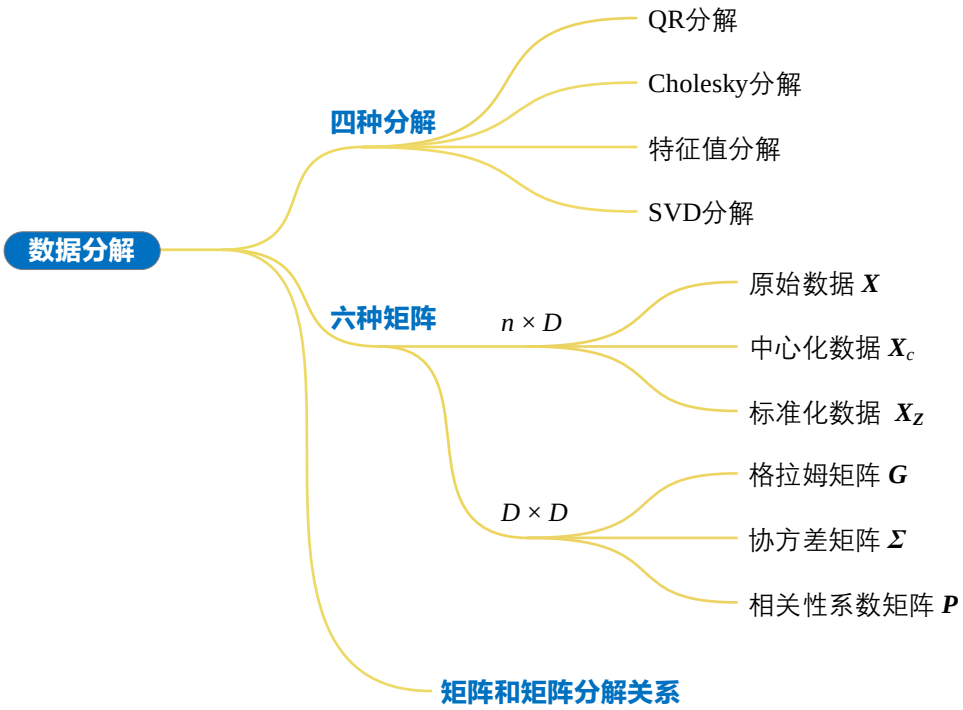
24 数据分解

Data Matrix Decomposition

从几何、空间、优化、统计视角解读

> 你不能教任何人任何东西，你只能帮助他在自己身上发现它。
>
> ***You cannot teach a man anything; you can only help him discover it in himself.***
>
> —— 伽利略·伽利莱 (Galilei Galileo) | 意大利物理学家、数学家及哲学家 | 1564 ~ 1642

- `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()` 完成 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



24.1 为什么要分解矩阵?

QR 分解、Cholesky 分解、特征值分解、SVD 分解，这四种常用的分解的目的是什么?

它们分解的对象到底是什么？有何应用条件?

每种矩阵分解结果是什么？有何特殊性质?

它们之间有哪些区分和联系?

灵魂拷问来了——我们到底为什么需要分解矩阵?

大家可能会反问，前文学都学完了，现在才问是不是太晚了?

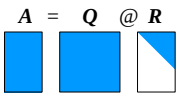
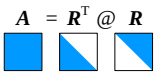
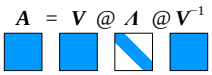
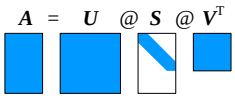
一点也不晚！矩阵分解是线性代数的核心中的核心，现在正是时候结合数据、几何、空间、优化、统计等视角观察比较这四种矩阵分解的最佳时机。

总结和比较

表 1 总结比较四种常用矩阵分解，请大家快速浏览这个表格，然后再开始本章的学习。也请大家在完成本章内容学习后，再回头仔细看一遍表格内容。如果对任何矩阵分解细节感到生疏的话，请翻看本书前文对应内容。

本章后续内容主要比较 Cholesky 分解、特征值分解、SVD 分解，它们三者是数据科学和机器学习中最常用的矩阵分解。

表 1. 比较四种常用矩阵分解

矩阵分解	QR 分解	Cholesky 分解	特征值分解	SVD 分解
前提	任何实数矩阵都可以 QR 分解	正定矩阵才能 Cholesky 分解	可对角化矩阵才能进行特征值分解	任何实数矩阵都可以 SVD 分解
示意图				
公式	$\boldsymbol{A}=\boldsymbol{QR}$	$\boldsymbol{A}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$ $\boldsymbol{A}=\boldsymbol{LL}^{\mathrm{T}}$	$\boldsymbol{A}=\boldsymbol{V\Lambda V}^{-1}$ $\boldsymbol{A}=\boldsymbol{V\Lambda V}^{\mathrm{T}}$ ($\boldsymbol{A}$ 为对称方阵时)	$\boldsymbol{A}=\boldsymbol{USV}^{\mathrm{T}}$ (注意 $\boldsymbol{V}$ 的转置运算)
结果	$\boldsymbol{Q}$ 是正交矩阵，意味着 $\boldsymbol{Q}$ 是规范正交基 $\boldsymbol{R}$ 是上三角矩阵	$\boldsymbol{L}$ 为下三角方阵 $\boldsymbol{R}$ 为上三角方阵	$\boldsymbol{\Lambda}$ 为对角方阵，对角线元素为特征值 如果 $\boldsymbol{A}$ 为对称方阵，$\boldsymbol{V}$ 为正交矩阵	$\boldsymbol{U}$ 为正交矩阵，它的列向量为左奇异向量 $\boldsymbol{S}$ 主对角线元素为奇异值 $\boldsymbol{V}$ 为正交矩阵，它的列向量为右奇异向量 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 都是规范正交基
结果唯一?	$\boldsymbol{A}$ 列满秩，且 $\boldsymbol{R}$ 的对角元素为正实数的情况下	当限定 $\boldsymbol{R}$ 的对角元素为正时，这时分解结	不唯一，但是本书的特征向量都是单位化，特征向量一般差在正负符	矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 不唯一

	结果唯一	果唯一	号上	
特殊类型	完全型 经济型	正定矩阵 埃尔米特矩阵 (复数矩阵，不在本书讨论范围之内)	对称矩阵 正规矩阵 ($\boldsymbol{A}$ 是埃尔米特矩阵，不在本书讨论范围之内)	完全型 经济型 缩略型 截断型
向量空间	$\boldsymbol{Q}$ 的列向量为规范正交基，$\boldsymbol{Q}$ 的第一列向量 $\boldsymbol{q}_1$ 是 $\boldsymbol{A}$ 的第一列向量 $\boldsymbol{a}_1$ 的单位化结果 $\boldsymbol{R}$ 的列向量相当于坐标值	如果 $\boldsymbol{A}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ (即 Gram 矩阵) 且正定，$\boldsymbol{A}$ 进行 Cholesky 分解得到上三角矩阵 $\boldsymbol{R}$，$\boldsymbol{R}$ 的列向量可以代表 $\boldsymbol{X}$ 列向量	如果 $\boldsymbol{A}$ 为对称方阵，$\boldsymbol{V}$ 为规范正交基 如果 $\boldsymbol{A}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 且 $\boldsymbol{X}$ 列满秩，$\boldsymbol{V}$ 是 $\boldsymbol{X}$ 的行空间 $R(\boldsymbol{X})$	完全型 SVD 分解获得四个空间：列空间 $C(\boldsymbol{X})$ 和左零空间 Null($\boldsymbol{X}^{\mathrm{T}}$)，行空间 $R(\boldsymbol{X})$ 和零空间 Null($\boldsymbol{X}$)
Numpy 函数	`numpy.linalg.qr()`	`numpy.linalg.cholesky()`	`numpy.linalg.eig()`	`numpy.linalg.svd()`
本章对象	原始数据矩阵 $\boldsymbol{X}$	格拉姆矩阵 $\boldsymbol{G}$ ($\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$) 协方差矩阵 $\boldsymbol{\Sigma}$ 相关性系数矩阵 $\boldsymbol{P}$	格拉姆矩阵 $\boldsymbol{G}$ ($\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$) 协方差矩阵 $\boldsymbol{\Sigma}$ 相关性系数矩阵 $\boldsymbol{P}$	原始数据矩阵 $\boldsymbol{X}$ 中心化数据矩阵 $\boldsymbol{X}_c$ 标准化数据矩阵 $\boldsymbol{Z}_{\boldsymbol{X}}$
本系列丛书主要应用	解线性方程组 最小二乘回归 施密特正交化	蒙特卡罗模拟，产生满足特定协方差矩阵要求的随机数 判断正定性	马尔科夫过程 主成分分析 瑞利商	求解伪逆矩阵 最小二乘回归 主成分分析 图像压缩

