Abstract

本节我们学习多元高斯分布的边缘概率和条件概率

Prior Knowledeg

在上一节中，我们推导了多元高斯分布的概率密度函数:

$$x \sim N(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu\right)^T \Sigma^{-1}(x-\mu)) \tag{1}$$

现在我们将随机变量 x 拆分为两部分:

$$x \in R^p \quad x_a \in R^m \quad x_b \in R^n \quad m+n=p \tag{2}$$

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \tag{3}$$

Theorem

$$X \sim N(\mu, \Sigma) \quad Y = AX + B \Longrightarrow Y \sim N(A\mu + B, A\Sigma A^T) \tag{4}$$

Derive marginal probability

$$x_a = \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} x_a \\ x_b \end{pmatrix} + 0$$

$$E[x_a] = \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = \mu_a$$

$$\begin{aligned} Var[x_a] &= \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} \\ &= \Sigma_{aa} \\ &\therefore x_a \sim N(\mu_a, \Sigma_{aa}) \end{aligned}$$

Derive conditional probability

Let's set:

$$\begin{cases} x_{b.a} = x_b - \Sigma_{ba}\Sigma_{aa}^{-1} x_a \\ \mu_{b.a} = \mu_b - \Sigma_{b.a}\Sigma_{aa}^{-1}\mu_a \\ \Sigma_{bb.a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \end{cases} \tag{5}$$

$$\begin{aligned} x_{b.a} &= x_b - \Sigma_{ba}\Sigma_{bb}^{-1} x_a \\ &= \begin{pmatrix} -\Sigma_{ba}\Sigma_{bb}^{-1} & I \end{pmatrix} \begin{pmatrix} x_a \\ x_b \end{pmatrix} + 0 \end{aligned} \tag{6}$$

$$
\begin{aligned}
E[x_{b.a}] &= \begin{pmatrix} -\Sigma_{ba}\Sigma_{aa}^{-1} & I \end{pmatrix} \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \\
&= \mu_b - \Sigma_{ba}\Sigma_{aa}^{-1}\mu_a \\
&= \mu_{b.a}
\end{aligned} \tag{7}
$$

$$
\begin{aligned}
Var[x_{b.a}] &= \begin{pmatrix} -\Sigma_{ba}\Sigma_{aa}^{-1} & I \end{pmatrix} \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \begin{pmatrix} -\Sigma_{aa}^{-1}\Sigma_{ba}^{T} \\ I \end{pmatrix} \\
&= \begin{pmatrix} 0 & \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \end{pmatrix} \begin{pmatrix} -\Sigma_{aa}^{-1}\Sigma_{ba}^{T} \\ I \end{pmatrix} \\
&= \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \\
&= \Sigma_{bb.a}
\end{aligned} \tag{8}
$$

$$
\therefore x_{b.a} \sim N(\mu_{b.a}, \Sigma_{bb.a}) \tag{9}
$$

$$
\begin{cases} \mu_{b.a} = \mu_b - \Sigma_{ba}\Sigma_{aa}^{-1}\mu_a \\ \Sigma_{bb.a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \end{cases} \tag{10}
$$

我们将式子稍作变形：

$$
\begin{aligned}
x_{b.a} &= x_b - \Sigma_{ba}\Sigma_{aa}^{-1}x_a \\
x_b|x_a &= x_{b.a} + \Sigma_{ba}\Sigma_{aa}^{-1}x_a \\
&= Ix_{b.a} + C
\end{aligned} \tag{11}
$$

此处，协方差矩阵的部分 Σ_{ba} Σ_{aa} 都是可以通过计算得到的，因此可以视为常数。

而我们此处要求的是 $x_b|x_a$ ，因此 x_a 也是已知的，因此上式的第二项可以视为常数。

因此：

$$
E[x_b|x_a] = IE[x_{b.a}] + C = \mu_{b.a} + \Sigma_{ba}\Sigma_{aa}^{-1}x_a \tag{12}
$$

$$
Var[x_b|x_a] = IVar[x_{b.a}]I^T = \Sigma_{bb.a} \tag{13}
$$

因此我们得出了高斯分布的条件概率:

$$
x_b|x_a \sim N(\mu_{b.a} + \Sigma_{ba}\Sigma_{aa}^{-1}x_a, \Sigma_{bb.a}) \tag{14}
$$