11 多元高斯分布

Multivariate Gaussian Distribution

几何、代数、概率统计的完美结合

在我看来，数学科学是一个不可分割的整体，一个有机体，其生命力取决于各部分的联系。

Mathematical science is in my opinion an indivisible whole, an organism whose vitality is conditioned upon the connection of its parts.

—— 大卫·希尔伯特 (David Hilbert) | 德国数学家 | 1862 ~ 1943

- `numpy.cov()` 计算协方差矩阵
- `numpy.diag()` 如果 A 为方阵，`numpy.diag(A)` 函数提取对角线元素，以向量形式输入结果；如果 a 为向量，`numpy.diag(a)` 函数将向量展开成方阵，方阵对角线元素为 a 向量元素
- `numpy.linalg.eig()` 特征值分解
- `numpy.linalg.inv()` 计算逆矩阵
- `numpy.linalg.norm()` 计算范数
- `numpy.linalg.svd()` 奇异值分解
- `scipy.spatial.distance.euclidean()` 计算欧氏距离
- `scipy.spatial.distance.mahalanobis()` 计算马氏距离
- `seaborn.heatmap()` 绘制热图
- `seaborn.kdeplot()` 绘制 KDE 核概率密度估计曲线
- `seaborn.pairplot()` 绘制成对分析图
- `sklearn.decomposition.PCA()` 主成分分析函数

11.1 矩阵角度：一元、二元、三元到多元

一元

本书第 9 章给出了一元高斯分布的 PDF 解析式，具体如下：

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{1}$$

图 1 (a) 所示为一元高斯分布 PDF 的图像。

二元

第 10 章中，我们看到二元高斯分布的 PDF 解析式：

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}}\times\exp\left(\frac{-1}{2}\frac{1}{\left(1-\rho_{X,Y}^2\right)}\underbrace{\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho_{X,Y}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)}_{\text{Ellipse}}\right) \tag{2}$$

图 1 (b) 所示为二元高斯分布 PDF 的图像。

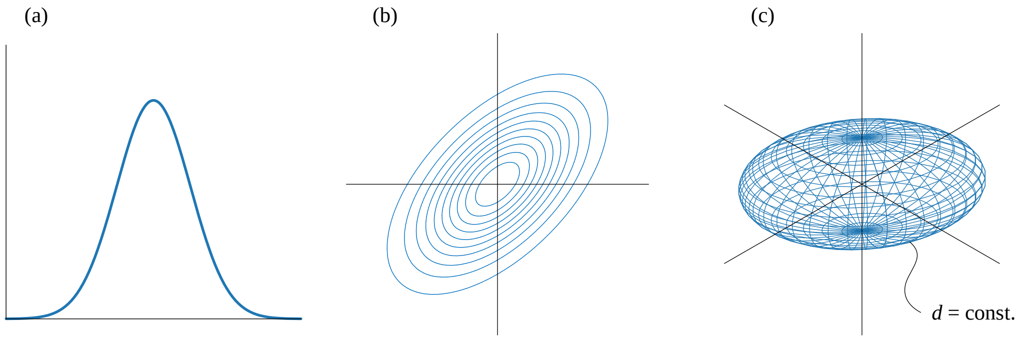

图 1. 一元、二元、三元高斯分布的几何形态

三元

(2) 已经很复杂，我们再看看三元高斯分布 PDF 解析式。在 $\sigma_1=\sigma_2=\sigma_3=1$，$\mu_1=\mu_2=\mu_3=0$ 条件下，三元高斯分布 PDF 解析式如下：

$$
f_{X1,X2,X3}\left(x_1,x_2,x_3\right)=\frac{\exp\left(\frac{-1}{2}d^2\right)}{\left(2\pi\right)^{\frac{3}{2}}\sqrt{1+2\rho_{1,2}\rho_{1,3}\rho_{2,3}-\left(\rho_{1,2}^2+\rho_{1,3}^2+\rho_{2,3}^2\right)}} \tag{3}
$$

其中

$$
d^2=\frac{x_1^2\left(\rho_{2,3}^2-1\right)+x_2^2\left(\rho_{1,3}^2-1\right)+x_3^2\left(\rho_{1,2}^2-1\right)+2\left[x_1x_2\left(\rho_{1,2}-\rho_{1,3}\rho_{2,3}\right)+x_1x_3\left(\rho_{1,3}-\rho_{1,2}\rho_{2,3}\right)+x_2x_3\left(\rho_{2,3}-\rho_{1,3}\rho_{2,3}\right)\right]}{\left(\rho_{1,2}^2+\rho_{1,3}^2+\rho_{2,3}^2-2\rho_{1,2}\rho_{1,3}\rho_{2,3}-1\right)} \tag{4}
$$

当 d 为确定值时，上式代表一个椭球 (ellipsoid)，如所示图 1 (c)。也就是说三元高斯分布 PDF 的几何图形是嵌套的椭球。

相信大家已经看到了三元高斯分布 PDF 解析式的复杂程度。更不用说，(3) 的解析式是在 $\sigma_1 = \sigma_2 = \sigma_3 = 1$，$\mu_1 = \mu_2 = \mu_3 = 0$ 这个极为特殊的条件下。

到了四元、五元、更高元高斯分布 PDF 解析式时，代数展开式已经完全不够用了。因此，对于多元高斯分布，我们需要矩阵算式。

多元

本书读者应该已经很熟悉多元正态分布 PDF，具体如下：

$$
f_{\chi}\left(\boldsymbol{x}\right)=\frac{\exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right)}{\left(2\pi\right)^{\frac{D}{2}}\left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}}} \tag{5}
$$

其中，$\boldsymbol{\chi}$、$\boldsymbol{x}$、$\boldsymbol{\mu}$ 均为列向量：

$$
\boldsymbol{\chi}=\begin{bmatrix}X_1\\X_2\\\vdots\\X_D\end{bmatrix},\quad \boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_D\end{bmatrix},\quad \boldsymbol{\mu}=\begin{bmatrix}\mu_1\\\mu_2\\\vdots\\\mu_D\end{bmatrix} \tag{6}
$$

向量 $\boldsymbol{\mu}$ 常常被称作质心 (centroid)，D 为高斯分布的特征数，比如二元高斯分布 $D = 2$。

协方差矩阵 $\boldsymbol{\Sigma}$ 为：

$$
\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_{1,1}&\sigma_{1,2}&\cdots&\sigma_{1,D}\\\sigma_{2,1}&\sigma_{2,2}&\cdots&\sigma_{2,D}\\\vdots&\vdots&\ddots&\vdots\\\sigma_{D,1}&\sigma_{D,2}&\cdots&\sigma_{D,D}\end{bmatrix}=\begin{bmatrix}\sigma_1^2&\rho_{1,2}\sigma_1\sigma_2&\cdots&\rho_{1,D}\sigma_1\sigma_D\\\rho_{1,2}\sigma_1\sigma_2&\sigma_2^2&\cdots&\rho_{2,D}\sigma_2\sigma_D\\\vdots&\vdots&\ddots&\vdots\\\rho_{1,D}\sigma_1\sigma_D&\rho_{2,D}\sigma_2\sigma_D&\cdots&\sigma_D^2\end{bmatrix} \tag{7}
$$