数据矩阵，衍生矩阵

本章用的数据还是大家很熟悉的鸢尾花数据集。

快速回顾一下，如图 1 所示，鸢尾花数据矩阵 $\boldsymbol{X}$ 的每一列分别代表鸢尾花的不同特征——花萼长度 (第 1 列，列向量 $\boldsymbol{x}_1$)、花萼宽度 (第 2 列，列向量 $\boldsymbol{x}_2$)、花瓣长度 (第 3 列，列向量 $\boldsymbol{x}_3$) 和花瓣宽度 (第 4 列，列向量 $\boldsymbol{x}_4$)。矩阵 $\boldsymbol{X}$ 的每一行代表一朵花的样本数据，每一行数据也是一个向量——行向量。图 1 不考虑鸢尾花分类。

图 1. 鸢尾花数据集行、列含义

图 2 所示为本章矩阵分解对象，它们都衍生自鸢尾花数据矩阵 $\boldsymbol{X}$，$\boldsymbol{X}$ 为细高长方形矩阵，形状为 $n \times D$。和本书第 10 章鸢尾花数据矩阵热图相比，图 2 中 $\boldsymbol{X}$ 的热图采用不同颜色范围。

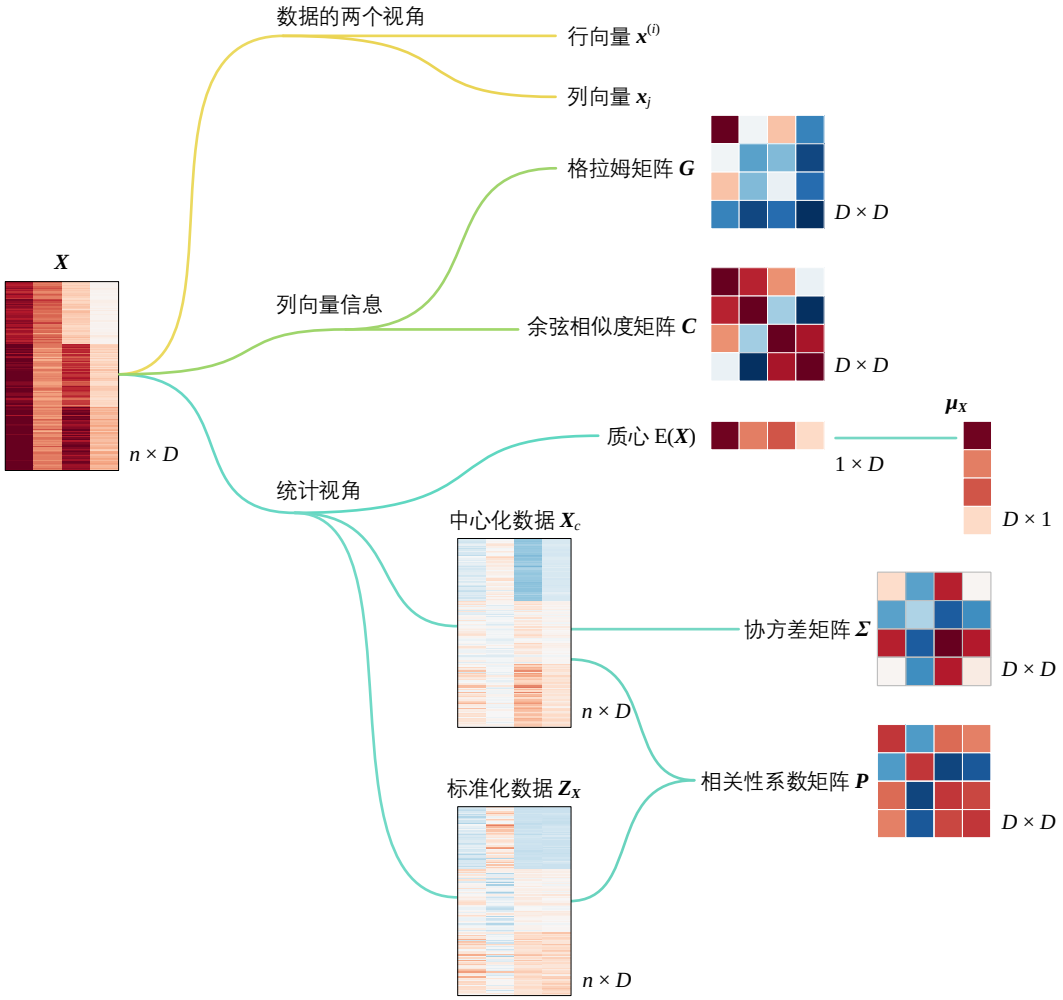


图 2. 鸢尾花数据衍生得到的几个矩阵

格拉姆矩阵 $\boldsymbol{G}$ 来自数据矩阵 $\boldsymbol{X}$，两者关系为 $\boldsymbol{G} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$。上一章提到，格拉姆矩阵 $\boldsymbol{G}$ 含有 $\boldsymbol{X}$ 列向量模、向量夹角两类重要信息。对于细高的长方形矩阵 $\boldsymbol{X}$，第二个格拉姆矩阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 不常用。

而余弦相似度矩阵 $\boldsymbol{C}$ 仅有 $\boldsymbol{X}$ 列向量两两夹角信息。余弦相似度的取值 $[-1, 1]$，因此不同余弦相似度具有可比性。这一点类似统计中的相关性系数。对 $\boldsymbol{X}$ 列向量先进行单位化，再求其格拉姆矩阵，得到的就是 $\boldsymbol{C}$。

在统计视角下，$\boldsymbol{X}$ 的两个重要信息——质心 E($\boldsymbol{X}$)、协方差矩阵 $\boldsymbol{\Sigma}$。E($\boldsymbol{X}$) 对应数据中心位置，$\boldsymbol{\Sigma}$ 描述数据分布。

⚠ 值得注意的是，本系列丛书定义 $E(\boldsymbol{X})$ 为行向量，$E(\boldsymbol{X})$ 的转置为列向量 $\boldsymbol{\mu}_X$。

和原始数据矩阵形状相同的矩阵有两个——中心化数据矩阵 $\boldsymbol{X}_c$、标准化数据矩阵 $\boldsymbol{Z}_X$。$\boldsymbol{X}$、$\boldsymbol{X}_c$、$\boldsymbol{Z}_X$ 的形状均为 $n \times D$。

$\boldsymbol{X}$ 每一列数据分别减去自己列的均值便得到中心化数据 $\boldsymbol{X}_c$，即 $\boldsymbol{X}_c = \boldsymbol{X} - E(\boldsymbol{X})$。这个式子用到了广播原则。请大家回顾如何采用矩阵运算计算 $\boldsymbol{X}_c$。

几何视角，对于 $\boldsymbol{X}$ 来说，它的数据质心位于 $\boldsymbol{\mu}_X$；而 $\boldsymbol{X}_c$ 的质心位于 $\boldsymbol{0}$。$\boldsymbol{X}$ 的列向量起点位于原点；而 $\boldsymbol{X}_c$ 列向量的起点相当于移动到了质心，向量终点不动。

标准化数据 $\boldsymbol{Z}_X$ 实际上就是 $\boldsymbol{X}$ 的 z 分数。几何视角，从 $\boldsymbol{X}$ 到 $\boldsymbol{Z}_X$ 经过了平移、缩放两步操作。

⚠ 注意，上一章创造了一个概念——标准差向量。标准差向量的模对应标准差大小，两个标准差向量的夹角余弦值对应相关性系数。

协方差矩阵 $\boldsymbol{\Sigma}$ 包含两类信息——标准差向量的模 (标准差)、两两夹角 (相关性系数)。协方差矩阵 $\boldsymbol{\Sigma}$ 完全类似格拉姆矩阵 $\boldsymbol{G}$。不同的是，协方差矩阵 $\boldsymbol{\Sigma}$ 数据中心化，而且存在缩放系数 $1/n$ (总体) 或 $1/(n-1)$ (样本)。

协方差矩阵 $\boldsymbol{\Sigma}$ 可以视作 $\boldsymbol{X}_c$ 的格拉姆矩阵，唯一差别就是缩放系数。

相关性系数矩阵 $\boldsymbol{P}$ 仅仅含有标准差向量夹角 (相关性系数) 信息。相关性系数矩阵 $\boldsymbol{P}$ 类似余弦相似度矩阵 $\boldsymbol{C}$。类似协方差矩阵 $\boldsymbol{\Sigma}$，相关性系数矩阵 $\boldsymbol{P}$ 也存在缩放系数 $1/n$ (总体) 或 $1/(n-1)$ (样本)。相关性系数矩阵 $\boldsymbol{P}$ 就是标准化数据 $\boldsymbol{Z}_X$ 的协方差矩阵。

$\boldsymbol{G}$、$\boldsymbol{C}$、$\boldsymbol{\Sigma}$、$\boldsymbol{P}$ 的形状均为 $D \times D$。

