Dive into Covariance Matrix

13 协方差矩阵

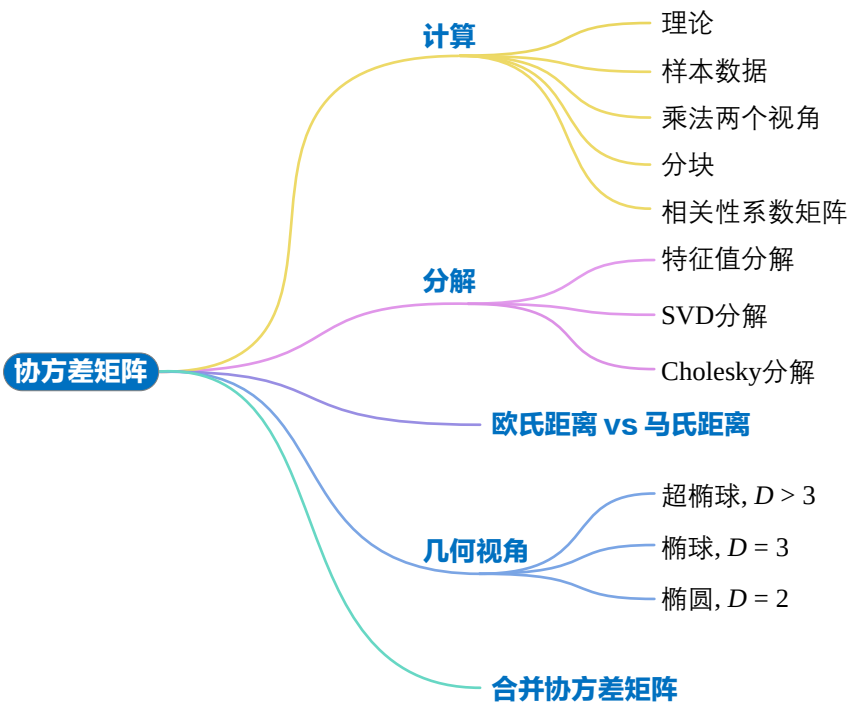
很多数学科学、机器学习算法的起点

科学的目标是寻求对复杂事实的最简单的解释。我们很容易误以为事实很简单，因为简单是我们追求的目标。每个自然哲学家生活中的指导格言都应该是——寻求简单而不相信它。

The aim of science is to seek the simplest explanations of complex facts. We are apt to fall into the error of thinking that the facts are simple because simplicity is the goal of our quest. The guiding motto in the life of every natural philosopher should be, seek simplicity and distrust it.

—— 阿尔弗雷德·怀特海 (Alfred Whitehead) | 英国数学家、哲学家 | 1861 ~ 1947

- `numpy.average()` 计算平均值
- `numpy.corrcoef()` 计算数据的相关性系数
- `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.cholesky()` Cholesky 分解
- `numpy.linalg.eig()` 特征值分解
- `numpy.linalg.inv()` 矩阵求逆
- `numpy.linalg.norm()` 计算范数
- `numpy.linalg.svd()` 奇异值分解
- `numpy.ones()` 创建全 1 向量或矩阵
- `numpy.sqrt()` 计算平方根



13.1 计算协方差矩阵：描述数据分布

协方差矩阵囊括多特征数据矩阵重要统计描述，在多元高斯分布中协方差矩阵扮演重要角色。不仅如此，数据科学和机器学习方法中随处可见，比如多元高斯分布、随机数生成器、OLS线性回归、主成分分析、正交回归、高斯过程、高斯朴素贝叶斯、高斯判别分析、高斯混合模型等。因此，我们有必要拿一章内容专门讨论协方差矩阵。

本系列丛书介绍的很多数学概念在协方差矩阵处达到完美融合，比如解析几何中的椭圆，概率统计中的高斯分布，线性代数中的线性变换、Cholesky 分解、特征值分解、正定性等。因此，本章也可以视作是对《矩阵力量》中重要的线性代数工具的梳理和应用。

形状

一般而言，协方差矩阵可视作方差和协方差两部分组成，方差是协方差矩阵对角线上的元素，协方差是协方差矩阵非对角线上的元素：

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

方差描述了某个特征上数据的离散度，而协方差则蕴含成对特征之间的相关性。

显而易见，协方差矩阵为对称矩阵：

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\mathrm{T}} \tag{2}$$

理论

定义随机变量的列向量 $\boldsymbol{\chi}$：

$$\boldsymbol{\chi} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_D \end{bmatrix} \tag{3}$$

$\boldsymbol{\chi}$ 的协方差矩阵可以通过下式计算得到：

$$\begin{aligned} \mathrm{var}(\boldsymbol{\chi}) &= \mathrm{cov}(\boldsymbol{\chi}, \boldsymbol{\chi}) = \mathrm{E}\left[\left(\boldsymbol{\chi} - \mathrm{E}(\boldsymbol{\chi})\right)\left(\boldsymbol{\chi} - \mathrm{E}(\boldsymbol{\chi})\right)^{\mathrm{T}} \right] \\ &= \mathrm{E}\left(\boldsymbol{\chi}\boldsymbol{\chi}^{\mathrm{T}}\right) - \mathrm{E}(\boldsymbol{\chi})\mathrm{E}(\boldsymbol{\chi})^{\mathrm{T}} \end{aligned} \tag{4}$$

⚠注意，为了方便表达，上式中列向量 χ 期望值向量 $\mathrm{E}(\chi)$ 也是列向量。$\mathrm{E}\left(\chi\chi^{\mathrm{T}}\right)$ 和 $\mathrm{E}(\chi)\mathrm{E}(\chi)^{\mathrm{T}}$ 的结果都是 $D \times D$ 方阵。

上式类似我们在本书第 4 章提到的计算方差和协方差的技巧，请大家类比：

$$\begin{aligned}\mathrm{var}(X) &= \underbrace{\mathrm{E}\left(X^2\right)}_{\text{Expectaton of } X^2} - \underbrace{\mathrm{E}(X)^2}_{\text{Square of } \mathrm{E}(X)} \\ \mathrm{cov}(X_1, X_2) &= \mathrm{E}(X_1 X_2) - \mathrm{E}(X_1)\mathrm{E}(X_2)\end{aligned} \tag{5}$$

样本数据

实践中，我们更常用的是样本数据的协方差矩阵。对于形状为 $n \times D$ 的样本数据矩阵 $\boldsymbol{X}$，$\boldsymbol{X}$ 的协方差矩阵 $\boldsymbol{\Sigma}$ 可以通过下式计算得到：

$$\boldsymbol{\Sigma} = \frac{\left(\underbrace{\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})}_{\text{Centered}}\right)^{\mathrm{T}}\left(\underbrace{\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})}_{\text{Centered}}\right)}{n-1} = \frac{\boldsymbol{X}_c{}^{\mathrm{T}}\boldsymbol{X}_c}{n-1} \tag{6}$$

其中，$\mathrm{E}(\boldsymbol{X})$ 为数据 $\boldsymbol{X}$ 质心，是行向量；利用广播原则，$\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})$ 得到去均值数据矩阵 $\boldsymbol{X}_c$。

⚠注意，式中分母为 $n-1$。

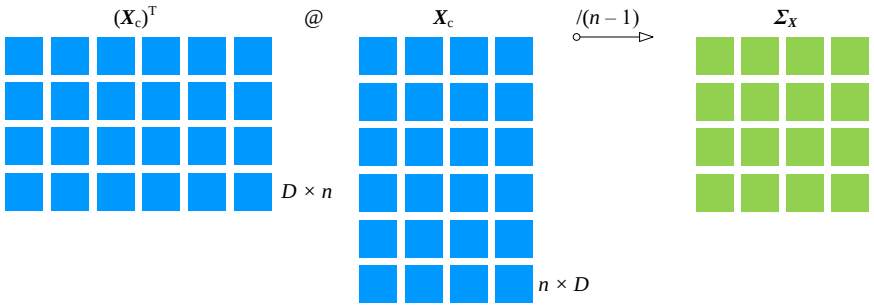


图 1. 计算 $\boldsymbol{X}$ 样本数据协方差矩阵 $\boldsymbol{\Sigma}_X$

(6) 可以写成：

$$\boldsymbol{\Sigma} = \frac{\left(\boldsymbol{X} - \boldsymbol{1}\mathrm{E}(\boldsymbol{X})\right)^{\mathrm{T}}\left(\boldsymbol{X} - \boldsymbol{1}\mathrm{E}(\boldsymbol{X})\right)}{n-1} \tag{7}$$

(7) 展开得到：

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \frac{\left(\boldsymbol{X}^{\mathrm{T}} - \mathrm{E}(\boldsymbol{X})^{\mathrm{T}}\boldsymbol{1}^{\mathrm{T}}\right)\left(\boldsymbol{X} - \boldsymbol{1}\mathrm{E}(\boldsymbol{X})\right)}{n-1} \\
&= \frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} - \underbrace{\mathrm{E}(\boldsymbol{X})^{\mathrm{T}}\boldsymbol{1}^{\mathrm{T}}\boldsymbol{X}}_{n\mathrm{E}(\boldsymbol{X})} - \underbrace{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{1}\mathrm{E}(\boldsymbol{X})}_{n\mathrm{E}(\boldsymbol{X})^{\mathrm{T}}} + \mathrm{E}(\boldsymbol{X})^{\mathrm{T}}\underbrace{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{1}}_{n}\mathrm{E}(\boldsymbol{X})}{n-1} \\
&= \overbrace{\frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}}{n-1}}^{\text{Gram matrix}} - \frac{n}{n-1}\mathrm{E}(\boldsymbol{X})^{\mathrm{T}}\mathrm{E}(\boldsymbol{X})
\end{aligned} \tag{8}
$$

观察 (8)，相信大家已经看到**格拉姆矩阵** (Gram matrix)。也就是说，协方差矩阵可以视作一种特殊的格拉姆矩阵。

此外，如果 n 足够大，可以用 n 替换 $n-1$，影响微乎其微。

把数据矩阵 $\boldsymbol{X}$ 展开成一组列向量 $[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D]$，$\mathrm{E}(\boldsymbol{X})$ 写成 $[\mu_1, \mu_2, \ldots, \mu_D]$，(6) 可以整理为：

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \frac{\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)^{\mathrm{T}}\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)}{n-1} \\
&= \frac{\begin{bmatrix}\boldsymbol{x}_1 - \mu_1 & \boldsymbol{x}_2 - \mu_2 & \cdots & \boldsymbol{x}_D - \mu_D\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\boldsymbol{x}_1 - \mu_1 & \boldsymbol{x}_2 - \mu_2 & \cdots & \boldsymbol{x}_D - \mu_D\end{bmatrix}}{n-1} \\
&= \frac{1}{n-1}\begin{bmatrix}
(\boldsymbol{x}_1 - \mu_1)^{\mathrm{T}}(\boldsymbol{x}_1 - \mu_1) & (\boldsymbol{x}_1 - \mu_1)^{\mathrm{T}}(\boldsymbol{x}_2 - \mu_2) & \cdots & (\boldsymbol{x}_1 - \mu_1)^{\mathrm{T}}(\boldsymbol{x}_D - \mu_D) \\
(\boldsymbol{x}_2 - \mu_2)^{\mathrm{T}}(\boldsymbol{x}_1 - \mu_1) & (\boldsymbol{x}_2 - \mu_2)^{\mathrm{T}}(\boldsymbol{x}_2 - \mu_2) & \cdots & (\boldsymbol{x}_2 - \mu_2)^{\mathrm{T}}(\boldsymbol{x}_D - \mu_D) \\
\vdots & \vdots & \ddots & \vdots \\
(\boldsymbol{x}_D - \mu_D)^{\mathrm{T}}(\boldsymbol{x}_1 - \mu_1) & (\boldsymbol{x}_D - \mu_D)^{\mathrm{T}}(\boldsymbol{x}_2 - \mu_2) & \cdots & (\boldsymbol{x}_D - \mu_D)^{\mathrm{T}}(\boldsymbol{x}_D - \mu_D)
\end{bmatrix}
\end{aligned} \tag{9}
$$

图 2 (a) 所示为鸢尾花四特征数据协方差矩阵 $\boldsymbol{\Sigma}$。

上一章讲解多元高斯分布时，讲过其概率密度函数 PDF 解析式中用到协方差矩阵的逆。而协方差矩阵的逆矩阵有自己的名字——**集中矩阵** (concentration matrix)。图 2 (b) 所示为协方差矩阵的逆 $\boldsymbol{\Sigma}^{-1}$。

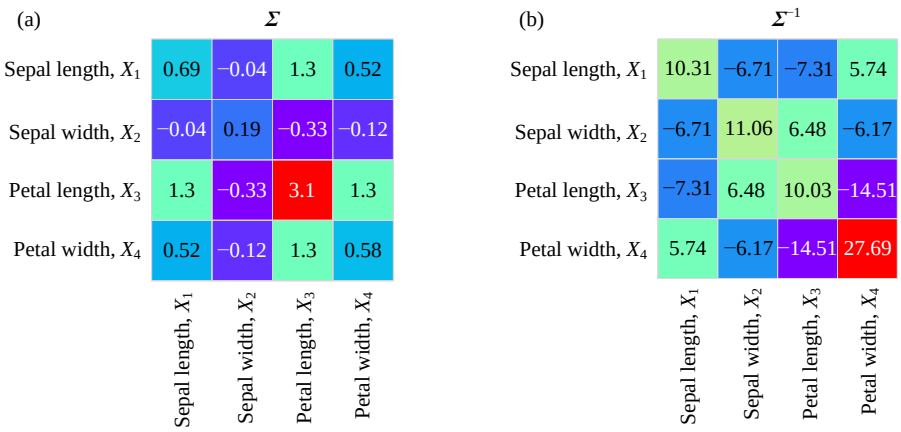


图 2. 鸢尾花四特征协方差矩阵、逆矩阵热图

四种椭圆

本书中常用椭圆代表协方差矩阵。χ 若服从多元高斯分布，$\chi \sim (\mu, \Sigma)$。如图 3 所示，当协方差矩阵形态不同时，对应的椭圆有四种类型。