特别需要大家注意的是，如果 (5) 成立，协方差矩阵 $\boldsymbol{\Sigma}$ 必须为正定矩阵。如果为 $\boldsymbol{\Sigma}$ 半正定，$\boldsymbol{\Sigma}$ 的行列式值为 0，而 (5) 分母不能为 0。$\boldsymbol{\Sigma}$ 半正定说明 $\boldsymbol{\chi}$ 存在线性相关。

一组随机变量构成的列向量 $\boldsymbol{\chi}$ 服从如 (5) 多元高斯分布，记做：

$$\boldsymbol{\chi}=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_D \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_D \end{bmatrix}, \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,D} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{D,1} & \sigma_{D,2} & \cdots & \sigma_{D,D} \end{bmatrix}\right) \tag{8}$$

或更简便地记做：

$$\boldsymbol{\chi} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \tag{9}$$

注意，这个语境下，$\boldsymbol{\chi}$ 为随机变量构成的列向量，每一行代表一个随机变量；而 $\boldsymbol{X}$ 代表数据矩阵，每一列对应一个随机变量的样本值。

多元 → 一元

$D = 1$ 时，质心为：

$$\boldsymbol{\mu} = \left[\mu\right] \tag{10}$$

协方差矩阵为：

$$\boldsymbol{\Sigma} = \left[\sigma^2\right] \tag{11}$$

(5) 分子中的二次式 (quadratic form) 可以展开为：

$$\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right) = \left(x-\mu\right)\sigma^{-2}\left(x-\mu\right) = \left(\frac{x-\mu}{\sigma}\right)^2 \tag{12}$$

我们看到的是 z 分数的平方。这和 (1) 解析式完全一致。

多元 → 二元

再以二元 ($D = 2$) 高斯分布为例，它的质心：

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{13}$$

二元高斯分布的协方差矩阵 $\boldsymbol{\Sigma}$ 具体为：

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{14}$$

协方差矩阵的行列式值 $|\boldsymbol{\Sigma}|$：

$$\left|\boldsymbol{\Sigma}\right| = \sigma_1^2\sigma_2^2 - \rho_{1,2}^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2\left(1-\rho_{1,2}^2\right) \tag{15}$$

再次强调，如果相关性系数为±1，行列式值为 0。相关性系数取值范围为 () 时，协方差矩阵的逆 $\boldsymbol{\Sigma}^{-1}$ 为：

$$\boldsymbol{\Sigma}^{-1}=\frac{1}{\sigma_1^2\sigma_2^2\left(1-\rho_{1,2}^2\right)}\begin{bmatrix}\sigma_2^2 & -\rho_{1,2}\sigma_1\sigma_2\\ -\rho_{1,2}\sigma_1\sigma_2 & \sigma_1^2\end{bmatrix}=\frac{1}{1-\rho_{1,2}^2}\begin{bmatrix}\dfrac{1}{\sigma_1^2} & \dfrac{-\rho_{1,2}}{\sigma_1\sigma_2}\\ \dfrac{-\rho_{1,2}}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2}\end{bmatrix} \tag{16}$$

对于二元高斯分布，(5) 分子中的二次式 (quadratic form) 可以展开写作：

$$\begin{aligned}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)&=\begin{bmatrix}x_1-\mu_1 & x_2-\mu_2\end{bmatrix}\frac{1}{1-\rho_{1,2}^2}\begin{bmatrix}\dfrac{1}{\sigma_1^2} & \dfrac{-\rho_{1,2}}{\sigma_1\sigma_2}\\ \dfrac{-\rho_{1,2}}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2}\end{bmatrix}\begin{bmatrix}x_1-\mu_1\\ x_2-\mu_2\end{bmatrix}\\ &=\frac{1}{1-\rho_{1,2}^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho_{1,2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\end{aligned} \tag{17}$$

分别将 (17) 和 (15) 代入 (5) 可以得到二元高斯分布 PDF 解析式。

随机变量独立

特别地，如果 (X_1, X_2) 服从二元高斯分布，并且随机变量 X_1 和 X_2 独立，这样 X_1 和 X_2 相关性系数 $\rho_{1,2}$ 为 0，协方差矩阵为：

$$\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix} \tag{18}$$

注意，这个协方差矩阵为对角阵。

根据上一章所学，我们知道 X_1 和 X_2 各自的边缘概率密度函数分别为：

$$\begin{aligned}f_{X1}\left(x_1\right)&=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(\frac{-1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)\\ f_{X2}\left(x_2\right)&=\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(\frac{-1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\end{aligned} \tag{19}$$

对应的如果 (X_1, X_2) 服从二元高斯函数，其概率密度函数可以写成两个边缘概率概率密度函数的乘积：

$$\begin{aligned}\underbrace{f_{X1,X2}\left(x_1,x_2\right)}_{\text{Joint}}&=\frac{1}{2\pi\sigma_1\sigma_2}\times\exp\left(\frac{-1}{2}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right)\\ &=\underbrace{\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(\frac{-1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)}_{\text{Marginal, }f_{X1}\left(x_1\right)}\times\underbrace{\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(\frac{-1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)}_{\text{Marginal, }f_{X2}\left(x_2\right)}\end{aligned} \tag{20}$$

这种情况，二元高斯分布 PDF 等高线为正椭圆。

11.2 高斯分布和椭圆

椭圆分布

高斯分布是椭圆分布 (elliptical distribution) 的一种特殊形式。而椭圆分布的 PDF 一般形式为：

$$f(\boldsymbol{x}) = k \cdot g\left[\underbrace{(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}_{\text{Ellipse}}\right] \tag{21}$$

本书第 7 章介绍的学生 t-分布、逻辑分布、拉普拉斯分布也都是椭圆分布家族成员。

对数

求 (5) 对数，得到：

$$\ln f_{\chi}(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}| \tag{22}$$

另外，(5) 可以写成：

$$f_{\chi}(\boldsymbol{x}) = \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c\right) \tag{23}$$

其中

$$\begin{aligned} &\boldsymbol{A} = \boldsymbol{\Sigma}^{-1} \\ &\boldsymbol{b} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ &c = \frac{-1}{2}\left[D \times \ln(2\pi) - \ln|\boldsymbol{A}| + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{b}\right] \end{aligned} \tag{24}$$

令，

$$G(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = d^2 \tag{25}$$

$G(\boldsymbol{x})$ 开平方得到 d 就是马氏距离 (Mahalanobis distance)。

椭圆结构

回顾上一章介绍的二元高斯分布的椭圆结构。如图 2 所示，椭圆中心对应质心 $\boldsymbol{\mu}$，椭圆和$\pm\sigma$标准差构成的正方形相切，四个切点分别为 A、B、C 和 D，对角切点两两相连得到两条直线 AC、BD。

AC 相当于在给定 X_2 条件下 X_1 的条件概率期望值；BD 相当于在给定 X_1 条件下 X_2 的条件概率期望值，这是本书第 13 章要讨论的话题。

在椭圆的学习中，我们很关注椭圆的长轴、短轴，对应图 2 中两条红线 EG、FH。EG 通过椭圆圆心 O 最长的线段，为椭圆长轴；FH 通过椭圆中心 O 最短的线段，为椭圆短轴。获得长轴、短轴的长度、角度需要用到特征值分解，这是本章后续要讨论的内容。而长轴就是主成分分析的第一主元方向，这是本书第 14、25 章要讨论的话题。

