CSE 150 5/7

- Learning CPTs from incomplete data

  Examples $t = 1, 2, \ldots, T$

  Hidden nodes $H^{(t)}$

  Visible nodes $V^{(t)}$

- EM algorithm

  E-step: compute $P(X_i = x, pa_i = \pi \mid V^{(t)})$

  M-step: update CPTs

- Update rule:

$$P(X_i = x \mid pa_i = \pi) \leftarrow \frac{\sum_t P(X_i = x, pa_i = \pi \mid V^{(t)})}{\sum_t P(pa_i = \pi \mid V^{(t)})}$$

- Iterate until convergence

  Monotonic improvement in log-likelihood

$$\mathcal{L} = \sum_t \log P(V^{(t)})$$

$\boxed{\text{Example}}$

$$A \longrightarrow B \longrightarrow C$$

incomplete data set $\{(a_t, c_t)\}_{t=1}^{T}$

- E-step: compute $P(B=b \mid A=a_t, C=c_t)$ for all examples $t = 1 \ldots T$
- M-step: $P(B=b \mid A=a) \leftarrow \dfrac{\sum_t P(A=a, B=b \mid A=a_t, C=c_t)}{\sum_t P(A=a \mid A=a_t, C=c_t)}$

Simplify:

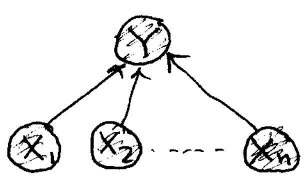
$$P(b|a) \leftarrow \frac{\sum_t I(a,a_t)\, P(b|a_t, c_t)}{\sum_t I(a, a_t)}$$

$$P(C=c\,|B=b) \leftarrow \frac{\sum_t P(B=b,\ C=c \,|\, A=a_t, C=c_t)}{\sum_t P(B=b\,|\,A=a_t, C=c_t)}$$

Simplify:

$$P(c|b) \leftarrow \frac{\sum_t I(c, c_t)\, P(b\,|\,a_t, c_t)}{\sum_t P(b\,|\,a_t, c_t)}$$

$\boxed{\text{Noisy-OR}}$



Symptom $y \in \{0,1\}$

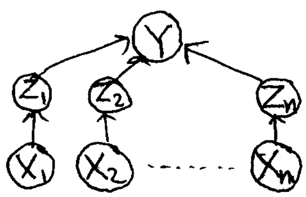
Diseases $X_i \in \{0,1\}$

$$P(Y=1\,|\,X_1, \dots, X_n) = 1 - \prod_{i=1}^{n} (1-p_i)^{X_i} \text{ with } p_i \in [0,1]$$

- From complete data $\{(\vec{X}_t, Y_t)\}_{t=1}^{T}$, how to estimate $p_i \in [0,1]$?

Note: noisy-OR is a "parametric" model of CPT. No simple, closed-form ML estimate of p_i.

- Alternative formulation



$$P(Y=1\,|\,z_1, \dots, z_n) = OR(z_1, z_2, \dots, z_n)$$

$$P(z_i=1\,|\,X=1) = p_i$$

$$P(z_i=1\,|\,X=0) = 0$$

Equivalently:

$$P(z_i=0\,|\,X_i) = (1-p_i)^{X_i} = \begin{cases} 1-p_i & \text{if } X_i=1 \\ 1 & \text{if } X_i=0 \end{cases}$$

- What is $P(Y=1|\vec{x})$ in this model?

$$P(Y=1|\vec{x}) = \sum_{\vec{z}\in\{0,1\}^n} P(Y=1,\vec{z}|\vec{x}) \quad \text{marginalization}$$

$$= \sum_{\vec{z}} P(Y=1|\vec{z},\vec{x})\, P(\vec{z}|x) \quad \text{product rule}$$

0 if $\vec{z}=\vec{0}$
1 otherwise

$$= \sum_{\vec{z}} P(Y=1|\vec{z})\, P(\vec{z}|\vec{x}) \quad \text{conditional independence}$$

$$= \sum_{\vec{z}\neq 0} P(\vec{z}|\vec{x})$$

$$= 1 - P(\vec{z}=0|\vec{x})$$

$$= 1 - \prod_{i=1}^{n} P(z_i=0|x_i)$$

$$= 1 - \prod_{i=1}^{n}(1-p_i)^{x_i}$$

$\boxed{\text{Same as original noisy-OR BN!}}$

- Posterior probability

$$P(z_i=1|\vec{x},y) = \frac{\overbrace{P(y|\vec{x},z_i=1)}^{\substack{1 \text{ if } y=1\\ 0 \text{ if } y=0}}\ \overbrace{P(z_i=1|\vec{x})}^{\begin{cases}p_i & \text{if } x_i=1\\ 0 & \text{if } x_i=0\end{cases}}}{P(y|\vec{x})}$$

$$= \begin{cases} 0 & \text{if } y=0 \\ \dfrac{p_i x_i}{1-\prod_{i=1}^{n}(1-p_i)^{x_i}} & \text{if } y=1 \end{cases} = \frac{y p_i x_i}{1-\prod_{i=1}^{n}(1-p_i)^{x_i}}$$

- (Conditional) Log-likelihood of (in)complete data $\{(x_t,y_t)\}_{t=1}^{T}$

$$\mathcal{L} = \sum_{t=1}^{T} \log P(y_t|\vec{x}_t)$$

$$= \sum_{t=1}^{T}\left[(1-y_t)\log P(Y=0|\vec{x}_t) + y_t \log P(Y=1|\vec{x}_t)\right]$$

$$= \sum_{t=1}^{T}\Big[(1-\underset{\text{data}}{y_t})\log\prod_{i=1}^{n}(1-\underset{\text{parameter}}{p_i})^{x_{it}} + \underset{\text{data}}{y_t}\log\Big[1-\prod_{i=1}^{n}(1-\underset{\text{parameter}}{p_i})^{\overset{\text{data}}{x_{it}}}\Big]\Big]$$

$$= \sum_{t=1}^{T}\left\{(1-y_t)\sum_{i=1}^{n} x_{it}\log(1-p_i) + y_t\log\left[1-\prod_{i=1}^{n}(1-p_i)^{x_{it}}\right]\right\}$$

Note: complicated expression to optimize with respect to p_i!

EM to the rescue!

Shorthand: let $T_i = \sum_{t=1}^{T} x_{it}$ count # times that $x_i = 1$ (or i^{th} disease is present)

- EM update rule

$$p_i = P(z_i=1 \mid x_i=1) \longleftarrow \frac{\sum_t P(z_i=1, x_i=1 \mid X=\vec{x}_t, Y=y_t)}{\sum_t P(X_i=1 \mid X=\vec{x}_t, Y=y_t)}$$

simplify

$$p_i \longleftarrow \frac{\sum_t I(x_{it},1)\, P(z_i=1 \mid \vec{x}_t, y_t)}{\sum_t I(x_{it},1)}$$

$P(z=1 \mid \vec{x}_t, y_t)$

$$p_i \longleftarrow \frac{1}{T_i} \sum_t I(x_{it},1) \left[\frac{y_t\, p_i x_{it}}{1 - \prod_{i=1}^{n} (1-p_i)^{x_{it}}} \right]$$

$$\boxed{p_i \longleftarrow \frac{1}{T_i} \sum_t \frac{y_t\, p_i x_{it}}{1 - \prod_{i=1}^{n} (1-p_i)^{x_{it}}}}$$

This update, applied in parallel to $\{p_i\}_{i=1}^{n}$, will monotonically increase $\mathcal{L} = \sum_t \log P(y_t \mid \vec{x}_t)$.