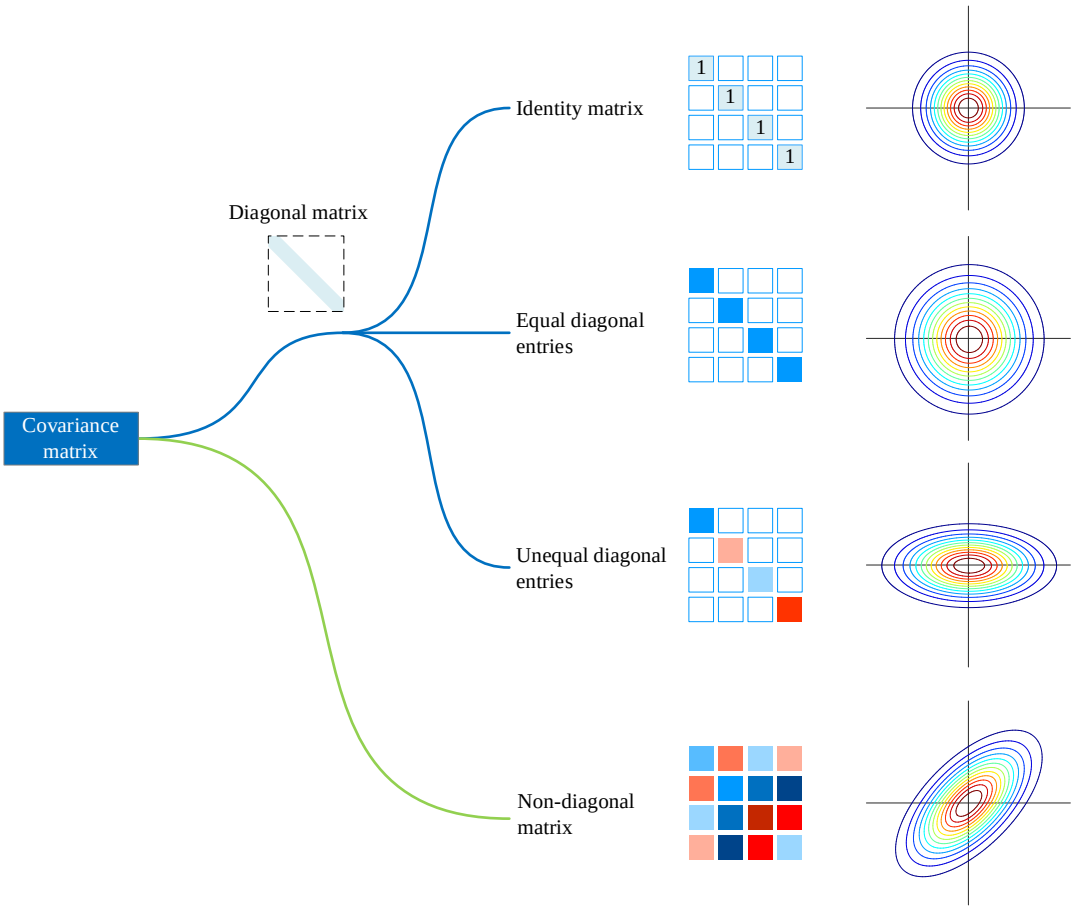


图 3. 协方差矩阵的形态影响高斯密度函数形状

当协方差矩阵为**单位矩阵** (identity matrix) 时，即 $\Sigma = I$，随机变量为 IID，每一个随机变量服从标准正态分布；因此，这种情况，我们用正圆代表其概率密度函数。准确来说是，概率密度函数对应的几何形状是多维空间的正球体。

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。

类似的，当协方差矩阵为 $\Sigma = kI$，这种情况对应的概率密度函数也是正圆，k 相当于缩放系数。当 Σ 为对角阵，对角线元素不同：

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{1,1} & 0 & \cdots & 0 \\ 0 & \sigma_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{D,D} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_D^2 \end{bmatrix} \tag{10}$$

这种情况，对应的概率密度函数形状为正椭圆。多元高斯分布的密度函数可以写成边际概率密度函数的累乘：

$$f_{\boldsymbol{x}}(\boldsymbol{x}) = \prod_{j=1}^{D} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(\frac{-1}{2} \left(\frac{x_j - \mu_j}{\sigma_j} \right)^2 \right) \tag{11}$$

$\boldsymbol{\Sigma}$不定时，高斯分布 PDF 形状为旋转椭圆。

本章最后将深入探讨协方差的几何视角。

给定标签为条件

当然，在计算协方差时，我们也可以考虑到数据标签。图 4 所示为三个不同标签数据各自的协方差矩阵 $\boldsymbol{\Sigma}_1$、$\boldsymbol{\Sigma}_2$、$\boldsymbol{\Sigma}_3$热图。

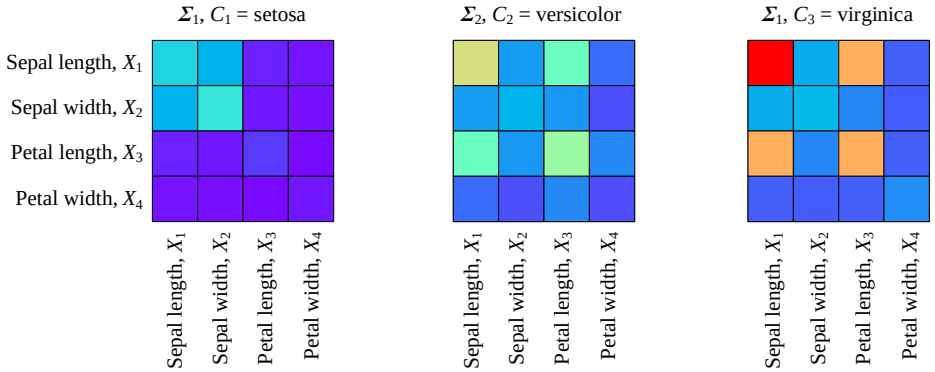


图 4. 协方差矩阵热图，考虑分类

质心位于原点

特别地，当所有均值都是 0 时，$[\mu_1, \mu_2, \ldots, \mu_D]^T = [0, 0, \ldots, 0]^T$，也就是说数据质心位于原点，并将 $\boldsymbol{X}$ 写成列向量，(9) 可以写成：

$$\boldsymbol{\Sigma} = \frac{\boldsymbol{X}^T\boldsymbol{X}}{n-1} = \frac{\boldsymbol{G}}{n-1} = \frac{1}{n-1} \begin{bmatrix} \boldsymbol{x}_1^T\boldsymbol{x}_1 & \boldsymbol{x}_1^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_1^T\boldsymbol{x}_D \\ \boldsymbol{x}_2^T\boldsymbol{x}_1 & \boldsymbol{x}_2^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_2^T\boldsymbol{x}_D \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_D^T\boldsymbol{x}_1 & \boldsymbol{x}_D^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D^T\boldsymbol{x}_D \end{bmatrix} \tag{12}$$

Gram matrix

用向量内积运算，(12) 可以写成：

$$\boldsymbol{\Sigma} = \frac{1}{n-1}\begin{bmatrix} \langle \boldsymbol{x}_1, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_1, \boldsymbol{x}_D \rangle \\ \langle \boldsymbol{x}_2, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_2, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_2, \boldsymbol{x}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{x}_D, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_D, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_D, \boldsymbol{x}_D \rangle \end{bmatrix} \tag{13}$$

上式是矩阵乘法的第一视角。

同样，当数据质心位于原点时，将 $\boldsymbol{X}$ 写成行向量，(9) 可以写成：

$$\begin{aligned} \boldsymbol{\Sigma} &= \overbrace{\frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}}{n-1}}^{\text{Gram matrix}} = \frac{1}{n-1}\begin{bmatrix} \boldsymbol{x}^{(1)\mathrm{T}} & \boldsymbol{x}^{(2)\mathrm{T}} & \cdots & \boldsymbol{x}^{(n)\mathrm{T}} \end{bmatrix}\begin{bmatrix} \boldsymbol{x}^{(1)} \\ \boldsymbol{x}^{(2)} \\ \vdots \\ \boldsymbol{x}^{(n)} \end{bmatrix} \\ &= \frac{1}{n-1}\left(\boldsymbol{x}^{(1)\mathrm{T}}\boldsymbol{x}^{(1)} + \boldsymbol{x}^{(2)\mathrm{T}}\boldsymbol{x}^{(2)} + \cdots + \boldsymbol{x}^{(n)\mathrm{T}}\boldsymbol{x}^{(n)}\right) = \frac{1}{n-1}\sum_{i=1}^{n}\boldsymbol{x}^{(i)\mathrm{T}}\boldsymbol{x}^{(i)} \end{aligned} \tag{14}$$

上式中，$\boldsymbol{x}^{(i)\mathrm{T}}\boldsymbol{x}^{(i)}$ 的形状为 $D \times D$。矩阵乘法写成了一系列形状大小相同 n 的矩阵层层叠加，这便是矩阵乘法的第二视角。

协方差矩阵分块

协方差矩阵还可以分块。比如，鸢尾花 4 × 4 协方差矩阵可以按照如下方式分块：

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{2,1} & \sigma_{2,2} & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_{3,3} & \sigma_{3,4} \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & \sigma_{4,4} \end{bmatrix} = \begin{bmatrix} \underbrace{\begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{bmatrix}}_{\boldsymbol{\Sigma}_{2\times 2}} & \underbrace{\begin{bmatrix} \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{2,3} & \sigma_{2,4} \end{bmatrix}}_{\boldsymbol{\Sigma}_{2\times(4-2)}} \\ \underbrace{\begin{bmatrix} \sigma_{3,1} & \sigma_{3,2} \\ \sigma_{4,1} & \sigma_{4,2} \end{bmatrix}}_{\boldsymbol{\Sigma}_{(4-2)\times 2}} & \underbrace{\begin{bmatrix} \sigma_{3,3} & \sigma_{3,4} \\ \sigma_{4,3} & \sigma_{4,4} \end{bmatrix}}_{\boldsymbol{\Sigma}_{(4-2)\times(4-2)}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{2\times 2} & \boldsymbol{\Sigma}_{2\times(4-2)} \\ \boldsymbol{\Sigma}_{(4-2)\times 2} & \boldsymbol{\Sigma}_{\Sigma_{(4-2)\times(4-2)}} \end{bmatrix} \tag{15}$$

4 × 4 协方差矩阵 $\boldsymbol{\Sigma}$ 被分为 4 块。注意，矩阵分块时切割线的交点位于主对角线上。

如图 5 所示，$\boldsymbol{\Sigma}_{2\times 2}$ 和 $\boldsymbol{\Sigma}_{(4-2)\times(4-2)}$ 都还是协方差矩阵，它俩的主对角线上还是方差。几何视角来看，$\boldsymbol{\Sigma}_{2\times 2}$ 和 $\boldsymbol{\Sigma}_{(4-2)\times(4-2)}$ 都是旋转椭圆。

而 $\boldsymbol{\Sigma}_{(4-2)\times 2}$ 和 $\boldsymbol{\Sigma}_{2\times(4-2)}$ 叫**互协方差矩阵** (cross-covariance matrix)。注意，互协方差矩阵中一般只含有协方差，没有方差。$\boldsymbol{\Sigma}_{(4-2)\times 2}$ 和 $\boldsymbol{\Sigma}_{2\times(4-2)}$ 互为转置矩阵，即 $\boldsymbol{\Sigma}_{(4-2)\times 2} = \boldsymbol{\Sigma}_{2\times(4-2)}^{\mathrm{T}}$。

丛书《数据有道》一册讲解**典型相关分析** (Canonical Correlation Analysis) 将会用到互协方差矩阵。

当然，协方差矩阵分块方式有很多，比如图 6。图 6 中 $\boldsymbol{\Sigma}_{3\times 3}$ 的几何形状为椭球。请大家自行分析图 6。

有关分块矩阵运算，建议大家回顾《矩阵力量》第 6 章。

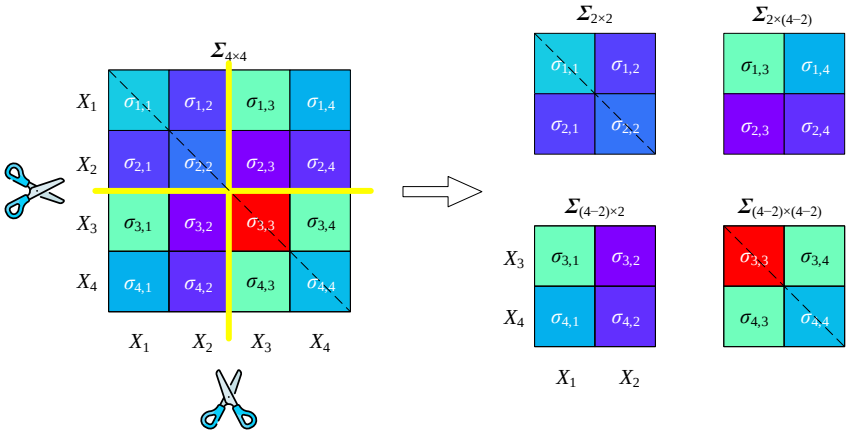


图 5. 协方差矩阵分块

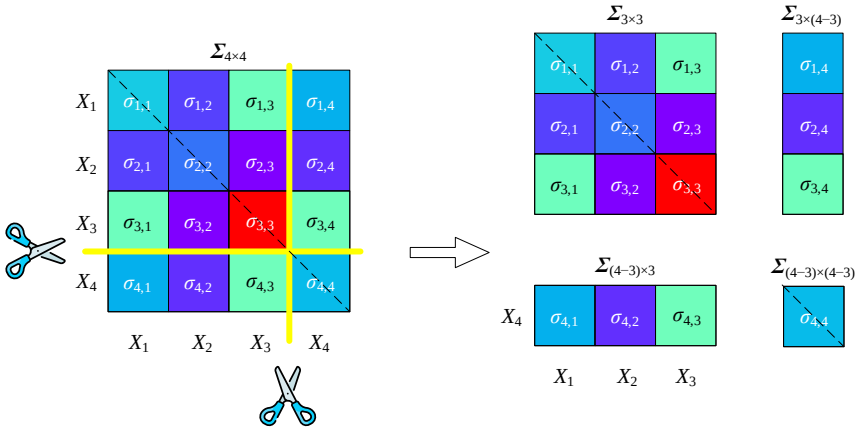


图 6. 协方差矩阵分块，第二种方式

13.2 相关性系数矩阵：描述 Z 分数分布

相关性系数矩阵 $\boldsymbol{P}$ 的定义为：

$$\boldsymbol{P}=\begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,D} \\ \rho_{1,2} & 1 & \cdots & \rho_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,D} & \rho_{2,D} & \cdots & 1 \end{bmatrix} \tag{16}$$

图 7 所示为鸢尾花数据相关性系数矩阵 $\boldsymbol{P}$。$\boldsymbol{P}$ 的对角线元素均为 1，对角线以外元素为成对相关性系数 $\rho_{i,j}$。类似协方差矩阵，相关性系数矩阵 $\boldsymbol{P}$ 当然也可以分块。

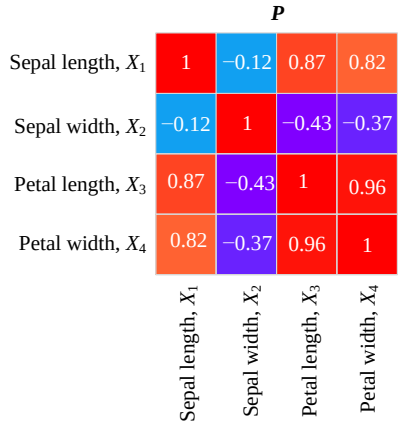


图 7. 鸢尾花数据相关性系数矩阵热图

协方差矩阵 vs 相关性系数矩阵

协方差矩阵 $\boldsymbol{\Sigma}$ 和相关性系数矩阵 $\boldsymbol{P}$ 关系如下：

$$\boldsymbol{\Sigma} = \boldsymbol{DPD} = \underbrace{\begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_D \end{bmatrix}}_{\boldsymbol{D}} \underbrace{\begin{bmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,D} \\ \rho_{1,2} & 1 & \cdots & \rho_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,D} & \rho_{2,D} & \cdots & 1 \end{bmatrix}}_{\text{Correlation matrix, } \boldsymbol{P}} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_D \end{bmatrix}}_{\boldsymbol{D}} \tag{17}$$