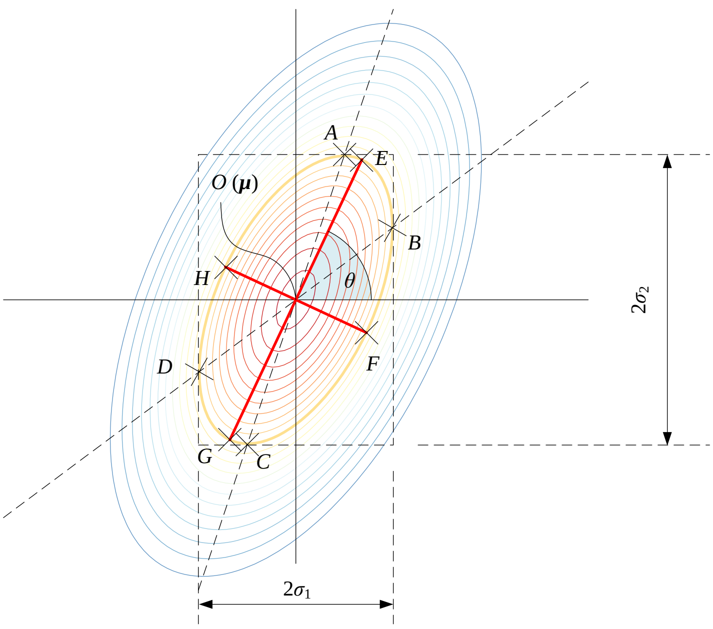

图 2. 椭圆和±σ 标准差长方形的关系

Bk5_Ch23_01.py 绘制图 2。

11.3 解剖多元高斯分布 PDF

《矩阵力量》第 20 章介绍过如何用“平移 → 旋转 → 缩放”解剖多元高斯分布，本节把其中重要的内容“抄”了过来。

特征值分解协方差矩阵

协方差矩阵 $\boldsymbol{\Sigma}$ 为对称矩阵，对 $\boldsymbol{\Sigma}$ 特征值分解得到：

$$\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} \tag{26}$$

其中，$\boldsymbol{V}$ 为正交矩阵，即满足 $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}$。上式也是谱分解。

利用 (26) 获得 $\boldsymbol{\Sigma}^{-1}$ 的特征值分解：

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{\mathrm{T}} \tag{27}$$

由此，将 $(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ 拆成 $\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu})$ 的“平方”：

$$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) = \left[\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{\mathrm{T}}\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) = \left\|\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu})\right\|_2^2 \tag{28}$$

平移 → 旋转 → 缩放

(28) 的几何解释是，旋转椭圆通过“平移 ($\boldsymbol{x}-\boldsymbol{\mu}$) → 旋转 ($\boldsymbol{V}^{\mathrm{T}}$) → 缩放 ($\boldsymbol{\Lambda}^{\frac{-1}{2}}$)”转换成单位圆，具体过程如图 3 所示。

图 3 (a) 中旋转椭圆代表多元高斯分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，随机数质心位于 $\boldsymbol{\mu}$，椭圆形状描述了协方差矩阵 $\boldsymbol{\Sigma}$。图 3 (a) 中散点是服从 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的随机数。

图 3 (a) 中散点经过平移得到 $\boldsymbol{x}_c = \boldsymbol{x} - \boldsymbol{\mu}$，这是一个去均值 (中心化过程)。图 3 (b) 中旋转椭圆代表多元高斯分布 $N(\boldsymbol{0}, \boldsymbol{\Sigma})$。随机数质心也随之平移到原点。

图 3 (b) 中椭圆旋转之后得到图 3 (c) 中正椭圆，对应：

$$\boldsymbol{y} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}_c = \boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) \tag{29}$$

协方差矩阵 $\boldsymbol{\Sigma}$ 通过特征值分解得到特征值矩阵 $\boldsymbol{\Lambda}$。而正椭圆的半长轴、半短轴长度蕴含在特征值矩阵 $\boldsymbol{\Lambda}$ 中，这算是拨开云雾的过程。图 3 (c) 中随机数服从 $N(\boldsymbol{0}, \boldsymbol{\Lambda})$。

最后一步是缩放，从图 3 (c) 到图 3 (d)，对应：

$$\boldsymbol{z} = \boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{y} = \boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) \tag{30}$$

图 3 (d) 中单位圆则代表多元标准分布 $N(\boldsymbol{0}, \boldsymbol{I})$。这意味着满足 $N(\boldsymbol{0}, \boldsymbol{I})$ 的随机变量为独立同分布。独立同分布 (Independent and identically distributed, IID) 是指一组随机变量中每个变量的概率分布都相同，且这些随机变量互相独立。

利用向量 $\boldsymbol{z}$，多元高斯分布 PDF 可以写成：

$$f_{\chi}(\boldsymbol{x}) = \frac{\exp\left(-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}\right)}{(2\pi)^{\frac{D}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} = \frac{\exp\left(-\frac{1}{2}\|\boldsymbol{z}\|_2^2\right)}{(2\pi)^{\frac{D}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \tag{31}$$

$\boldsymbol{z}$ 的模 $\|\boldsymbol{z}\|$ 实际上代表“整体”z 分数。

缩放 → 旋转 → 平移

反向来看，$x = V\Lambda^{\frac{1}{2}}z + \mu$ 代表通过“缩放 → 旋转 → 平移”把单位圆转换成中心在 μ 的旋转椭圆。也就是把 $N(\mathbf{0}, \mathbf{I})$ 转换成 $N(\mu, \Sigma)$。从数据角度来看，我们可以通过“缩放 → 旋转 → 平移”，把服从 $N(\mathbf{0}, \mathbf{I})$ 的随机数转化为服从 $N(\mu, \Sigma)$ 的随机数。

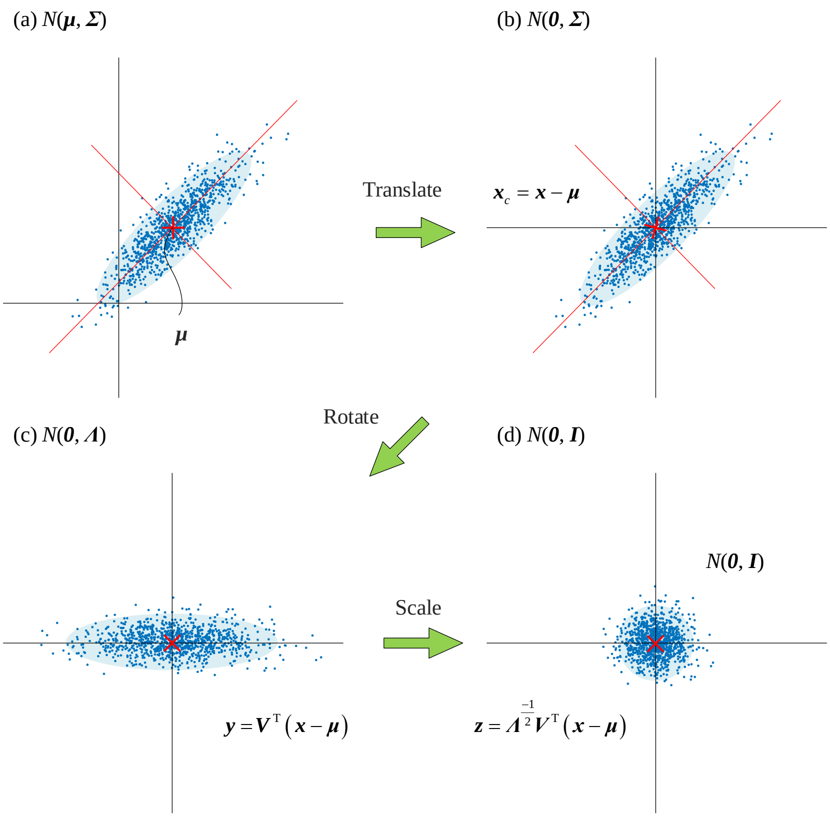

图 3. 平移 → 旋转 → 缩放，图片来自《矩阵力量》

马氏距离

马氏距离可以写成：

$$d = \sqrt{(x-\mu)^{T}\Sigma^{-1}(x-\mu)} = \left\| \Lambda^{\frac{-1}{2}}V^{T}(x-\mu) \right\| = \|z\| \tag{32}$$