如果大家对这部分内容感到陌生，请回顾本书第 22 章。大家必须对矩阵分解的对象有充分的认识，才能开始本章后续内容学习。

矩阵 + 矩阵分解

搭配六种不同矩阵和三种矩阵分解，会碰撞出什么？

表 2 给出了答案。这张表总结了不同矩阵和矩阵分解之间的关系，本章后续内容将主要以表格中内容展开。

表 2. 矩阵和矩阵分解之间的关系

对象		Cholesky 分解	特征值分解	SVD 分解
$n \times D$	$\boldsymbol{X}$	不适用	不适用	$\boldsymbol{X} = \boldsymbol{U}_X\boldsymbol{S}_X\boldsymbol{V}_X^T$
	$\boldsymbol{X}_c = \boldsymbol{X} - E(\boldsymbol{X})$	不适用	不适用	$\boldsymbol{X}_c = \boldsymbol{U}_c\boldsymbol{S}_c\boldsymbol{V}_c^T$
	$\boldsymbol{Z}_X = (\boldsymbol{X} - E(\boldsymbol{X}))\boldsymbol{S}^{-1}$ $\boldsymbol{S} = \text{diag}(\text{diag}(\boldsymbol{\Sigma}))^{\frac{1}{2}}$	不适用	不适用	$\boldsymbol{Z}_X = \boldsymbol{U}_Z\boldsymbol{S}_Z\boldsymbol{V}_Z^T$

$D \times D$	$\boldsymbol{G} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$	正定矩阵 $\boldsymbol{G} = \boldsymbol{R}_{\boldsymbol{X}}{}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{X}}$	$\boldsymbol{G} = \boldsymbol{V}_{\boldsymbol{X}}\boldsymbol{\Lambda}_{\boldsymbol{X}}\boldsymbol{V}_{\boldsymbol{X}}{}^{\mathrm{T}} = \boldsymbol{V}_{\boldsymbol{X}}\boldsymbol{S}_{\boldsymbol{X}}{}^{\mathrm{T}}\boldsymbol{S}_{\boldsymbol{X}}\boldsymbol{V}_{\boldsymbol{X}}{}^{\mathrm{T}}$	$\boldsymbol{G} = \boldsymbol{V}_{\boldsymbol{X}}\boldsymbol{\Lambda}_{\boldsymbol{X}}\boldsymbol{V}_{\boldsymbol{X}}{}^{\mathrm{T}}$
	样本：$\boldsymbol{\Sigma} = \frac{(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))^{\mathrm{T}}(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))}{n-1}$ 总体：$\boldsymbol{\Sigma} = \frac{(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))^{\mathrm{T}}(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))}{n}$	正定矩阵 $\boldsymbol{\Sigma} = \boldsymbol{R}_{\boldsymbol{C}}{}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{C}}$	样本：$\boldsymbol{\Sigma} = \boldsymbol{V}_{\boldsymbol{C}}\boldsymbol{\Lambda}_{\boldsymbol{C}}\boldsymbol{V}_{\boldsymbol{C}}{}^{\mathrm{T}} = \boldsymbol{V}_{\boldsymbol{C}}\boldsymbol{S}_{\boldsymbol{C}}{}^{\mathrm{T}}\boldsymbol{S}_{\boldsymbol{C}}/(n-1)\boldsymbol{V}_{\boldsymbol{C}}{}^{\mathrm{T}}$ 总体：$\boldsymbol{\Sigma} = \boldsymbol{V}_{\boldsymbol{C}}\boldsymbol{\Lambda}_{\boldsymbol{C}}\boldsymbol{V}_{\boldsymbol{C}}{}^{\mathrm{T}} = \boldsymbol{V}_{\boldsymbol{C}}\boldsymbol{S}_{\boldsymbol{C}}{}^{\mathrm{T}}\boldsymbol{S}_{\boldsymbol{C}}/n\boldsymbol{V}_{\boldsymbol{C}}{}^{\mathrm{T}}$	$\boldsymbol{\Sigma} = \boldsymbol{V}_{\boldsymbol{C}}\boldsymbol{\Lambda}_{\boldsymbol{C}}\boldsymbol{V}_{\boldsymbol{C}}{}^{\mathrm{T}}$
	$\boldsymbol{P} = \boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}$ $\boldsymbol{S} = \mathrm{diag}(\mathrm{diag}(\boldsymbol{\Sigma}))^{\frac{1}{2}}$	正定矩阵 $\boldsymbol{\Sigma} = \boldsymbol{R}_{\boldsymbol{Z}}{}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{Z}}$	样本：$\boldsymbol{P} = \boldsymbol{V}_{\boldsymbol{Z}}\boldsymbol{\Lambda}_{\boldsymbol{Z}}\boldsymbol{V}_{\boldsymbol{Z}}{}^{\mathrm{T}} = \boldsymbol{V}_{\boldsymbol{Z}}\boldsymbol{S}_{\boldsymbol{Z}}{}^{\mathrm{T}}\boldsymbol{S}_{\boldsymbol{Z}}/(n-1)\boldsymbol{V}_{\boldsymbol{Z}}{}^{\mathrm{T}}$ 总体：$\boldsymbol{P} = \boldsymbol{V}_{\boldsymbol{Z}}\boldsymbol{\Lambda}_{\boldsymbol{Z}}\boldsymbol{V}_{\boldsymbol{Z}}{}^{\mathrm{T}} = \boldsymbol{V}_{\boldsymbol{Z}}\boldsymbol{S}_{\boldsymbol{Z}}{}^{\mathrm{T}}\boldsymbol{S}_{\boldsymbol{Z}}/n\boldsymbol{V}_{\boldsymbol{Z}}{}^{\mathrm{T}}$	$\boldsymbol{P} = \boldsymbol{V}_{\boldsymbol{Z}}\boldsymbol{\Lambda}_{\boldsymbol{Z}}\boldsymbol{V}_{\boldsymbol{Z}}{}^{\mathrm{T}}$

Bk4_Ch24_01.py 中 Bk4_Ch24_01_A 部分计算得到图 2 所有矩阵，请读者根据前文所学自行绘制本章所有热图。

24.2 QR 分解：获得正交系

QR 分解不是本章重点，我们仅仅蜻蜓点水回顾一下。

如图 3 所示，对矩阵 $\boldsymbol{X}$ 进行缩略型 QR 分解，得到 $\boldsymbol{Q}$ 和 $\boldsymbol{R}$。$\boldsymbol{Q}$ 是正交矩阵的一部分，也就是说 $\boldsymbol{Q}$ 的列向量 $[\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \boldsymbol{q}_4]$ 是规范正交基。$[\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \boldsymbol{q}_4]$ 相当于 $[\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4]$ 正交化的结果。

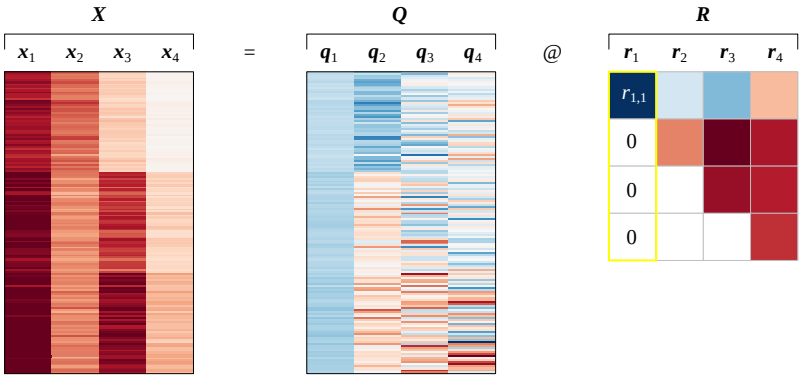


图 3. QR 分解热图

如图 4 所示，从空间角度来讲，如果 $\boldsymbol{x}_1$、$\boldsymbol{x}_2$、$\boldsymbol{x}_3$、$\boldsymbol{x}_4$ 线性无关，则 $\mathrm{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4) = \mathrm{span}(\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \boldsymbol{q}_4)$。

请大家特别关注如下关系：

$$\boldsymbol{x}_1 = r_{1,1}\boldsymbol{q}_1 \tag{1}$$

也就是说 $\boldsymbol{x}_1$ 和 $\boldsymbol{q}_1$ 平行。取决于 $r_{1,1}$ 正负，$\boldsymbol{x}_1$ 和 $\boldsymbol{q}_1$ 可以同向或反向。

$(r_{1,1}, 0, 0, 0)$ 是 $\boldsymbol{x}_1$ 在 $[\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \boldsymbol{q}_4]$ 的坐标。

⚠ 此外请大家注意，QR 分解和格拉姆-施密特正交化 (Gram–Schmidt process) 之间联系。

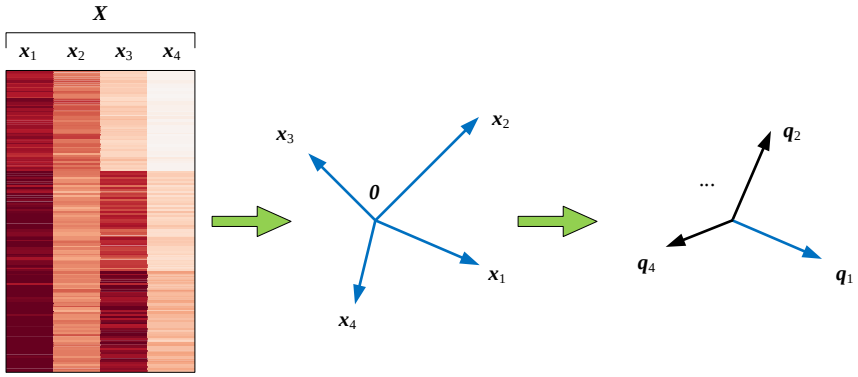


图 4. $[\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \boldsymbol{q}_4]$ 是规范正交基