从几何角度来看，上式中对角方阵 $\boldsymbol{S}$ 起到的是缩放作用。

图 8 所示为协方差矩阵和相关性矩阵关系热图。

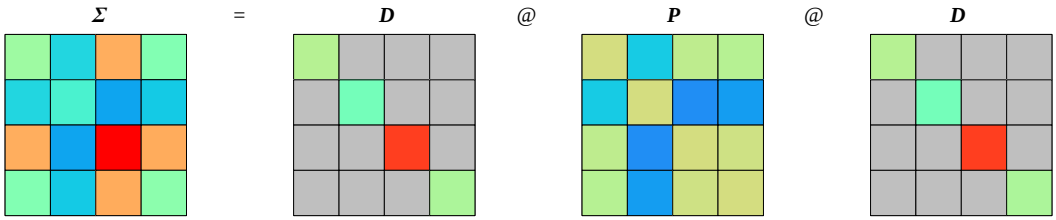


图 8. 协方差矩阵和相关性矩阵关系热图

从 $\boldsymbol{\Sigma}$ 反求相关性系数矩阵 $\boldsymbol{P}$

$$\boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{\Sigma}\boldsymbol{D}^{-1} \tag{18}$$

其中

$$\boldsymbol{D}^{-1}=\operatorname{diag}\left(\operatorname{diag}\left(\boldsymbol{\Sigma}\right)\right)^{\frac{-1}{2}}=\begin{bmatrix}1/\sigma_1 & 0 & \cdots & 0\\ 0 & 1/\sigma_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1/\sigma_D\end{bmatrix} \tag{19}$$

考虑标签

图 9 为考虑分类标签条件下的协方差矩阵热图，我们管它们叫条件协方差矩阵。

大家是否立刻想到，既然协方差可以用椭圆代表，图 9 中的三个条件协方差矩阵也肯定有它们各自的椭圆！这是本章最后要介绍的内容。

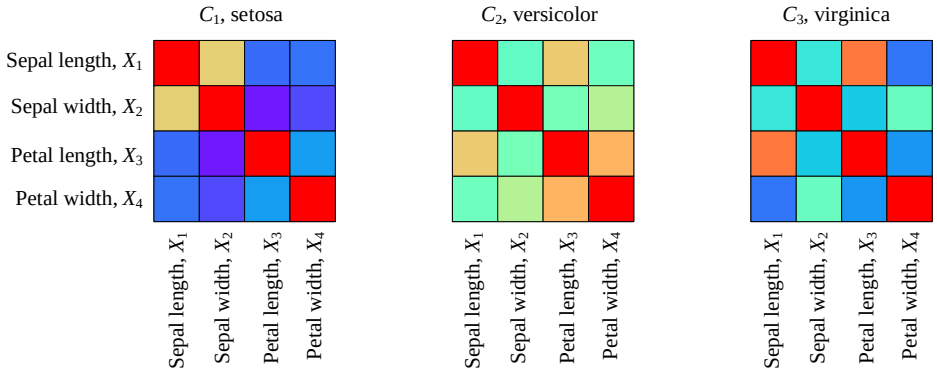


图 9. 相关性系数矩阵热图，考虑分类标签

13.3 特征值分解：找到旋转、缩放

对协方差矩阵 $\boldsymbol{\Sigma}$ 特征值分解：

$$\boldsymbol{\Sigma}=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} \tag{20}$$

其中，特征值矩阵 $\boldsymbol{\Lambda}$ 为对角方阵：

$$\boldsymbol{\Lambda}=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_D\end{bmatrix} \tag{21}$$

由于 $\boldsymbol{\Sigma}$ 为对称矩阵，所以对协方差矩阵特征值分解是谱分解：

$$\boldsymbol{\Sigma}=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} \tag{22}$$

图 10 所示为鸢尾花数据协方差矩阵 $\boldsymbol{\Sigma}$ 的特征值分解运算热图。

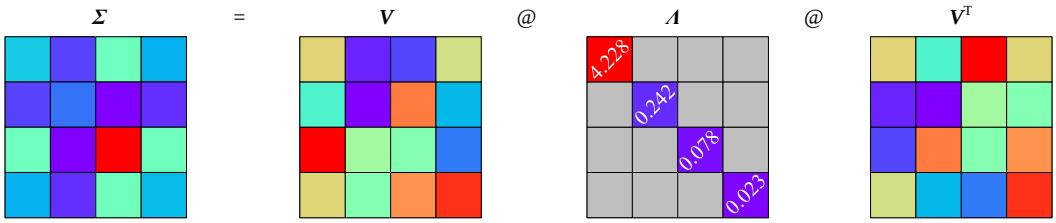


图 10. 协方差矩阵特征值分解

矩阵 $\boldsymbol{V}$ 为正交矩阵：

$$\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I} \tag{23}$$

图 11 运算热图对应 (23)。

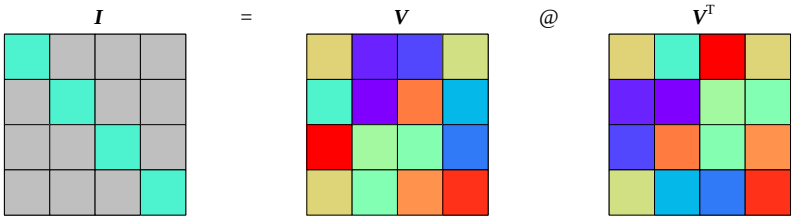


图 11. 矩阵 $\boldsymbol{V}$ 为正交矩阵

谱分解：外积展开

将 (22) 展开来写得到：

$$\begin{aligned}\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} &= \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D\end{bmatrix}\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}}\end{bmatrix} \\ &= \lambda_1\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}} + \lambda_2\boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots + \lambda_D\boldsymbol{v}_D\boldsymbol{v}_D^{\mathrm{T}} = \sum_{j=1}^{D}\lambda_j\boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{T}}\end{aligned} \tag{24}$$

这便是《矩阵力量》第 5 章介绍的矩阵乘法第二视角——外积展开，将矩阵乘法展开写成加法。

用向量张量积来写 (24) 得到：

$$\boldsymbol{\Sigma} = \lambda_1\boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \lambda_2\boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \cdots + \lambda_D\boldsymbol{v}_D \otimes \boldsymbol{v}_D = \sum_{j=1}^{D}\lambda_j\boldsymbol{v}_j \otimes \boldsymbol{v}_j \tag{25}$$

注意，$\boldsymbol{v}_j$为单位向量，无量纲，即没有单位。几何角度来看，$\boldsymbol{v}_j$仅仅提供投影的方向，而真正提供缩放大小的是特征值λ_j。图 12 所示为协方差矩阵谱分解展开热图。虽然$\lambda_1\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}}$的秩为 1，但是$\lambda_1\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}}$已经几乎“还原”$\boldsymbol{\Sigma}$。

此外，几何视角来看，$\lambda_1\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}}$代表向量投影，即《矩阵力量》第 10 章中讲过的“二次投影”，建议大家回顾。

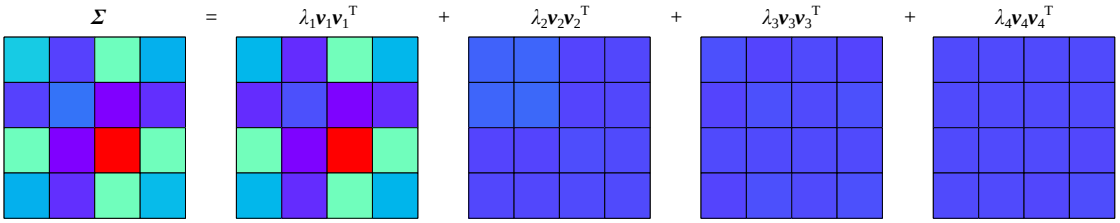


图 12. 协方差矩阵谱分解展开热图

迹

一个值得注意的性质是，协方差矩阵$\boldsymbol{\Sigma}$的迹——方阵对角线元素之和——等于 (21) 特征值之和：

$$\begin{aligned}\operatorname{trace}(\boldsymbol{\Sigma}) &= \sigma_1^2+\sigma_2^2+\cdots+\sigma_D^2=\sum_{j=1}^{D}\sigma_j^2 \\ &= \lambda_1+\lambda_2+\cdots\lambda_D=\sum_{j=1}^{D}\lambda_j\end{aligned} \tag{26}$$

协方差矩阵$\boldsymbol{\Sigma}$对角线元素之和，相当于所有特征的方差之和，即数据整体的方差。$\boldsymbol{V}$相当于旋转，而旋转操作不改变数据整体方差。本章后文将介绍理解上式的几何视角。

图 13 所示为鸢尾花数据矩阵$\boldsymbol{X}$中每一列数据的方差σ_j^2对整体方差$\sum_{j=1}^{D}\sigma_j^2$的贡献。

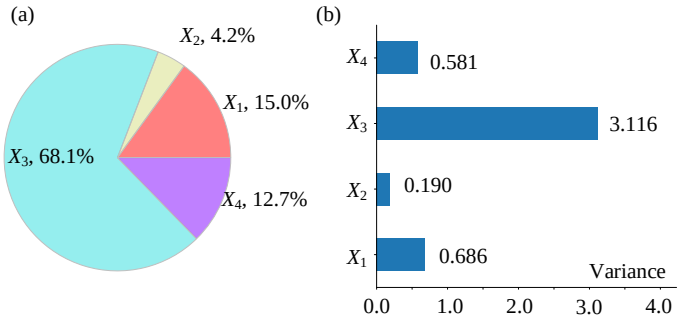


图 13. 协方差矩阵$\boldsymbol{\Sigma}$的主对角线成分，即$\boldsymbol{X}$的方差

投影视角

利用我们已经学过的有关特征值分解的几何视角，中心化数据矩阵 $\boldsymbol{X}_c$ 在 $\boldsymbol{V}$ 投影得到数据 $\boldsymbol{Y}$：

$$\boldsymbol{Y} = \boldsymbol{X}_c \boldsymbol{V} = \left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)\boldsymbol{V} \tag{27}$$

求数据矩阵 $\boldsymbol{Y}$ 的协方差矩阵：

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{Y}} &= \frac{\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}}{n-1} = \frac{\left(\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)\boldsymbol{V}\right)^{\mathrm{T}}\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)\boldsymbol{V}}{n-1} \\ &= \boldsymbol{V}^{\mathrm{T}}\frac{\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)^{\mathrm{T}}\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)}{n-1}\boldsymbol{V} \\ &= \boldsymbol{V}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{V} = \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_D \end{bmatrix}\end{aligned} \tag{28}$$

观察 (28) 的矩阵 $\boldsymbol{Y}$ 的协方差矩阵，可以发现投影得到的数据列向量相互正交特征值从小到大排列，即 $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_D$，矩阵 $\boldsymbol{Y}$ 第一列 $\boldsymbol{y}_1$ 的方差最大。

如图 14 所示，以鸢尾花数据投影结果为例，$\boldsymbol{y}_1$ 的方差对整体方差贡献超过 90%。这便是主成分分析的思路，本书第 25 章将继续这一话题的探讨。

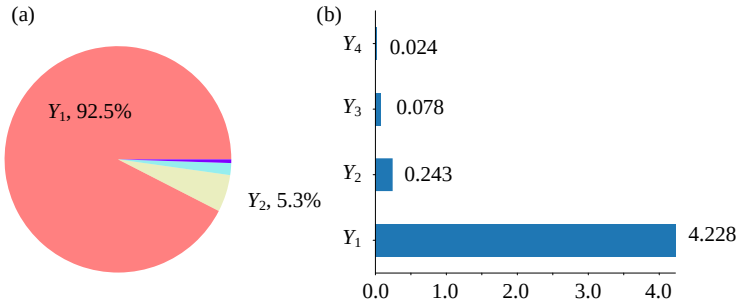


图 14. $\boldsymbol{\Sigma}_{\boldsymbol{Y}}$ 的主对角线成分，$\boldsymbol{Y}$ 的方差

协方差的“投影”

举个例子，数据矩阵 $\boldsymbol{X}_c$ 在 $\boldsymbol{v}_1$ 方向投影结果为 $\boldsymbol{y}_1$：

$$\boldsymbol{y}_1 = \boldsymbol{X}_c \boldsymbol{v}_1 \tag{29}$$

由于 $\boldsymbol{X}_c$ 的质心在原点，所以 $\boldsymbol{y}_1$ 的期望值为 0。而 $\boldsymbol{y}_1$ 的方差为：

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}_1} = \sigma^2_{\boldsymbol{Y}_1} = \frac{\boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{y}_1}{n-1} = \frac{\left(\boldsymbol{X}_c\boldsymbol{v}_1\right)^{\mathrm{T}}\boldsymbol{X}_c\boldsymbol{v}_1}{n-1} = \boldsymbol{v}_1^{\mathrm{T}}\frac{\boldsymbol{X}_c^{\mathrm{T}}\boldsymbol{X}_c}{n-1}\boldsymbol{v}_1 = \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{v}_1 \tag{30}$$

将 (25) 代入上式得到：

$$\boldsymbol{\Sigma}_{Y_1} = \boldsymbol{v}_1^{\mathrm{T}} \left(\lambda_1 \boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \lambda_2 \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \cdots + \lambda_D \boldsymbol{v}_D \otimes \boldsymbol{v}_D \right) \boldsymbol{v}_1 = \lambda_1 \tag{31}$$

上式相当于 $\boldsymbol{\Sigma}$ 在 $\boldsymbol{v}_1$ 方向上“投影”的结果。

类似地，$\boldsymbol{\Sigma}$ 在 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ “投影”的结果为：

$$\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} \boldsymbol{\Sigma} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \tag{32}$$

本书下一章将深入探讨这一话题。

开平方

用特征值分解结果，可以对协方差矩阵 $\boldsymbol{\Sigma}$ 开平方：

$$\boldsymbol{\Sigma} = \boldsymbol{V} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V} \boldsymbol{\Lambda}^{\frac{1}{2}} \left(\boldsymbol{V} \boldsymbol{\Lambda}^{\frac{1}{2}} \right)^{\mathrm{T}} \tag{33}$$

请大家利用本章代码自行绘制上式热图。

行列式值

协方差矩阵 $\boldsymbol{\Sigma}$ 的行列式值为其特征值乘积：

$$\left| \boldsymbol{\Sigma} \right| = \left| \boldsymbol{\Lambda} \right| = \prod_{j=1}^{D} \lambda_j \tag{34}$$

本章后文会探讨上式的几何内涵。

$\boldsymbol{\Sigma}$ 行列式值的平方根为：

$$\left| \boldsymbol{\Sigma} \right|^{\frac{1}{2}} = \left| \boldsymbol{\Lambda} \right|^{\frac{1}{2}} = \sqrt{\prod_{j=1}^{D} \lambda_j} \tag{35}$$

$\boldsymbol{\Sigma}^{-1}$ 行列式值的平方根为：

$$\left| \boldsymbol{\Sigma} \right|^{\frac{-1}{2}} = \left| \boldsymbol{\Lambda} \right|^{\frac{-1}{2}} = \frac{1}{\sqrt{\prod_{j=1}^{D} \lambda_j}} \tag{36}$$

注意，上式只有在特征值均不为 0 时才存在，也就是说此时 $\boldsymbol{\Sigma}$ 为正定。

逆的特征值分解

如果协方差矩阵正定，对协方差矩阵的逆矩阵进行特征值分解，得到：

$$\boldsymbol{\Sigma}^{-1} = \left(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\right)^{-1} = \left(\boldsymbol{V}^{\mathrm{T}}\right)^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{\mathrm{T}} \tag{37}$$

