Math Basis 02

Chen Gong

19 October 2019

本节的主要目的是从概率的角度来分析高斯分布，包括马氏距离和高斯分布的几何表示，以及高斯分布的局限性和解决的方法等等。对于多变量的高斯分布 $X \sim \mathcal{N}(\mu, \Sigma)$，概率密度函数为：

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

其中，$x \in \mathbb{R}^p$，

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \tag{2}$$

其中，Σ 一般为正定矩阵或者为半正定矩阵。

1 什么是马氏距离

在高斯分布中，$\sqrt{(x-\mu)^T \Sigma^{-1} (x-\mu)}$ 的计算结果是一个数，这个数被称为马氏距离。设 $z_1 = (z_{11}, z_{12})^T$，$z_2 = (z_{21}, z_{22})^T$。那么 z_1 和 z_2 之间的马氏距离的平方为：

$$(z_1 - z_2)^T \Sigma^{-1} (z_1 - z_2) = \begin{pmatrix} z_{11} - z_{12} & z_{21} - z_{22} \end{pmatrix} \Sigma^{-1} \begin{pmatrix} z_{11} - z_{12} \\ z_{21} - z_{22} \end{pmatrix} \tag{3}$$

显然，当 $\Sigma^{-1} = I$ 时，马氏距离等于欧式距离 $(z_1 - z_2)^T \Sigma^{-1} (z_1 - z_2) = (z_{11} - z_{12})^2 + (z_{21} - z_{22})^2$。

2 对 $(x-\mu)^T \Sigma^{-1} (x-\mu)$ 的值进行推导

由于 Σ 为实对称矩阵，那么可以对 Σ 进行特征分解，那么有 $\Sigma = U \Lambda U^T$，并且 $UU^T = U^T U = I$，所以 $U^{-1} = U^T$，$\Lambda = diag(\lambda_i) \quad (i = 1, 2, \cdots, N)$，并且 $U = (u_1, u_2, \cdots, u_p)_{p \times p}$。

$$
\begin{aligned}
\Sigma =& U\Lambda U^T && (4)\\
=&(u_1,u_2,\cdots,u_p)\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p\end{pmatrix}\begin{pmatrix}u_1^T\\ u_2^T\\ \vdots\\ u_p^T\end{pmatrix} && (5)\\
=&(u_1\lambda_1,u_2\lambda_2,\cdots,u_p\lambda_p)\begin{pmatrix}u_1^T\\ u_2^T\\ \vdots\\ u_p^T\end{pmatrix} && (6)\\
=&\sum_{i=1}^{p}u_i\lambda_i u_i^T && (7)
\end{aligned}
$$

而 Σ^{-1} 的求解过程如下所示：

$$
\Sigma^{-1}=(U\Lambda U^T)^{-1}=(U^T)^{-1}\Lambda^{-1}U^{-1}=U\Lambda^{-1}U^T \tag{8}
$$

代入可以解得：

$$
\Sigma^{-1}=\sum_{i=1}^{p}u_i\frac{1}{\lambda_i}u_i^T \tag{9}
$$

那么，

$$
\begin{aligned}
(x-\mu)^T\Sigma^{-1}(x-\mu) =&(x-\mu)^T\sum_{i=1}^{p}u_i\frac{1}{\lambda_i}u_i^T(x-\mu) && (10)\\
=&\sum_{i=1}^{p}(x-\mu)^T u_i\frac{1}{\lambda_i}u_i^T(x-\mu) && (11)
\end{aligned}
$$

令 $y_i=(x-\mu)^Tu_i$，这是一个典型的投影算法，其中 u_i 是 Σ 的特征值为 λ_i 的特征向量，那么

$$
(x-\mu)^T\Sigma^{-1}(x-\mu)=\sum_{i=1}^{p}y_i\frac{1}{\lambda_i}y_i^T=\sum_{i=1}^{p}\frac{y_i^2}{\lambda_i} \tag{12}
$$

3 高斯分布的几何意义

如何令 $p=2$，则有 $\Delta=\frac{y_1^2}{\lambda_1}+\frac{y_2^2}{\lambda_2}$，这实际上就是一个椭圆，如果 Δ 取不同的值，就会像等高线一样，一圈圈的环绕，因为每一个 Δ 都对应着一个二次型，即对应着概率密度函数的一个取值，同样的 Δ 取值对应着同样的概率密度。那么最终的概率密度等高线表示为：

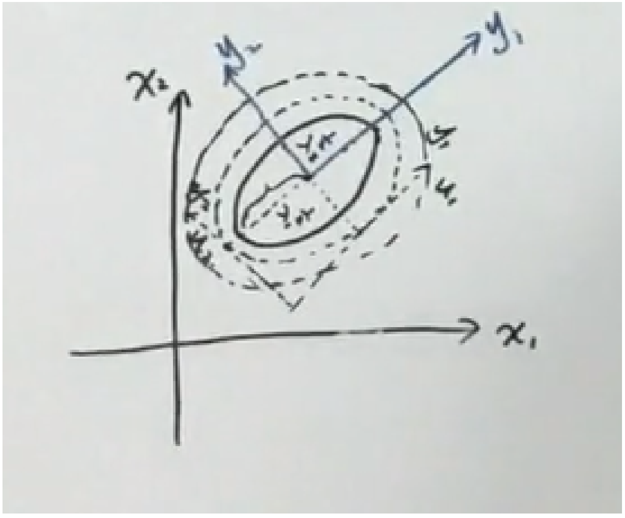

图 1: 二维高斯分布的可视化表示图

4 高斯分布中遇到的困难

4.1 维度灾难 (curse of dimension)

由于 Σ 是一个 $p \times p$ 的矩阵，矩阵中一共有 $\frac{p(p+1)}{2}$ 个参数，算法的复杂度为 $O(p^2)$。一旦输入维度过大，这个矩阵的计算会变得很复杂。所以，在某些时候，将 Σ 矩阵简化成对角矩阵将算法复杂度降低到 $O(p)$。进一步简化，可以令 $\lambda_1 = \lambda_2 = \cdots = \lambda_p$。这时，高斯分布的可视化表示即为一个中心在原点的同心圆。这样的高斯分布，被我们称为“各向同性”的高斯分布。

4.2 高斯分布的表达的局限性

很多时候，高斯分布的表达有局限性，这时，学者提出了混合高斯模型 (GMM) 来拟合现实情况中复杂多样的分布。具体有关于混合高斯模型 (GMM) 的内容将在后续部分进行详细解释，