`Bk4_Ch24_01.py` 中 `Bk4_Ch24_01_B` 部分完成矩阵 $\boldsymbol{X}$ 的 QR 分解。

24.3 Cholesky 分解：找到列向量的坐标

格拉姆矩阵

数据矩阵 $\boldsymbol{X}$ 的每一列可以看做一个向量，而 Cholesky 分解能够找到它们的坐标。

⚠ 注意，这里存在一个前提——$\boldsymbol{X}$ 列满秩。只有这样 $\boldsymbol{X}$ 的格拉姆矩阵 $\boldsymbol{G}$ 才正定，才能进行 Cholesky 分解。

假设 $\boldsymbol{G}$ 正定，对 $\boldsymbol{G}$ 进行 Cholesky 分解：

$$\boldsymbol{G} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} \tag{2}$$

其中，$\boldsymbol{R}$ 为上三角矩阵。

⚠ 注意，(2) 中的 $\boldsymbol{R}$ 不同于上一节 QR 分解的 $\boldsymbol{R}$。

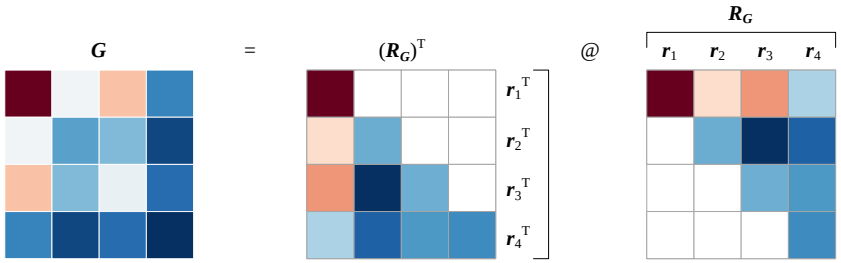


图 5. 对格拉姆矩阵 $\boldsymbol{G}$ 进行 Cholesky 分解矩阵运算热图

如图 5 所示，将 $\boldsymbol{R}$ 写成 $[\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_D]$，(2) 可以写成向量标量积形式，并建立它们和 $[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D]$ 的联系：

$$
\begin{aligned}
\boldsymbol{G} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} &= \begin{bmatrix} \langle \boldsymbol{r}_1, \boldsymbol{r}_1 \rangle & \langle \boldsymbol{r}_1, \boldsymbol{r}_2 \rangle & \cdots & \langle \boldsymbol{r}_1, \boldsymbol{r}_D \rangle \\ \langle \boldsymbol{r}_2, \boldsymbol{r}_1 \rangle & \langle \boldsymbol{r}_2, \boldsymbol{r}_2 \rangle & \cdots & \langle \boldsymbol{r}_2, \boldsymbol{r}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{r}_D, \boldsymbol{r}_1 \rangle & \langle \boldsymbol{r}_D, \boldsymbol{r}_2 \rangle & \cdots & \langle \boldsymbol{r}_D, \boldsymbol{r}_D \rangle \end{bmatrix} = \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} \\
&= \begin{bmatrix} \|\boldsymbol{r}_1\|\|\boldsymbol{r}_1\|\cos\theta_{1,1} & \|\boldsymbol{r}_1\|\|\boldsymbol{r}_2\|\cos\theta_{2,1} & \cdots & \|\boldsymbol{r}_1\|\|\boldsymbol{r}_D\|\cos\theta_{1,D} \\ \|\boldsymbol{r}_2\|\|\boldsymbol{r}_1\|\cos\theta_{1,2} & \|\boldsymbol{r}_2\|\|\boldsymbol{r}_2\|\cos\theta_{2,2} & \cdots & \|\boldsymbol{r}_2\|\|\boldsymbol{r}_D\|\cos\theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \|\boldsymbol{r}_D\|\|\boldsymbol{r}_1\|\cos\theta_{1,D} & \|\boldsymbol{r}_D\|\|\boldsymbol{r}_2\|\cos\theta_{2,D} & \cdots & \|\boldsymbol{r}_D\|\|\boldsymbol{r}_D\|\cos\theta_{D,D} \end{bmatrix} \\
&= \begin{bmatrix} \|\boldsymbol{x}_1\|\|\boldsymbol{x}_1\|\cos\theta_{1,1} & \|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\|\cos\theta_{2,1} & \cdots & \|\boldsymbol{x}_1\|\|\boldsymbol{x}_D\|\cos\theta_{1,D} \\ \|\boldsymbol{x}_2\|\|\boldsymbol{x}_1\|\cos\theta_{1,2} & \|\boldsymbol{x}_2\|\|\boldsymbol{x}_2\|\cos\theta_{2,2} & \cdots & \|\boldsymbol{x}_2\|\|\boldsymbol{x}_D\|\cos\theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \|\boldsymbol{x}_D\|\|\boldsymbol{x}_1\|\cos\theta_{1,D} & \|\boldsymbol{x}_D\|\|\boldsymbol{x}_2\|\cos\theta_{2,D} & \cdots & \|\boldsymbol{x}_D\|\|\boldsymbol{x}_D\|\cos\theta_{D,D} \end{bmatrix}
\end{aligned} \tag{3}
$$

$[\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_D]$ 的每个列向量的模分别等于 $[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D]$ 列向量的模；$[\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_D]$ 中两两向量夹角等于 $[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D]$ 中对应列向量夹角。

⚠ 再次强调，不考虑向量的具体数值时，两个特征就可以确定向量——模、向量间两两夹角。而格拉姆矩阵集成了这两部分信息。

也就是说，我们可以用上三角矩阵 $\boldsymbol{R}$ 的列向量 $[\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_D]$ 代表 $\boldsymbol{X}$ 列向量 $[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D]$。而 $\boldsymbol{R}$ 的形状为 $D \times D$，$\boldsymbol{X}$ 的形状为 $n \times D$。以鸢尾花数据为例，$\boldsymbol{R}$ 的形状为 4×4，$\boldsymbol{X}$ 的形状为 150×4。显然，$\boldsymbol{R}$ 远比 $\boldsymbol{X}$“小巧”的多。

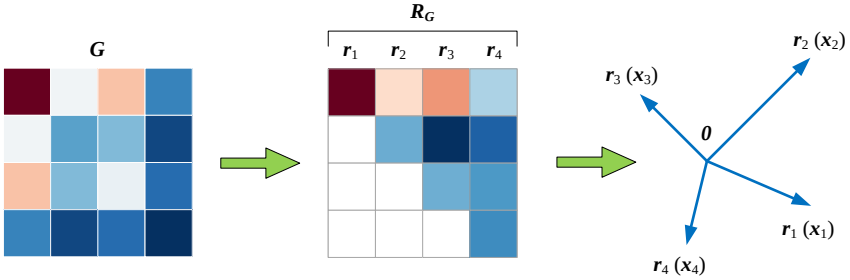


图 6. $[\boldsymbol{x}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_D]$ 和 $[\boldsymbol{r}_1, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_D]$ 等价

协方差矩阵

类似地，对协方差矩阵 $\boldsymbol{\Sigma}$ 进行 Cholesky 分解，具体如图 7 所示。将 $\boldsymbol{R}_c$ 写成“标准差向量” $[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots, \boldsymbol{\sigma}_D]$，整理得到：

$$\boldsymbol{\Sigma} = \boldsymbol{R}_c^{\mathrm{T}} \boldsymbol{R}_c = \begin{bmatrix} \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1 \rangle & \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \rangle & \cdots & \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_D \rangle \\ \langle \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_1 \rangle & \langle \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_2 \rangle & \cdots & \langle \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{\sigma}_D, \boldsymbol{\sigma}_1 \rangle & \langle \boldsymbol{\sigma}_D, \boldsymbol{\sigma}_2 \rangle & \cdots & \langle \boldsymbol{\sigma}_D, \boldsymbol{\sigma}_D \rangle \end{bmatrix} = \begin{bmatrix} \mathrm{cov}(X_1, X_1) & \mathrm{cov}(X_1, X_2) & \cdots & \mathrm{cov}(X_1, X_D) \\ \mathrm{cov}(X_2, X_1) & \mathrm{cov}(X_2, X_2) & \cdots & \mathrm{cov}(X_2, X_D) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_D, X_1) & \mathrm{cov}(X_D, X_2) & \cdots & \mathrm{cov}(X_D, X_D) \end{bmatrix} \tag{4}$$

$\boldsymbol{R}_c$ 将会用在蒙特卡洛模拟中，用来生成满足协方差矩阵 $\boldsymbol{\Sigma}$ 要求的随机数组，这是本系列丛书《概率统计》要将的内容。

