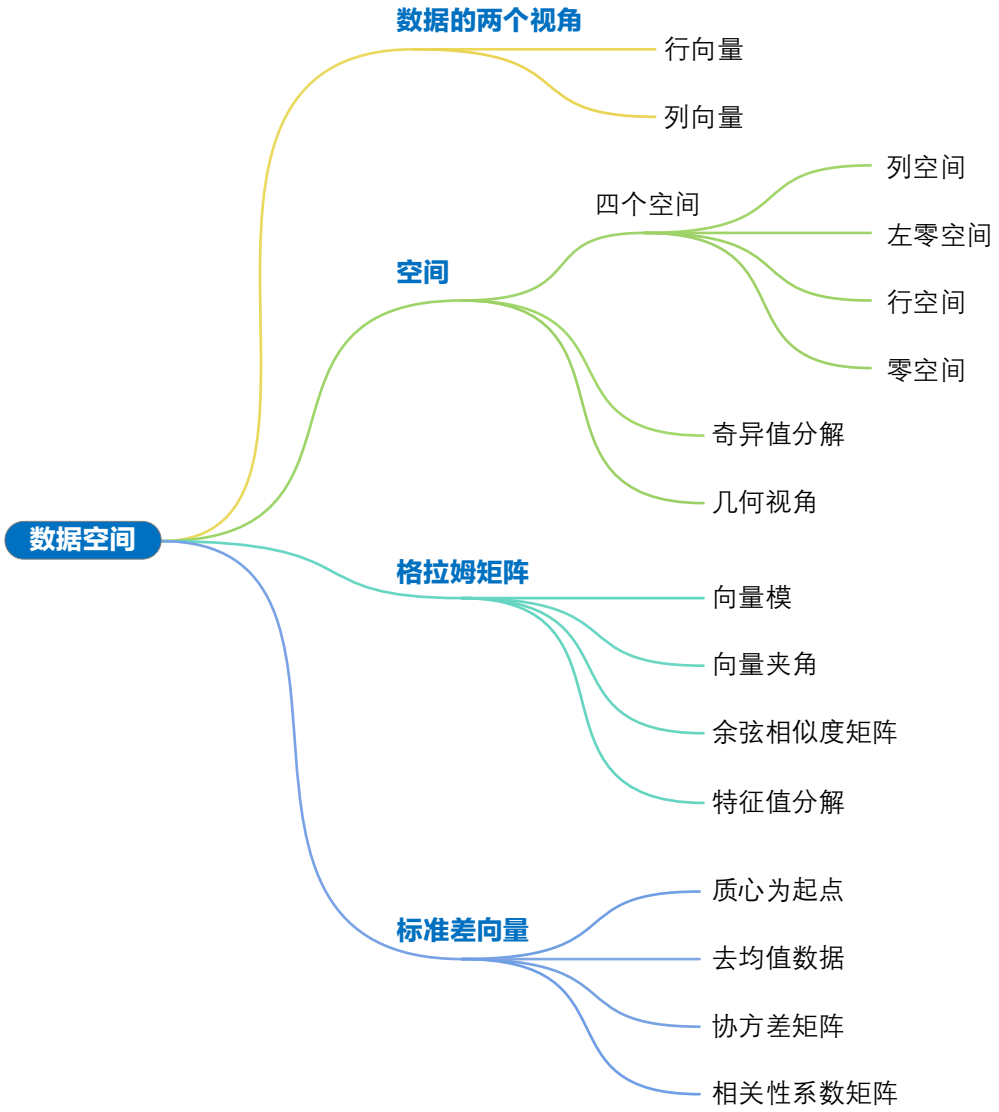
23 Four Vector Spaces
数据空间

用 SVD 分解寻找数据矩阵的四个空间

> 智慧的真正标志不是知识，而是想象力。
>
> ***The true sign of intelligence is not knowledge but imagination.***
>
> —— 阿尔伯特·爱因斯坦 (Albert Einstein) | 理论物理学家 | 1879 ~ 1955

- `numpy.cov()` 计算协方差矩阵
- `numpy.corr()` 计算相关性系数矩阵
- `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
- `seaborn.heatmap()` 绘制热图



23.1 从数据矩阵 X 说起

本书最后三章叫“数据三部曲”，这三章一方面从数据、空间、几何角度总结全书前文核心内容，另外一方面介绍这些数学工具在数据科学和机器学习领域的应用。

毫不夸张地说，没有线性代数就没有现代计算，大家将会在本系列丛书《数据科学》和《机器学习》两册书的每个角落看到矩阵运算。

“多重视角”仍然是这三章的特色。线性代数中向量、空间、投影、矩阵、矩阵分解等数学工具天然地弥合代数、几何、数据之间的鸿沟。

本章是“数据三部曲”的第一章，将以数据矩阵为切入点，主要通过奇异值分解和大家探讨四个重要的空间定义和用途。

数据矩阵

数据矩阵 (data matrix) 不过就是以表格形式存储的数据。

除了表格功能，矩阵更重要的功能是——**线性映射** (linear mapping)。而矩阵乘法是线性映射的核心。矩阵分解不过是矩阵连乘，将一个复杂的几何变换拆解成容易理解的成分，比如缩放、旋转、投影、剪切等等。

本书最开始便介绍过，数据矩阵可以从两个角度观察。数据矩阵 $\boldsymbol{X}$ 的每一行是一个行向量，代表一个样本观察值；$\boldsymbol{X}$ 的每一列为一个列向量，代表某个特征上的样本数据。

行向量

回顾前文，为了区分数据矩阵中的行向量和列向量，本书中数据矩阵的行向量序号采用上标加括号记法，比如：

$$\boldsymbol{X}_{n\times D}=\begin{bmatrix}\boldsymbol{x}^{(1)}\\ \boldsymbol{x}^{(2)}\\ \vdots\\ \boldsymbol{x}^{(n)}\end{bmatrix} \tag{1}$$

其中，第 i 行行向量 D 个元素为：

$$\boldsymbol{x}^{(i)}=\begin{bmatrix}x_{i,1} & x_{i,2} & \cdots & x_{i,D}\end{bmatrix} \tag{2}$$

图 1 所示为从行向量角度观察数据矩阵，每一个行向量 $\boldsymbol{x}^{(i)}$ 代表坐标系中一个点。所有数据散点构成坐标系中的“云”。

实际上，行向量也是具有方向和大小的向量，也可以看成是箭头，因此也有自己的空间。这是本书马上要探讨的内容。

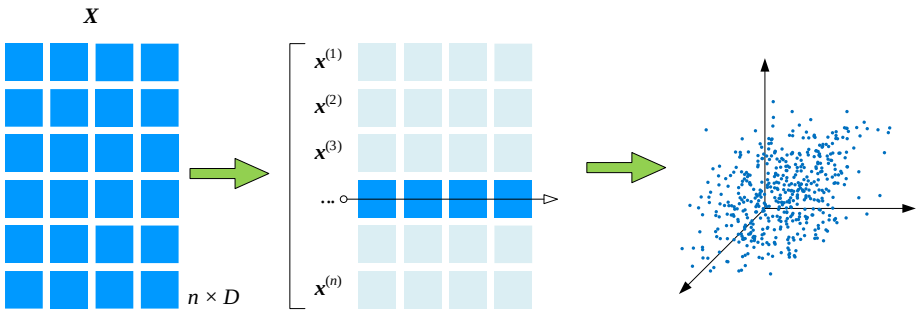


图 1. 从行向量角度观察数据矩阵

列向量

数据矩阵的列向量序号采用下标记法，比如：

$$\boldsymbol{X}_{n\times D}=\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D\end{bmatrix} \tag{3}$$

其中，第 j 列列向量 n 个元素为：

$$\boldsymbol{x}_j=\begin{bmatrix}x_{1,j}\\ x_{2,j}\\ \vdots\\ x_{n,j}\end{bmatrix} \tag{4}$$

如图 2 所示，从几何角度，数据矩阵 $\boldsymbol{X}$ 的所有列向量 (蓝色箭头) 的起始点均在原点 $\boldsymbol{0}$。$[\boldsymbol{x}_1, \boldsymbol{x}_1, \ldots \boldsymbol{x}_D]$ 这些向量的长度和方向信息均包含在格拉姆矩阵 $\boldsymbol{G}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 之中。

