Gaussian Graphical Models

- Until now, we've worked with $X_i \in \mathcal{X}$, $|\mathcal{X}| < \infty$
- Any factor $f_a(x_1, \dots x_\ell)$ as a $|\mathcal{X}|^\ell$ table
- Alg's only used $+$, $\otimes$, lookups

Let's start discussing continuous RVs in $\mathbb{R}^n$.

- A parametric family allow us to [store factors / compute msgs] efficiently.

Def $X = (X_1, \dots, X_n)$ is Gaussian $\sim \mathcal{N}(\mu, \Sigma)$

if

$$P(X) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \cdot \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

Probability density

-n-

where Σ is positive definite.

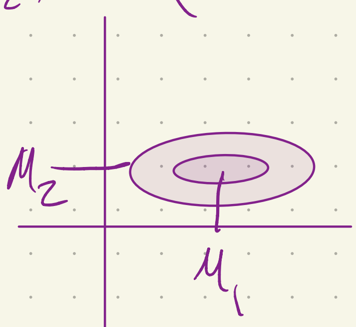
$$x^T \Sigma x > 0 \quad \forall x \neq 0.$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

$$\mu = \mathbb{E}[X] \in \mathbb{R}^n$$

$$\Sigma = \mathrm{Cov}(X) = \mathbb{E}\left[(X-\mu)(X-\mu)^T\right]$$

$$\sigma_{\min}(\Sigma) > 0$$

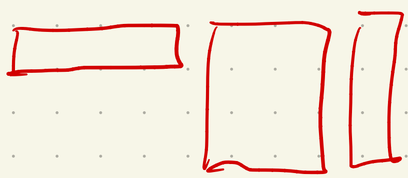


Def The covariance form of a multivariate gaussian is $X \sim N(\mu, \Sigma)$, $\Sigma \to$ PD

$$\hookrightarrow P(x) = \frac{1}{Z} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

$$\propto \; -\frac{1}{2}x^T\Sigma^{-1}x + \frac{1}{2}x^T\Sigma^{-1}\mu + \frac{1}{2}\mu^T\Sigma^{-1}x + c$$

$$= -\frac{1}{2}x^T\Sigma^{-1}x + \underbrace{x^T\Sigma^{-1}\mu}_{=(\Sigma^{-1}\mu)^T x} + c$$



The **information form** of a MV gaussian $X \sim N^{-1}(h, J)$, $J \hookrightarrow$ PD

$$P(X) = \frac{1}{Z_{h,J}} \exp\left(-\frac{1}{2}x^T J x + h^T x\right)$$

Claim/observation

$$N(\mu, \Sigma) = N^{-1}(h, J) \quad \text{iff} \quad \begin{cases} \Sigma^{-1} = J \\ \Sigma^{-1}\mu = h \end{cases}$$

(1) Marginalization (easier in covariance form)
(2) Conditioning (easier in information form)

Marginalization

$$\begin{bmatrix} x_1 \in \mathbb{R}^n \\ x_2 \in \mathbb{R}^m \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array} \right] \right)$$

What does $P(x_1)$ look like?

Claim $x_1 \sim N(\mu_1, \Sigma_{11})$.

Pf (sketch)

$$P(x_1) = \int P(x_1, x_2)\, dx_2$$

$$\mathbb{E}[x_1] = \mu_1$$

$$\mathbb{E}[(x_1 - \mu_1)^T (x_1 - \mu_1)] = \Sigma_{11}$$

What about marginalization when

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}^{-1}\left(\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \left[\begin{array}{c|c} J_{11} & J_{12} \\ \hline J_{21} & J_{22} \end{array}\right]\right)?$$

how do we write $P(x_1)$ in terms of h_1, h_2, J_{ij}?

$$x_1 \sim \mathcal{N}^{-1}\left(h_1 - J_{12} J_{22}^{-1} h_2,\ J_{11} - J_{12} J_{22}^{-1} J_{21}\right)$$

(2) What about conditioning $P(x_1 | x_2)$?