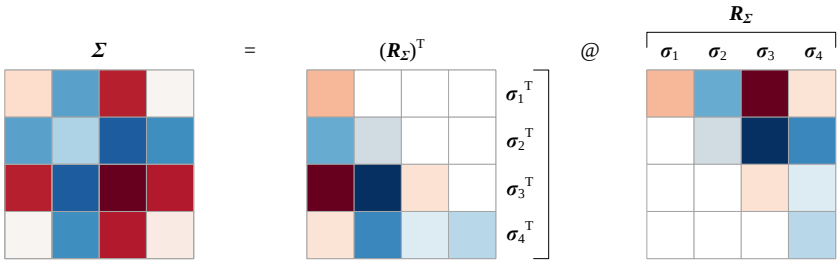


图 7. 对协方差矩阵 $\boldsymbol{\Sigma}$ 进行 Cholesky 分解矩阵运算热图

向量 $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots, \boldsymbol{\sigma}_D$ 的模分别对应 $\boldsymbol{x}_1$ (X_1), $\boldsymbol{x}_2$ (X_2), …, $\boldsymbol{x}_D$ (X_D) 的标准差，向量 $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots, \boldsymbol{\sigma}_D$ 两两夹角余弦值对应 $\boldsymbol{x}_1$ (X_1), $\boldsymbol{x}_2$ (X_2), …, $\boldsymbol{x}_D$ (X_D) 的两两线性相关系数。也就是说，协方差矩阵 $\boldsymbol{\Sigma}$ 集成了标准差和线性相关系数这两类信息。

如图 8 所示，$[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots, \boldsymbol{\sigma}_D]$ 相当于以数据 $\boldsymbol{X}$ 质心为中心一组非正交基。数据 $\boldsymbol{X}$ 的很多统计学运算和分析都是依托这个空间完成的。

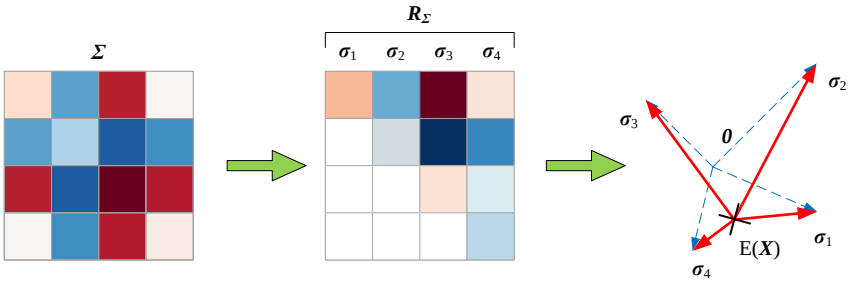
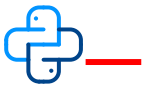


图 8. $[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots, \boldsymbol{\sigma}_D]$ 相当于以 $\boldsymbol{X}$ 质心为中心张成一个空间

当然，我们也可以对线性相关系数矩阵 $\boldsymbol{P}$ 进行 Cholesky 分解。实践中，对协方差矩阵 $\boldsymbol{\Sigma}$ 的 Cholesky 分解最常见。



`Bk4_Ch24_01.py` 中 `Bk4_Ch24_01_C` 部分完成对格拉姆矩阵 $\boldsymbol{G}$ 和协方差矩阵 $\boldsymbol{\Sigma}$ 的 Cholesky 分解。

24.4 特征值分解：获得行空间和零空间

本节要进行三个特征值分解，为了区分，我们在分解结果加了下角标。

格拉姆矩阵

图 9 所示为格拉姆矩阵 $\boldsymbol{G} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 进行特征值分解。因为 $\boldsymbol{G}$ 为对称矩阵，所以 $\boldsymbol{V}_X$ 为正交矩阵，即满足 $\boldsymbol{V}_X^{-1} = \boldsymbol{V}_X^{\mathrm{T}}$。从而，$\boldsymbol{G}$ 的特征值分解可以写成 $\boldsymbol{G} = \boldsymbol{V}_X\boldsymbol{\Lambda}_X\boldsymbol{V}_X^{\mathrm{T}}$。

根据上一章内容，$\boldsymbol{V}_X$ 的列向量 $[\boldsymbol{v}_{X\_1}, \boldsymbol{v}_{X\_2}, \ldots, \boldsymbol{v}_{X\_D}]$ 是一组规范正交基。$[\boldsymbol{v}_{X\_1}, \boldsymbol{v}_{X\_2}, \ldots, \boldsymbol{v}_{X\_D}]$ 张成 $\mathbb{R}^D$ 空间，它是矩阵 $\boldsymbol{X}$ 的行空间和零空间的合体。

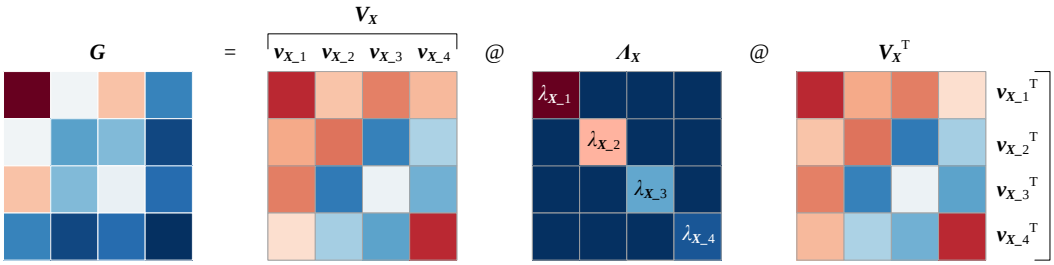


图 9. 对格拉姆矩阵进行特征值分解

如图 10 所示，从 $[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D]$ 到 $[\boldsymbol{v}_{X\_1}, \boldsymbol{v}_{X\_2}, \ldots, \boldsymbol{v}_{X\_D}]$ 相当于对 $[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D]$ 正交化。

⚠ 值得注意的是，本章矩阵 $\boldsymbol{X}$ 为鸢尾花数据，每一列数据单位都是厘米 (cm)。格拉姆矩阵 $\boldsymbol{G}$ 中数值的单位为平方厘米 cm^2。$\boldsymbol{V}_X$ 中每一列都是单位向量，仅仅表达方向，不含有单位。而特征值 λ_X 的单位为平方厘米 cm^2。从几何角度来看，特征值含有椭圆 (椭球) 的大小形状信息，而 $\boldsymbol{V}$ 仅仅提供空间旋转操作。

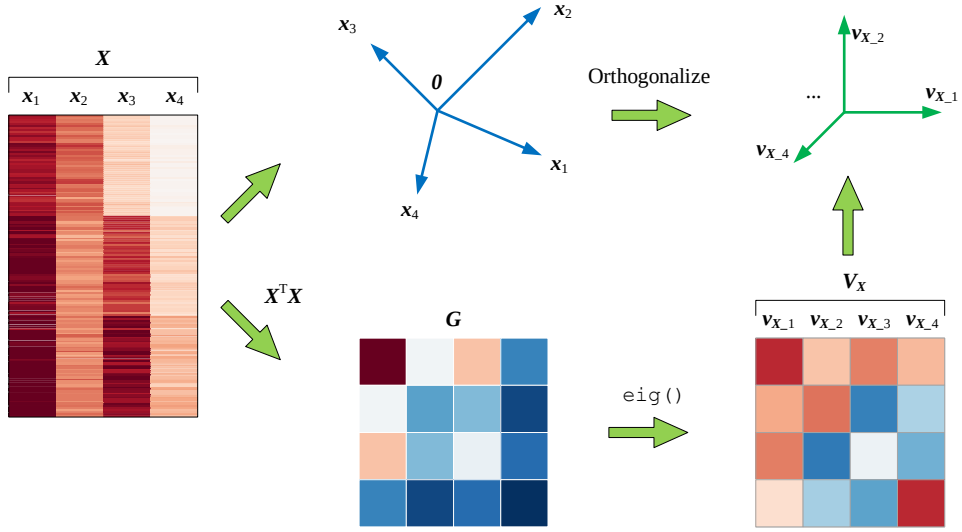


图 10. 特征值分解 $\boldsymbol{G}$ 获得规范正交基 $[\boldsymbol{v}_{X\_1}, \boldsymbol{v}_{X\_2}, \ldots, \boldsymbol{v}_{X\_D}]$

优化视角

本书第 18 章讲过，获得规范正交基 $[\boldsymbol{v}_{X\_1}, \boldsymbol{v}_{X\_2}, \ldots, \boldsymbol{v}_{X\_D}]$ 有着特定的几何目标。下面，我们简要回顾一下。

矩阵 $\boldsymbol{X}$ 在 $\boldsymbol{v}$ 方向投影得到 $\boldsymbol{y}$:

$$\boldsymbol{X}\boldsymbol{v} = \boldsymbol{y} \tag{5}$$

而 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{v}$ 可以写成:

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{v} = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v} = \left(\boldsymbol{X}\boldsymbol{v}\right)^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v} = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} = \left\|\boldsymbol{y}\right\|_2^2 \tag{6}$$