向量长度的表现形式为向量的模，即 L^2 范数。

向量方向是两两向量之间的相对夹角。更具体地说，是两两向量夹角余弦值。

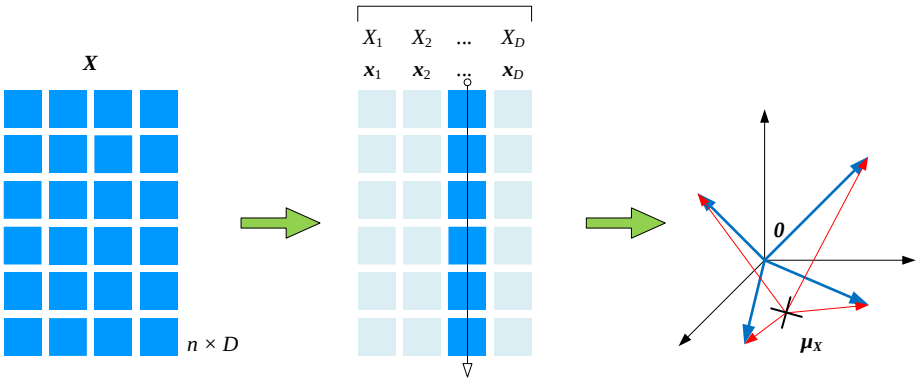


图 2. 从列向量角度观察数据矩阵

如果将图 2 向量起点移动到数据质心 $\boldsymbol{\mu}_X$ (即 E($\boldsymbol{X}$) 的转置)，这时向量 (红色箭头) 的长度可以看做是标准差 (的若干倍)，而向量之间夹角为随机变量之间的线性相关系数。

从统计角度来看，将向量起点移动到 $\boldsymbol{\mu}_X$ 实际上就是数据矩阵 $\boldsymbol{X}$ 去均值，即中心化，对应运算为 $\boldsymbol{X}_c = \boldsymbol{X} - \mathrm{E}(\boldsymbol{X})$。本章后文还将深入介绍这一重要视角。

协方差矩阵 $\boldsymbol{\Sigma}$ 相当于是 $\boldsymbol{X}_c$ 的格拉姆矩阵。准确来说，对于样本数据，$\boldsymbol{X}_c^{\mathrm{T}}\boldsymbol{X}_c = (n-1)\boldsymbol{\Sigma}$。协方差矩阵 $\boldsymbol{\Sigma}$ 包含了样本标准差和线性相关系数等信息。

区分符号

现在有必要再次强调本系列丛书的容易混淆的代数、线性代数和概率统计符号。

粗体、斜体、小写 $\boldsymbol{x}$ 为列向量。从概率统计的角度，$\boldsymbol{x}$ 可以代表随机变量 X 采样得到的样本数据，偶尔也代表 X 总体数据。随机变量 X 样本数据集合为 $X = \left\{x^{(1)}, x^{(2)}, ..., x^{(n)}\right\}$。

粗体、斜体、小写、加下标序号的 $\boldsymbol{x}_1$ 为列向量，下角标仅仅是序号，以便区分 $\boldsymbol{x}_1$、$\boldsymbol{x}_2$、$\boldsymbol{x}_j$、$\boldsymbol{x}_D$ 等等。从概率统计的角度，$\boldsymbol{x}_1$ 可以代表随机变量 X_1 样本数据，也可以表达 X_1 总体数据。

行向量 $\boldsymbol{x}^{(1)}$ 代表一个具有多个特征的样本点。

从代数角度，斜体、小写、非粗体 x_1 代表变量，下角标代表变量序号。这种记法常用在函数解析式中，比如线性回归解析式 $y = x_1 + x_2$。

$x^{(1)}$ 代表变量 x 的一个取值，或代表随机变量 X 的一个取值。

而 $x_1^{(1)}$ 代表变量 x_1 的一个取值，或代表随机变量 X_1 的一个取值，比如 $X_1 = \left\{x_1^{(1)}, x_1^{(2)}, ..., x_1^{(n)}\right\}$。

粗体、斜体、大写 $\boldsymbol{X}$ 则专门用来表达多行、多列的数据矩阵，$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D]$。数据矩阵 $\boldsymbol{X}$ 中第 i 行、第 j 列元素则记做 $\underline{x_{i,j}}$。多元线性回归中，$\boldsymbol{X}$ 也叫**设计矩阵** (design matrix)。

我们还会用粗体、斜体、小写希腊字母 $\boldsymbol{\chi}$ (chi，读作/ˈkaɪ/) 代表 D 维随机变量构成的列向量，$\boldsymbol{\chi} = [X_1, X_2, \ldots, X_D]^{\mathrm{T}}$。希腊字母 $\boldsymbol{\chi}$ 主要用在多元概率统计中。

23.2 向量空间：从 SVD 分解角度理解

这一节介绍 $\boldsymbol{X}$ 列向量和行向量张成的四个空间以及它们之间关系。

列向量：列空间、左零空间

由 $\boldsymbol{X}$ 的**列向量** $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_j, \ldots, \boldsymbol{x}_D$ 张成的子空间 $\text{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$ 为 $\boldsymbol{X}$ 的**列空间** (column space)，记做 $C(\boldsymbol{X})$。很多书上也将**列空间**记做 $\text{Col}(\boldsymbol{X})$，或 $\text{Col}\boldsymbol{X}$。

与 $C(\boldsymbol{X})$ 相对应的是**左零空间** (left null space)，记做 $\text{Null}(\boldsymbol{X}^{\text{T}})$。$C(\boldsymbol{X})$ 和 $\text{Null}(\boldsymbol{X}^{\text{T}})$ 构成了 $\mathbb{R}^n$。$\boldsymbol{X}$ 的**列向量**元素个数为 n，因此需要匹配空间 $\mathbb{R}^n$，才能“装下”$\boldsymbol{X}$ 的**列向量**。

而 $C(\boldsymbol{X})$ 和 $\text{Null}(\boldsymbol{X}^{\text{T}})$ 分别都是 $\mathbb{R}^n$ 的子空间，两者的维度之和为 n，即 $\dim(C(\boldsymbol{X})) + \dim(\text{Null}(\boldsymbol{X}^{\text{T}})) = n$。

$C(\boldsymbol{X})$ 和 $\text{Null}(\boldsymbol{X}^{\text{T}})$ 互为**正交补** (orthogonal complement)，即：

$$C(\boldsymbol{X})^{\perp} = \text{Null}(\boldsymbol{X}^{\text{T}}) \tag{5}$$

行向量：行空间、零空间

由 $\boldsymbol{X}$ 的**行向量** $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(j)}, \ldots, \boldsymbol{x}^{(n)}$ 张成的子空间 $\text{span}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(n)})$ 为 $\boldsymbol{X}$ 的**行空间** (row space)，记做 $R(\boldsymbol{X})$。很多书上也记做 $\text{Row}(\boldsymbol{X})$ 或 $\text{Row}\boldsymbol{X}$。

与 $R(\boldsymbol{X})$ 相对应的是**零空间** (null space)，也叫**右零空间** (right null space)，记做 $\text{Null}(\boldsymbol{X})$。

$\boldsymbol{X}$ 的**行向量**元素数量为 D，空间 $\mathbb{R}^D$ 才能“装下”$\boldsymbol{X}$ 的**行向量**。$R(\boldsymbol{X})$ 和 $\text{Null}(\boldsymbol{X})$ 构成了 $\mathbb{R}^D$。$R(\boldsymbol{X})$ 和 $\text{Null}(\boldsymbol{X})$ 分别都是 $\mathbb{R}^D$ 的子空间。

$R(\boldsymbol{X})$ 和 $\text{Null}(\boldsymbol{X})$ 互为正交补，即：

$$R(\boldsymbol{X})^{\perp} = \text{Null}(\boldsymbol{X}) \tag{6}$$

$R(\boldsymbol{X})$的维度为 $\dim(R(\boldsymbol{X})) = \text{rank}(\boldsymbol{X})$。$R(\boldsymbol{X})$ 和 $\text{Null}(\boldsymbol{X})$ 的维度之和为 D，即 $\dim(R(\boldsymbol{X})) + \dim(\text{Null}(\boldsymbol{X})) = D$。也就是说，只有 $\boldsymbol{X}$ 非满秩，$\text{Null}(\boldsymbol{X})$ 维数才不为 0。

怎么理解这四个空间?

