Abstract

本期我们将从概率视角观察多元高斯分布。

Prior Knowledge

$$
\begin{aligned}
x &\sim N(\mu,\sigma^2) && (1)\\
\mu \in R^p&, \sigma \in R^p && (2)\\
x_i &\sim N(\mu_i,\sigma_i) && (3)\\
p(x_i) &= \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right) && (4)
\end{aligned}
$$

Derivation

首先我们假设每个 x_i 之间是 $iid(independent \quad identically \quad distribution)$ 独立同分布的。

即:

$$
\begin{aligned}
p(x) &= \prod_{i=1}^{p} p(x_i)\\
&= \frac{1}{(2\pi)^{\frac{p}{2}}\prod_{i=1}^{p}\sigma_i}\exp\left(-\frac{1}{2}\sum_{i=1}^{p}(\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right))\\
&= \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}}\exp[-\frac{1}{2}(x_1-\mu_1 \quad x_2-\mu_2 \quad \dots \quad x_p-\mu_p)\begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & \dots & 0 & \frac{1}{\sigma_p^2}\end{pmatrix}\begin{pmatrix} x_1-\mu_1\\ .\\ .\\ x_p-\mu_p\end{pmatrix}]\\
&= \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}}\exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)]
\end{aligned} \tag{5}
$$

以上为多元高斯分布的概率密度函数。

而我们知道 Σ 为半正定矩阵，因此可以进行奇异值分解。所以我们有:

$$
\begin{aligned}
\Sigma &= UVU^T\\
&= (u_1 \quad \dots \quad u_p)\begin{pmatrix} \lambda_1 & 0 & \dots & 0\\ \dots & 0 & \dots & \dots\\ 0 & \dots & \dots & \lambda_p\end{pmatrix}\begin{pmatrix} u_1^T\\ .\\ .\\ u_p^T\end{pmatrix}\\
&= (u_1\lambda_1 \quad \dots \quad u_p\lambda_p)\begin{pmatrix} u_1^T\\ .\\ .\\ u_p^T\end{pmatrix}\\
&= \sum_{i=1}^{p} u_i\lambda_i u_i^T
\end{aligned} \tag{6}
$$

因此

$$\begin{aligned}\Sigma^{-1} &= (UVU^T)^{-1} \\ &= (U^T)^{-1}V^{-1}U^{-1} \\ &= UV^{-1}U^T \\ &= \sum_{i=1}^{p} u_i \frac{1}{\lambda_i} u_i^T\end{aligned} \tag{7}$$

下面我们令 $\Delta = (x-\mu)^T\Sigma^{-1}(x-\mu)$

将上面推导的结果代入:

$$\begin{aligned}\Delta &= (x-\mu)^T\Sigma^{-1}(x-\mu) \\ &= (x-\mu)^T\sum_{i=1}^{p} u_i \frac{1}{\lambda_i} u_i^T (x-\mu) \\ &= \sum_{i=1}^{p}(x-\mu)^T u_i \frac{1}{\lambda_i} u_i^T (x-\mu)\end{aligned} \tag{8}$$

下面我们令 $y_i = (x-\mu)^T u_i$

这里 y_i 代表 x 经过中心化后投影到新的正交基 u_i 的坐标值。

所以:

$$\Delta = \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} \tag{9}$$

下面我们再看多元高斯分布的概率密度函数:

$$p(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}}\exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)] \tag{10}$$

可以看到式子里与变量 x 相关的只有指数。前面的系数是为了使概率和为 1。

因此高斯分布的概率与 Δ 的值直接相关。

我们假设 $p=2$ ，即:

$$\frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} = \Delta \tag{11}$$

我们惊喜地发现，这与椭圆方程很像。而 Δ 的值是不固定的，因此对于不同的 x ，这些样本点于平面内形成了一个个同心的椭圆。而这就是高斯分布的性质之一。