上式利用到对称矩阵特征值分解，$\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = 1$ 这个性质。

$\boldsymbol{\Sigma}^{-1}$ 的特征值矩阵为：

$$\boldsymbol{\Lambda}^{-1} = \begin{bmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\lambda_D \end{bmatrix} \tag{38}$$

图 15 所示为 $\boldsymbol{\Sigma}^{-1}$ 的特征值分解运算热图。

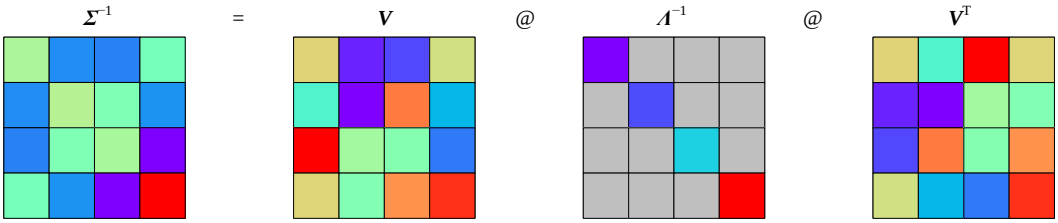


图 15. 协方差矩阵的逆的特征值分解运算热图

相关性系数矩阵的特征值分解

大家肯定能够想到，既然协方差矩阵可以特征值分解，相关性系数矩阵当然也可以进行协方差矩阵分解！图 16 所示为相关性系数矩阵的特征值分解，也是谱分解。

对 $\boldsymbol{X}$ 的每一列求 z 分数得到 $\boldsymbol{Z}_{\boldsymbol{X}}$，相关性系数矩阵是 $\boldsymbol{Z}_{\boldsymbol{X}}$ 的协方差矩阵。也就是说，如图 16 所示，$\boldsymbol{Z}_{\boldsymbol{X}}$ 的整体方差为 4。比较图 10 和图 16，容易发现两个正交矩阵不同，本书第 25 章将会继续这一话题。

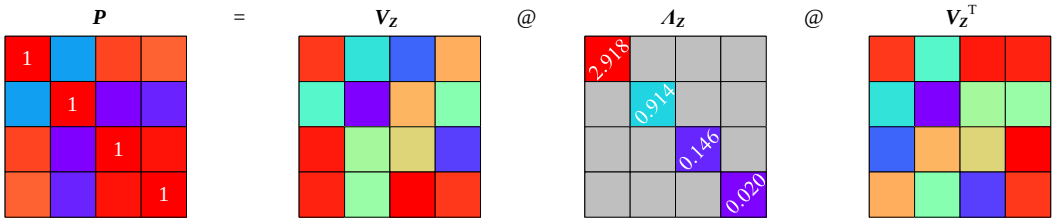


图 16. 相关性系数矩阵的特征值分解

13.4 SVD 分解：分解数据矩阵

《矩阵力量》一册反复提过特征值分解 EVD 和奇异值分解 SVD 的关系。本节探讨对中心化 $\boldsymbol{X}_c$ 矩阵 SVD 分解结果和本章前文介绍的特征值分解结果之间的关系。

回顾 SVD 分解

如图 17 所示，对中心化数据矩阵 $\boldsymbol{X}_c$ 进行经济型 SVD 分解得到：

$$\boldsymbol{X}_c = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}} \tag{39}$$

经济型 SVD 分解中，$\boldsymbol{U}$ 的形状和 $\boldsymbol{X}_c$ 完全相同，都是 $n \times D$。$\boldsymbol{U}$ 的列向量两两正交，即满足 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{I}_{D\times D}$，但是不满足 $\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{I}_{n\times n}$。

完全型 SVD 分解中，$\boldsymbol{U}$ 的形状为 $n \times n$。$\boldsymbol{U}$ 为正交矩阵，满足 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{I}_{n\times n}$。

经济型 SVD 分解中，$\boldsymbol{S}$ 为对角方阵，对角元素为奇异值 s_i。

经济型 SVD 分解中，$\boldsymbol{V}$ 的形状为 $D \times D$。$\boldsymbol{V}$ 为正交矩阵，满足 $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}_{D\times D}$。$\boldsymbol{V}$ 为规范正交基。注意，本书后文为了区分不同规范正交基，会把 (39) 中的 $\boldsymbol{V}$ 写成 $\boldsymbol{V}_c$。

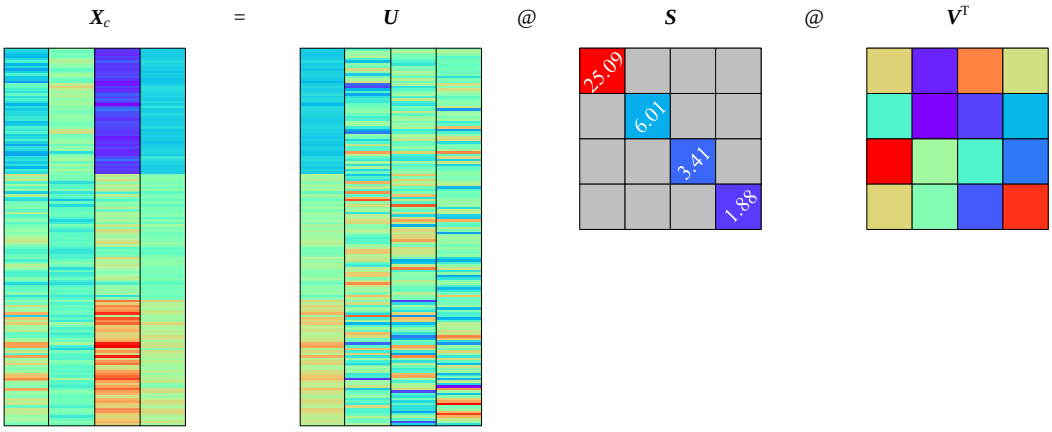


图 17. 矩阵 $\boldsymbol{X}_c$ 进行经济型 SVD 分解

$\boldsymbol{X}_c$ 投影到 $\boldsymbol{V}$

如图 18 所示，将中心化矩阵 $\boldsymbol{X}_c$ 投影到 $\boldsymbol{V}$ 得到 $\boldsymbol{Y}_c$，

$$\boldsymbol{Y}_c = \boldsymbol{X}_c\boldsymbol{V} \tag{40}$$

$\boldsymbol{Y}_c$ 的形状和 $\boldsymbol{X}_c$ 一致。

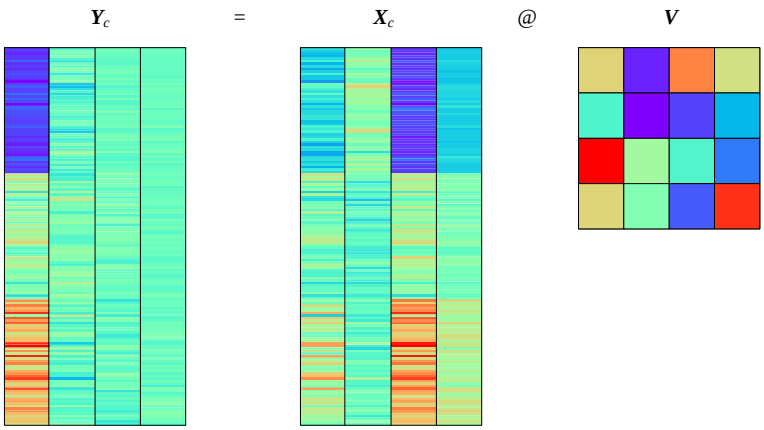


图 18. 矩阵 X_c 投影到 V

X_c 的质心位于原点，Y_c 的质心也位于原点，即：

$$\mathrm{E}(\boldsymbol{Y}_c) = \mathrm{E}(\boldsymbol{X}_c\boldsymbol{V}) = \mathrm{E}(\boldsymbol{X}_c)\boldsymbol{V} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}\boldsymbol{V} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} \tag{41}$$

本章前文提过，Y_c 的协方差为：

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_D \end{bmatrix} \tag{42}$$

而原数据矩阵 X 的质心位于 $\mathrm{E}(X)$。X_c 和 X 的协方差矩阵完全相同。

几何视角来看，X 到 X_c 是质心从 $\mathrm{E}(X)$ 平移到原点。数据本身的分布“形状”相对于质心来说没有任何改变，而协方差矩阵描述的就是分布形状。

X 投影到 V

$V = [v_1, v_2, v_3, v_4]$ 是个 $\mathbb{R}^4$ 规范正交基，不但 X_c 可以投影到 V 中，原始数据 X 也可以投影到 V 中。将 X 投影到 V 得到 Y，

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{V} \tag{43}$$

Y 的质心显然不在原点，$\mathrm{E}(Y)$ 具体位置为：

$$\mathrm{E}(\boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X})\boldsymbol{V} = \begin{bmatrix} 5.843 & 3.057 & 3.758 & 1.199 \end{bmatrix}\boldsymbol{V} = \begin{bmatrix} 5.502 & -5.326 & 0.631 & -0.033 \end{bmatrix} \tag{44}$$

Y 的协方差矩阵则和 Y_c 完全相同，这一点请大家自己证明，并用代码验证。

奇异值 vs 特征值

将 (39) 代入 (6) 得到：

$$\begin{aligned}\boldsymbol{\Sigma} &= \frac{\boldsymbol{X}_c^{\mathrm{T}}\boldsymbol{X}_c}{n-1} = \frac{\left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\right)^{\mathrm{T}}\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}}{n-1} = \frac{\boldsymbol{V}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}}{n-1} \\ &= \boldsymbol{V}\frac{\boldsymbol{S}^2}{n-1}\boldsymbol{V}^{\mathrm{T}}\end{aligned} \tag{45}$$

对比 (45) 和 (20)，可以建立对 $\boldsymbol{\Sigma}$ 特征值分解和对 $\boldsymbol{X}_c$ 进行 SVD 分解的关系：

$$\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}\frac{\boldsymbol{S}^2}{n-1}\boldsymbol{V}^{\mathrm{T}} \tag{46}$$

注意，等式左右两侧的 $\boldsymbol{V}$ 都是正交矩阵，虽然代码计算得到的结果在正负号上会存在差别。

从 (46) 中我们还可以看到 $\boldsymbol{\Sigma}$ 特征值和 $\boldsymbol{X}_c$ 奇异值之间的量化关系：

$$\underbrace{\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D\end{bmatrix}}_{\boldsymbol{\Lambda}} = \frac{1}{n-1}\underbrace{\begin{bmatrix}s_1^2 & & & \\ & s_2^2 & & \\ & & \ddots & \\ & & & s_D^2\end{bmatrix}}_{\boldsymbol{S}^2} \tag{47}$$

即

$$\lambda_j = \frac{1}{n-1}s_j^2 \tag{48}$$

图 19 所示为鸢尾花协方差矩阵特征值和中心化数据奇异值之间的关系。

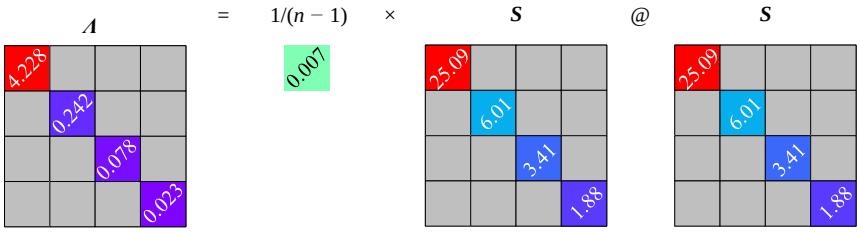


图 19. 特征值和奇异值的关系

有读者可能会问对原数据矩阵 $\boldsymbol{X}$ 直接 SVD 分解，和对 $\boldsymbol{X}_c$ 进行 SVD 分解，两者的区别在哪？这是本书第 25 章要探讨的内容。

矩阵乘法第二视角

如图 20 所示，利用矩阵乘法第二视角，(39) 可以展开写成：

$$\begin{aligned}\boldsymbol{X}_c &= \underbrace{\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_D\end{bmatrix}}_{\boldsymbol{U}} \underbrace{\begin{bmatrix}s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D\end{bmatrix}}_{\boldsymbol{S}} \underbrace{\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_1^{\mathrm{T}}\end{bmatrix}}_{\boldsymbol{V}^{\mathrm{T}}} \\ &= s_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + s_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots + s_D\boldsymbol{u}_D\boldsymbol{v}_D^{\mathrm{T}} = \sum_{j=1}^{D} s_j\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}\end{aligned} \tag{49}$$

同样，$\boldsymbol{u}_j$、$\boldsymbol{v}_j$仅仅提供投影方向，s_j决定重要性。

利用向量张量积，上式可以写成：

$$\boldsymbol{X}_c = s_1\boldsymbol{u}_1 \otimes \boldsymbol{v}_1 + s_2\boldsymbol{u}_2 \otimes \boldsymbol{v}_2 + \cdots + s_D\boldsymbol{u}_D \otimes \boldsymbol{v}_D = \sum_{j=1}^{D} s_j\boldsymbol{u}_j \otimes \boldsymbol{v}_j \tag{50}$$

这种分解类似图 12。

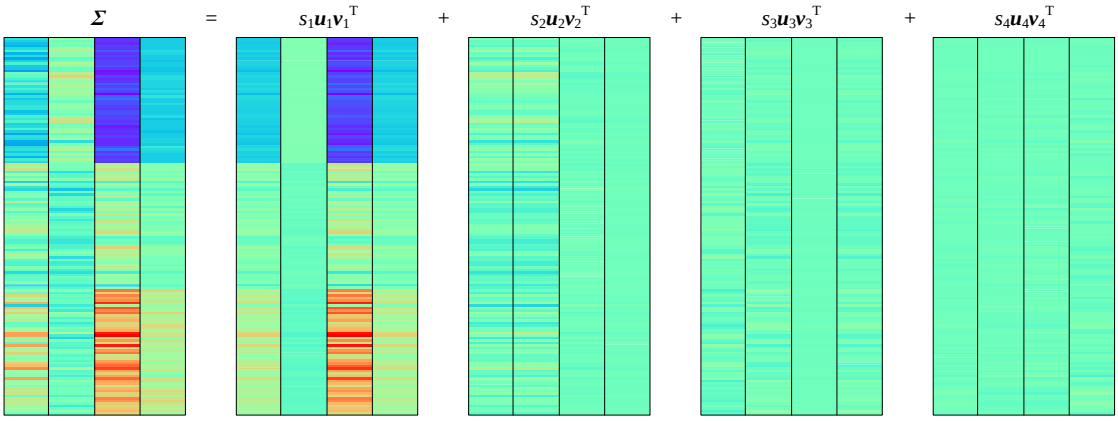


图 20. 利用矩阵乘法第二视角展开 SVD 分解

第二种展开方式

《矩阵力量》第 10 章还介绍过“二次投影”的展开方式，具体如下：

$$\begin{aligned}\boldsymbol{X}_c &= \boldsymbol{X}_c\boldsymbol{I} = \boldsymbol{X}_c\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{X}_c \underbrace{\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D\end{bmatrix}}_{\boldsymbol{V}} \underbrace{\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_1^{\mathrm{T}}\end{bmatrix}}_{\boldsymbol{V}^{\mathrm{T}}} \\ &= \boldsymbol{X}_c\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{X}_c\boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots + \boldsymbol{X}_c\boldsymbol{v}_D\boldsymbol{v}_D^{\mathrm{T}} = \sum_{j=1}^{D} \boldsymbol{X}_c\boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{T}} = \boldsymbol{X}_c\left(\sum_{j=1}^{D} \boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{T}}\right)\end{aligned} \tag{51}$$