相信大家读完本节前文这四个空间定义已经晕头转向，云里雾里不知所云。

的确，这四个空间的定义让很多人望而却步。很多线性代数教材多是从线性方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 角度讲解这四个空间，而作者认为这个视角并没有降低理解这四个空间的难度。

下面，我们从数据和几何两个角度来理解这四个空间，并且介绍如何将它们和本书前文介绍的向量内积、格拉姆矩阵、向量空间、子空间、秩、特征值分解、SVD 分解、数据质心、协方差矩阵等线性代数概念联系起来。

从完全型 SVD 分解说起

对“细长”矩阵 $\boldsymbol{X}$ 进行完全型 SVD 分解，得到等式：

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}} \tag{7}$$

图 3 所示为 $\boldsymbol{X}$ 完全型 SVD 分解示意图。

⚠ 请大家注意几个矩阵形状。完全型 SVD 分解，$\boldsymbol{X}$ 和 $\boldsymbol{S}$ 为一般为细高型，$\boldsymbol{U}$ 和 $\boldsymbol{V}$ 为方阵。

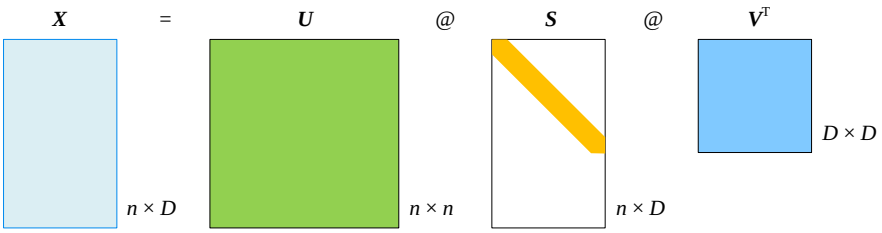


图 3. $\boldsymbol{X}$ 进行完全型 SVD 分解

根据前文所学，大家应该很清楚 $\boldsymbol{U}$ 为 $n \times n$ 正交矩阵，也就是说 $\boldsymbol{U}$ 列向量 $[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 特点是两两正交 (向量内积为 0)，且向量模均为 1。

$[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 为张成 $\mathbb{R}^n$ 空间的一组规范正交基。

同理，$\boldsymbol{V}$ 为 $D \times D$ 正交矩阵，因此 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$ 是张成 $\mathbb{R}^D$ 空间的一组规范正交基。

如图 4 所示，$\boldsymbol{U}$ 和 $\boldsymbol{V}$ 之间的联系为 $\boldsymbol{US} = \boldsymbol{XV}$。

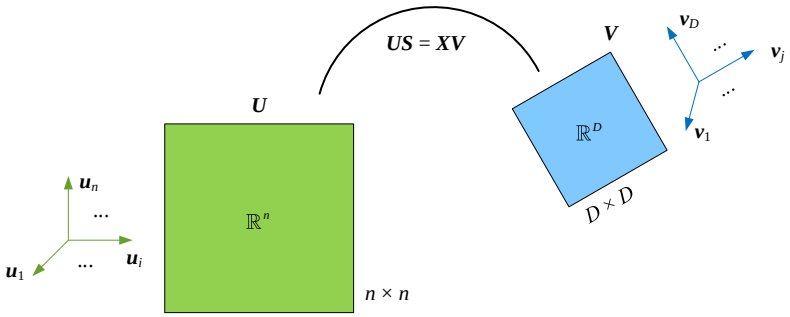


图 4. 对于矩阵 $\boldsymbol{X}$ 来说，$\mathbb{R}^n$ 空间和 $\mathbb{R}^D$ 空间关系

另外，对“粗短”$\boldsymbol{X}^{\mathrm{T}}$矩阵进行完全型 SVD 分解，就是对 (7) 转置：

$$\boldsymbol{X}^{\mathrm{T}} = \left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\right)^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} \tag{8}$$

图 5 所示为 $\boldsymbol{X}^{\mathrm{T}}$进行完全型 SVD 分解示意图。后面，我们会用到这一分解。

⚠ 注意，对于完全型 SVD 分解，奇异值矩阵 $\boldsymbol{S}$ 虽然是对角阵，但不是方阵，因此 $\boldsymbol{S}^{\mathrm{T}} \neq \boldsymbol{S}$。

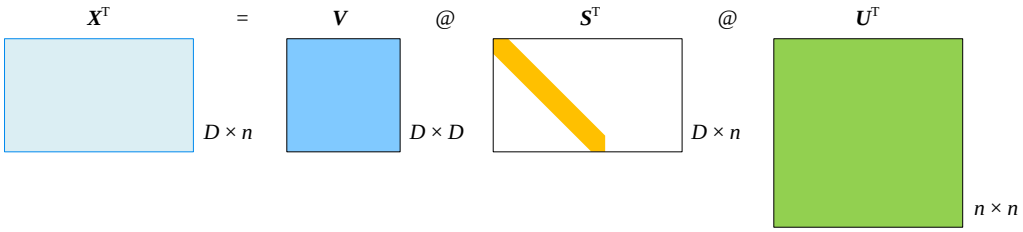


图 5. $\boldsymbol{X}^{\mathrm{T}}$进行完全型 SVD 分解

23.3 紧凑型 SVD 分解：剔除零空间

紧凑型 SVD 分解

在讲解奇异值分解时，我们特别介绍了紧凑型 SVD 分解。紧凑型 SVD 分解对应的情况为 $\mathrm{rank}(\boldsymbol{X}) = r < D$。奇异值矩阵 $\boldsymbol{S}$ 可以分成四个子块：

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{r\times r} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} \tag{9}$$

上式中，矩阵 $\boldsymbol{S}_{r\times r}$对角线元素为非 0 奇异值。

图 6 所示为 $\boldsymbol{X}$ 进行紧凑型 SVD 分解示意图。本书第 16 章介绍过，分块矩阵乘法中，图 6 中阴影部分对应的分块矩阵可以全部消去。

正交矩阵 $\boldsymbol{U}$ 保留 $[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r]$ 子块，消去 $[\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n]$。

正交矩阵 $\boldsymbol{V}$ 保留 $[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r]$ 子块，消去 $[\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_D]$。

$[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r]$ 是 $\boldsymbol{X}$ 的**列空间** $C(\boldsymbol{X})$ 基底。而 $[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r]$ 是 $\boldsymbol{X}$ 的**行空间** $R(\boldsymbol{X})$ 基底。

⚠ 注意，图 6 中 $\boldsymbol{V}$ 存在转置运算。

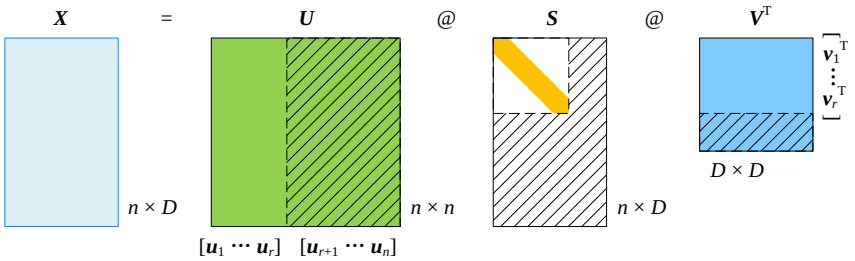


图 6. $\boldsymbol{X}$ 进行紧凑型 SVD 分解

实际上，$\boldsymbol{U}$ 和 $\boldsymbol{V}$ 矩阵中消去的子块和上一节说到的**零空间**有直接联系。先给出图 7 这幅图，我们马上展开讲解。