这就是特征值分解对应的优化问题——找到一个单位向量 $\boldsymbol{v}$，使得 $\boldsymbol{X}$ 在 $\boldsymbol{v}$ 上投影结果 $\boldsymbol{y}$ 的模最大。这个 $\boldsymbol{v}$ 就是 $\boldsymbol{v}_{X\_1}$，对应 $\boldsymbol{y}$ 的最大的模为 $\sqrt{\lambda_{X\_1}}$。

解决这个优化问题采用的方法可以是瑞利商，也可以是拉格朗日乘子法。

有了 $\boldsymbol{v}_{X\_1}$，寻找 $\boldsymbol{v}_{X\_2}$ 时，首先让 $\boldsymbol{v}_{X\_2}$ 垂直 $\boldsymbol{v}_{X\_1}$ (约束条件)，且 $\boldsymbol{X}$ 在 $\boldsymbol{v}_{X\_2}$ 上投影结果 $\boldsymbol{y}$ 的模最大。以此类推得到所有特征向量。

特征值

前文介绍过，特征值分解得到的特征值之和，等于原矩阵对角线元素之和，即:

$$\lambda_{X\_1} + \lambda_{X\_2} + \lambda_{X\_3} + \lambda_{X\_4} = \mathrm{sum}\left(\mathrm{diag}\left(\boldsymbol{G}\right)\right) = \left\|\boldsymbol{x}_1\right\|_2^2 + \left\|\boldsymbol{x}_2\right\|_2^2 + \left\|\boldsymbol{x}_3\right\|_2^2 + \left\|\boldsymbol{x}_4\right\|_2^2 \tag{7}$$

格拉姆矩阵 G 中蕴含着 X 列向量的模和方向，对 G 特征值分解得到第一特征向量 v_{X_1}，相较其他所有可能的单位向量，解释了 G 中最多的差异，因此 λ_{X_1} 在 (7) 中占比最大。

每个特征值占特征值总和的比例是主成分分析中重要的一项分析指标，这是本系列丛书《数据科学》一册要介绍的内容。

协方差矩阵

第二个例子是对协方差矩阵 Σ 进行特征值分解，图 11 所示为对应热图。下角标用 $_c$ 的原因是对协方差矩阵特征值分解结果和中心化 (去均值) 矩阵 X_c 直接相关。

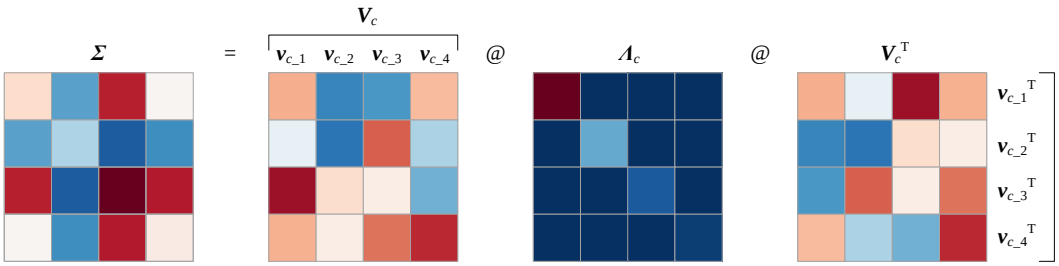


图 11. 对协方差矩阵进行特征值分解

前文提到过，Σ 囊括标准差向量 $[\sigma_1, \sigma_2, \sigma_3, \sigma_4]$ 所有信息——模 (标准差) 和夹角余弦值 (线性相关系数)。上一章提到 $[\sigma_1, \sigma_2, \sigma_3, \sigma_4]$ 的起始点为数据质心。

如图 12 所示，对 Σ 特征值分解得到的特征向量 $[v_{c\_1}, v_{c\_2}, v_{c\_3}, v_{c\_4}]$ 也是一组规范正交基。$[v_{c\_1}, v_{c\_2}, v_{c\_3}, v_{c\_4}]$ 相当于对 $[\sigma_1, \sigma_2, \sigma_3, \sigma_4]$ 的正交化，它显然不同于 $[v_{X\_1}, v_{X\_2}, \ldots, v_{X\_D}]$。

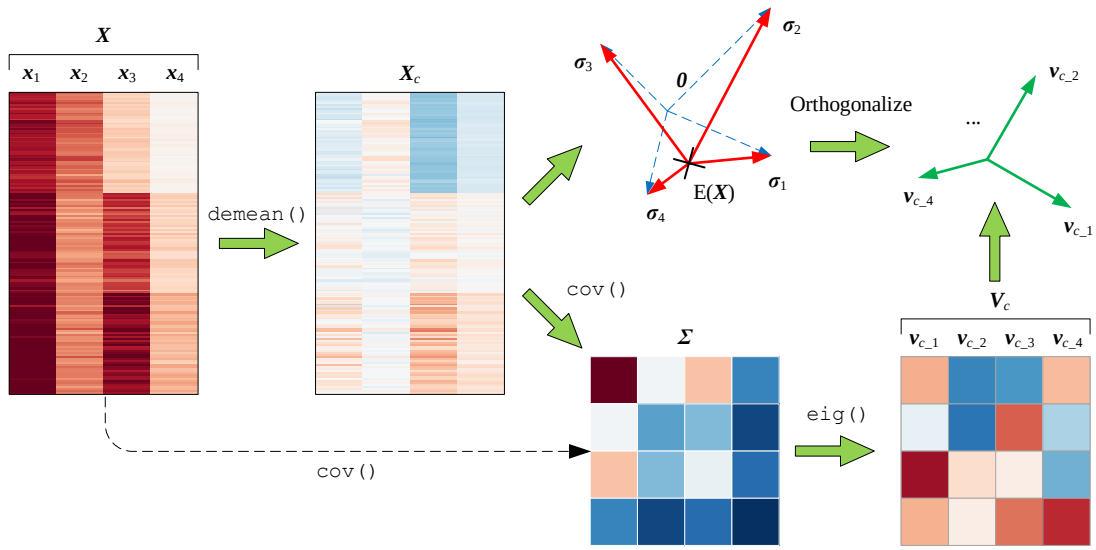


图 12. 特征值分解 Σ 获得规范正交基 $[v_{c\_1}, v_{c\_2}, v_{c\_3}, v_{c\_4}]$

优化视角

采用和本节前文一样的优化角度分析对协方差矩阵的特征值分解。

中心化数据矩阵 $\boldsymbol{X}_c$ 向 $\boldsymbol{v}$ 投影得到 $\boldsymbol{y}$：

$$\boldsymbol{X}_c\boldsymbol{v}=\boldsymbol{y} \tag{8}$$

而 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{v}$ 可以写成：

$$\boldsymbol{v}^{\mathrm{T}}(n-1)\boldsymbol{\Sigma}\boldsymbol{v}=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}_c{}^{\mathrm{T}}\boldsymbol{X}_c\boldsymbol{v}=(\boldsymbol{X}_c\boldsymbol{v})^{\mathrm{T}}\boldsymbol{X}_c\boldsymbol{v}=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}=\|\boldsymbol{y}\|_2^2=(n-1)\mathrm{var}(\boldsymbol{y}) \tag{9}$$

上式告诉我们，对协方差矩阵特征值分解，就是要找到一个单位向量 $\boldsymbol{v}$，使得中心化数据 $\boldsymbol{X}_c$ 在 $\boldsymbol{v}$ 上投影结果 $\boldsymbol{y}$ 的方差最大。我们要找的这个 $\boldsymbol{v}$ 就是图 11 中的 $\boldsymbol{v}_{c\_1}$，对应的特征值为 λ_{c_1}。

大家可能会问，(9) 是如何把协方差矩阵和 $\boldsymbol{y}$ 的方差联系起来的？这是我们下一章要探讨的内容。

再次注意单位问题，对于鸢尾花数据，协方差矩阵中的数值单位都是平方厘米 cm^2。其特征值 λ_c 的单位也是平方厘米 cm^2，而 $\boldsymbol{v}_c$ 是无单位的。

$\boldsymbol{\Sigma}$ 的特征值之和，等于 $\boldsymbol{X}$ 的每列数据方差之和，即：

$$\lambda_{\boldsymbol{\Sigma}\_1}+\lambda_{\boldsymbol{\Sigma}\_2}+\lambda_{\boldsymbol{\Sigma}\_3}+\lambda_{\boldsymbol{\Sigma}\_4}=\mathrm{diag}(\boldsymbol{\Sigma})=\sigma_1^2+\sigma_2^2+\sigma_3^2+\sigma_4^2 \tag{10}$$

$\boldsymbol{\Sigma}$ 的第一特征向量 $\boldsymbol{v}_{c\_1}$，解释了最多的方差，这便是主成分分析中重要的思路。

相关性系数矩阵

本节的第三个例子是对相关性系数矩阵 $\boldsymbol{P}$ 进行特征值分解，图 13 所示为对应热图。相关性系数矩阵 $\boldsymbol{P}$ 可以视作 $\boldsymbol{Z}_X$ ($\boldsymbol{X}$ 的 z 分数矩阵) 的协方差矩阵。

