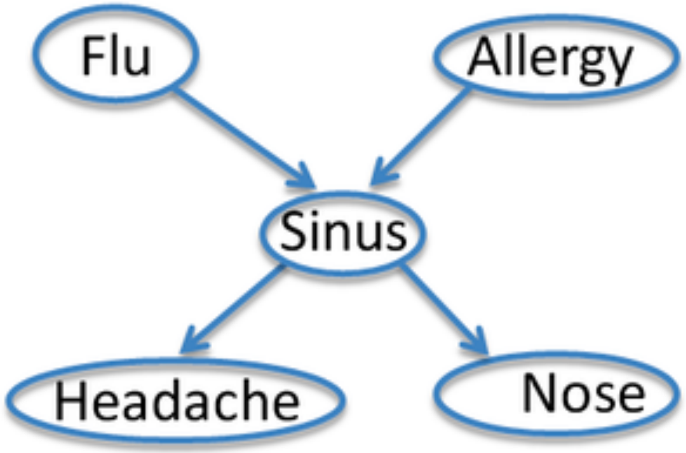
Expectation Maximiation

INSTRUCTOR: HONGJIE CHEN

JUNE 14TH 2022

MLE of $\theta_{s|ij}$ from Partially Observed Data

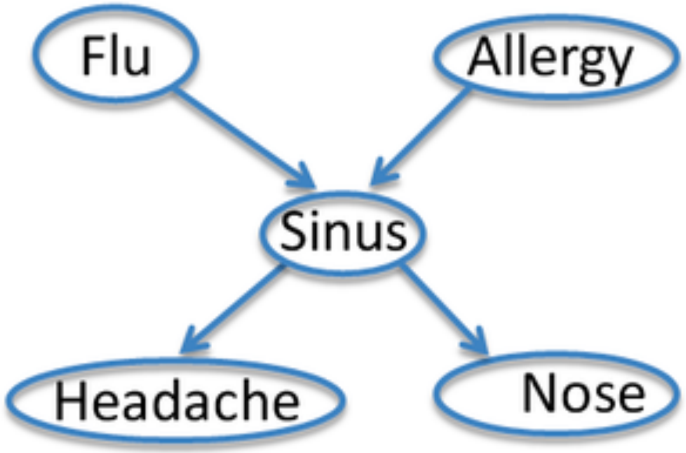
- If data are partially observed
  - For example, S is not observed



- $\theta \leftarrow \text{argmax}_\theta \log P(D|\theta)$
  - Let X be all observed variables
  - Let Z be all unobserved variables
  - $\theta \leftarrow \text{argmax}_\theta \log P(X, Z|\theta)$
- Can't calculate since Z is unknown, solution?
- Expectation Maximation

Expectation-Maximization

- EM seeks to estimate $P_\theta(Z|X)$
- $\theta \leftarrow \text{argmax}_\theta \mathbb{E}_{Z|X,\theta}\left[\log P(X,Z|\theta)\right]$
  - $X = \{F, A, H, N\}$ **Observed**
  - $Z = \{S\}$ **Unobserved**



$$\log P(X,Z|\theta) = \sum_{k=1}^{K} \log P(f_k) + \log P(a_k) + \log P(s_k|f_k a_k) + \log P(h_k|s_k) + \log P(n_k|s_k)$$

$$\mathbb{E}[x] = \sum_i P(X=i)i$$

$$\mathbb{E}_{P(Z|X,\theta)}\left[\log P(X,Z|\theta)\right] = \sum_{k=1}^{K}\sum_{i=0}^{1} P(s_k = i | f_k, a_k, h_k, n_k)$$

$$\left[\log P(f_k) + \log P(a_k) + \log P(s_k|f_k a_k) + \log P(h_k|s_k) + \log P(n_k|s_k)\right]$$

Expectation-Maximization Algorithm

- EM is an iterative method to estimate parameters from partly observed data that contains two step
- Given observed variable set X, unobserved variable set Z
- Initialize parameters θ
- Iterate until converged
  - Expectation step:

    Estimate the values of unobserved Z conditioned on X with θ
  - Maximization step: *argmax*