马氏距离的独特之处在于，它通过引入协方差矩阵在计算距离时考虑了数据的分布。此外，马氏距离**无量纲量** (unitless 或 dimensionless)，它将各个特征数据标准化。本书第 23 章将专门讲解马氏距离及其应用。

高斯函数

将 (32) 中马氏距离 d 代入多元高斯分布概率密度函数，得到：

$$f_{\chi}(\boldsymbol{x})=\frac{\exp\left(-\frac{1}{2}d^{2}\right)}{(2\pi)^{\frac{D}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \tag{33}$$

上式，我们看到高斯函数 exp(−1/2 ●) 把“距离度量”转化成“亲近度”。图 4 所示为马氏距离图像。大家可以发现这个曲面为开口朝上的锥面，等高线为旋转椭圆。一般来说，离质心 $\boldsymbol{\mu}$ 越远，马氏距离相对较大。但是，这也不是绝对的。

图 4 (b) 中白色虚线正圆代表距离质心 $\boldsymbol{\mu}$ 欧氏距离为 1 的等高线。欧氏距离是最自然的距离度量。而马氏距离则引入协方差矩阵 $\boldsymbol{\Sigma}$，计算距离时考虑数据的分布情况。本书第 23 章将区分欧氏距离和马氏距离。

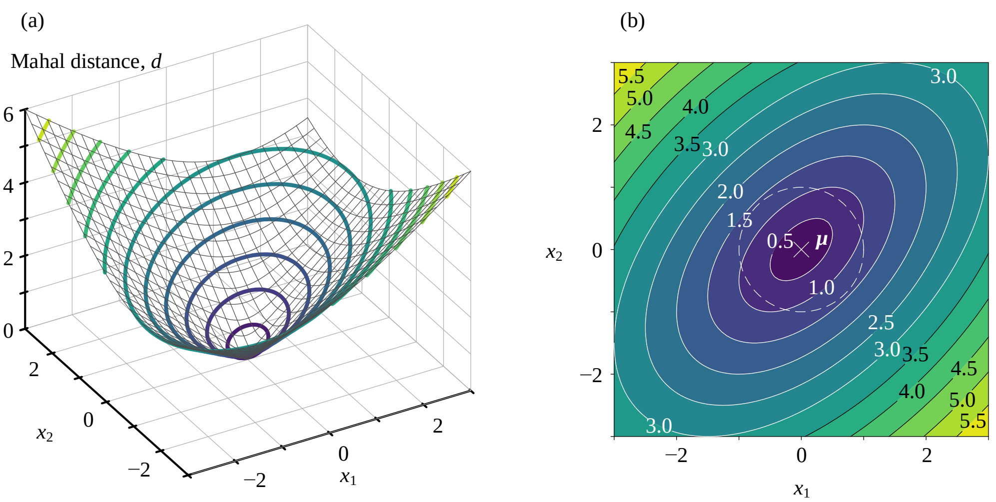

图 4. 马氏距离椭圆等高线

将具体马氏距离 d 值代入上式，可以得到高斯概率密度值。也就是说，图 4 每一个椭圆都对应一个概率密度值。这就是图 5 中等高线的含义。

本书中代表高斯分布会用到这两种不同的可视化方案。请大家注意区分，等高线代表马氏距离，还是概率密度值。

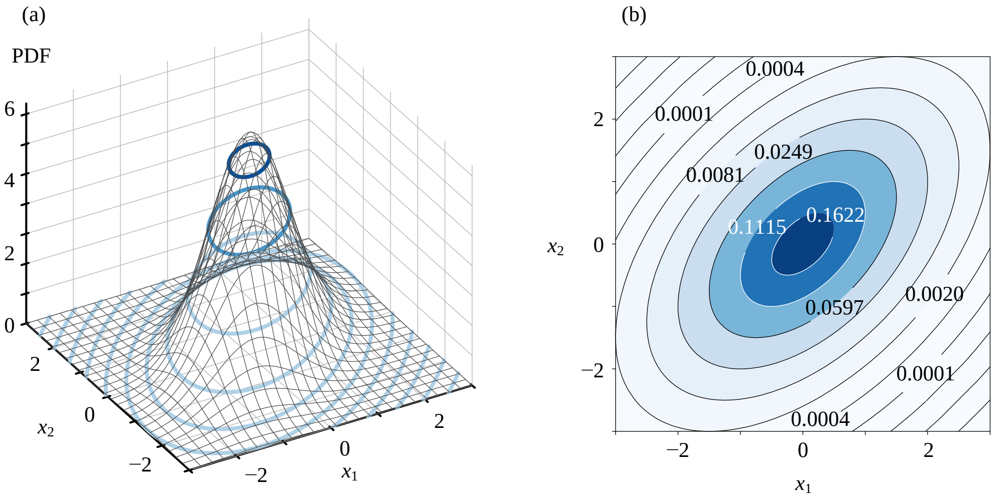

图 5. 高斯放分布 PDF 椭圆等高线

分母：行列式值

把 $|\boldsymbol{\Sigma}|^{\frac{1}{2}}$ 从 (5) 分母移到分子可以写成 $|\boldsymbol{\Sigma}|^{\frac{-1}{2}}$ 。而 $\boldsymbol{\Sigma}^{\frac{-1}{2}}$ 相当于：

$$\boldsymbol{\Sigma}^{\frac{-1}{2}} \sim \boldsymbol{\Lambda}^{\frac{-1}{2}} \boldsymbol{V}^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}) \tag{34}$$

从体积角度来看，“平移 → 旋转 → 缩放”几何变换带来的面积/体积缩放系数便是 $|\boldsymbol{\Sigma}|^{\frac{-1}{2}}$ 。准确来说，只有“缩放”才影响面积/体积，因此 $\|\boldsymbol{\Sigma}\|^{\frac{-1}{2}} = \|\boldsymbol{\Lambda}\|^{\frac{-1}{2}}$ 。

分母：体积归一化

从几何角度来看，(5) 分母中 $(2\pi)^{\frac{D}{2}}$ 一项起到归一化作用，为了保证概率密度函数曲面和整个水平面包裹的体积为 1，即概率为 1。