Covariance form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]\right)$$

$$P(x_1 | x_2) \sim \mathcal{N}\left(\mu_1 + \Sigma_{12} \Sigma_{22}^{-1}(x_2 - \mu_2),\ \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}\right)$$

Information form of $P(x_1 \mid x_2)$

$$P(x_1 \mid x_2) \sim \mathcal{N}^{-1}(h_1 - J_{12} x_2, J_{11})$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}^{-1}\left(\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \left[\begin{array}{c|c} J_{11} & J_{12} \\ \hline J_{21} & J_{22} \end{array}\right] \right)$$

$$P[x_1 \mid x_2] \propto \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} h_1^T & h_2^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right)$$

$$\propto \exp\left(-\frac{1}{2} x_1^T J_{11} x_1 + \left(-\frac{1}{2} J_{12} x_2 - \frac{1}{2} J_{12} x_2 + h_1 \right)^T x_1 \right)$$

Remark:

Independence is simpler in cov form ($\Sigma_{12} = 0$)

$x_i \perp\!\!\!\perp x_j \mid x_{rest}$ is simpler in info form $J_{12} = 0$

Def | undirected gaussian graphical models

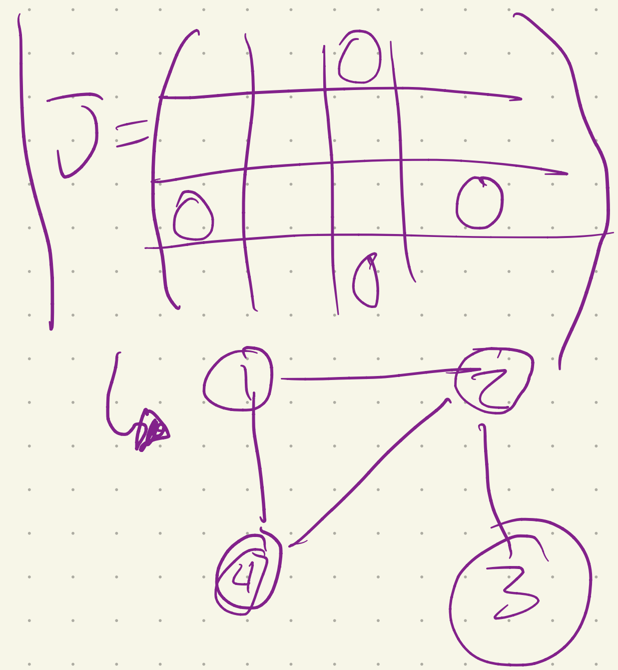
$$X \sim N^{-1}(h, J)$$

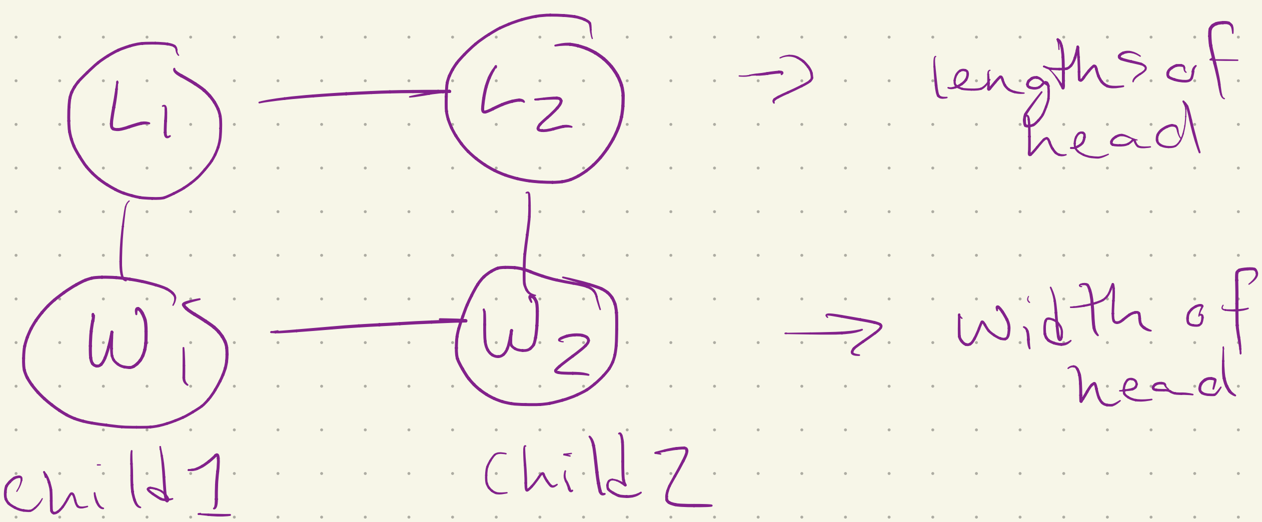
$$P(X) = \frac{1}{Z} \prod_{(i,j)\in E} e^{-x_i^T J_{ij} x_j} \times \prod_{i \in V} e^{-\frac{1}{2} x_i^T J_{ii} x_i + h_i^T x_i}$$