    Use observed values and E-step estimates to derive a better θ

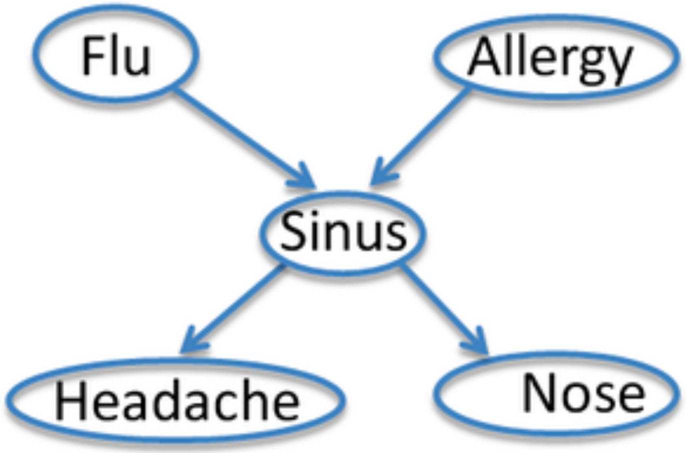
EM Algorithm with Maths

- Iterative until convergence
  - E step: Calculate $P(Z \mid X, \theta)$
  - M step: $\theta' \leftarrow \text{argmax}_{\theta'} Q(\theta' \mid \theta)$
- Guaranteed to find local maximum
  - Each iteration increases $\mathbb{E}_{P(Z|X,\theta)}\left[\log P(X, Z \mid \theta')\right]$

VIRGINIA TECH

E step

- $X = \{F, A, H, N\}$
- $Z = \{S\}$



$$P(S_k = 1 | f_k, a_k, h_k, n_k, \theta) = \frac{P(S_k = 1, f_k, a_k, h_k, n_k | \theta)}{P(S_k = 1, f_k, a_k, h_k, n_k | \theta) + P(S_k = 0, f_k, a_k, h_k, n_k | \theta)}$$

$$\rightarrow \mathbb{E}[s_k]$$

M step

- Recall fully observed case

$$\theta_{s|ij} = \frac{\sum_{k=1}^{K} \delta(f_k = i, a_k = j, s_k = 1)}{\sum_{k=1}^{K} \delta(f_k = i, a_k = j)}$$

- Partially observed case

$$\theta_{s|ij} \leftarrow \frac{\sum_{k=1}^{K} \delta(f_k = i, a_k = j)\mathbb{E}[s_k]}{\sum_{k=1}^{K} \delta(f_k = i, a_k = j)}$$

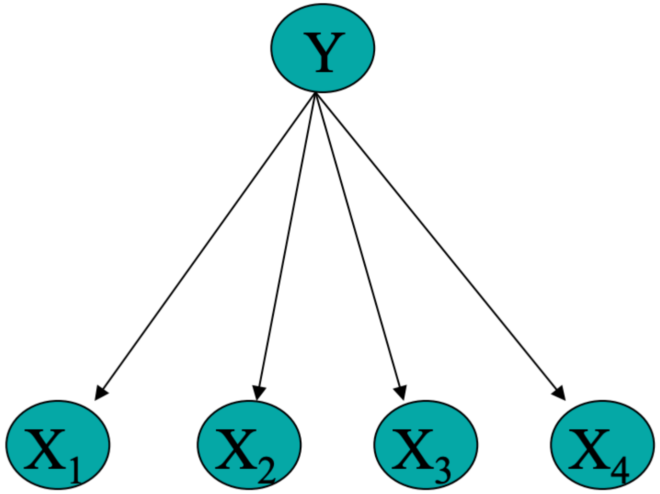
EM Description in a High Level

- E step:
  - Calulcate the expectation for each unobserved variables
- M step:
  - Calculate ‘MLE’ except the actual count (which is unknown) is replaced by its expectation count

Another example

- Train Naïve Bayes with unlabled data

Learn P(Y|X)



Y	X1	X2	X3	X4
1	0	0	1	1
0	0	1	0	0
0	0	0	1	0
?	0	1	1	0
?	0	1	0	1

EM with the Example

- E step:
  - Calulcate the expectation for each unobserved variables

$$\mathbb{E}_{P(Y|X_1\ldots X_N)}[y(k)] = P(y(k)=1\,|\,x_1(k),\ldots,x_N(k);\theta) = \frac{P(y(k)=1)\prod_i P(x_i(k)\,|\,y(k)=1)}{\sum_{j=0}^{1} P(y(k)=j)\prod_i P(x_i(k)\,|\,y(k)=j)}$$

- M step:
  - Calculate MLE except the actual count (which is unknown) is replaced by its expectation count

Expectation

$$\theta_{ij|m} = \hat{P}(X_i=j\,|\,Y=m) = \frac{\sum_k P(y(k)=m\,|\,x_1(k)\ldots x_N(k))\,\delta(x_i(k)=j)}{\sum_k P(y(k)=m\,|\,x_1(k)\ldots x_N(k))}$$

Actual

Which substitutes the term in MLE $\hat{P}(X_i=j\,|\,Y=m) = \frac{\sum_k \delta((y(k)=m)\wedge(x_i(k)=j))}{\sum_k \delta(y(k)=m)}$