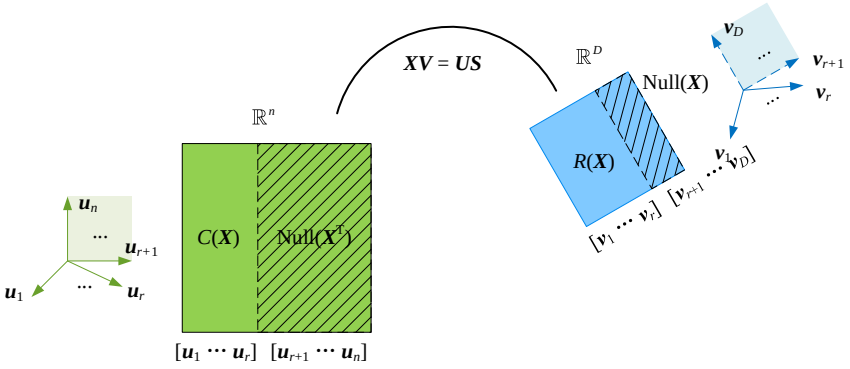


图 7. $\mathbb{R}^n$ 空间和 $\mathbb{R}^D$ 空间关系，考虑**列空间**、**左零空间**、**行空间**、**零空间**

列空间，左零空间

$[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r]$ 是 $\boldsymbol{X}$ 的**列空间** $C(\boldsymbol{X})$ 基底。而 $[\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n]$ 是**左零空间** $\text{Null}(\boldsymbol{X}^{\text{T}})$ 基底。

如图 8 所示，将 $\boldsymbol{S}^{\text{T}}$ 左右分块，右侧分块矩阵为 $\boldsymbol{O}$ 矩阵。$\boldsymbol{X}^{\text{T}}$ 向**左零空间** $\text{Null}(\boldsymbol{X}^{\text{T}})$ $[\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n]$ 投影的结果为全 0 矩阵 $\boldsymbol{O}$。

白话说，$\mathbb{R}^n$ 用来装 $\boldsymbol{X}$ 的**列向量**，绝对“杀鸡用牛刀”。$[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r]$ 张起的子空间就“刚刚好”够装下 $\boldsymbol{X}$ 的**列向量**。而 $\mathbb{R}^n$ 中没有被用到的部分就是 $[\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n]$ 张起的**左零空间** $\text{Null}(\boldsymbol{X}^{\text{T}})$。

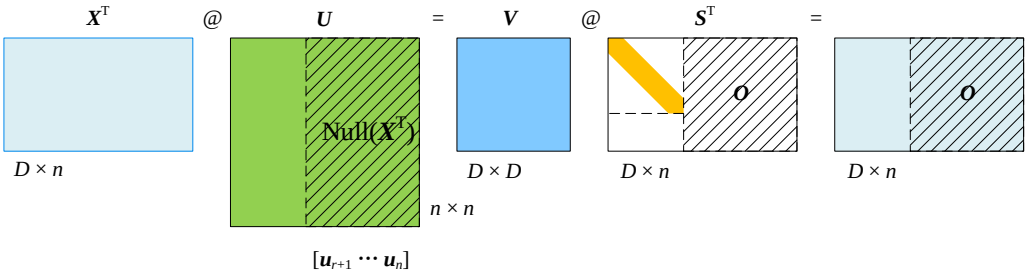


图 8. $\boldsymbol{X}^{\mathrm{T}}$ 向 Null($\boldsymbol{X}^{\mathrm{T}}$) $[\boldsymbol{u}_{r+1}, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 投影的结果为 $\boldsymbol{O}$

这就是为什么 Null($\boldsymbol{X}^{\mathrm{T}}$) 被称作**左“零”空间**的原因，因为投影结果为零矩阵。而且，我们也同时在图 8 中投影运算中 $\boldsymbol{X}^{\mathrm{T}}$ 看到了“转置”，这就解释了为什么**列空间** $C(\boldsymbol{X})$ 对应 Null($\boldsymbol{X}^{\mathrm{T}}$)。

多说一句，(8) 可以写成：

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{V}\boldsymbol{S}^{\mathrm{T}} \tag{10}$$

上式正交投影中，矩阵 $\boldsymbol{X}^{\mathrm{T}}$ 对应的投影矩阵是 $\boldsymbol{U}$，$\boldsymbol{X}^{\mathrm{T}}$ 的每一行代表一个散点，对应 $\boldsymbol{X}$ 的**列向量**。大家在这句话中看到**列空间** $C(\boldsymbol{X})$ 和 Null($\boldsymbol{X}^{\mathrm{T}}$) 中“**列**”和“$\boldsymbol{X}^{\mathrm{T}}$”这两个字眼了吧！

行空间，零空间

而 $[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r]$ 是 $\boldsymbol{X}$ 的**行空间** $R(\boldsymbol{X})$ 基底。$[\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_D]$ 是**零空间** Null($\boldsymbol{X}$) 的基底。

白话说，$\mathbb{R}^D$ 来装 $\boldsymbol{X}$ 的**行向量**，可能大材小用，也可能大小合适。$[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r]$ 张起的子空间就刚刚好够装下 $\boldsymbol{X}$ 的**行向量**。富余的部分就是 $[\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_D]$ 张起的**零空间** Null($\boldsymbol{X}$)。rank($\boldsymbol{X}$) = r = D 时，$\mathbb{R}^D$ 装 $\boldsymbol{X}$ 的**行向量**后没有任何余量。

也用正交投影视角来看，将 (8) 写成：

$$\boldsymbol{X}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S} \tag{11}$$

矩阵 $\boldsymbol{X}$ 对应的投影矩阵是 $\boldsymbol{V}$，$\boldsymbol{X}$ 的每一行代表一个散点，对应 $\boldsymbol{X}$ 的**行向量**。如图 9 所示，将 $\boldsymbol{S}$ 左右分块，右侧分块矩阵为 $\boldsymbol{O}$ 矩阵。$\boldsymbol{X}$ 向 Null($\boldsymbol{X}$) 投影的结果为 $\boldsymbol{Z}$ 的右侧零矩阵 $\boldsymbol{O}$。

图 9 解释了为什么 Null($\boldsymbol{X}$) 被称作“零”空间，而**行空间** $R(\boldsymbol{X})$ 对应**零空间** Null($\boldsymbol{X}$)。

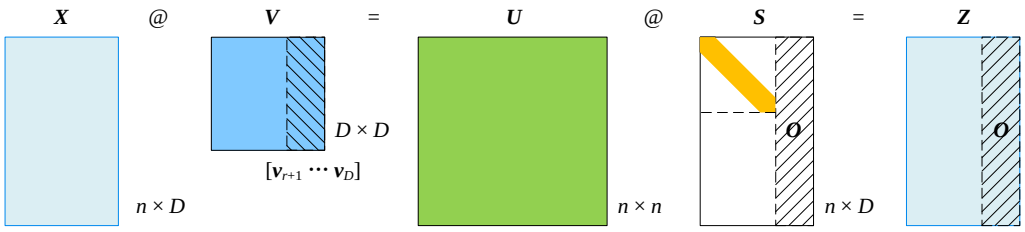


图 9. $\boldsymbol{X}$ 向 Null($\boldsymbol{X}$) 投影的结果为 $\boldsymbol{O}$

复盘一下

$\boldsymbol{U}$ 是正交矩阵，即 $[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 中列向量两两垂直。基底 $[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 张起了 $\mathbb{R}^n$。

将 $[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 划分成两块—— $C(\boldsymbol{X}) = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r]$、$\text{Null}(\boldsymbol{X}^{\text{T}}) = [\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n]$。**列空间** $C(\boldsymbol{X})$ 和**左零空间** $\text{Null}(\boldsymbol{X}^{\text{T}})$ 互为正交补。

“正交”两字，来自于 $[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 中列向量两两垂直。“补”字，可以理解为“补齐”，也就是说 $C(\boldsymbol{X})$ 和 $\text{Null}(\boldsymbol{X}^{\text{T}})$ 补齐了 $\mathbb{R}^n$。