同样用向量张量积，上式可以写成：

$$\boldsymbol{X}_c = \boldsymbol{X}_c\boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \boldsymbol{X}_c\boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \cdots + \boldsymbol{X}_c\boldsymbol{v}_D \otimes \boldsymbol{v}_D = \boldsymbol{X}_c\left(\sum_{j=1}^{D}\boldsymbol{v}_j \otimes \boldsymbol{v}_j\right) \tag{52}$$

请大家自行绘制上式的矩阵运算热图。

13.5 Cholesky 分解：列向量坐标

对协方差矩阵 $\boldsymbol{\Sigma}$ 进行 Cholesky 分解，得到的结果是下三角矩阵 $\boldsymbol{L}$ 和上三角矩阵 $\boldsymbol{L}^{\mathrm{T}}$ 乘积：

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} \tag{53}$$

其中，$\boldsymbol{R}$ 为上三角矩阵，$\boldsymbol{R} = \boldsymbol{L}^{\mathrm{T}}$。

图 21 所示为协方差矩阵 Cholesky 分解运算热图。

建议大家回顾《矩阵力量》第 12、24 章，从几何角度、数据角度理解 Cholesky 分解，本节不再重复。

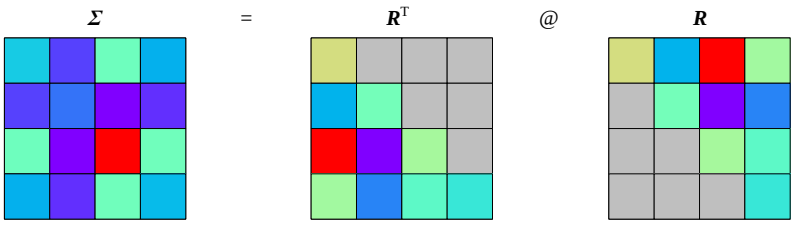


图 21. 对协方差矩阵 Cholesky 分解运算热图

给定数据矩阵 $\boldsymbol{Z}$，$\boldsymbol{Z}$ 的每个随机变量均服从标准正态分布，且相互独立，也就是 IID；$\boldsymbol{Z}$ 的协方差矩阵为单位矩阵 $\boldsymbol{I}$，

$$\boldsymbol{\Sigma}_z = \frac{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}}{n-1} = \boldsymbol{I} \tag{54}$$

令

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{R} + \mathrm{E}(\boldsymbol{X}) \tag{55}$$

从 (55) 推导 $\boldsymbol{X}$ 的协方差矩阵：

$$\boldsymbol{\Sigma}_x = \frac{\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)^{\mathrm{T}}\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)}{n-1} = \frac{(\boldsymbol{Z}\boldsymbol{R})^{\mathrm{T}}(\boldsymbol{Z}\boldsymbol{R})}{n-1} = \frac{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{R}}{n-1} = \boldsymbol{L}\underbrace{\frac{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}}{n-1}}_{\boldsymbol{I}}\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} \tag{56}$$

以上内容对于产生满足特定相关性随机数特别重要，本书第 15 章将展开讲解。

13.6 距离：欧氏距离 vs 马氏距离

协方差矩阵还出现在距离度量运算中，比如马氏距离。本节比较欧氏距离和马氏距离，并引出下一节内容。

欧氏距离

从矩阵运算角度来看，欧氏距离的平方就是《矩阵力量》第 5 章介绍的**二次型** (quadratic form)。比如，空间中任意一点 $\boldsymbol{x}$ 到质心 $\boldsymbol{\mu}$ 的欧氏距离为：

$$d^2 = (\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}) = \|\boldsymbol{x}-\boldsymbol{\mu}\|_2^2 = \sum_{j=1}^{D}(x_j-\mu_j)^2 \tag{57}$$

如图 22 (a) 所示，如果 $\boldsymbol{x}$ 有 2 个特征，即 $D=2$，$d=\|\boldsymbol{x}-\boldsymbol{\mu}\|=1$ 代表圆心位于质心 $\boldsymbol{\mu}$、半径为 1 的正圆。图 22 (a) 中正圆的解析式为：

$$(x_1-\mu_1)^2+(x_2-\mu_2)^2=1 \tag{58}$$

如图 22 (b) 所示，如果 $\boldsymbol{x}$ 有 3 个特征，即 $D=3$，$d=\|\boldsymbol{x}-\boldsymbol{\mu}\|=1$ 代表圆心位于质心 $\boldsymbol{\mu}$、半径为 1 的正球体，对应的解析式为：

$$(x_1-\mu_1)^2+(x_2-\mu_2)^2+(x_3-\mu_3)^2=1 \tag{59}$$

当 $D>3$ 时，$d=\|\boldsymbol{x}-\boldsymbol{\mu}\|=1$ 代表空间中的超球体。

换个角度，$D=2$，当 d 取不同值时，欧氏距离等距线则是一层层同心圆，具体如图 22 (c) 所示。$D=3$，当 d 取不同值时，欧氏距离等距线变成了一层层同心正球体。

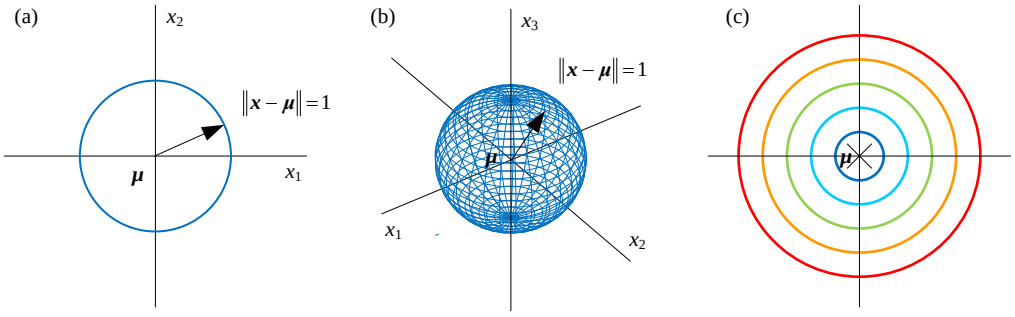


图 22. 正圆、正球体、同心圆

以鸢尾花数据为例，它的质心位于：

$$\boldsymbol{\mu} = \begin{bmatrix} 5.843 \\ 3.057 \\ 3.758 \\ 1.199 \end{bmatrix} \tag{60}$$

原点 $\boldsymbol{0}$ 和质心 $\boldsymbol{\mu}$ 的欧氏距离为：

$$\|\boldsymbol{0}-\boldsymbol{\mu}\| = \sqrt{(0-5.843)^2+(0-3.057)^2+(0-3.758)^2+(0-1.199)^2} \approx 7.684 \tag{61}$$

⚠ 注意，上式中欧氏距离的单位为厘米。

欧氏距离

马氏距离的平方也是二次型：

$$d^2 = (\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) = \left\| \boldsymbol{\Lambda}^{\frac{-1}{2}} \boldsymbol{V}^{\mathrm{T}} (\boldsymbol{x}-\boldsymbol{\mu}) \right\|_2^2 \tag{62}$$

如图 22 (a) 所示，$D = 2$ 时，$d = \left\| \boldsymbol{\Lambda}^{\frac{-1}{2}} \boldsymbol{V}^{\mathrm{T}} (\boldsymbol{x}-\boldsymbol{\mu}) \right\| = 1$ 代表圆心位于质心 $\boldsymbol{\mu}$ 的椭圆。

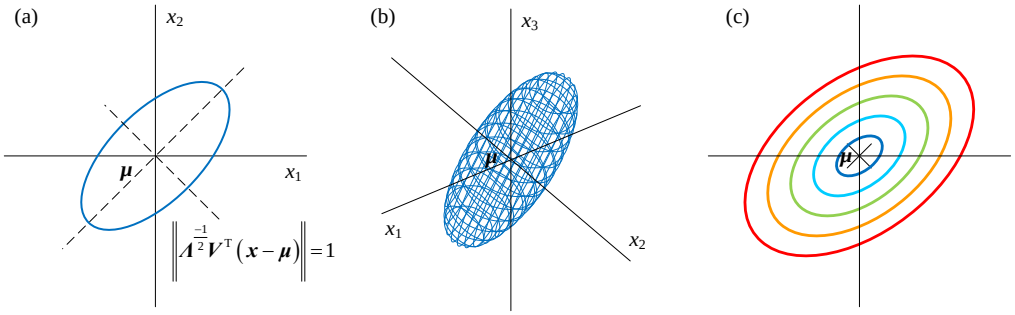


图 23. 椭圆、椭球、同心椭圆

特别地，如果协方差矩阵 $\boldsymbol{\Sigma}$ 为：

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{bmatrix}, \quad \sigma_1 > \sigma_2 \tag{63}$$

马氏距离 $d = 1$ 对应椭圆的解析式为：

$$\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} = 1 \tag{64}$$

这个椭圆显然是正椭圆，圆心位于 (μ_1, μ_2)，半长轴为 σ_1，半短轴为 σ_2。

对于一般的协方差矩阵 $\boldsymbol{\Sigma}_{2\times 2}$，想知道旋转椭圆的半长轴、半短轴长度，则需要利用特征值分解得到其特征值矩阵：

$$\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2\\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix} \overset{\text{EVD}}{\Rightarrow} \boldsymbol{\Lambda}=\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix} \tag{65}$$

这个旋转椭圆的圆心位于 (μ_1, μ_2)，半长轴为 $\sqrt{\lambda_1}$，半短轴为 $\sqrt{\lambda_2}$。特征值分解得到的特征向量 $\boldsymbol{v}_1$、$\boldsymbol{v}_2$ 则告诉我们椭圆长轴、短轴方向。

如图 22 (b) 所示，如果 $\boldsymbol{x}$ 有 3 个特征，即 $D=3$，$d=\left\|\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right\|=1$ 代表圆心位于质心 $\boldsymbol{\mu}$ 的椭球体。

同样，如果协方差矩阵 $\boldsymbol{\Sigma}$ 为：

$$\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \sigma_3^2\end{bmatrix},\quad \sigma_1>\sigma_2>\sigma_3 \tag{66}$$

马氏距离 $d=1$ 对应椭球的解析式为：

$$\frac{\left(x_1-\mu_1\right)^2}{\sigma_1^2}+\frac{\left(x_2-\mu_2\right)^2}{\sigma_2^2}+\frac{\left(x_3-\mu_3\right)^2}{\sigma_3^2}=1 \tag{67}$$

σ_1、σ_2、σ_3 都是椭球的半主轴 (principal semi-axis) 长度，我们管它们分别叫第一、第二、第三半主轴长度。

同理，对于更一般的协方差矩阵 $\boldsymbol{\Sigma}_{3\times 3}$，需要通过特征值分解找到半主轴长度 $\sqrt{\lambda_1}$、$\sqrt{\lambda_2}$、$\sqrt{\lambda_3}$。三个主轴的方向则分别对应三个特征向量 $\boldsymbol{v}_1$、$\boldsymbol{v}_2$、$\boldsymbol{v}_3$。

当 $D>3$ 时，$d=\left\|\boldsymbol{\Lambda}^{\frac{-1}{2}}\boldsymbol{V}^{\mathrm{T}}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right\|=1$ 代表空间中的超椭球。

$D=2$，当 d 取不同值时，马氏距离等距线则是一层层同心椭圆，如图 22 (c) 所示。

还是以鸢尾花数据为例，如图 24 所示，原点 $\boldsymbol{0}$ 和质心 $\boldsymbol{\mu}$ 的马氏距离平方值为：

$$d^2=\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}-\begin{bmatrix}5.843\\3.057\\3.758\\1.199\end{bmatrix}\right)^{\mathrm{T}}\left(\begin{bmatrix}0.69 & -0.042 & 1.3 & 0.52\\ -0.042 & 0.19 & -0.33 & -0.12\\ 1.3 & -0.33 & 3.1 & 1.3\\ 0.52 & -0.12 & 1.3 & 0.58\end{bmatrix}\right)^{-1}\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}-\begin{bmatrix}5.843\\3.057\\3.758\\1.199\end{bmatrix}\right)=129.245 \tag{68}$$

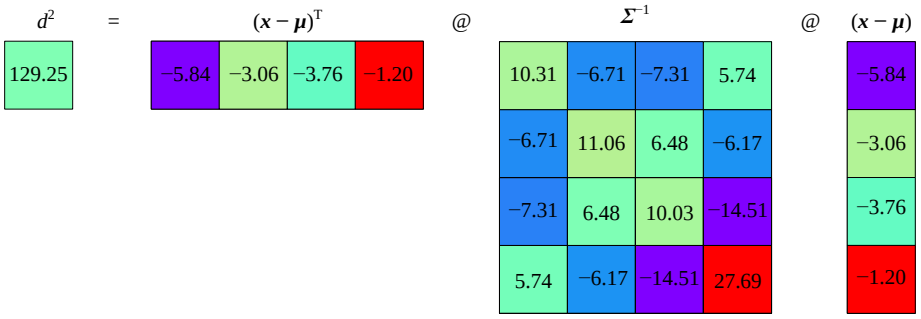


图 24 计算 d^2 的矩阵运算热图

(68) 开平方得到原点 $\boldsymbol{0}$ 和质心 $\boldsymbol{\mu}$ 的马氏距离为：

$$d = \sqrt{129.245} = 11.3686 \tag{69}$$

马氏距离没有单位。更准确地说，马氏距离的单位是标准差，比如 $d = 11.3686$ 代表马氏距离为“11.3686 个均方差”。

本书第 23 章还会继续探讨马氏距离。

有了本节内容铺垫，下一节深入探讨协方差的几何内涵。

Bk5_Ch13_01.py 绘制本章前文大部分矩阵运算热图。

13.7 几何视角：超椭球、椭球、椭圆

“旋转”超椭球

根据上一节所学，如果 $D = 4$，$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = 1$ 代表四维空间 $\mathbb{R}^4$ 圆心位于原点的超椭球。我们知道，在 $\mathbb{R}^4$ 中超椭球的圆心位于 $\mathrm{E}(\boldsymbol{X})$，即：

$$\mathrm{E}(\boldsymbol{X}) = \begin{bmatrix} 5.843 & 3.057 & 3.758 & 1.199 \end{bmatrix} \tag{70}$$

根据图 10 中所示的对 $\boldsymbol{\Sigma}$ 特征值分解，我们知道超椭球的四个半主轴长度分别为：

$$\begin{aligned}\sqrt{\lambda_1} &\approx \sqrt{4.228} \approx 2.056 \text{ cm}\\ \sqrt{\lambda_2} &\approx \sqrt{0.242} \approx 0.492 \text{ cm}\\ \sqrt{\lambda_3} &\approx \sqrt{0.078} \approx 0.279 \text{ cm}\\ \sqrt{\lambda_4} &\approx \sqrt{0.023} \approx 0.154 \text{ cm}\end{aligned} \tag{71}$$

$\mathbb{R}^4$ 中超椭球四个主轴所在方向对应图 10 中 $\boldsymbol{V}$ 的四个列向量，即：

$$\boldsymbol{V} = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4\end{bmatrix} = \begin{bmatrix} 0.751 & 0.284 & 0.502 & 0.321 \\ 0.380 & 0.547 & -0.675 & -0.317 \\ 0.513 & -0.709 & -0.059 & -0.481 \\ 0.168 & -0.344 & -0.537 & 0.752 \end{bmatrix} \tag{72}$$

