

Abstract

本节我们学习高斯联合分布。

Given

$$x \sim N(x|\mu, \Lambda^{-1}) \quad (20)$$

$$y|x \sim N(y|Ax + b, L^{-1}) \quad (21)$$

Inference

$$y = Ax + b + \epsilon, \quad \epsilon \sim N(0, L^{-1}), \quad x \perp \epsilon \quad (22)$$

To solve

$$\begin{cases} p(y) \\ p(x|y) \end{cases} \quad (23)$$

Derive

Derive $p(y)$

$$E[y] = AE[x] + b + E[\epsilon] = A\mu + b \quad (24)$$

$$Var[y] = A\Lambda^{-1}A^T \quad (25)$$

$$\therefore y \sim N(A\mu + b, A\Lambda^{-1}A^T) \quad (26)$$

Derive $p(x|y)$

construct dist z

此处我们构造一个分布:

$$z = \begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{bmatrix} \mu \\ A\mu + b \end{bmatrix}, \begin{bmatrix} \Lambda^{-1} & \Delta \\ \Delta & A\Lambda^{-1}A^T \end{bmatrix}\right) \quad (27)$$

$$\begin{aligned} \Delta &= cov(x, y) \\ &= E[(x - E[x])(y - E[y])^T] \\ &= E[(x - \mu)(y - A\mu - b)^T] \\ &= E[(x - \mu)(Ax + b + \epsilon - A\mu - b)^T] \\ &= E[(x - \mu)(Ax - A\mu + \epsilon)^T] \\ &= E[(x - \mu)(x - \mu)^T A^T + (x - \mu)\epsilon^T] \\ &= E[(x - \mu)(x - \mu)^T] A^T + E[(x - \mu)\epsilon^T] \\ &\therefore x \perp \epsilon \\ &\therefore = E[(x - \mu)(x - \mu)^T] A^T \\ &= \Lambda^{-1} A^T \end{aligned} \quad (28)$$

$$\therefore z = \begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{bmatrix} \mu \\ A\mu + b \end{bmatrix}, \begin{bmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ \Lambda^{-1}A^T & A\Lambda^{-1}A^T \end{bmatrix}\right) \quad (29)$$

construct dist $x.y$

let's set

$$\begin{aligned} x.y &= x - \Sigma_{xy}\Sigma_{yy}^{-1}y \\ &= x - (\Lambda^{-1}A^T)(A\Lambda^{-1}A^T)^{-1}y \\ &= x - A^{-1}y \\ &= (I \quad -A^{-1}) \begin{pmatrix} x \\ y \end{pmatrix} \end{aligned} \quad (30)$$

$$\begin{aligned} E[x.y] &= E[x] - A^{-1}E[y] \\ &= \mu - A^{-1}(A\mu + b) \\ &= -A^{-1}b \end{aligned} \quad (31)$$

$$\begin{aligned} Var[x.y] &= (I \quad -A^{-1})Var[z] \begin{pmatrix} I \\ -(A^{-1})^T \end{pmatrix} \\ &= (I \quad -A^{-1}) \begin{pmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ \Lambda^{-1}A^T & A\Lambda^{-1}A^T \end{pmatrix} \begin{pmatrix} I \\ -(A^{-1})^T \end{pmatrix} \\ &= (\Lambda^{-1} - A^{-1}\Lambda^{-1}A^T \quad 0) \begin{pmatrix} I \\ -(A^{-1})^T \end{pmatrix} \\ &= \Lambda^{-1} - A^{-1}\Lambda^{-1}A^T \end{aligned} \quad (32)$$

$$\therefore x.y \sim N(-A^{-1}b, \Lambda^{-1} - A^{-1}\Lambda^{-1}A^T) \quad (33)$$

construct $x|y$

we got

$$x|y = x.y + A^{-1}y \quad (34)$$

here, we can see $A^{-1}y$ as constant C .

then:

$$x|y = x.y + C \quad (35)$$

$$E[x|y] = A^{-1}y - A^{-1}b \quad (36)$$

$$Var[x|y] = Var[x.y] \quad (37)$$

$$\therefore x|y \sim N(A^{-1}y - A^{-1}b, \Lambda^{-1} - A^{-1}\Lambda^{-1}A^T) \quad (38)$$

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。