11.4 平移 → 旋转

本节以二元高斯分布 PDF 为例，利用特征值分解这个工具进一步深入理解多元高斯分布。

特征值分解

形状为 2×2 协方差矩阵 $\boldsymbol{\Sigma}$，它的特征值和特征向量关系为：

$$\begin{cases} \boldsymbol{\Sigma} \boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1 \\ \boldsymbol{\Sigma} \boldsymbol{v}_2 = \lambda_2 \boldsymbol{v}_2 \end{cases} \tag{35}$$

(35) 可以写成：

$$\boldsymbol{\Sigma} \underbrace{\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix}}_{\boldsymbol{V}} = \underbrace{\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix}}_{\boldsymbol{V}} \underbrace{\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}}_{\boldsymbol{\Lambda}} \tag{36}$$

即，

$$\boldsymbol{\Sigma V} = \boldsymbol{V \Lambda} \tag{37}$$

将 $\boldsymbol{\Sigma}$ 具体值代入 (35) 得到：两个特征值对应的特征向量如下：

$$\begin{aligned} \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1 \\ \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \boldsymbol{v}_2 = \lambda_2 \boldsymbol{v}_2 \end{aligned} \tag{38}$$

两个特征值可以通过下式求得：

$$\begin{aligned} \lambda_1 = \frac{\sigma_1^2 + \sigma_2^2}{2} + \sqrt{\left(\rho_{1,2}\sigma_1\sigma_2\right)^2 + \left(\frac{\sigma_1^2 - \sigma_2^2}{2}\right)^2} \\ \lambda_2 = \frac{\sigma_1^2 + \sigma_2^2}{2} - \sqrt{\left(\rho_{1,2}\sigma_1\sigma_2\right)^2 + \left(\frac{\sigma_1^2 - \sigma_2^2}{2}\right)^2} \end{aligned} \tag{39}$$

只有当 $\rho_{1,2} = 0$ 且 $\sigma_1 = \sigma_2$ 时，(39) 中两个特征值相等。这种条件下，概率密度等高线为正圆。

长轴、短轴

大家已经清楚，二元高斯分布的 PDF 函数平面等高线而椭圆。如图 6 所示，$\sqrt{\lambda_1}$ 就是椭圆半长轴长度，$\sqrt{\lambda_2}$ 就是半短轴长度：

$$\begin{aligned} EO = GO = \sqrt{\lambda_1} = \sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{2} + \sqrt{\left(\rho_{X,Y}\sigma_X\sigma_Y\right)^2 + \left(\frac{\sigma_X^2 - \sigma_Y^2}{2}\right)^2}} \\ FO = HO = \sqrt{\lambda_2} = \sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{2} - \sqrt{\left(\rho_{X,Y}\sigma_X\sigma_Y\right)^2 + \left(\frac{\sigma_X^2 - \sigma_Y^2}{2}\right)^2}} \end{aligned} \tag{40}$$

图 6 中，$\boldsymbol{v}_1$ 对应的就是椭圆半长轴方向，$\boldsymbol{v}_2$ 对应半短轴方向。在主成分分析中，$\boldsymbol{v}_1$ 就是第一主元方向。$\boldsymbol{v}_2$ 便是第二主元方向。

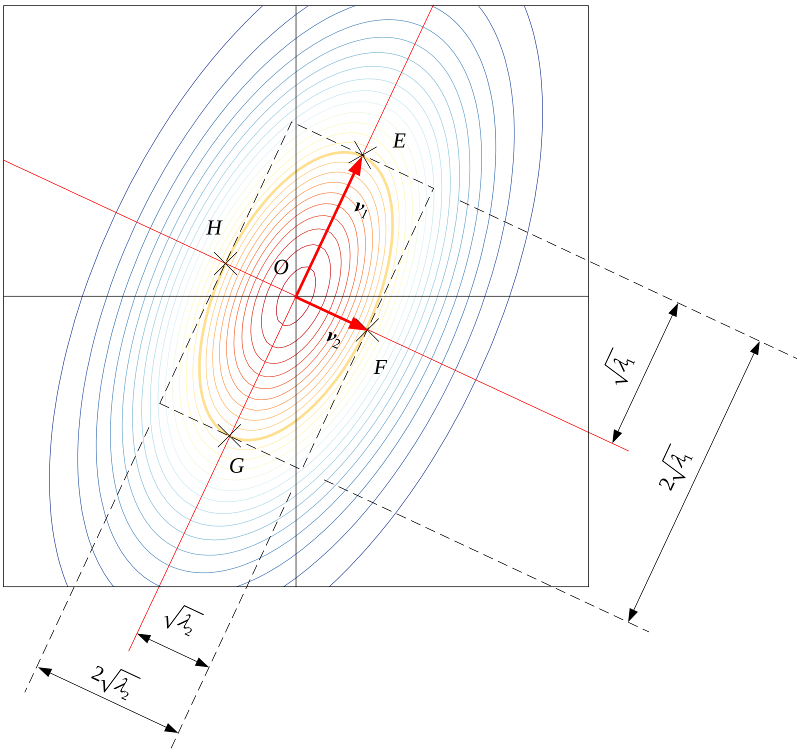

图 6. 椭圆的长轴、短轴

实际上，将 (X_1, X_2) 投影到 $\boldsymbol{v}_1$ 得到的随机变量的方差就是 λ_1，对应的标准差为 $\sqrt{\lambda_1}$。将 (X_1, X_2) 投影到 $\boldsymbol{v}_2$ 得到的随机变量的方差为 λ_2，其标准差为 $\sqrt{\lambda_2}$。

$\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 具体值为：

$$
\begin{aligned}
\boldsymbol{v}_1 &= \begin{bmatrix} \dfrac{\dfrac{\sigma_1^2-\sigma_2^2}{2} - \sqrt{\left(\rho_{1,2}\sigma_1\sigma_2\right)^2 + \left(\dfrac{\sigma_1^2-\sigma_2^2}{2}\right)^2}}{\rho_{1,2}\sigma_1\sigma_2} \\ 1 \end{bmatrix} \\
\boldsymbol{v}_2 &= \begin{bmatrix} \dfrac{\dfrac{\sigma_1^2-\sigma_2^2}{2} + \sqrt{\left(\rho_{1,2}\sigma_1\sigma_2\right)^2 + \left(\dfrac{\sigma_1^2-\sigma_2^2}{2}\right)^2}}{\rho_{1,2}\sigma_1\sigma_2} \\ 1 \end{bmatrix}
\end{aligned}
\tag{41}
$$

Bk5_Ch11_01.py 绘制图 6。

随机变量的线性变换

从另外一个角度来看，如图 7 所示，某个满足二元高斯分布随机变量 (X_1, X_2) 朝若干方向投影。我们先给出结论，这些方向中，向 $\boldsymbol{v}_1$ 投影得到的随机变量方差最大，向 $\boldsymbol{v}_2$ 投影得到的随机变量方差最小。