同理，将 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$ 划分成两块—— $R(\boldsymbol{X}) = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r]$、$\text{Null}(\boldsymbol{X}) = [\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_D]$。而**行空间** $R(\boldsymbol{X})$ 和**零空间** $\text{Null}(\boldsymbol{X})$ 互为正交补，两者“补齐”得到 $\mathbb{R}^D$。

在图 7 基础上，考虑这两对正交关系，加上 $\mathbb{R}^n$ 空间和 $\mathbb{R}^D$ 空间，我们用图 10 可视化这六个空间。图中加阴影的部分对应**左零空间**和**零空间**。

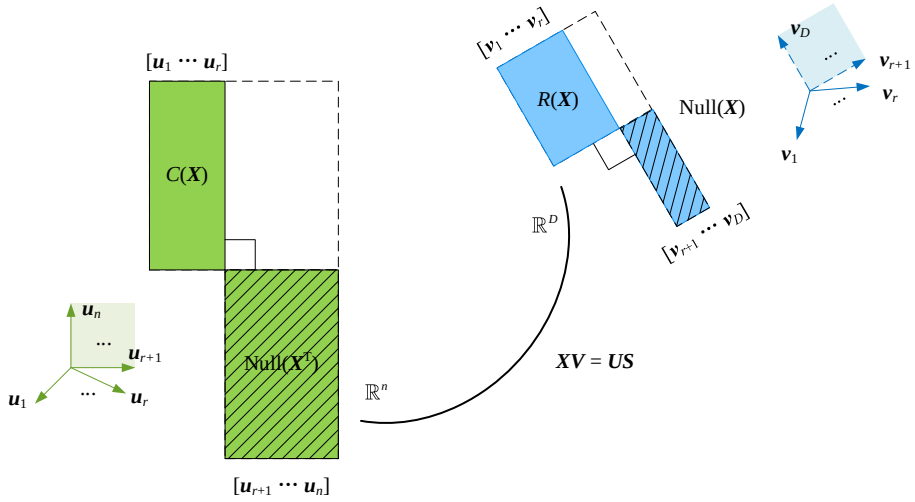


图 10. $\mathbb{R}^n$ 空间和 $\mathbb{R}^D$ 空间关系，考虑**列空间**、**左零空间**、**行空间**、**零空间**的正交关系

四个空间：因 $\boldsymbol{X}$ 而生

格外强调，$\mathbb{R}^n$ 空间和 $\mathbb{R}^D$ 空间是“永恒”存在的，是“铁打的庙”。但是，能张成这两个空间的规范正交基有无数组，都是“流水的和尚”。

$[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]$，即 $C(\boldsymbol{X}) + \text{Null}(\boldsymbol{X}^{\text{T}})$，是张成 $\mathbb{R}^n$ 空间无数组规范正交基中的一组。

$[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_D]$，即 $R(\boldsymbol{X}) + \text{Null}(\boldsymbol{X})$，是张成 $\mathbb{R}^D$ 空间无数组规范正交基中的一组。

值得强调的是，在矩阵 $\boldsymbol{X}$ 眼中，$C(\boldsymbol{X})$、Null($\boldsymbol{X}^{\mathrm{T}}$)、$R(\boldsymbol{X})$ 和 Null($\boldsymbol{X}$) 是独一无二的存在，因为它们都是为矩阵 $\boldsymbol{X}$ 而生！

也就是说，数据矩阵 $\boldsymbol{X}$ 稍有变化，不管是元素、还是形状变化，这四个空间就会随之变化。

而获得这四个空间最便捷的方法就是堪称宇宙第一矩阵分解的奇异值分解。

怎么记忆？

如果大家还是分不清这四个空间，我还有一个小技巧！

大家只需要记住 $\boldsymbol{XV} = \boldsymbol{US}$ 这个式子。

$\boldsymbol{U}$ 和 $\boldsymbol{X}$ 等长，即**列向量**行数相等，因此 $\boldsymbol{U}$ 一定包含**列空间**。

$\boldsymbol{U}$ 在矩阵乘积 $\boldsymbol{US}$ 左边，因此包含**“左”零空间**。

$\boldsymbol{V}$ 和 $\boldsymbol{X}$ 等宽，即**行向量**列数相等，且 $\boldsymbol{XV}$ 中的 $\boldsymbol{V}$ 是 $\boldsymbol{X}$ **行向量**投影方向，因此 $\boldsymbol{V}$ 包含**行空间**。

$\boldsymbol{V}$ 在矩阵乘积 $\boldsymbol{XV}$ 右边，因此包含**“右”零空间**。而**右零空间**，就简称**零空间**。因为**右零空间**最常用，所以独占了“**零空间**”这个更简洁的头衔。

问题来了，要是记不住 $\boldsymbol{XV} = \boldsymbol{US}$，怎么办？

就一句话——我们永远 15 岁！

$\boldsymbol{US}$ 代表“我们”，$\boldsymbol{XV}$ 是罗马数字的 15。

23.4 几何视角说空间

下面我们用具体数值从几何视角再强化理解上节介绍的几个空间。

举个例子

给定矩阵 $\boldsymbol{X}$ 如下：

$$\boldsymbol{X} = \begin{bmatrix} 1 & -1 \\ -\sqrt{3} & \sqrt{3} \\ 2 & -2 \end{bmatrix} \tag{12}$$

一眼就能看出来，$\boldsymbol{X}$ 的两个**列向量**线性相关，因为：

$$\boldsymbol{x}_1 = -\boldsymbol{x}_2 \tag{13}$$

也就是说 $\boldsymbol{X}$ 的秩为 1，即 $\mathrm{rank}(\boldsymbol{X}) = r = 1$。

列向量

为了可视化 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 这两个**列向量**，我们需要三维直角坐标系 $\mathbb{R}^3$，如图 11 (a) 所示。

白话说，$\mathbb{R}^3$ 才能装下长度为 3 的**列向量** $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$。

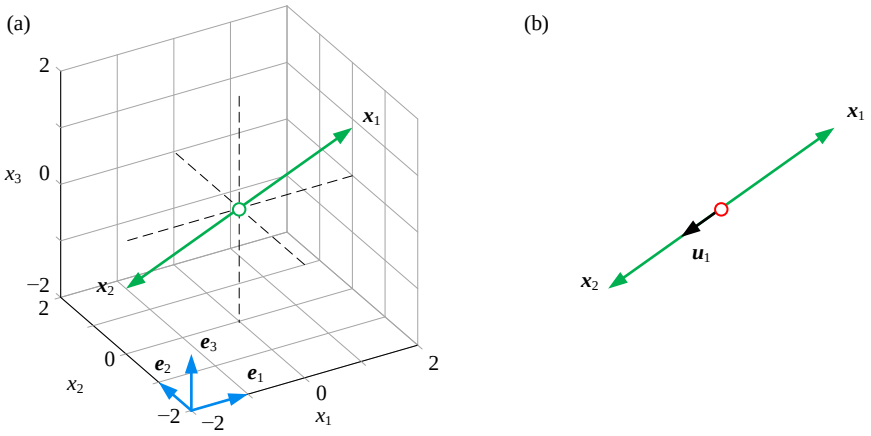


图 11. 从三维空间到一维空间

但是我们发现，实际上，图 11 (b) 告诉我们有了 $\boldsymbol{u}_1$ 这个单位向量，我们就可以把 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 写成：

$$\boldsymbol{x}_1 = a\boldsymbol{u}_1, \quad \boldsymbol{x}_2 = b\boldsymbol{u}_1 \tag{14}$$

也就是说，$\mathbb{R}^3$ 中一维子空间 span($\boldsymbol{u}_1$) 就足够装下 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$，这就是为什么 rank($\boldsymbol{X}$) = 1。