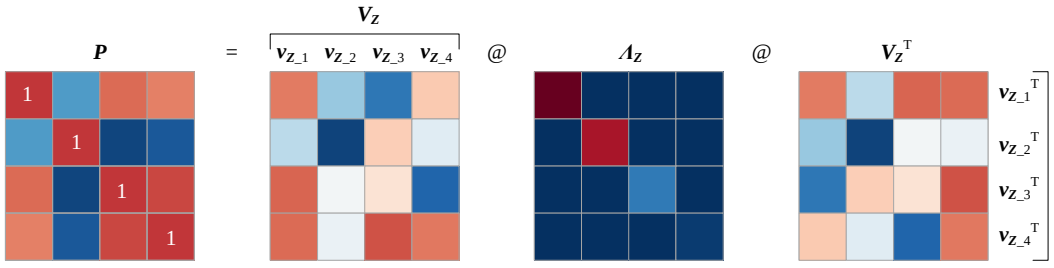


图 13. 对相关性系数矩阵进行特征值分解

矩阵 $\boldsymbol{Z}_X$ 的特点是，每列均值都是 0，数据已经标准化。

矩阵 Z_X 的列向量可以看成是 $\left[\frac{\sigma_1}{\|\sigma_1\|},\frac{\sigma_2}{\|\sigma_2\|},\frac{\sigma_2}{\|\sigma_2\|},\frac{\sigma_2}{\|\sigma_2\|}\right]$。

从相关性系数矩阵 P 对角线元素可以看出来，Z_X 每个特征贡献的方差为 1。

⚠ 注意这个 1，没有单位；因为数据标准化的过程，单位已经去掉。

如图 14 所示，对 P 进行特征值分解得到的特征向量 $[v_{Z\_1}, v_{Z\_2}, v_{Z\_3}, v_{Z\_4}]$ 也是一组规范正交基。一般情况，$[v_{Z\_1}, v_{Z\_2}, v_{Z\_3}, v_{Z\_4}]$ 不同于 $[v_{C\_1}, v_{C\_2}, v_{C\_3}, v_{C\_4}]$。

利用对相关性系数矩阵特征值分解进行主成分分析也是常见技术路线。这种技术路线可以解决 X 中某些特征的方差异常 (过大或过小) 的问题。

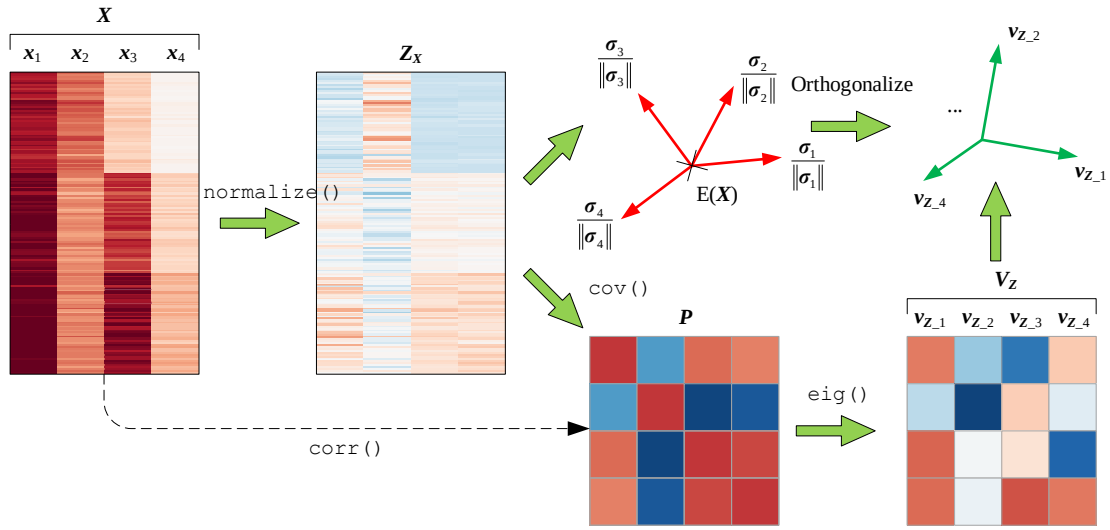


图 14. 特征值分解 P 获得规范正交基 $[v_{Z\_1}, v_{Z\_2}, v_{Z\_3}, v_{Z\_4}]$

Bk4_Ch24_01.py 中 Bk4_Ch24_01_D 部分完成本节介绍的三个特征值分解。

24.5 SVD 分解：获得四个空间

SVD 分解可谓矩阵分解之集大成者，本书前文花了很多笔墨从各个角度探讨 SVD 分解。本节对比原始数据矩阵 X、中心化矩阵 X_c、标准化矩阵 Z_X 等三个矩阵 SVD 分解。

原始数据矩阵

图 15 所示为矩阵 $\boldsymbol{X}$ 进行 SVD 分解矩阵运算热图。图中的 $\boldsymbol{V}_X$ 实际上和图 9 中的 $\boldsymbol{V}_X$ 等价，两者可能会在某些向量的正负号存在反号的情况。图 15 中矩阵 $\boldsymbol{U}_X$ 来自对 $\boldsymbol{X}\boldsymbol{X}^T$ 特征值分解。

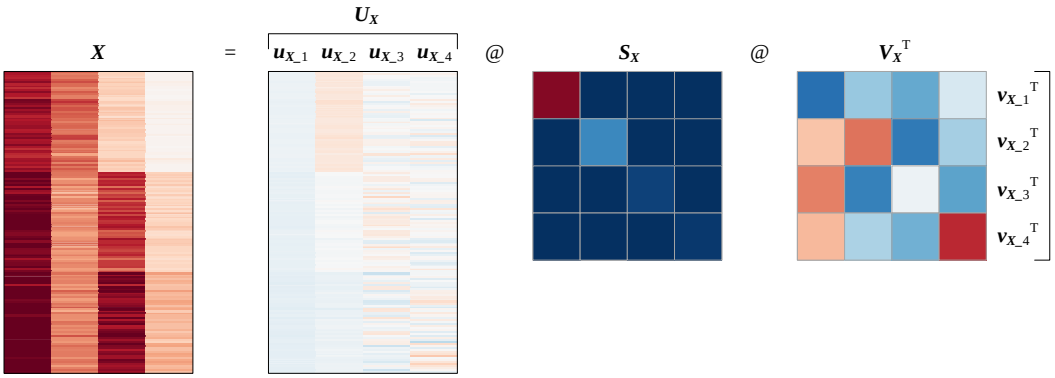


图 15. 对矩阵 $\boldsymbol{X}$ 进行 SVD 分解

前文提到多次，SVD 分解的结果包含了两个特征值分解结果。此外，SVD 分解不丢失原始数据 $\boldsymbol{X}$ 的任何信息。某种程度上说，SVD 分解包含了特征值分解，比特征值分解更“高阶”。

另外，请大家注意图 15 中奇异值和图 9 中特征值之间的关系：

$$\boldsymbol{S}_X^2 = \boldsymbol{\Lambda}_X \tag{11}$$

中心化数据

图 16 所示为中心化数据矩阵 $\boldsymbol{X}_c$ 进行 SVD 分解矩阵运算热图。

图中的 $\boldsymbol{V}_c$ 和图 11 中的 $\boldsymbol{V}_c$ 等价，两者若干位置列向量也可能存在符号相反情况。

有些读者可能会问，既然 $\boldsymbol{V}_c$ 也是规范正交基，那么将原始数据 $\boldsymbol{X}$ 在 $\boldsymbol{V}_c$ 上投影会有怎样的结果？下一章会给大家一些理论基础，本系列丛书《概率统计》一册会专门回答这个问题。

