0.1 联合分布

0.1.1 摘要

本节我们学习高斯联合分布。

0.1.2 已知

$$x \sim N(x|\mu, \Lambda^{-1})$$
$$y|x \sim N(y|Ax+b, L^{-1})$$

隐含条件

$$y = Ax + b + \epsilon, \quad \epsilon \sim N(0, L^{-1}), \quad x \perp \epsilon$$

0.1.3 所求

$$\begin{cases} p(y) \\ \\ p(x|y) \end{cases}$$

0.1.4 推导

推导 $p(y)$

$$E[y] = AE[x] + b + E[\epsilon] = A\mu + b$$
$$Var[y] = A\Lambda^{-1}A^T$$
$$\therefore y \sim N(A\mu + b, A\Lambda^{-1}A^T)$$

推导 $p(x|y)$

构造分布 z

此处我们构造一个分布:

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \sim N(\begin{bmatrix} \mu \\ A\mu + b \end{bmatrix}, \begin{bmatrix} \Lambda^{-1} & \Delta \\ \Delta & A\Lambda^{-1}A^T \end{bmatrix})$$

$$
\begin{aligned}
\Delta &= cov(x,y) \\
&= E[(x-E[x])(y-E[y])^T] \\
&= E[(x-\mu)(y-A\mu-b)^T] \\
&= E[(x-\mu)(Ax+b+\epsilon-A\mu-b)^T)] \\
&= E[(x-\mu)(Ax-A\mu+\epsilon)^T] \\
&= E[(x-\mu)(x-\mu)^T A^T + (x-\mu)\epsilon^T] \\
&= E[(x-\mu)(x-\mu)^T]A^T + E[(x-\mu)\epsilon^T] \\
&\because x \perp \epsilon \\
&\therefore = E[(x-\mu)(x-\mu)^T]A^T \\
&= \Lambda^{-1}A^T \\
\therefore z = \begin{pmatrix} x \\ y \end{pmatrix} &\sim N(\begin{bmatrix} \mu \\ A\mu+b \end{bmatrix}, \begin{bmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ \Lambda^{-1}A^T & A\Lambda^{-1}A^T \end{bmatrix})
\end{aligned}
$$

构造分布 $x.y$
我们设:

$$
\begin{aligned}
x.y &= x - \Sigma_{xy}\Sigma_{yy}^{-1}y \\
&= x - (\Lambda^{-1}A^T)(A\Lambda^{-1}A^T)^{-1}y \\
&= x - A^{-1}y \\
&= \begin{pmatrix} I & -A^{-1} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
E[x.y] &= E[x] - A^{-1}E[y] \\
&= \mu - A^{-1}(A\mu+b) \\
&= -A^{-1}b
\end{aligned}
$$

$$
\begin{aligned}
Var[x.y] &= \begin{pmatrix} I & -A^{-1} \end{pmatrix} Var[z] \begin{pmatrix} I \\ -(A^{-1})^T \end{pmatrix} \\
&= \begin{pmatrix} I & -A^{-1} \end{pmatrix} \begin{pmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ \Lambda^{-1}A^T & A\Lambda^{-1}A^T \end{pmatrix} \begin{pmatrix} I \\ -(A^{-1})^T \end{pmatrix} \\
&= \begin{pmatrix} \Lambda^{-1} - A^{-1}\Lambda^{-1}A^T & 0 \end{pmatrix} \begin{pmatrix} I \\ -(A^{-1})^T \end{pmatrix} \\
&= \Lambda^{-1} - A^{-1}\Lambda^{-1}A^T \\
&\therefore x.y \sim N(-A^{-1}b, \Lambda^{-1} - A^{-1}\Lambda^{-1}A^T)
\end{aligned}
$$

构造分布 $x|y$
我们有:

$$
x|y = x.y + A^{-1}y
$$

这里，我们可以将 $A^{-1}y$ 视为常数 C.
那么：

$$x|y = x.y + C$$
$$E[x|y] = A^{-1}y - A^{-1}b$$
$$Var[x|y] = Var[x.y]$$
$$\therefore x|y \sim N(A^{-1}y - A^{-1}b, \Lambda^{-1} - A^{-1}\Lambda^{-1}A^T)$$

现在，我们根据一个边缘分布和条件分布，通过构造联合分布，求出了另一个边缘分布和条件分布。