高斯分布：从一般到特殊

Jianqi Huang

October 2022

1 一维高斯分布

对于一个一维高斯，从一开始需要表示一个对称的分布，同时从 y 轴开始向两侧不断递减的形式，也就是说需要满足以下的几个条件：

$$X \to \infty, f(X) \to 0 \tag{1}$$

$$f'(X) < 0, X > 0 \tag{2}$$

$$F'(X) > 0, X \leq 0 \tag{3}$$

而因此一开始从一个指数函数开始：

$$exp\{-\frac{1}{2}x^2\} \tag{4}$$

其中规定指数的系数为 $\frac{1}{2}$，目的在于为后来的积分、求导的处理的方便考虑.

进一步的，我们需要将其标准化，也就是其积分之后为 1，显然目前函数形式的积分并不等于 1，需要通过增加常数系数实现：先进行积分之后再进行除以积分量．先令

$$I = \int_{-\infty}^{+\infty} exp\{-x^2\} \tag{5}$$

实际积分的函数的 x 系数与划一之后的系数可建立等式关系，先划一处理.

$$\begin{aligned}
I^2 &= (\int_R f(x)dx)^2 \\
&= \int_{-\infty}^{+\infty} exp\{-x^2\} \int_{-\infty}^{+\infty} exp\{-y^2\} \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} exp\{-(x^2+y^2)\} \\
&= \int_0^{2\pi} d\theta \int_0^{\infty} rexp\{-r^2\}dr \\
&= 2\pi \times (-\frac{1}{2})e^{-r^2}|_0^{\infty} \\
&= \pi
\end{aligned}$$

所以最后会得到 (2) 式 $I = \int_{-\infty}^{+\infty} exp\{-x^2\} = \sqrt{\pi}$ 因此一个标准正态（高斯）分布：

$$f(x) = \frac{1}{\sqrt{2\pi}} exp\{-\frac{1}{2}x^2\} \tag{6}$$

若均值和方差不是 $(0,1)$，则进行相应的标准化，标准化的方式以 z-scores 实现：

$$z = \frac{x-\mu}{\sigma} \tag{7}$$

由此

$$f'(x) = \frac{1}{\sqrt{2\pi\sigma}} exp\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\} \tag{8}$$

2 多维高斯分布

从一个相互独立的高斯分布来引起，所谓的相互独立，在事件 A 和 B 同时发生的概率是等于两者单独发生的乘积．以数学公式的表示：

$$P(AB) = P(A) \times P(B) \tag{9}$$

由此公式可推出对于一个多维的高斯分布的密度函数可以表示为多个一维高斯分布的乘积.

$$\begin{aligned}
f(X) &= f(x_1) \cdot f(x_2) \cdots f(x_n) \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} exp\{-\frac{1}{2}(\frac{x_1-\mu_1}{\sigma_1})\} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} exp\{-\frac{1}{2}(\frac{x_2-\mu_2}{\sigma_2})\} \cdots \frac{1}{\sqrt{2\pi}\sigma_d} exp\{-\frac{1}{2}(\frac{x_d-\mu_d}{\sigma_d})\}
\end{aligned}$$

其中 $X = (x_1, x_2, \cdots, x_n)$

同时在考虑指数的结合上，令：

$$
\begin{aligned}
d^2(x,\mu) &= (\frac{x_1-\mu_1}{\sigma_1})^2 + (\frac{x_2-\mu_2}{\sigma_2})^2 + \cdots + (\frac{x_d-\mu_d}{\sigma_d})^2 \\
&= [x_1-\mu_1, x_2-\mu_2, \cdots, x_d-\mu_d] \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{2}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_d^2} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ \vdots \\ x_d-\mu_d \end{bmatrix} \\
&= (X-\mu)^T \Sigma^{-1} (X-\mu)
\end{aligned}
$$

2.1 向量之间的不独立

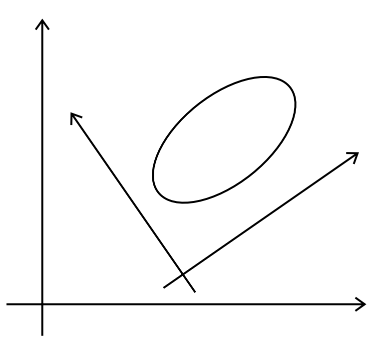
一般来说，相关性是独立性的必要非充分条件，但对于一组高斯分布来说，其不相关性可与独立性等价.

证明.

$$
\begin{aligned}
\sigma_{ij} &= E(X_i-\mu_i)(X_j-\mu_j) \\
&= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (x_i-\mu_i)(x_j-\mu_j) f(x_1, \cdots x_q) \cdot f(x_{q+1}, \cdots, x_p) dx_1 \cdots dx_p \\
&= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (x_i-\mu_i) f(x_1, \cdots x_q) dx \cdot \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (x_j-\mu_j) f(x_{q+1}, \cdots, x_p) dx \\
&= 0
\end{aligned}
$$

因为 $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}, \sigma_i \neq 0, \sigma_j \neq 0$，因此对于条件 $\sigma_{ij} = 0$ 等价于 $\rho_{ij} = 0$．因此不相关可以推出独立 □

核心思想：从一个二维开始，将两个非独立变量通过线性变换转变为一个独立不相关的一对变量.