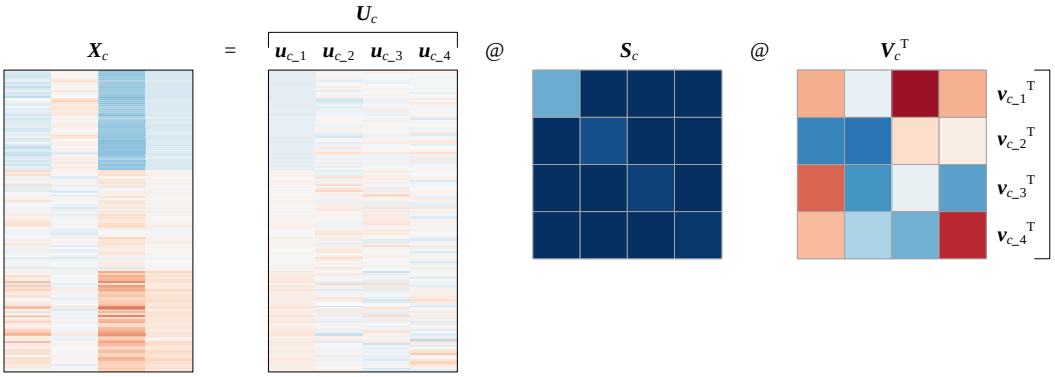


图 16. 对矩阵 X_c 进行 SVD 分解

标准化数据

图 17 所示为标准化数据矩阵 Z_X 进行 SVD 分解矩阵运算热图。图中的 V_Z 和图 13 中的 V_Z 等价，两者某些列向量也可能存在符号相反情况。

类似之前的一个问题，原始数据 X 在 V_Z 上投影会有怎样的结果？请大家带着这个问题学习《概率统计》一册。

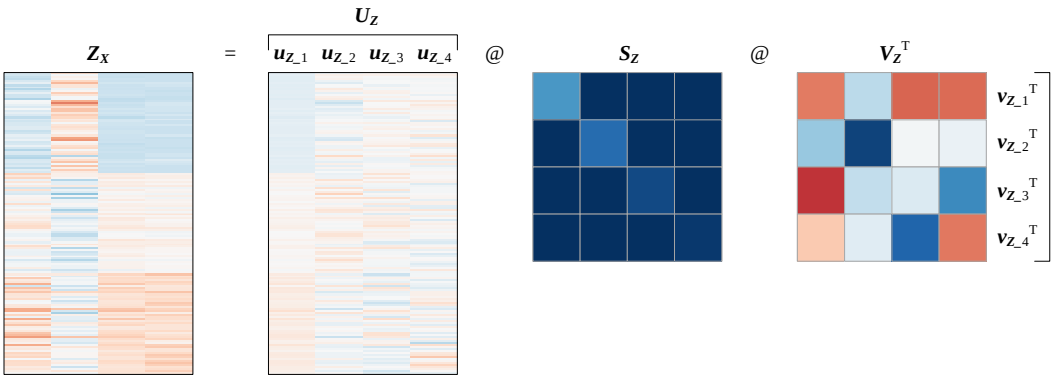


图 17. 对矩阵 Z_X 进行 SVD 分解

Bk4_Ch24_01.py 中 Bk4_Ch24_01_E 部分完成本节三个 SVD 分解运算。

本章最后用一幅图总结本章和上一章内容。

图 18 这幅图是本书中非常重要的几幅图之一，这幅图总结了整本书中和数据矩阵 $\boldsymbol{X}$ 有关的向量、矩阵、矩阵分解、空间等概念。

这幅图的数据分为两个部分：第一部分以 $\boldsymbol{X}$ 为核心，向量以 $\boldsymbol{0}$ 为起点；第二部分是统计视角，以去均值数据 $\boldsymbol{X}_c$ 为核心，向量以质心为起点。

下面，我们聊一下图 18 中关键细节。

对 $\boldsymbol{X}$ 进行 SVD 分解可以得到四个空间。

格拉姆矩阵 $\boldsymbol{G}$ 含有 $\boldsymbol{X}$ 列向量模、向量夹角两类重要信息。余弦相似度矩阵 $\boldsymbol{C}$ 仅仅还有向量夹角信息。对格拉姆矩阵 $\boldsymbol{G}$ 进行特征值分解只能获得两个空间。

对格拉姆矩阵 $\boldsymbol{G}$ 进行 Cholesky 分解得到上三角矩阵 $\boldsymbol{R}$ 代表 $\boldsymbol{X}$ 列向量坐标。

⚠ 再次强调，只有正定矩阵才能进行 Cholesky 分解。

在统计视角下，$\boldsymbol{X}$ 有两个重要信息——质心、协方差矩阵。质心确定数据中心位置，协方差矩阵描述数据分布。协方差矩阵 $\boldsymbol{\Sigma}$ 同样含有“标准差向量”的模 (标准差大小)、向量夹角 (余弦值为相关性系数) 两类重要信息。相关性系数矩阵 $\boldsymbol{P}$ 仅仅含有向量夹角 (相关性系数) 信息。

⚠ 值得格外注意的是，质心和协方差是多元高斯分布的两个参数，因此需要大家注意协方差矩阵和椭圆的联系。

$\boldsymbol{X}_c$ 是中心化数据矩阵，即每一列数据都去均值。协方差矩阵 $\boldsymbol{\Sigma}$ 相当于 $\boldsymbol{X}_c$ 的格拉姆矩阵。在几何视角下，$\boldsymbol{X}_c$ 到 $\boldsymbol{X}$ 相当于“平移”。而 $\boldsymbol{X}_c$ 到标准化数据 $\boldsymbol{Z}_X$ 相当于“缩放”。

对 $\boldsymbol{X}_c$ 进行 SVD 分解也可以得到四个空间。

此外，请大家格外注意不同矩阵的单位！以鸢尾花数据为例，$\boldsymbol{X}$ 的每一列数据单位恰好都是 cm，$\boldsymbol{X}_c$ 的单位也都是 cm，而 $\boldsymbol{Z}_X$ 没有单位 (或者说，单位是标准差)；$\boldsymbol{G}$ 的单位是 cm^2，$\boldsymbol{\Sigma}$ 的单位也是 cm^2，$\boldsymbol{P}$ 没有单位。

但是多数时候数据矩阵列向量单位比较丰富，比如高度、质量、时间、温度、密度、百分比、股价、收益率、GDP 等等。这就是为什么我们需要标准化数据 (去单位化) 的原因之一。

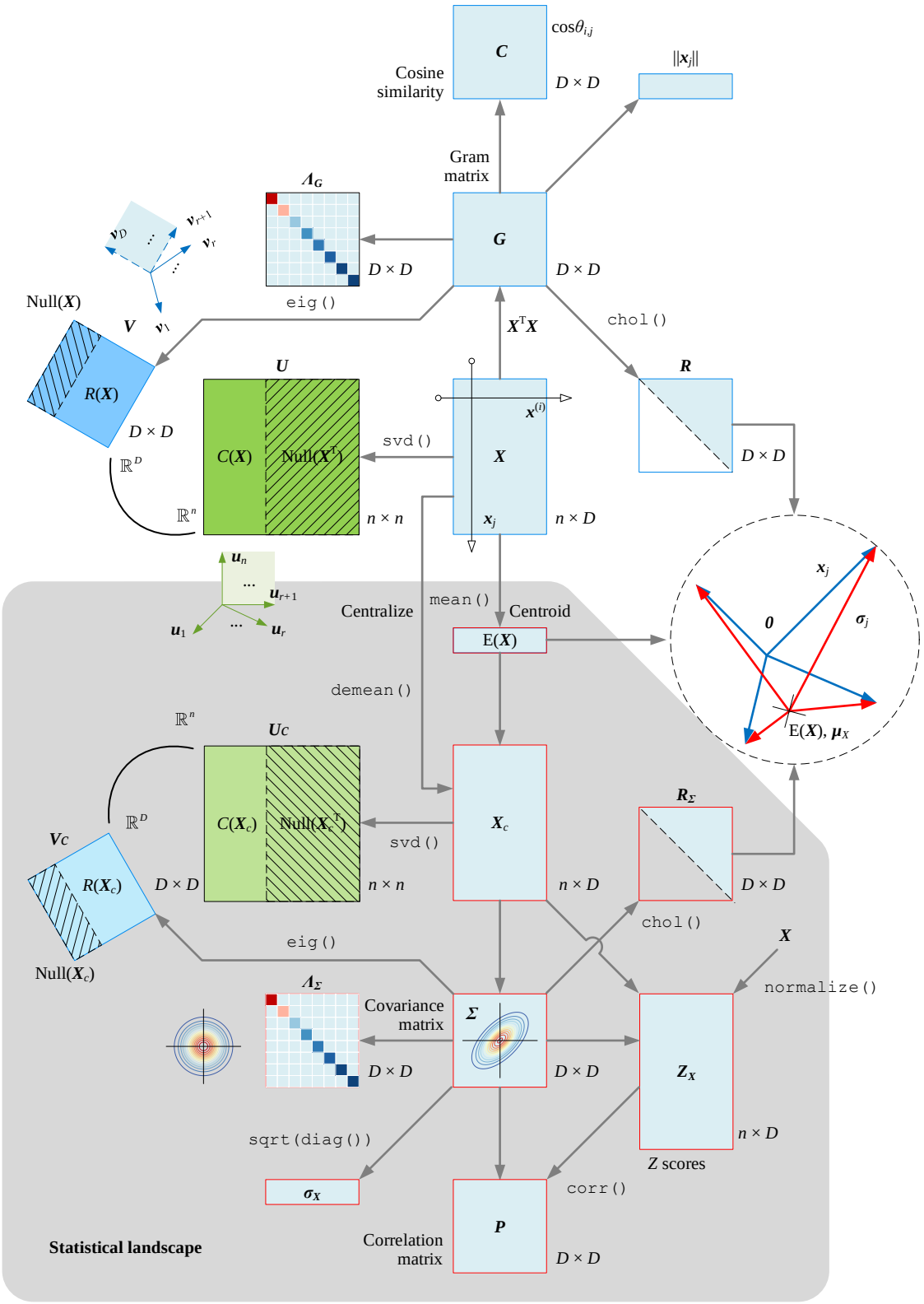


图 18. 总结本章内容的一幅图