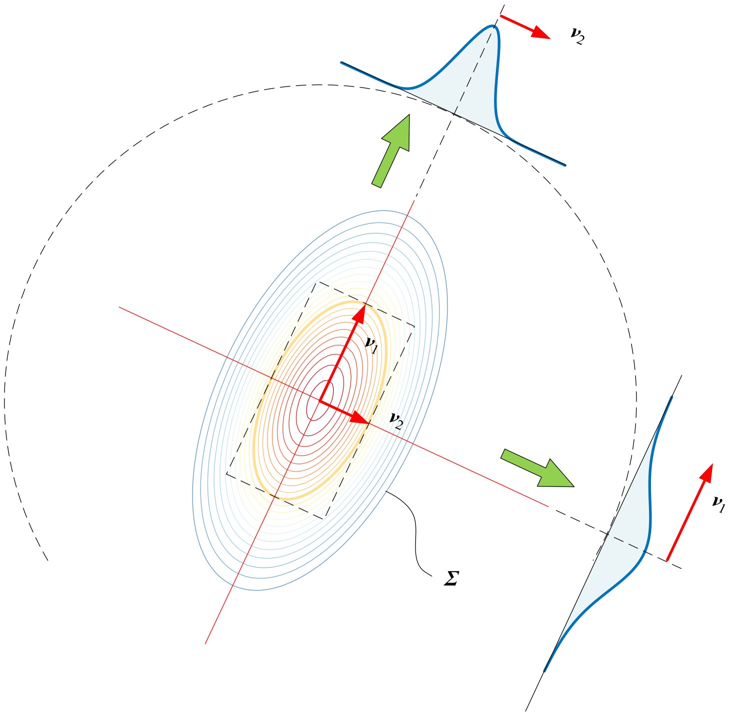

图 7. 二元高斯分布朝不同方向投影

假设二元随机变量列向量 $\boldsymbol{\chi} = [X_1, X_2]^{\mathrm{T}}$ 是满足图 7 这个二元高斯分布，$\boldsymbol{\chi}$ 先中心化，再向 $\boldsymbol{v}_1$ 投影得到 Y_1：

$$Y_1 = \left(\boldsymbol{\chi} - \boldsymbol{\mu}_{\chi}\right)^{\mathrm{T}} \boldsymbol{v}_1 = \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\right)^{\mathrm{T}} \begin{bmatrix} v_{1,1} \\ v_{2,1} \end{bmatrix} = \left(X_1 - \mu_1\right) v_{1,1} + \left(X_2 - \mu_2\right) v_{2,1} \tag{42}$$

从数据角度，上述过程如图 8 所示。

对 Y_1 求方差：

$$\begin{aligned}\operatorname{var}(Y_1) &= \mathrm{E}\left[\left(Y_1-\mu_{Y1}\right)^2\right] = \mathrm{E}\left[\left(\left(\chi-\mu_\chi\right)^{\mathrm{T}} v_1\right)^{\mathrm{T}}\left(\chi-\mu_\chi\right)^{\mathrm{T}} v_1\right] \\ &= \boldsymbol{v}_1^{\mathrm{T}} \overbrace{\mathrm{E}\left[\left(\left(\chi-\boldsymbol{\mu}_\chi\right)^{\mathrm{T}}\right)\left(\chi-\boldsymbol{\mu}_\chi\right)^{\mathrm{T}}\right]}^{\Sigma_\chi} \boldsymbol{v}_1 \\ &= \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{\Sigma}_\chi \boldsymbol{v}_1\end{aligned} \tag{43}$$

因为 Y_1 已经中心化，所以上式中 $\mu_{Y1} = 0$。

将 $\boldsymbol{\Sigma}_\chi$ 的特征值分解代入 (43) 得到：

$$\begin{aligned}\operatorname{var}(Y_1) = \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{\Sigma}_\chi \boldsymbol{v}_1 = \boldsymbol{v}_1^{\mathrm{T}}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}}\end{bmatrix}\boldsymbol{v}_1 \\ = \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}\begin{bmatrix}1 \\ 0\end{bmatrix} = \lambda_1\end{aligned} \tag{44}$$

实际上就是随机变量的线性变换，我们将会在本书第 14 章继续这一话题。

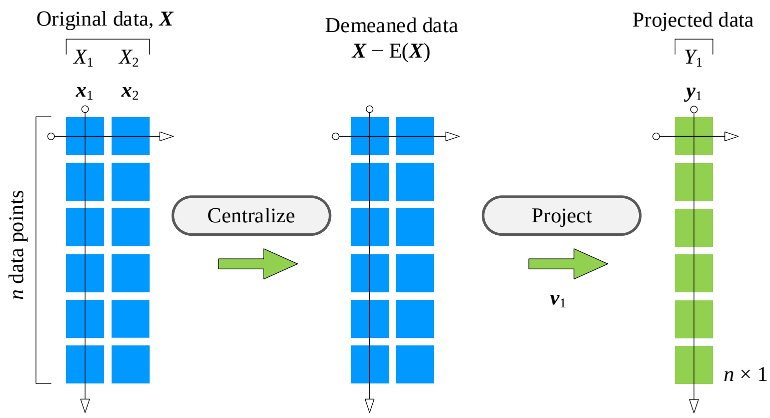

图 8. $\boldsymbol{X}$ 先中心化，再向 $\boldsymbol{v}_1$ 投影得到 $\boldsymbol{y}_1$

椭圆旋转

椭圆旋转角度 θ：

$$\theta = \frac{1}{2}\arctan\left(\frac{2\rho_{1,2}\sigma_1\sigma_2}{\sigma_1^2-\sigma_2^2}\right) \tag{45}$$

图 9 所示为在 σ_X、σ_Y 大小不同，ρ 取值不同对椭圆旋转的影响。

观察 (45)，发现椭圆的旋转角度和 σ_X、σ_Y、$\rho_{X,Y}$ 有关。

特别地，当 $\sigma_X = \sigma_Y$ 时，如果 $\rho_{X,Y}$ 为小于 1 的正数，椭圆的旋转角度为 45°；如果 $\rho_{X,Y}$ 为大于−1 的负数，椭圆的旋转角度为−45°。

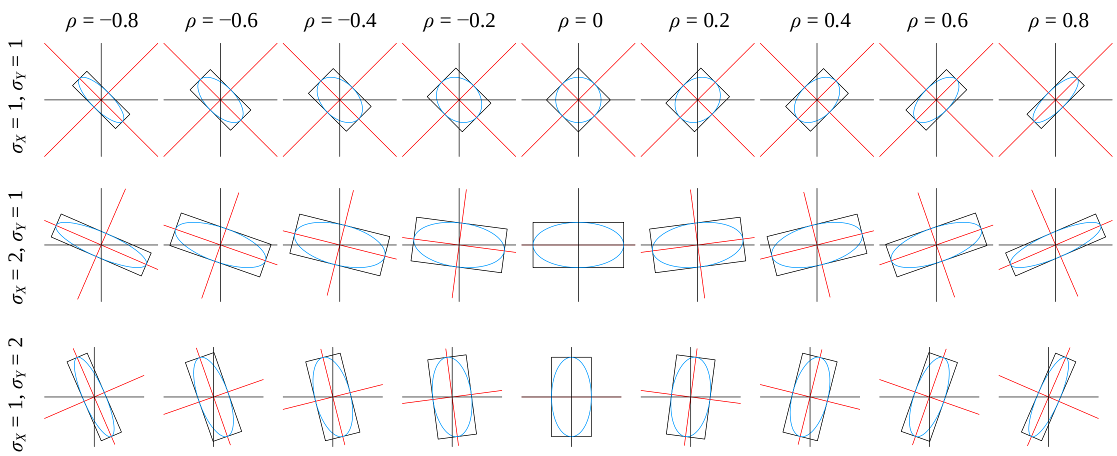

图 9. 在 σ_X、σ_Y 大小不同，ρ 取值不同对椭圆旋转的影响

Bk5_Ch23_02.py 绘制图 9。

特征值之和

可以发现 (39) 中两个特征值之和，等于协方差矩阵 $\boldsymbol{\Sigma}$ 的两个方差之和：

$$\lambda_1 + \lambda_2 = \sigma_1^2 + \sigma_2^2 \tag{46}$$

这正是《矩阵力量》讲到的特征值分解中，原矩阵迹等于特征值矩阵的迹。建议大家回顾特征值分解的优化视角。

特征值之积

两个特征值乘积为：

$$\begin{aligned} \lambda_1 \lambda_2 &= \left(\frac{\sigma_1^2 + \sigma_2^2}{2} \right)^2 - \left(\left(\rho_{1,2} \sigma_1 \sigma_2 \right)^2 + \left(\frac{\sigma_1^2 - \sigma_2^2}{2} \right)^2 \right) \\ &= \sigma_1^2 \sigma_2^2 - \rho_{1,2}^2 \sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 \left(1 - \rho_{1,2}^2 \right) \end{aligned} \tag{47}$$