显然，在纸面上很难可视化一个四维空间的超椭球，因此我们选择用投影的办法将超椭球投影在不同三维空间、二维平面上。

“旋转”超椭球投影到三维空间

图 25 (a) 所示为四维空间超椭球在 $x_1x_2x_3$ 这个三维空间的投影，结果是个圆心位于质心的椭球。

为了获得这个椭球的解析式，我们先将 4 × 4 协方差矩阵 $\boldsymbol{\Sigma}$“投影”到图 25 (a) 这个三维空间中，我们把这个新的协方差记做：

$$\boldsymbol{\Sigma}_{1,2,3} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 0.686 & -0.042 & 1.274 & 0.516 \\ -0.042 & 0.190 & -0.330 & -0.122 \\ 1.274 & -0.330 & 3.116 & 1.296 \\ 0.516 & -0.122 & 1.296 & 0.581 \end{bmatrix}}_{\boldsymbol{\Sigma}} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} = \begin{bmatrix} 0.686 & -0.042 & 1.274 \\ -0.042 & 0.190 & -0.330 \\ 1.274 & -0.330 & 3.116 \end{bmatrix} \tag{73}$$

$\boldsymbol{\Sigma}$ 消去了第 4 行和第 4 列得到 $\boldsymbol{\Sigma}_{1,2,3}$。

从数据角度来看，原始数据矩阵 $\boldsymbol{X}_{4\times 150}$ 先投影得到 $\boldsymbol{X}_{1,2,3}$：

$$\boldsymbol{X}_{1,2,3} = \underbrace{\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 & \boldsymbol{x}_4\end{bmatrix}}_{\boldsymbol{X}} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \\ & & \end{bmatrix} = \begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3\end{bmatrix} \tag{74}$$

如上运算相当于，保留了 $\boldsymbol{X}$ 的前三列数据。$\boldsymbol{X}_{1,2,3}$ 再算协方差矩阵结果就是 $\boldsymbol{\Sigma}_{1,2,3}$。

单位矩阵 $\boldsymbol{I}_{4\times 4}$ 是 $\mathbb{R}^4$ 的标准正交系，可以写成：

$$\boldsymbol{I}_{4\times 4} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} = \begin{bmatrix}\boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 & \boldsymbol{e}_4\end{bmatrix} \tag{75}$$

(73) 相当于 $\boldsymbol{X}$ 在 $[\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3]$ 基底中投影。

四维空间的超椭球的圆心 $\mathrm{E}(\boldsymbol{X})$ 在图 25 (a) 这个三维空间的位置很容易计算：

$$\mathrm{E}(\boldsymbol{X})\begin{bmatrix}\boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3\end{bmatrix} = \begin{bmatrix}5.843 & 3.057 & 3.758 & 1.199\end{bmatrix}\begin{bmatrix}1 & & \\ & 1 & \\ & & 1 \\ & & \end{bmatrix} = \begin{bmatrix}5.843 & 3.057 & 3.758\end{bmatrix} \tag{76}$$

如果想要调换图 25 (a) 中 x_1 和 x_2 的顺序，只需要 $[\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3]$ 乘上如下的**置换矩阵** (permutation matrix)：

$$\begin{bmatrix}\boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3\end{bmatrix}\begin{bmatrix} & 1 & \\ 1 & & \\ & & 1\end{bmatrix} = \begin{bmatrix}\boldsymbol{e}_2 & \boldsymbol{e}_1 & \boldsymbol{e}_3\end{bmatrix} \tag{77}$$

《矩阵力量》第 5 章讲过置换矩阵，大家可以回顾。

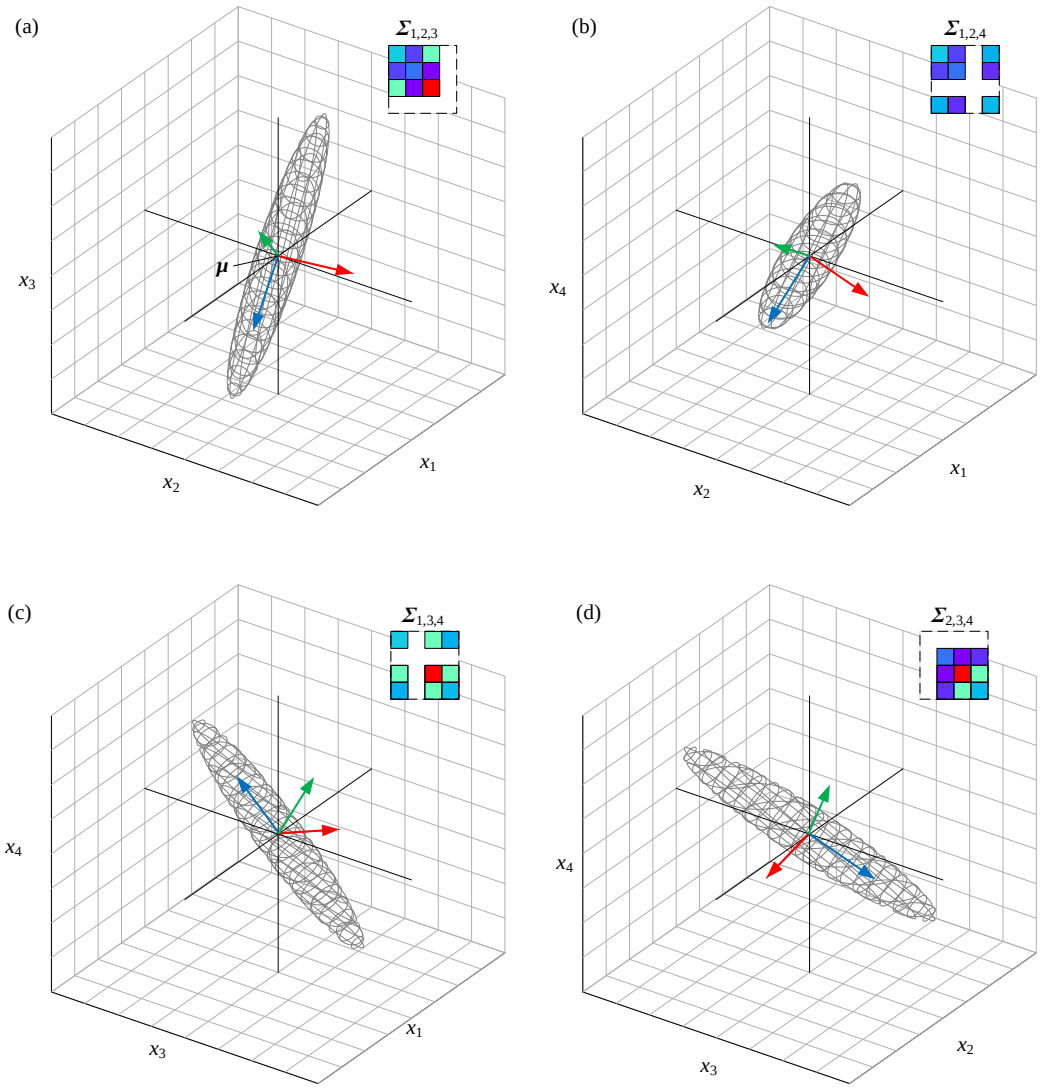


图 25 四维空间的“旋转”超椭球在三维空间中的四个投影

图 25 (a) 中的蓝、红、绿箭头分别代表三维椭球的第一、第二、第三主轴方向。这三个主轴方向需要特征值分解 (73) 中协方差矩阵：

$$
\begin{aligned}
\boldsymbol{\Sigma}_{1,2,3} &= \begin{bmatrix} 0.686 & -0.042 & 1.274 \\ -0.042 & 0.190 & -0.330 \\ 1.274 & -0.330 & 3.116 \end{bmatrix} \\
&= \begin{bmatrix} -0.389 & 0.662 & 0.639 \\ 0.091 & -0.663 & 0.743 \\ -0.916 & -0.347 & -0.198 \end{bmatrix} \begin{bmatrix} 3.691 & & \\ & 0.059 & \\ & & 0.241 \end{bmatrix} \begin{bmatrix} -0.389 & 0.662 & 0.639 \\ 0.091 & -0.663 & 0.743 \\ -0.916 & -0.347 & -0.198 \end{bmatrix}^{\mathrm{T}}
\end{aligned} \tag{78}
$$

由此，我们知道图 25 (a) 中椭球的三个半主轴的长度为：

$$\begin{aligned}\sqrt{3.691} &\approx 1.921 \text{ cm}\\ \sqrt{0.059} &\approx 0.243 \text{ cm}\\ \sqrt{0.241} &\approx 0.491 \text{ cm}\end{aligned} \tag{79}$$

(78) 的特征值分解也帮我们求得椭球的三个主轴方向。

注意，图 25 (a) 中的蓝、红、绿箭头显然不是 (72) 中 $\boldsymbol{V}$ 在 $\mathbb{R}^3$ 中投影，原因很简单 $\boldsymbol{V}$ 在 $\mathbb{R}^3$ 中应该有四个“影子”，而不是三个。这一点在图 26 中看得更明显。此外，大家在本书第 25 章可以看到 $\boldsymbol{V}$ 在六个平面上的投影结果。

只有 $\boldsymbol{V}$ 在沿着 $\boldsymbol{v}_j$ 方向投影 (注意，不是在 $\boldsymbol{v}_j$ 方向投影)，$\boldsymbol{v}_j$ 的分量才会消失。这就好比，正午阳光下，一根柱子相当于“没有”影子。

请大家自行分析图 25 剩余三幅子图，并写出对应的投影运算。

“旋转”椭球投影到二维平面

图 26 所示为图 25 (a) 中椭球进一步投影到三个二维平面上。

以 x_1x_2 平面为例，先将 4×4 协方差矩阵 $\boldsymbol{\Sigma}$ 投影 x_1x_2 平面，结果为：

$$\boldsymbol{\Sigma}_{1,2} = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 0.686 & -0.042 & 1.274 & 0.516 \\ -0.042 & 0.190 & -0.330 & -0.122 \\ 1.274 & -0.330 & 3.116 & 1.296 \\ 0.516 & -0.122 & 1.296 & 0.581 \end{bmatrix}}_{\boldsymbol{\Sigma}} \begin{bmatrix} 1 & \\ & 1 \\ & \\ & \end{bmatrix} = \begin{bmatrix} 0.686 & -0.042 \\ -0.042 & 0.190 \end{bmatrix} \tag{80}$$

请大家自己写出数据投影对应的矩阵运算。

为了计算 (80) 协方差对应的椭圆，需要对其特征值分解：

$$\boldsymbol{\Sigma}_{1,2} = \begin{bmatrix} 0.686 & -0.042 \\ -0.042 & 0.190 \end{bmatrix} = \begin{bmatrix} 0.996 & 0.084 \\ -0.084 & 0.996 \end{bmatrix} \begin{bmatrix} 0.689 & \\ & 0.186 \end{bmatrix} \begin{bmatrix} 0.996 & 0.084 \\ -0.084 & 0.996 \end{bmatrix}^{\mathrm{T}} \tag{81}$$

通过上述特征值分解，我们知道在 x_1x_2 平面上椭圆的半长轴、半短轴长度分别为 0.830、0.431。单位都是厘米 cm。

此外，请大家注意图 25 (a) 中 x_1x_2 平面上这个椭圆中背景蓝色的矩形，这是本节后续要讨论的内容。

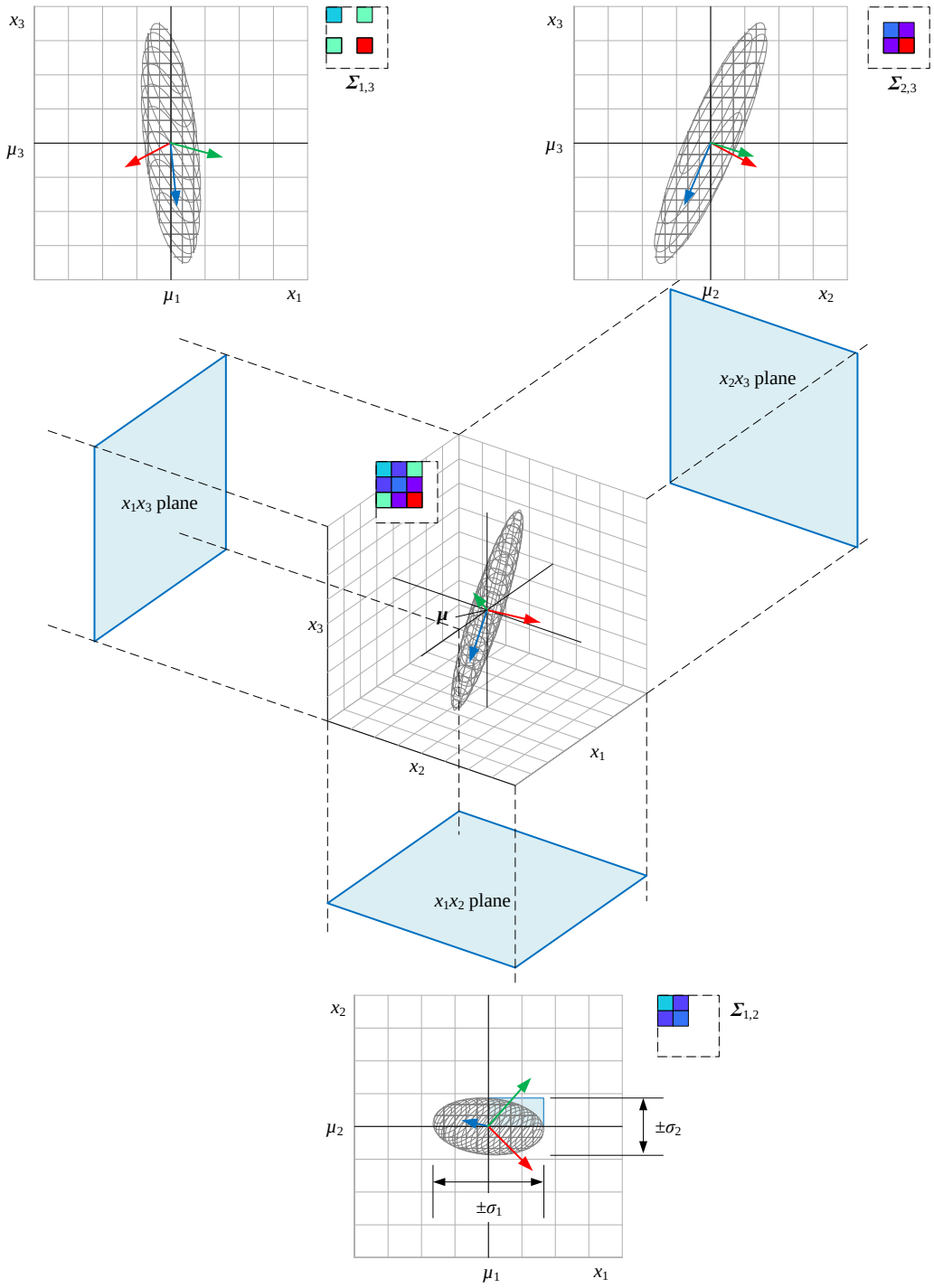


图 26. “旋转”椭球投影到三个二维平面

不考虑 x_i、x_j $(i \neq j)$ 顺序的话，$\mathbb{R}^4$ 中超椭球朝 x_ix_j 面投影，一共可以获得 6 个不同平面上的椭圆投影结果，具体如图 27 所示。请大家自行分析图 27 中这 6 幅子图。