where $E = \{(i,j) \mid J_{ij} \neq 0\}$.

J is called the precision matrix

h the information vector



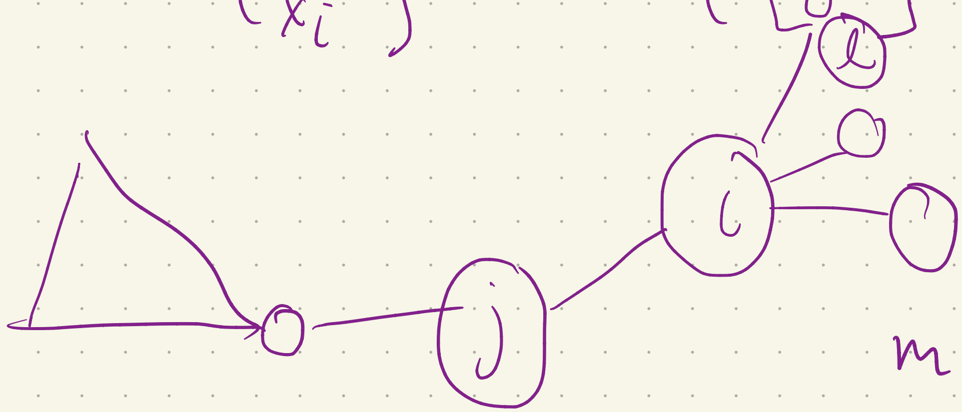


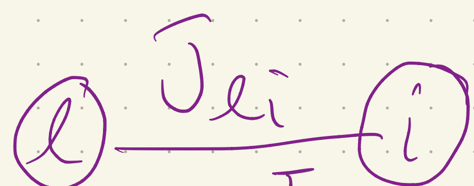
Example Exam scores for 5 related classes

Mechanics, Vectors, Algebra, Analysis, Stats

Gaussian belief propagation

Given $\begin{pmatrix} x_\ell \\ x_i \end{pmatrix} \sim N^{-1}\left(\begin{bmatrix} h_\ell \\ 0 \end{bmatrix}, \begin{bmatrix} J_{\ell,\ell} & J_{\ell i} \\ J_{i\ell} & 0 \end{bmatrix} \right)$





$$m_{\ell\to i}(x_i) \propto \exp\left(-\tfrac{1}{2} x_i^T J_{\ell\to i} x_i + h_i^T x_i\right)$$

$$\triangleq (h_{\ell\to i}, J_{\ell\to i}) \longleftarrow$$

$(h_\ell, J_{\ell\ell})$ $\qquad$ (h_i, J_{ii})

$$x_i \sim \mathcal{N}^{-1}(h_{\ell \to i}, J_{\ell \to i})$$

$$-J_{i\ell} J_{\ell\ell}^{-1} h_\ell \qquad -J_{i\ell} J_{\ell\ell}^{-1} J_{\ell i}$$

$J \in \mathbb{R}^{dn \times dn}$

Inversion takes $O(d^3 n^3)$ time

T iterations of BP take $(d^3 \cdot |E| \cdot T)$