这和协方差矩阵 $\boldsymbol{\Sigma}$ 行列式值相等：

$$\left| \boldsymbol{\Sigma} \right| = \sigma_1^2 \sigma_2^2 - \rho_{1,2}^2 \sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 \left(1 - \rho_{1,2}^2 \right) \tag{48}$$

谱分解

$\boldsymbol{\Sigma}$ 对称矩阵，因此：

$$\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{I} \tag{49}$$

从而，$\boldsymbol{\Sigma}$ 的谱分解可以进一步写成：

$$\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \end{bmatrix} = \lambda_1 \boldsymbol{v}_1 \boldsymbol{v}_1^{\mathrm{T}} + \lambda_2 \boldsymbol{v}_2 \boldsymbol{v}_2^{\mathrm{T}} \tag{50}$$

本书下一章还会继续这一话题。

平移 → 旋转

令

$$\boldsymbol{y} = \boldsymbol{V}^{\mathrm{T}}\left(\boldsymbol{x} - \boldsymbol{\mu}\right) \tag{51}$$

发现上式 $\boldsymbol{V}^{\mathrm{T}}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)$ 相当于 $\boldsymbol{x}$ 经过平移 $(\boldsymbol{x} - \boldsymbol{\mu})$、旋转 $(\boldsymbol{V}^{\mathrm{T}})$ 两步操作得到 $\boldsymbol{y}$。整个过程如图 10 所示。

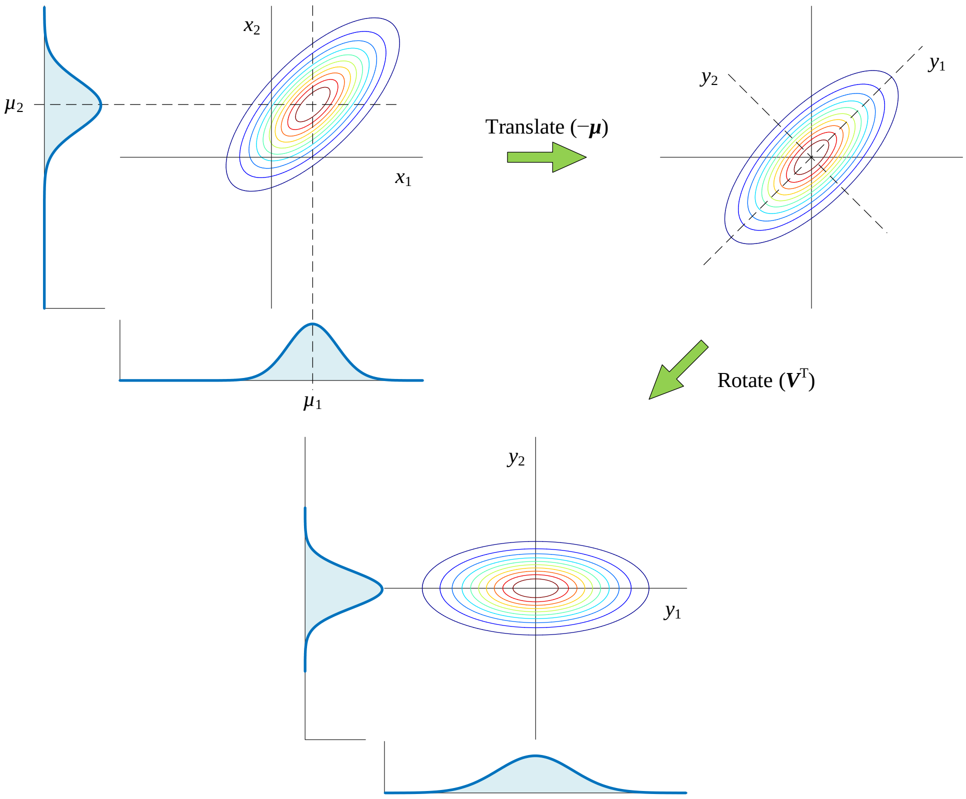

图 10. 椭圆先平移再旋转

将 (51) 代入 (57)，得到：

$$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{y}=\begin{bmatrix} y_1 & y_2 & \cdots & y_q \end{bmatrix}\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_q \end{bmatrix}^{-1}\begin{bmatrix} y_1 & y_2 & \cdots & y_q \end{bmatrix}^{\mathrm{T}}=\sum_{j=1}^{D}\frac{y_j^2}{\lambda_j} \tag{52}$$

其中，$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_D$。上式代表着一个多维空间正椭球体。

平移 $(\boldsymbol{x}-\boldsymbol{\mu})$、旋转 $(\boldsymbol{V}^{\mathrm{T}})$ 两步两步几何变换只改变椭球的空间位置和旋转角度，不改变椭球本身的几何尺寸。也就是说，$|\boldsymbol{\Sigma}|=|\boldsymbol{\Lambda}|$。

特别地，当 $D=2$ 时，令 $(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ 为 1，(52) 可以写成平面椭圆：

$$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=\frac{y_1^2}{\lambda_1}+\frac{y_2^2}{\lambda_2}=1 \tag{53}$$

显然这个椭圆中心位于原点，同样这就解释了为什么图 6 中椭圆的半长轴为 $\sqrt{\lambda_1}$，半短轴为 $\sqrt{\lambda_2}$。

反过来，$\boldsymbol{y}$ 先经过旋转、再平移得到 $\boldsymbol{x}$：

$$\boldsymbol{x}=\boldsymbol{V}\boldsymbol{y}+\boldsymbol{\mu} \tag{54}$$

独立

二元随机变量 (Y_1, Y_2) 对应的二元高斯分布 PDF 为：

$$\begin{aligned} f_{Y1,Y2}(y_1,y_2)&=\frac{1}{2\pi\sqrt{\lambda_1\lambda_2}}\times\exp\left(\frac{-1}{2}\left(\frac{y_1^2}{\lambda_1}+\frac{y_2^2}{\lambda_1}\right)\right)\\ &=\underbrace{\frac{1}{\sqrt{2\pi}\sqrt{\lambda_1}}\exp\left(\frac{-1}{2}\frac{y_1^2}{\lambda_1}\right)}_{f_{Y1}(y_1)}\times\underbrace{\frac{1}{\sqrt{2\pi}\sqrt{\lambda_2}}\exp\left(\frac{-1}{2}\frac{y_2^2}{\lambda_1}\right)}_{f_{Y2}(y_2)} \end{aligned} \tag{55}$$

