

Learning Parameters of Gaussian Bayesian Networks

Sargur Srihari
srihari@cedar.buffalo.edu

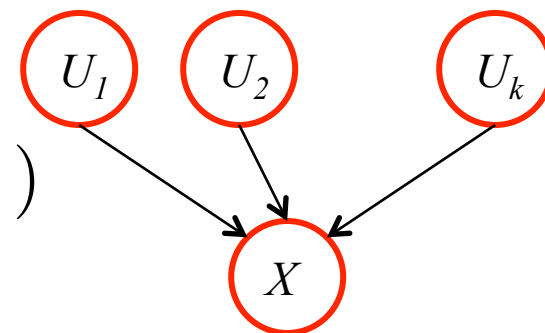
Topics

1. Linear Gaussian Model
2. Maximum Likelihood Solution

Learning parameters of Gaussian BN

- All variables are Gaussian

$$P(X|u_1, \dots, u_k) \sim N(\beta_0 + \beta_1 u_1 + \dots + \beta_k u_k; \sigma^2)$$



- Task is to learn $k+2$ parameters

$$\theta_{X|U} = \{\beta_0, \beta_1, \dots, \beta_k, \sigma^2\}$$

- To find ML estimates, define log-likelihood

$$\log L_X(\theta_{X|U} : D) = \sum_m \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\beta_0 + \beta_1 u_1[m] + \dots + \beta_k u_k[m] - x[m])^2 \right]$$

- which is likelihood of a Gaussian with mean

$$\beta_0 + \beta_1 u_1 + \dots + \beta_k u_k$$

- summation is over samples $m=1, \dots, M$

- Take derivatives wrt β_i , σ , set equal to zero and solve for parameters

Solving for the parameters

- Gradient of log-likelihood wrt β_0

$$-\frac{1}{\sigma^2} \left(M\beta_0 + \beta_1 \sum_m u_1[m] + \dots + \beta_k \sum_m u_k[m] - x[m] \right)$$

- Equating to zero and rearranging we get

$$\frac{1}{M} \sum_m x[m] = \beta_0 + \beta_1 \frac{1}{M} \sum_m u_1[m] + \dots + \beta_k \frac{1}{M} \sum_m u_k[m]$$

- All the summations can be obtained from data thus giving us a linear equation
- Similarly we get k more linear equations by taking derivatives wrt β_i
- Standard linear algebra techniques are used to solve $k+1$ simultaneous equations ⁴

Solving Linear Equations

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array}$$

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

1. Matrix Solution: $Ax=b$
Therefore $x=A^{-1}b$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

2. Gaussian Elimination

$$\begin{array}{l} x + 3y - 2z = 5 \\ 3x + 5y + 6z = 7 \\ 2x + 4y + 3z = 8 \end{array}$$

$$\begin{aligned} \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 3 & 5 & 6 & 7 \\ 2 & 4 & 3 & 8 \end{array} \right] &\sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & -4 & 12 & -8 \\ 2 & 4 & 3 & 8 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & -4 & 12 & -8 \\ 0 & -2 & 7 & -2 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & 1 & -3 & 2 \\ 0 & -2 & 7 & -2 \end{array} \right] \\ &\sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & 1 & -3 & 2 \\ 0 & 0 & 1 & 2 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -2 & 5 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 2 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 3 & 0 & 9 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 2 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & -15 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 2 \end{array} \right] \end{aligned}$$

Estimating variance

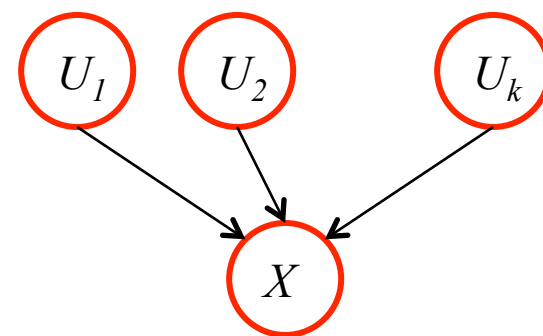
- Taking derivative of likelihood and setting to zero, we get

$$\sigma^2 = Cov_D[X; X] - \sum_i \sum_j \beta_i \beta_j Cov_D[U_i; U_j]$$

– where

$$Cov_D[X; Y] = E_D[X \cdot Y] - E_D[X] \cdot E_D[Y]$$

$$E_D[X \cdot Y] = \frac{1}{M} \sum_m x[m]y[m] \quad E_D[X] = \frac{1}{M} \sum_m x[m]$$



- First term is the empirical variance of X
- Other terms are empirical covariances of inputs