一对目标变换方向的向量以 $u_1 = \begin{bmatrix} u_1^1 \\ u_1^2 \end{bmatrix}, u_2 = \begin{bmatrix} u_2^1 \\ u_2^2 \end{bmatrix}$ 形式写出，则数据的投影到上面的长度可

以用

$$u_1^T \cdot X = [u_1^1, u_1^2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{10}$$

最终的投影表示的是在 μ_1 上的坐标.

$$Y = \begin{bmatrix} u_1^T \cdot X \\ u_2^T \cdot X \end{bmatrix} = U^T X \tag{11}$$

其中 U 是一个正交矩阵 $U^T = U^{-1}$

如之前类似：先对于数据进行标准化：

$$\begin{aligned} Z = \begin{bmatrix} \frac{Y_1 - \mu_{Y_1}}{\sigma_{Y_1}} \\ \frac{Y_1 - \mu_{Y_1}}{\sigma_{Y_1}} \end{bmatrix} &= \begin{bmatrix} 1/\sigma_{Y_1}, 0 \\ 0, 1/\sigma_{Y_2} \end{bmatrix} \begin{bmatrix} Y_1 - \mu_{Y_1} \\ Y_2 - \mu_{Y_2} \end{bmatrix} \\ &= D(Y - \mu_Y) = D(U^T - U^T \mu_X) \\ &= DU^T(X - \mu_X) \end{aligned}$$

计算相对距离

$$d^2(z, \mu) = z_1^2 + z_2^2 = [z_1, z_2] \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = Z^T Z \tag{12}$$

$$d^2(x, \mu) = Z^T Z = (DU^T(X - \mu_X))^T (DU^T(X - \mu_X)) \quad = (X - \mu_X)^T U D^T D U^T (X - \mu_X) \tag{13}$$

$$\Sigma_Y^{-1} = D^T D = \begin{bmatrix} 1/\sigma_{Y1}^2, 0 \\ 0, 1/\sigma_{Y2}^2 \end{bmatrix} \tag{14}$$

从协方差 $\Sigma_Y = \begin{bmatrix} 1/\sigma_{Y1}^2, 0 \\ 0, 1/\sigma_{Y2}^2 \end{bmatrix}$ 的定义和与 X 之间的等式关系可进一步推得：

$$\begin{aligned} \Sigma_Y &= E[(Y - \mu_Y)(Y - \mu_Y)^T] \\ &= E[U^T(X - \mu_X)(X - \mu_X)^T U] \\ &= U^T E[(X - \mu_X)(X - \mu_X)^T] U = U^T \Sigma_X U \end{aligned}$$

即

$$\Sigma_Y = U^T \Sigma_X U \tag{15}$$

在其中的 $UD^TDU^T = U\Sigma_Y^{-1}U^T$ 而结合上等式可得

$$UD^TDU^T = \Sigma_X^{-1} \tag{16}$$

最终可将距离转换为只与 X 向量自身相关的函数:

$$d^2(x,\mu) = (X-\mu_X)^T\Sigma_X^{-1}(X-\mu_X) \tag{17}$$

在标准化中 $Z = DU^T(X-\mu_X)$，因此

$$|DU^T| = \sqrt{|\Sigma_X^{-1}|} = |\Sigma_X|^{-\frac{1}{2}} \tag{18}$$

因此最终的概率密度表达式

$$f(X) = \frac{1}{\sqrt{2\pi}}exp\{-z^2/2\} = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_X|^{\frac{1}{2}}}exp\{-\frac{1}{2}(X-\mu_X)^T\Sigma_X^{-1}(X-\mu_X)\} \tag{19}$$

2.2 从代数角度理解

我们同样站在推论高斯分布下独立与相关的等价下，利用代数知识推导高斯分布. 因为 $X \sim \mathcal{N}(\mu,\Sigma)$，其中 Σ 为一个半正定矩阵，由此存在满秩矩阵 $B \in R^n$ 使得

$$Z = B^{-1}(X-\mu) \tag{20}$$

$$Z \sim \mathcal{N}(0,I) \tag{21}$$

同时 B 矩阵满足

$$B = B^T \tag{22}$$

由雅可比矩阵 $J(Z \to X)$，其中

$$J(Z \to X) = \left|B^{-1}\right| = |B|^{-1} = |B|^{-\frac{1}{2}}\left|B^T\right|^{-\frac{1}{2}} = \left|BB^T\right|^{-1/2} \tag{23}$$

其中 $g(x) = f(x)J(Z \to X)$

$$g(x) = f(x)J(Z \to X) = \int_{\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{(2\pi)^{n/2}|BB^T|^{\frac{1}{2}}}e^{-1/2}[(X-\mu)(BB^T)^{-1}(X-\mu)^T]dx_1\cdots dx_n \tag{24}$$

再对于之前的 Σ 进行推导，可以看到

$$\begin{aligned}\Sigma &= E[(X-\mu)(X-\mu)^T] \\ &= E[(BZ-0)(BZ-0)^T] \\ &= Cov(BZ, BZ) \\ &= B \cdot Cov(Z, Z^T)B^T \\ &= BB^T\end{aligned}$$

再代入 (24) 式中可以得到与 (19) 相同的结论：

$$g(X) = \frac{1}{\sqrt{2\pi}} exp\{-z^2/2\} = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_X|^{\frac{1}{2}}} exp\{-\frac{1}{2}(X-\mu_X)^T\Sigma_X^{-1}(X-\mu_X)\} \tag{25}$$