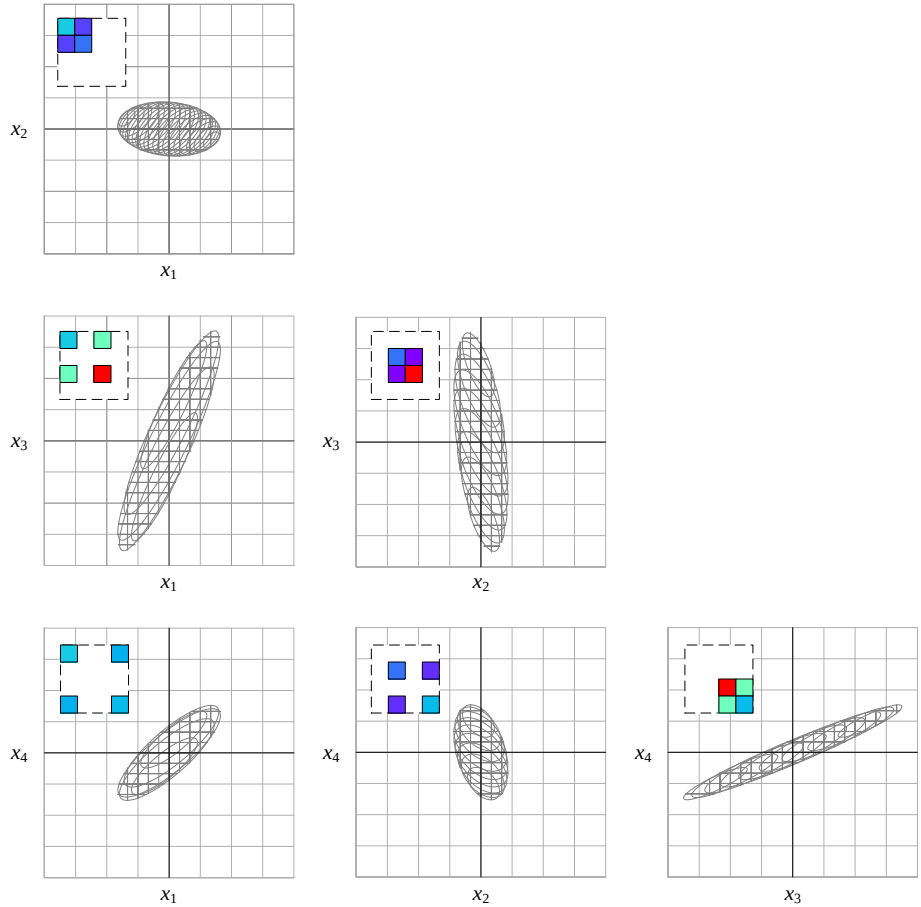


图 27. “旋转”超椭球在 6 个平面上的投影结果

矩形的面积、对角线长度

如图 28 (a) 所示，椭圆相切于矩形的四条边。该矩形的四个顶点分别是 $(\mu_1-\sigma_1, \mu_2-\sigma_2)$、$(\mu_1-\sigma_1, \mu_2+\sigma_2)$、$(\mu_1+\sigma_1, \mu_2+\sigma_2)$、$(\mu_1+\sigma_1, \mu_2-\sigma_2)$。图 28 (c) 中矩形的四个顶点分别为 (μ_1, μ_2)、$(\mu_1, \mu_2+\sigma_2)$、$(\mu_1+\sigma_1, \mu_2+\sigma_2)$、$(\mu_1+\sigma_1, \mu_2)$。

图 28 (a) 所示矩形的面积为 $4\sigma_1\sigma_2$，而图 28 (c) 中矩形为图 28 (a) 矩形的 1/4，对应面积为 $\sigma_1\sigma_2$。

图 28 (c) 中 1/4 矩形对角线长度为 $\sqrt{\sigma_1^2+\sigma_2^2}$，这个值是其协方差迹的平方根，即：

$$\sqrt{\sigma_1^2+\sigma_2^2}=\sqrt{\mathrm{tr}\left(\boldsymbol{\Sigma}_{2\times 2}\right)} \tag{82}$$

图 28 (b) 所示矩形也和椭圆相切于四条边，两组对边分别平行于 $\boldsymbol{v}_1$、$\boldsymbol{v}_2$。这个矩形的面积为 $4\sqrt{\lambda_1\lambda_2}$。而图 28 (d) 中矩形为图 28 (b) 矩形的 1/4，对应面积为 $\sqrt{\lambda_1\lambda_2}$。

$\sqrt{\lambda_1\lambda_2}$ 这个值协方差行列式值的平方根：

$$\sqrt{\lambda_1\lambda_2}=\sqrt{\left|\boldsymbol{\Lambda}_{2\times 2}\right|}=\sqrt{\left|\boldsymbol{\Sigma}_{2\times 2}\right|} \tag{83}$$

图 28 (d) 中 1/4 矩形对角线长度为 $\sqrt{\lambda_1+\lambda_2}$，和图 28 (c) 中矩形对角线长度相同，即：

$$\sqrt{\sigma_1^2+\sigma_2^2}=\sqrt{\mathrm{tr}\left(\boldsymbol{\Sigma}_{2\times 2}\right)}=\sqrt{\mathrm{tr}\left(\boldsymbol{\Lambda}_{2\times 2}\right)}=\sqrt{\lambda_1+\lambda_2} \tag{84}$$

这是本书下一章要讨论协方差的重要几何性质之一。

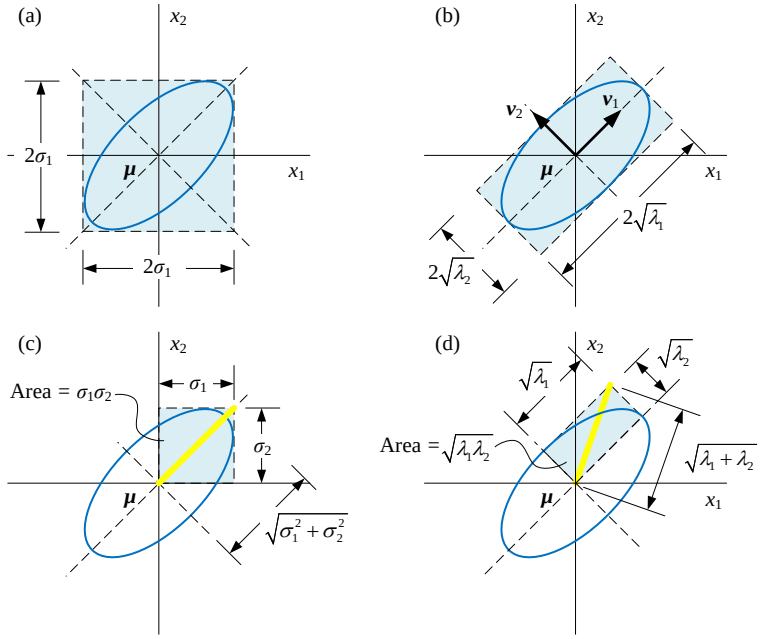


图 28. 和椭圆相切矩形的面积

“正”超椭球投影到三维空间

本节前文的“旋转”超椭球经过旋转之后得到“正”超椭球，这个“正”超椭球对应的协方差矩阵为 $\boldsymbol{\Lambda}$，具体值为：

$$\boldsymbol{\Lambda}=\begin{bmatrix}4.228 & & & \\ & 0.242 & & \\ & & 0.078 & \\ & & & 0.023\end{bmatrix} \tag{85}$$

这个“正”超椭球的解析式为：

$$\frac{y_1^2}{4.228}+\frac{y_2^2}{0.242}+\frac{y_3^2}{0.078}+\frac{y_4^2}{0.023}=1 \tag{86}$$

图 29 所示为“正”超椭球在四个三维空间中投影得到的椭球。其中，图 29 (a) 所示为“正”超椭球在 $y_1y_2y_3$ 这个三维空间的投影，对应的解析式为：

$$\frac{y_1^2}{4.228}+\frac{y_2^2}{0.242}+\frac{y_3^2}{0.078}=1 \tag{87}$$

图 29 (a) 中蓝、红、绿色箭头对应为上述“正”超椭球的第一、第二、第三主轴方向。

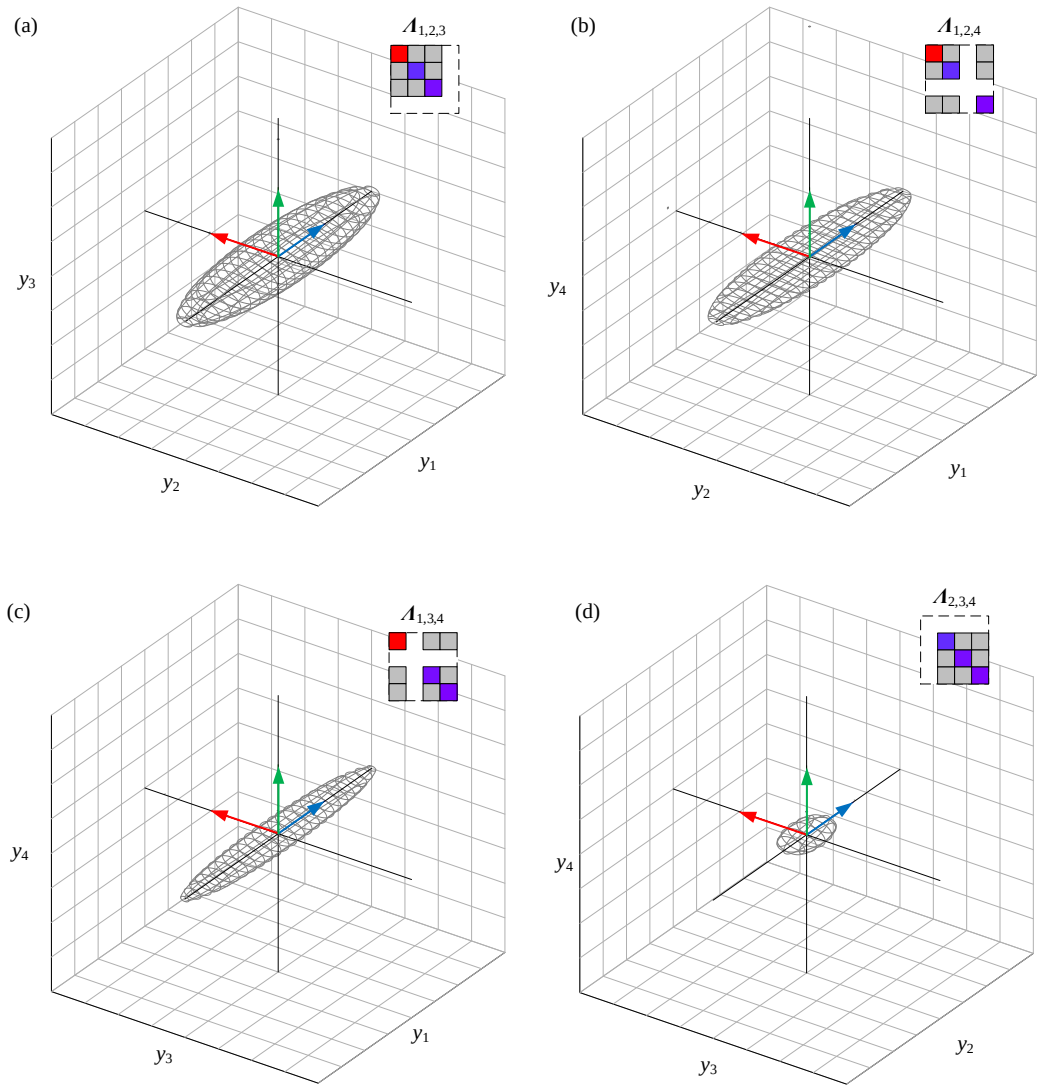


图 29 四维空间的“正”超椭球在三维空间中的四个投影

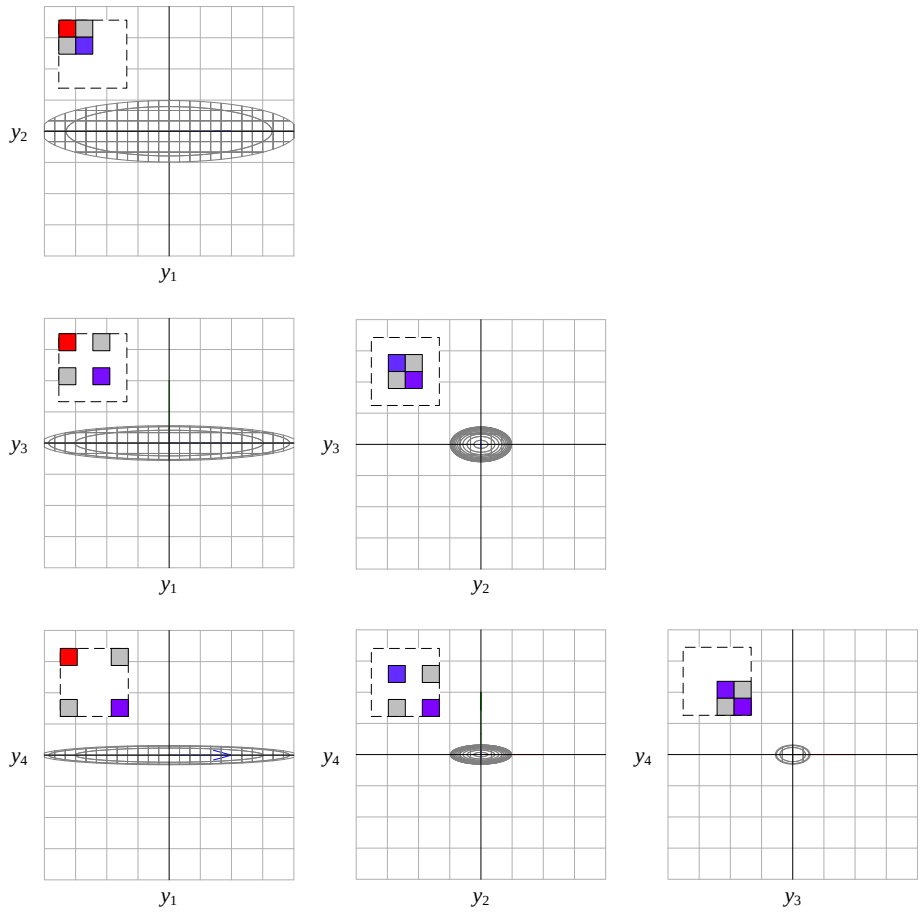


图 30. “正”超椭球在 6 个平面上的投影结果

相关性系数矩阵

大家是否立刻想到，相关性系数矩阵 $\boldsymbol{P}$ 也可以做特征值分解，也就是说 $\boldsymbol{P}$ 也可以有类似的几何解释。

根据图 16，相关性系数矩阵 $\boldsymbol{P}$ 对应的超椭球的半主轴长度分别为 $\sqrt{2.918}=1.708$、$\sqrt{0.914}=0.956$、$\sqrt{0.146}=0.383$、$\sqrt{0.021}=0.143$。

