X = X_m)]$$

π(X) is a vector that represents all support that node X receives from its non-descendents.

$$\pi(X) = [P(X = X_1 \mid \vec{e}_n)\ P(X = X_2 \mid \vec{e}_n) \\ \ldots P(X = X_m \mid \vec{e}_n)]$$


H
Burglary
S
Alarm Sound
E1
E2
Neighbor2
Neighbor1


U
V
λ_X (u)
π_X (u)
W
X
λ_{Y_1} (x)
π_{Y_2} (x)
π_{Y_1} (x)
λ_{Y_2} (x)
Y_1
Y_2