1 Graph

1.1 Gaussian Graph

Take the Gaussian graph as an example. Suppose $p(\boldsymbol{x}, \boldsymbol{\xi}) = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\xi} \end{pmatrix}; \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}^{-1}\right)$, then $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \mathbf{0}, \boldsymbol{\Sigma}_{11}) = \mathcal{N}(\boldsymbol{x}; \mathbf{0}, (\boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21})^{-1})$, and the conditional distribution can be expressed as

$$p(\boldsymbol{\xi}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{\xi})}{p(\boldsymbol{x})} = \frac{\exp\left(-\frac{d}{2}\log(2\pi) + \frac{1}{2}\log\det\begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} - \frac{1}{2}(\boldsymbol{x}^\top\boldsymbol{\Omega}_{11}\boldsymbol{x} + \boldsymbol{\xi}^\top\boldsymbol{\Omega}_{22}\boldsymbol{\xi} + \boldsymbol{x}^\top\boldsymbol{\Omega}_{12}\boldsymbol{\xi} + \boldsymbol{\xi}^\top\boldsymbol{\Omega}_{21}\boldsymbol{x})\right)}{\exp\left(-\frac{d-1}{2}\log(2\pi) + \frac{1}{2}\log\det(\boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}) - \frac{1}{2}\boldsymbol{x}^\top(\boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21})\boldsymbol{x}\right)}$$

$$= \exp\left(-\frac{1}{2}\log(2\pi) + \frac{1}{2}\log\det(\boldsymbol{\Omega}_{22}) - \frac{1}{2}(\boldsymbol{\xi}^\top\boldsymbol{\Omega}_{22}\boldsymbol{\xi} + \boldsymbol{x}^\top\boldsymbol{\Omega}_{12}\boldsymbol{\xi} + \boldsymbol{\xi}^\top\boldsymbol{\Omega}_{21}\boldsymbol{x} + \boldsymbol{x}^\top\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}\boldsymbol{x})\right) = \mathcal{N}(\boldsymbol{\xi}; -\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}\boldsymbol{x}, \boldsymbol{\Omega}_{22}^{-1}).$$

If there is conditional independence between $\boldsymbol{\xi}$ and $\boldsymbol{x}$, then $\boldsymbol{\Omega}_{12}$ is sparse, and $p(\boldsymbol{\xi}|\boldsymbol{x})$ will only depend on a subset of $\boldsymbol{x}$. However, $p(\boldsymbol{x}, \boldsymbol{\xi})$ and $p(\boldsymbol{x})$ still depend on the whole support of $\boldsymbol{x}$.

1.2 Logconcave Graph

Now consider the log-concave graph. In the conditional likelihood maximization approach, the objective function is

$$\max\ \frac{1}{n}\sum_{i=1}^{n}\log p(\xi_i|\boldsymbol{x}_i) = \frac{1}{n}\sum_{i=1}^{n}\left(\log p(\boldsymbol{x}_i, \xi_i) - \log\int p(\boldsymbol{x}_i, \xi)d\xi\right)\ \text{s.t.}\ p(\boldsymbol{x}, \xi)\ \text{is a logconcave density.}$$

We need to compute the marginal density on $\boldsymbol{x} \in \mathbb{R}^{(d-1)}$ with one feature dimension ξ being integrated out, i.e.

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x}, \xi)d\xi \triangleq \int \exp(f(\boldsymbol{x}, \xi))d\xi.$$

Here $f(\boldsymbol{x}, \xi)$ is the joint log density, which is an affine function (hyperplane) on each simplex (triangle):

$$f(\boldsymbol{x}, \xi) = (f(\boldsymbol{x}_{j_0}, \xi_{j_0}), f(\boldsymbol{x}_{j_1}, \xi_{j_1}), \cdots, f(\boldsymbol{x}_{j_d}, \xi_{j_d}))\boldsymbol{w}; \quad \begin{pmatrix} 1 \\ \boldsymbol{x} \\ \xi \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{x}_{j_0} & \boldsymbol{x}_{j_1} & \cdots & \boldsymbol{x}_{j_d} \\ \xi_{j_0} & \xi_{j_1} & \cdots & \xi_{j_d} \end{pmatrix}\boldsymbol{w}; \quad \boldsymbol{w} \geq 0; \quad (\boldsymbol{x}, \xi) \in C_j;$$

where $(j_0, j_1, \cdots, j_d)$ is the vertex set for simplex C_j, or equivalently

$$f(\boldsymbol{x}, \xi) = (f(\boldsymbol{x}_{j_0}, \xi_{j_0}), f(\boldsymbol{x}_{j_1}, \xi_{j_1}), \cdots, f(\boldsymbol{x}_{j_d}, \xi_{j_d}))\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{x}_{j_0} & \boldsymbol{x}_{j_1} & \cdots & \boldsymbol{x}_{j_d} \\ \xi_{j_0} & \xi_{j_1} & \cdots & \xi_{j_d} \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ \boldsymbol{x} \\ \xi \end{pmatrix}; \quad (\boldsymbol{x}, \xi) \in C_j; \tag{1}$$

$$p(\boldsymbol{x}) = \int \exp(f(\boldsymbol{x}, \xi))d\xi = \sum_j \int_{(\boldsymbol{x},\xi)\in C_j} \exp\left((f(\boldsymbol{x}_{j_0}, \xi_{j_0}), f(\boldsymbol{x}_{j_1}, \xi_{j_1}), \cdots, f(\boldsymbol{x}_{j_d}, \xi_{j_d}))\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{x}_{j_0} & \boldsymbol{x}_{j_1} & \cdots & \boldsymbol{x}_{j_d} \\ \xi_{j_0} & \xi_{j_1} & \cdots & \xi_{j_d} \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ \boldsymbol{x} \\ \xi \end{pmatrix}\right)d\xi$$

The expression in (1) can be equivalently expressed as minimum of affine functions:

$$f(\boldsymbol{x}, \xi) = \min_j \{a_j + \boldsymbol{b}_j^\top\boldsymbol{x} + c_j^\top\xi\}$$

and the j corresponding to the minimum defines the simplex C_j. Then

$$p(\xi|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \xi)}{p(\boldsymbol{x})} = \frac{\exp\left(\min_j \{a_j + \boldsymbol{b}_j^\top\boldsymbol{x} + c_j^\top\xi\}\right)}{\int \exp\left(\min_j \{a_j + \boldsymbol{b}_j^\top\boldsymbol{x} + c_j^\top\xi\}\right)d\xi}$$

To infer conditional independence, it is tempting to impose group sparsity on $\boldsymbol{b}_j$. However, this might be inappropriate, since $p(\boldsymbol{x}, \boldsymbol{\xi})$ may still depend on the whole support of $\boldsymbol{x}$, as illustrated in the Gaussian case above.