那么问题来了，我们如何找到 $\boldsymbol{u}_1$ 这个单位向量？

根据前文所学，我们知道有至少有两种办法：a) SVD 分解；b) 特征值分解。

SVD 分解

对 $\boldsymbol{X}$ 进行 SVD 分解得到：

$$\boldsymbol{X} = \begin{bmatrix} 1 & -1 \\ -\sqrt{3} & \sqrt{3} \\ 2 & -2 \end{bmatrix} = \underbrace{\begin{bmatrix} -0.3536 & -0.9297 & 0.1034 \\ 0.6124 & -0.1465 & 0.7769 \\ -0.7071 & 0.3380 & 0.6211 \end{bmatrix}}_{\boldsymbol{U}} \underbrace{\begin{bmatrix} 4 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\boldsymbol{S}} \left(\underbrace{\begin{bmatrix} -0.7071 & -0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}}_{\boldsymbol{V}} \right)^{\mathrm{T}} \tag{15}$$

其中，矩阵 $\boldsymbol{U}$ 的第一**列向量**就是我们要找的 $\boldsymbol{u}_1$，而这个 $\boldsymbol{u}_1$ 便独立张成**列空间** $C(\boldsymbol{X})$。

也就是说，$C(\boldsymbol{X})$ 对应 $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3]$ 线性无关的成分。

顺藤摸瓜，有意思的是完全型 SVD 分解中，我们顺路还得到了 $\boldsymbol{u}_2$ 和 $\boldsymbol{u}_3$，基底 $[\boldsymbol{u}_2, \boldsymbol{u}_3]$ 张起了**左零空间** $\text{Null}(\boldsymbol{X}^T)$。

而规范正交基 $[\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3]$ 则是张成 $\mathbb{R}^3$ 无数规范正交基中的一个。$[\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3]$ 这个独特存在全靠矩阵 $\boldsymbol{X}$。

而 $[\boldsymbol{u}_1]$ 和 $[\boldsymbol{u}_2, \boldsymbol{u}_3]$ 补齐得到 $\mathbb{R}^3$。显然，$\boldsymbol{u}_1$ 垂直于 $\boldsymbol{u}_2$ 和 $\boldsymbol{u}_3$ 张成的平面 $\text{span}(\boldsymbol{u}_2, \boldsymbol{u}_3)$。所示，$[\boldsymbol{u}_1]$ 和 $[\boldsymbol{u}_2, \boldsymbol{u}_3]$ 互为正交补。

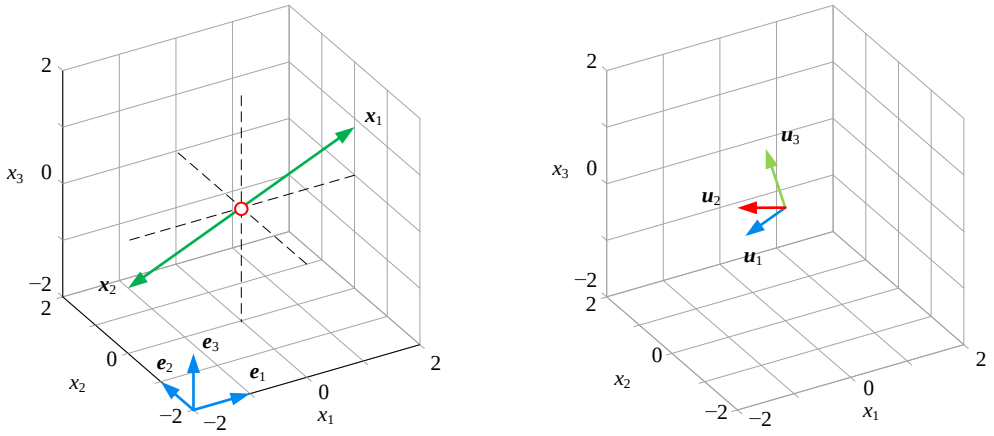


图 12. 矩阵 $\boldsymbol{X}$ 的**列空间** $C(\boldsymbol{X})$ 和**左零空间** $\text{Null}(\boldsymbol{X}^T)$

投影

把**列向量** $\boldsymbol{x}_1$ 投影到 $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3]$ 中得到：

$$\boldsymbol{x}_1^T\boldsymbol{U} = \begin{bmatrix} 1 & -\sqrt{3} & 2 \end{bmatrix} \begin{bmatrix} \underbrace{\begin{bmatrix} -0.3536 \\ 0.6124 \\ -0.7071 \end{bmatrix}}_{\boldsymbol{u}_1} & \underbrace{\begin{bmatrix} -0.9297 \\ -0.1465 \\ 0.3380 \end{bmatrix}}_{\boldsymbol{u}_2} & \underbrace{\begin{bmatrix} 0.1034 \\ 0.7769 \\ 0.6211 \end{bmatrix}}_{\boldsymbol{u}_3} \end{bmatrix} = \begin{bmatrix} -2.8284 & 0 & 0 \end{bmatrix} \tag{16}$$

也就是说，$\boldsymbol{x}_1$ 在 $[\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3]$ 这个标准正交基中的坐标为 $(-2.8282, 0, 0)$。

大家可以看到 $\boldsymbol{x}_1$ 在 $\boldsymbol{u}_2$ 和 $\boldsymbol{u}_3$ 上投影结果均为 0，这就是为什么 $\boldsymbol{u}_2$ 和 $\boldsymbol{u}_3$ 上构成**左零空间** $\text{Null}(\boldsymbol{X}^T)$。

(16) 中的 $\boldsymbol{x}_1$ 转置运算也解释了 $\text{Null}(\boldsymbol{X}^T)$ 括号里面为什么是 $\boldsymbol{X}^T$。

同理，把 $\boldsymbol{x}_2$ 投影到 $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ 中得到 $\boldsymbol{x}_1$ 在 $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ 的坐标为 $(2.8282, 0, 0)$，对应矩阵运算具体为：

$$\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{U}=\begin{bmatrix}-1 & \sqrt{3} & -2\end{bmatrix}\begin{bmatrix}\underbrace{\begin{bmatrix}-0.3536\\0.6124\\-0.7071\end{bmatrix}}_{\boldsymbol{u}_1} & \underbrace{\begin{bmatrix}-0.9297\\-0.1465\\0.3380\end{bmatrix}}_{\boldsymbol{u}_2} & \underbrace{\begin{bmatrix}0.1034\\0.7769\\0.6211\end{bmatrix}}_{\boldsymbol{u}_3}\end{bmatrix}=\begin{bmatrix}2.8284 & 0 & 0\end{bmatrix} \tag{17}$$

特征值分解

当然，我们也可以用特征值分解得到 $\boldsymbol{U}$。首先计算格拉姆矩阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$:

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}=\begin{bmatrix}1 & -1\\-\sqrt{3} & \sqrt{3}\\2 & -2\end{bmatrix}\begin{bmatrix}1 & -1\\-\sqrt{3} & \sqrt{3}\\2 & -2\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}2 & -3.4641 & 4\\-3.4641 & 6 & -6.9282\\4 & -6.9282 & 8\end{bmatrix} \tag{18}$$

对 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 特征值分解可以得到 $\boldsymbol{U}$:

$$\begin{aligned}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}&=\begin{bmatrix}2 & -3.4641 & 4\\-3.4641 & 6 & -6.9282\\4 & -6.9282 & 8\end{bmatrix}\\&=\underbrace{\begin{bmatrix}-0.3536 & -0.9297 & 0.1034\\0.6124 & -0.1465 & 0.7769\\-0.7071 & 0.3380 & 0.6211\end{bmatrix}}_{\boldsymbol{U}}\begin{bmatrix}16 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}\underbrace{\begin{bmatrix}-0.3536 & 0.6124 & -0.7071\\-0.9297 & -0.1465 & 0.3380\\0.1034 & 0.7769 & 0.6211\end{bmatrix}}_{\boldsymbol{U}^{\mathrm{T}}}\end{aligned} \tag{19}$$

图 13 所示为矩阵 $\boldsymbol{X}$ 的**列空间** $C(\boldsymbol{X})$ 和**左零空间** $\mathrm{Null}(\boldsymbol{X}^{\mathrm{T}})$ 之间关系。