可以发现随机变量 Y_1 和 Y_2 独立。如图 10 所示，随机变量 Y_1 对应的方差为 λ_1，标准差为 $\sqrt{\lambda_1}$；随机变量 Y_2 对应的方差为 λ_2，标准差为 $\sqrt{\lambda_2}$。

11.5 平移 → 旋转 → 缩放

利用 $\boldsymbol{\Sigma}$ 的谱分解，$\boldsymbol{\Sigma}$ 求逆为：

$$\boldsymbol{\Sigma}^{-1} = \left(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\right)^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{\mathrm{T}} \tag{56}$$

将 (56) 代入 (25)，整理得到：

$$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \left[\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{\mathrm{T}}\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{\Lambda}^{\frac{-1}{2}}\left[\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu})\right] = \left(\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu})\right)^{2} \tag{57}$$

这就是前文讲到的“开方”。

令：

$$\boldsymbol{z} = \boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) \tag{58}$$

上式相当于 $\boldsymbol{x}$ 经过平移、旋转和缩放，最后得到 $\boldsymbol{z}$，整个过程如图 11 所示。

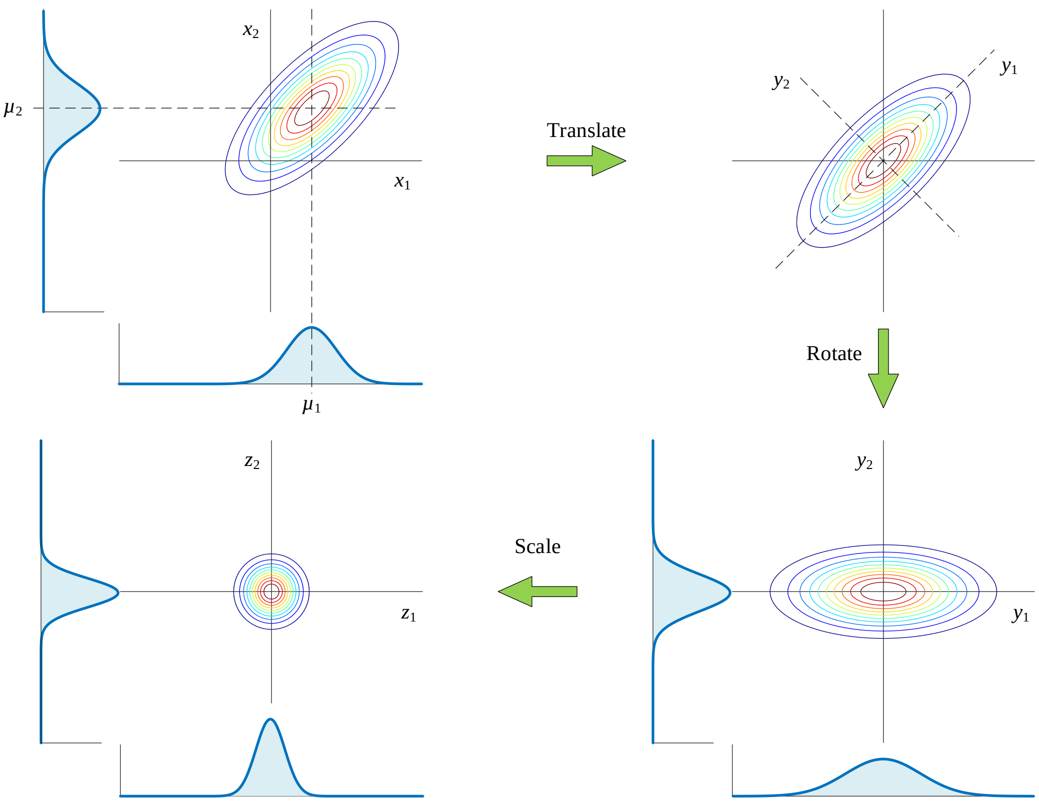

图 11. 椭圆先平移、再旋转，最后缩放，得到单位圆

单位球体

将 (58) 代入 (57)，得到的解析式是多维空间的单位球体：

$$\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)=\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}=z_1^2+z_2^2+\cdots z_D^2=\sum_{j=1}^{D}z_j^2 \tag{59}$$

反过来，也可以利用 $\boldsymbol{z}$ 通过缩放、旋转、平移，反求 $\boldsymbol{x}$：

$$\boldsymbol{x}=\underbrace{\boldsymbol{V}}_{\text{Rotate}}\underbrace{\boldsymbol{D}}_{\text{Scale}}\boldsymbol{z}+\underbrace{\boldsymbol{\mu}}_{\text{Translate}} \tag{60}$$

图 12 展示 (60) 对应的几何变换。

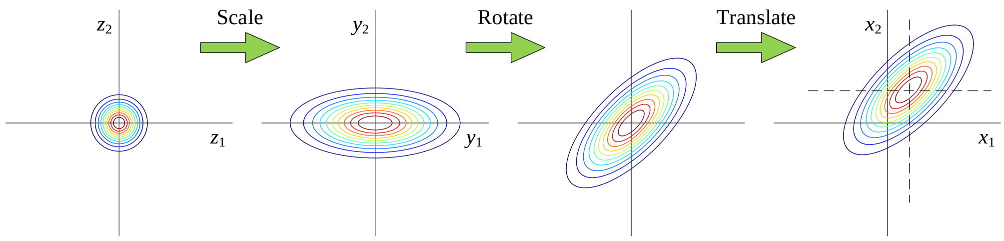

图 12. 单位圆先缩放，再旋转，最后平移

数据视角

类似 (42)，从数据角度来看，如果数据矩阵 $\boldsymbol{X}$ 服从 $N(\mathrm{E}(\boldsymbol{X}), \boldsymbol{\Sigma}_X)$。对 $\boldsymbol{X}$ 先中心化，再向 $\boldsymbol{V}$ 投影，最后缩放得到 $\boldsymbol{Z}$：

$$\boldsymbol{Z}=\left(\boldsymbol{X}-\mathrm{E}\left(\boldsymbol{X}\right)\right)\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}} \tag{61}$$

得到 $\boldsymbol{Z}$ 的协方差矩阵为单位矩阵 $\boldsymbol{I}$：

$$\begin{aligned}\boldsymbol{\Sigma}_Z&=\frac{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}}{n-1}=\frac{\left(\left(\boldsymbol{X}-\mathrm{E}\left(\boldsymbol{X}\right)\right)\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}}\right)^{\mathrm{T}}\left(\left(\boldsymbol{X}-\mathrm{E}\left(\boldsymbol{X}\right)\right)\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}}\right)}{n-1}\\&=\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}\overbrace{\frac{\left(\boldsymbol{X}-\mathrm{E}\left(\boldsymbol{X}\right)\right)^{\mathrm{T}}\left(\boldsymbol{X}-\mathrm{E}\left(\boldsymbol{X}\right)\right)}{n-1}}^{\boldsymbol{\Sigma}_X}\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}}\\&=\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{\Sigma}_X\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{-1}{2}}=\boldsymbol{I}\end{aligned} \tag{62}$$

也就是，如果 $\boldsymbol{X}$ 服从多维高斯分布的话，$\boldsymbol{Z}$ 的每一个维度都服从 IID 标准正态分布。