

写在前面

需熟悉常见的矩阵微积分表示方法及其具体含义。

1 变分法

$F[y]$ 泛函。

$$F(y(x) + \epsilon\eta(x)) = F(y(x)) + \eta \int \frac{\partial F}{\partial y(x)} \eta(x) dx + O(\epsilon^2)$$

$$\int_x \frac{\partial F}{\partial y(x)} \eta(x) dx = 0$$

2 高斯分布的矩

一阶原点矩

$$E(x) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \int \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} x dx$$

做换元 $z = x - \mu$ 后，有：

$$E(x) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \int \exp \left\{ -\frac{1}{2} z^T \Sigma^{-1} z \right\} (z + \mu) dz$$

而由对称性：

$$\int \exp \left\{ -\frac{1}{2} z^T \Sigma^{-1} z \right\} (z) dz = 0$$

故有：

$$E(x) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \int \exp \left\{ -\frac{1}{2} z^T \Sigma^{-1} z \right\} (z + \mu) dz = \mu$$

二阶原点矩 类似地，做换元 $z = x - \mu$ 后，有：

$$E(x) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \int \exp \left\{ -\frac{1}{2} z^T \Sigma^{-1} z \right\} (z + \mu)(z + \mu)^T dz$$

将 $(z + \mu)(z + \mu)^T$ 展开, 考虑对称性, 有:

$$E(x) = \mu\mu^T + \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \int \exp\left\{-\frac{1}{2}z^T \Sigma^{-1}z\right\} zz^T dz$$

考虑 $I = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \int \exp\left\{-\frac{1}{2}z^T \Sigma^{-1}z\right\} zz^T dz$ 又有: $z = U^{-1}y = \Sigma y_j u_j$, 即 $y = Uz$

$$I = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \sum_{i,j} u_i u_j^T \int 1 \cdot \exp\left\{-\sum_{k=1}^D \frac{y_k^2}{2\lambda_k}\right\} y_i y_j d\vec{y}$$

再次利用对称性, 当且仅当 $i = j$ 时, 有:

$$\sum_{i,j} u_i u_j^T \int 1 \cdot \exp\left\{-\sum_{k=1}^D \frac{y_k^2}{2\lambda_k}\right\} y_i y_j d\vec{y} \neq 0$$

所以:

$$\begin{aligned} I &= \sum_{i=1}^D D u_i u_i^T \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \int 1 \cdot \exp\left\{-\sum_{k=1}^D \frac{y_k^2}{2\lambda_k}\right\} y_i y_j d\vec{y} \\ &= \sum_{i=1}^D D u_i u_i^T \lambda_i = \Sigma \end{aligned}$$

二阶中心矩

$$\begin{aligned} \text{Var}(x) &= E((X - E(X))(X - E(X))^T) \\ &= E(xx^T) - E(x)(E(x))^T = \Sigma \end{aligned}$$

3 条件高斯分布

划分向量

$$\begin{bmatrix} a_1 \\ \vdots \\ a_m \\ b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}, \mu = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$$

精度矩阵

$$\Lambda = \Sigma^{-1} = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$$

值得注意的是: $\Lambda_{ij} \neq \Sigma_{ij}$ (分块的性质)

4 Remark. Schur 补

对于矩阵 $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$, 若 $D \in \text{GLn}(\mathbb{F})$, 则:

称 $A - BD^{-1}C$ 为 D 关于 M 的 Schur 补;

称 $D - CA^{-1}B$ 为 A 关于 M 的 Schur 补。

求解 Schur 补的过程如下:

$$\begin{aligned} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} A & B \\ 0 & \Delta_A \end{bmatrix} \\ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} &= \begin{bmatrix} I & 0 \\ C & \Delta_A \end{bmatrix} \end{aligned}$$

故有:

$$\begin{aligned} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} &= \begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix} \\ \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} \end{aligned}$$

所以:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}$$

Rmk. 从上面的过程可以推出有 $\Delta_A = D - CA^{-1}B \in \text{GLn}(\mathbb{F})$ (做法类似高斯消元)
进而:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B\Delta_A CA^{-1} & A^{-1}B\Delta_A \\ -\Delta_A CA^{-1} & \Delta_A \end{bmatrix}$$

5 边缘高斯分布

若联合分布 $p(x_a, x_b)$ 是高斯分布, 则 $p(x_a|x_b)$ 也是高斯分布。现在讨论分布 $p(x_a) = \int p(x_a, x_b)dx_b$, 即边缘概率分布。

在联合分布 d.f. 中选出关于 x_b 的项：

$$\begin{aligned} & -\frac{1}{2}x_b^T \Lambda_{bb} x_b + x_b^T m \\ & = -\frac{1}{2}(x_b - \Lambda_{bb}^{-1} m)^T \Lambda_{bb} + \frac{1}{2}m^T \Lambda_{bb}^{-1} m \end{aligned}$$

其中， $m = \Lambda_{bb}\mu_b - \Lambda_{ba}(x_a - \mu_a)$ 。（回忆： Λ_{bb} 是精度矩阵中的分块）