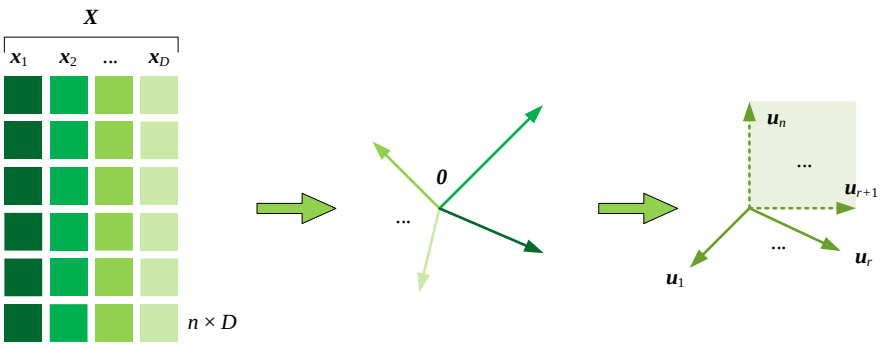


图 13. 矩阵 $\boldsymbol{X}$ 的**列空间** $C(\boldsymbol{X})$ 和**左零空间** $\mathrm{Null}(\boldsymbol{X}^{\mathrm{T}})$

行向量

下面，我们聊一下 $\boldsymbol{X}$ 矩阵的**行向量**。

很明显 $\boldsymbol{X}$ 的三个**行向量**也是线性相关:

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} 1 & -1 \end{bmatrix}, \quad \boldsymbol{x}^{(2)} = \begin{bmatrix} -\sqrt{3} & \sqrt{3} \end{bmatrix}, \quad \boldsymbol{x}^{(3)} = \begin{bmatrix} 2 & -2 \end{bmatrix} \tag{20}$$

如图 14 (a) 所示，为了装下**行向量** $\boldsymbol{x}^{(1)}$、$\boldsymbol{x}^{(2)}$ 和 $\boldsymbol{x}^{(3)}$，我们需要二维直角坐标系 $\mathbb{R}^2$。而图 14 (b) 告诉我们，用 $\boldsymbol{v}_1$ 这个单位向量就足以描述 $\boldsymbol{x}^{(1)}$、$\boldsymbol{x}^{(2)}$ 和 $\boldsymbol{x}^{(3)}$，因为 $\boldsymbol{x}^{(1)}$、$\boldsymbol{x}^{(2)}$ 和 $\boldsymbol{x}^{(3)}$可以写成：

$$\boldsymbol{x}^{(1)} = a\boldsymbol{v}_1^{\mathrm{T}}, \quad \boldsymbol{x}^{(2)} = b\boldsymbol{v}_1^{\mathrm{T}}, \quad \boldsymbol{x}^{(3)} = c\boldsymbol{v}_1^{\mathrm{T}} \tag{21}$$

白话说，$\mathbb{R}^2$ 中一维子空间 span($\boldsymbol{v}_1$) 就足够装下 $\boldsymbol{X}$ 的三个**行向量**。

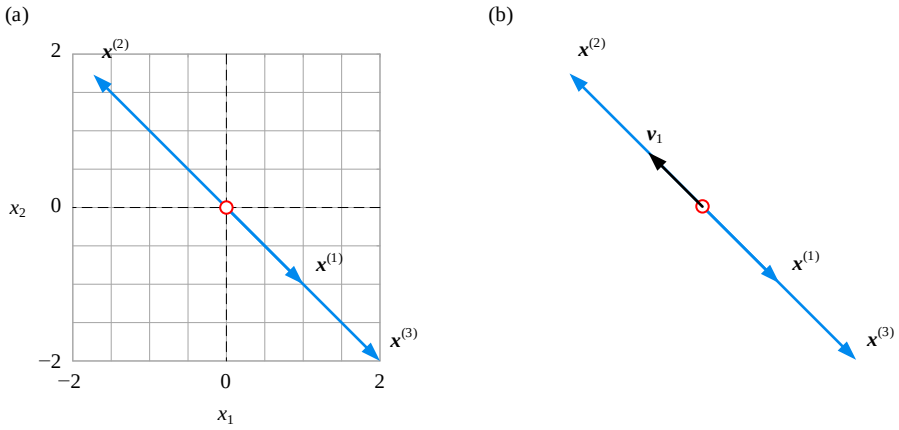


图 14. 从二维空间到一维空间

(15) 给出的 SVD 分解结果已经帮我们找到了 $\boldsymbol{v}_1$。拔出萝卜带出泥，我们也计算得到 $\boldsymbol{v}_2$。$\boldsymbol{v}_1$ 张成**行空间** $R(\boldsymbol{X})$ = span($\boldsymbol{v}_1$)，$\boldsymbol{v}_2$ 张成**零空间** Null($\boldsymbol{X}$) = span($\boldsymbol{v}_2$)。

而规范正交基 [$\boldsymbol{v}_1$, $\boldsymbol{v}_2$] 则是张成 $\mathbb{R}^2$ 无数规范正交基中的一个。[$\boldsymbol{v}_1$, $\boldsymbol{v}_2$] 是因 $\boldsymbol{X}$ 而来。

如图 15 (b) 所示，显然 $R(\boldsymbol{X})$ = span($\boldsymbol{v}_1$) 垂直于 Null($\boldsymbol{X}$) = span($\boldsymbol{v}_2$)，即互为正交补。

⚠ 格外注意，大家不要留下错误印象，$\boldsymbol{x}_1$ 或 $\boldsymbol{x}^{(1)}$ 就是 $\boldsymbol{u}_1$ 或 $\boldsymbol{v}_1$ 的方向重合。一般情况这种重合关系不存在，本例中产生重合的原因是 rank($\boldsymbol{X}$) = 1。

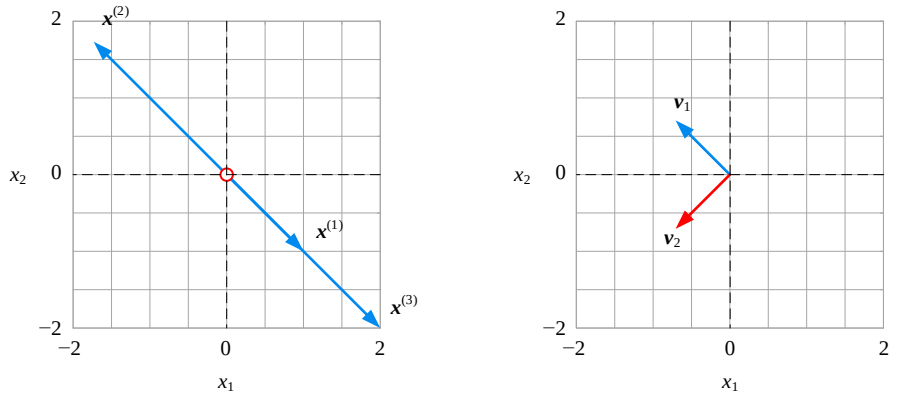


图 15. 矩阵 $\boldsymbol{X}$ 的**行空间** $R(\boldsymbol{X})$ 和**零空间** Null($\boldsymbol{X}$)

把**行向量** $\boldsymbol{x}^{(1)}$投影到 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 中得到：

$$\boldsymbol{x}^{(1)}\boldsymbol{V} = \begin{bmatrix} 1 & -1 \end{bmatrix} \left[\underbrace{\begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix}}_{\boldsymbol{v}_1} \underbrace{\begin{bmatrix} -0.7071 \\ -0.7071 \end{bmatrix}}_{\boldsymbol{v}_2} \right] = \begin{bmatrix} -1.4142 & 0 \end{bmatrix} \tag{22}$$

也就是说，$\boldsymbol{x}^{(1)}$ 在 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 这个规范正交基中的坐标为 $(-1.4142, 0)$。请大家自己计算 $\boldsymbol{x}^{(2)}$和 $\boldsymbol{x}^{(3)}$ 投影到 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 结果。

总结来说，Null($\boldsymbol{X}$) 是 $\boldsymbol{X}$ 的**零空间**是因为 $\boldsymbol{X}$ 投影到这个空间的结果都是 0。而 Null($\boldsymbol{X}^{\mathrm{T}}$) 是 $\boldsymbol{X}$ 的**左零空间**是因为，$\boldsymbol{X}^{\mathrm{T}}$ 投影到这个空间的结果都是 0。