图 31 所示为相关系数矩阵所代表的四维空间的“旋转”超椭球在三维空间中的四个投影。图 32 所示为这个超椭圆在六个平面的投影。请大家自行分析这两幅图，特别是方差、标准差。注意，相关性系数矩阵可以视作 z 分数的协方差矩阵。

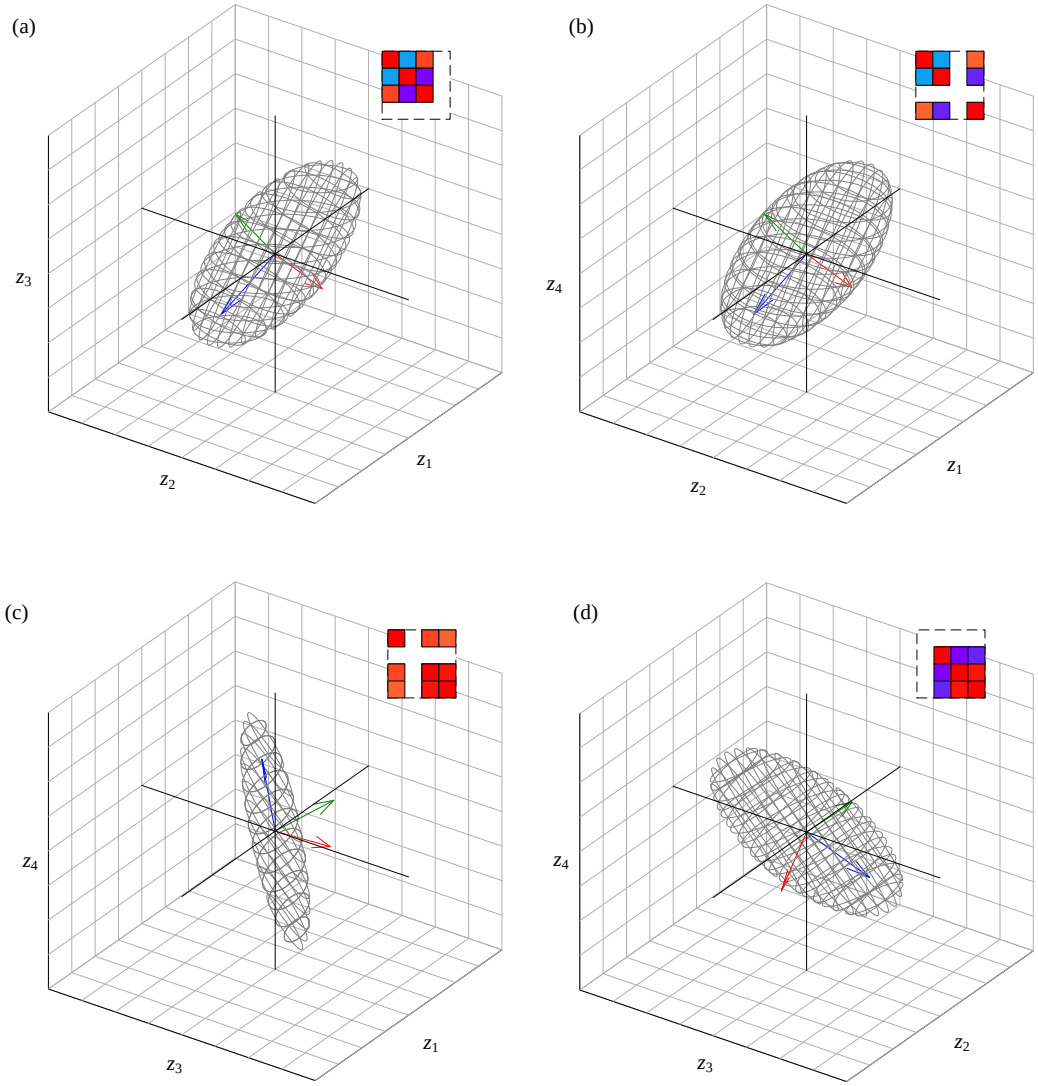


图 31 四维空间的“旋转”超椭球在三维空间中的四个投影，相关系数矩阵

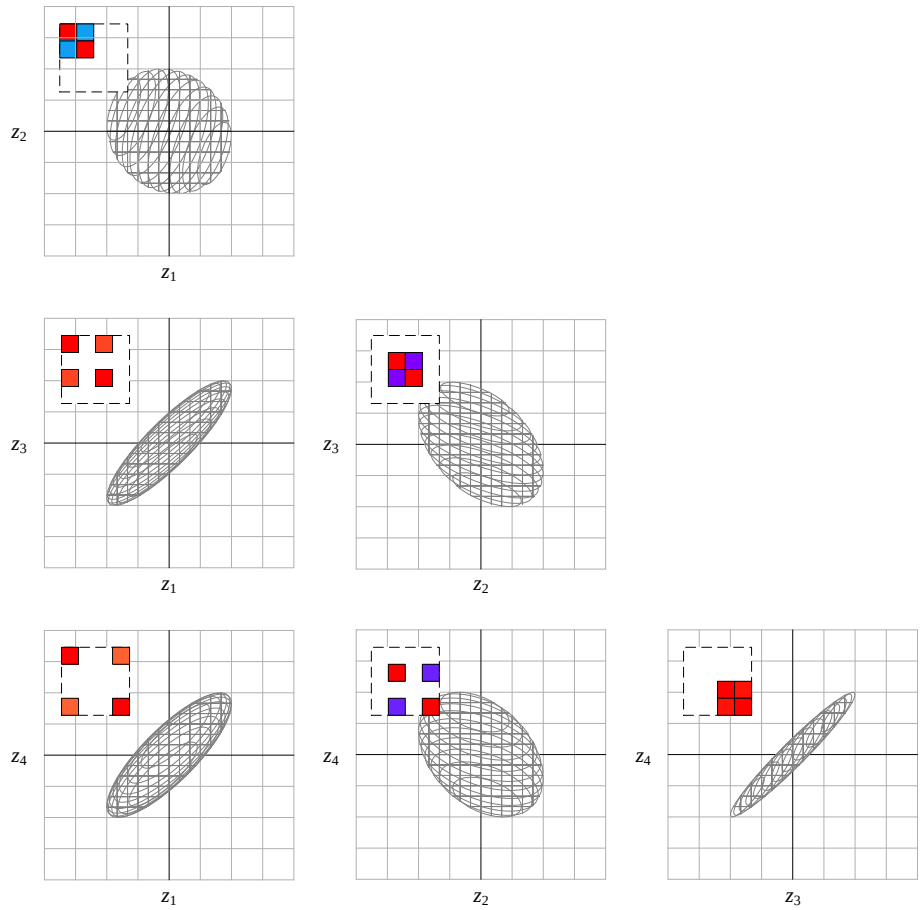


图 32. “旋转”超椭球在 6 个平面上的投影结果，相关性系数矩阵

13.8 合并协方差矩阵

本节介绍一个概念——**合并协方差矩阵** (pooled covariance matrix)，定义为：

$$\boldsymbol{\Sigma}_{\text{pooled}} = \frac{1}{\sum_{k=1}^{K}(n_k - 1)} \sum_{k=1}^{K}(n_k - 1)\boldsymbol{\Sigma}_k = \frac{1}{n-K} \sum_{k=1}^{K}(n_k - 1)\boldsymbol{\Sigma}_k \tag{88}$$

其中，n 代表总体样本数，n_k 为标签为 C_k 的样本数，K 为标签数量。$\boldsymbol{\Sigma}_i$ 是标签为 C_k 的样本数据协方差矩阵。

上式相当于加权平均，这么做是为了保证整体协方差矩阵的无偏性，因为每个组内的样本数可能不同，直接将所有样本合并起来计算协方差矩阵可能会导致估计偏差。

如果假设分类质心重叠，合并协方差矩阵可以用来估算样本整体方差。合并协方差矩阵可以用来比较不同子集的协方差之间的差异，也就是不同分类标签数据的分布情况。此外，我们会在“鸢尾花书”《数据有道》主成分分析中看到合并协方差的应用。

以鸢尾花数据矩阵为例，总体样本数为 $n = 150$，一共有三种 ($K = 3$) 标签 C_1、C_2、C_3，分别对应的样本数为 $n_1 = 50$、$n_2 = 50$、$n_3 = 50$。合并协方差矩阵为：

$$\boldsymbol{\Sigma}_{\text{pooled}} = \frac{1}{150-3}\sum_{k=1}^{3}(50-1)\boldsymbol{\Sigma}_k = \frac{49}{147}\times(\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2+\boldsymbol{\Sigma}_3) \tag{89}$$

图 33 中三个彩色的椭圆代表 $\boldsymbol{\Sigma}_1$、$\boldsymbol{\Sigma}_2$、$\boldsymbol{\Sigma}_3$，对应马氏距离为 1。注意，图 33 中并没有展示 $\boldsymbol{\Sigma}_{\text{pooled}}$。

图 33 中 $\boldsymbol{\Sigma}$ 代表整体数据协方差矩阵。$\boldsymbol{\Sigma}$ 完全不同于 $\boldsymbol{\Sigma}_{\text{pooled}}$。也可以说，$\boldsymbol{\Sigma}_{\text{pooled}}$ 只是 $\boldsymbol{\Sigma}$ 的一部分。$\boldsymbol{\Sigma}_{\text{pooled}}$ 仅仅考虑子集内部的数据分布，没有考虑子集之间的分布差异 (分类质心的差异)。因此，图 33 中 $\boldsymbol{\Sigma}$ 对应的旋转椭圆远大于 $\boldsymbol{\Sigma}_1$、$\boldsymbol{\Sigma}_2$、$\boldsymbol{\Sigma}_3$。

换个角度来看，合并协方差矩阵相当于，全方差定理中的条件方差的期望，缺少的成分是条件期望的方差。

为了方便比较不同分类协方差矩阵，我们可以将所有椭圆中心重合，得到图 34。$\boldsymbol{\Sigma}_{\text{pooled}}$ 的对应图 34 中的黑色划线椭圆。比较彩色椭圆和黑色划线椭圆，可以知道不同标签数据分布之间差异。

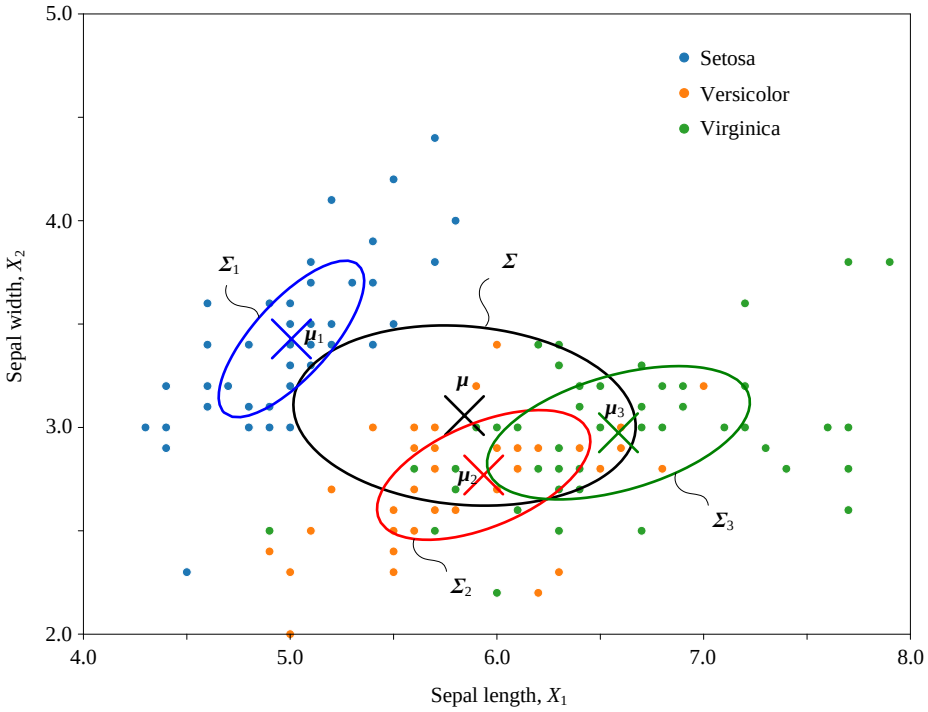


图 33. 分类协方差矩阵、整体协方差矩阵马氏距离为 1 椭圆，花萼长度、花萼宽度

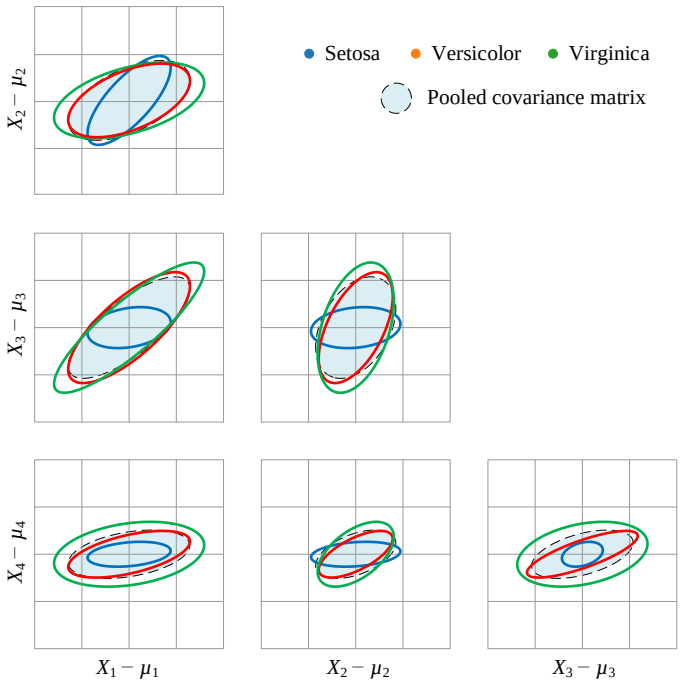


图 34. 马氏距离 1 椭圆，Σ_1、Σ_2、Σ_3和合并协方差矩阵 Σ_{pooled}

这一章结束了本书“高斯”这一板块。这个板块以高斯分布为主线，分别介绍了一元、二元、多元、条件高斯分布，最后介绍了多元高斯分布中的主角——协方差矩阵。相信通过这几章的学习，大家已经看到了线性代数工具在多元统计中的重要作用。

多元统计数据通常表示为向量或矩阵形式，线性代数提供了处理和计算这些对象的基本工具。例如，我们可以使用矩阵运算来计算协方差矩阵、进行线性变换、求解线性方程组等。

在多元统计中，特征值和特征向量是非常重要的概念。通过计算特征值和特征向量，我们可以识别出数据中的主要方向和结构，从而进行降维、聚类、分类等任务。

奇异值分解被广泛用于主成分分析 (PCA)、矩阵分解、压缩和图像处理等任务中。此外，我们可以使用特征值分解或奇异值分解来分析数据的主要结构和变化模式，使用矩阵迹、行列式等概念来计算协方差矩阵的性质，使用矩阵乘法、转置等运算来进行矩阵变换等。

在多元统计中，很多问题可以被视为一个优化问题。线性代数提供了很多优化方法和技巧，例如梯度下降、牛顿法、共轭梯度法等，可以用来解决最小化误差、最大化似然等问题。大家会在本书后续看到更多线性代数在多元统计、数据分析、机器学习领域的应用。

协方差估计的方法还有很多，请大家参考：

https://scikit-learn.org/stable/modules/covariance.html

有关合并协方差矩阵，请大家参考：

https://arxiv.org/pdf/1805.05756.pdf