特征值分解

下面，我们再用特征值分解求解 $\boldsymbol{V}$。也是先计算格拉姆矩阵 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$：

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \begin{bmatrix} 1 & -1 \\ -\sqrt{3} & \sqrt{3} \\ 2 & -2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 1 & -1 \\ -\sqrt{3} & \sqrt{3} \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 8 & -8 \\ -8 & 8 \end{bmatrix} \tag{23}$$

对 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 进行特征值分解，便得到 $\boldsymbol{V}$：

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \begin{bmatrix} 8 & -8 \\ -8 & 8 \end{bmatrix} = \underbrace{\begin{bmatrix} -0.7071 & -0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}}_{\boldsymbol{V}} \begin{bmatrix} 16 & 0 \\ 0 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} -0.7071 & 0.7071 \\ -0.7071 & -0.7071 \end{bmatrix}}_{\boldsymbol{V}^{\mathrm{T}}} \tag{24}$$

图 16 所示为矩阵 $\boldsymbol{X}$ 的**行空间** $R(\boldsymbol{X})$ 和**零空间** Null($\boldsymbol{X}$) 之间的关系。

此外，值得大家注意的是，比较 (19) 和 (24)，大家容易发现，两个特征值分解都得到了 16 这个特征值。为什么会出现这种情况？下一节将给出答案。

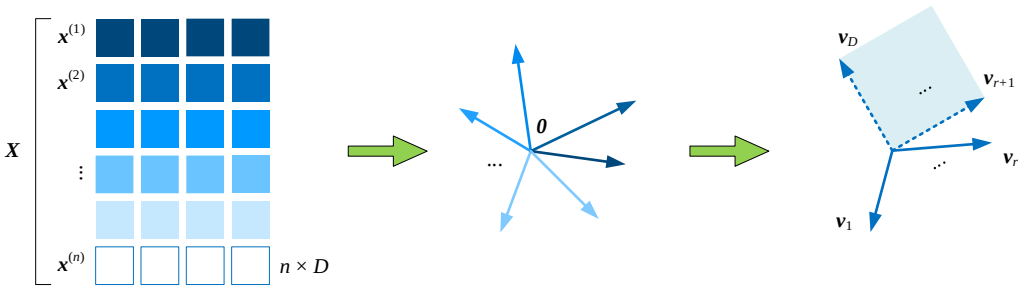


图 16. 矩阵 $\boldsymbol{X}$ 的**行空间** $R(\boldsymbol{X})$ 和**零空间** Null($\boldsymbol{X}$)

通过以上分析，希望大家能从几何角度理解六个空间之间的关系。此外，大家也看到奇异值分解的强大之处——任何实数矩阵都可以进行奇异值分解。

23.5 格拉姆矩阵：向量模、夹角余弦值的集合体

我们可以把矩阵 $\boldsymbol{X}$ 的每一行或每一列分别视作向量。而对于一个向量而言，最能概括它的性质的基本信息莫过于——长度和方向。

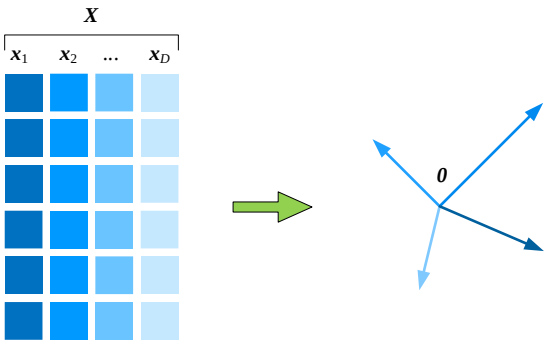


图 17. 矩阵 $\boldsymbol{X}$ 列向量几何化为空间向量

向量长度不难确定，向量模 (L^2 范数) 就是向量长度。

然而，向量的方向该怎么量化？我们目前接触到几何形体定位最常用的手段是平面或三维直角坐标系，直角坐标系在量化位置、长度、方向具有天然优势。

但是对于图 17 所示向量，随着维度不断升高，直角坐标系显得有点力有不逮。

极坐标系

于是，我们想到利用极坐标量化方向。

如图 18 所示，极坐标中定位需要长度和角度，恰巧对应向量的两个重要的元素。唯一的问题是，极坐标系中量化向量和极轴的夹角，即绝对角度。我们接触最多的是向量两两夹角，即相对角度值。

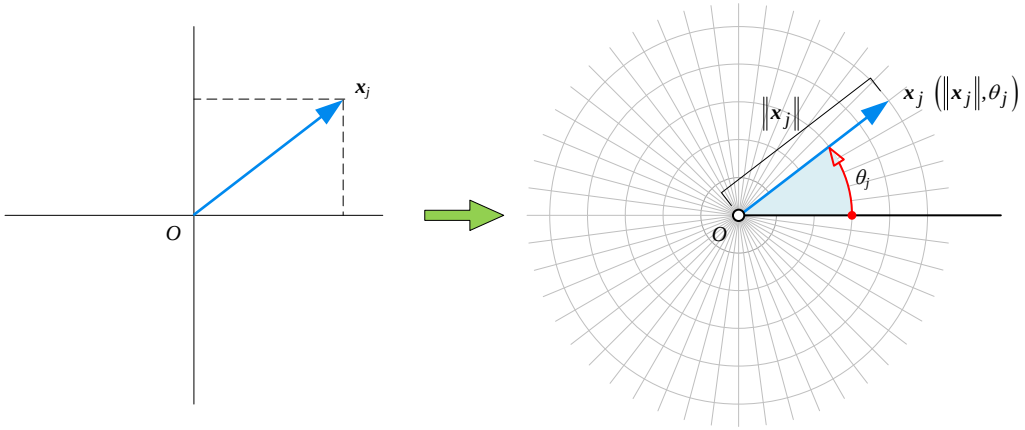


图 18. 从平面直角坐标系到极坐标系，图片参考《数学要素》

此外，向量两两夹角数量也是个问题。数据矩阵 $\boldsymbol{X}$ 有 D 个列向量，这意味着我们可以得到 D 个向量模，以及 $D(D-1)/2$ (C_D^2) 个向量两两夹角余弦值。按照怎样规则保存这些结果？我们反复提到的格拉姆矩阵就是解决方案。而且，本书第 12 章介绍的 Cholesky 分解则帮我们找到这些向量的“绝对位置”。

长度、相对夹角

给定一个 $n \times D$ 数据矩阵 $\boldsymbol{X}$，形状细高，也就是 $n > D$，它的格拉姆矩阵 $\boldsymbol{G}$ 为：

$$\boldsymbol{G} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \tag{25}$$

如图 19 所示，$\boldsymbol{G}$ 为对称方阵，形状为 $D \times D$。

用向量内积来表达 $\boldsymbol{G}$:

$$\boldsymbol{G} = \begin{bmatrix} \langle \boldsymbol{x}_1, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_1, \boldsymbol{x}_D \rangle \\ \langle \boldsymbol{x}_2, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_2, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_2, \boldsymbol{x}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{x}_D, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_D, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_D, \boldsymbol{x}_D \rangle \end{bmatrix} = \begin{bmatrix} \|\boldsymbol{x}_1\|\|\boldsymbol{x}_1\|\cos\theta_{1,1} & \|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\|\cos\theta_{2,1} & \cdots & \|\boldsymbol{x}_1\|\|\boldsymbol{x}_D\|\cos\theta_{1,D} \\ \|\boldsymbol{x}_2\|\|\boldsymbol{x}_1\|\cos\theta_{1,2} & \|\boldsymbol{x}_2\|\|\boldsymbol{x}_2\|\cos\theta_{2,2} & \cdots & \|\boldsymbol{x}_2\|\|\boldsymbol{x}_D\|\cos\theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \|\boldsymbol{x}_D\|\|\boldsymbol{x}_1\|\cos\theta_{1,D} & \|\boldsymbol{x}_D\|\|\boldsymbol{x}_2\|\cos\theta_{2,D} & \cdots & \|\boldsymbol{x}_D\|\|\boldsymbol{x}_D\|\cos\theta_{D,D} \end{bmatrix} \tag{26}$$

可以发现，$\boldsymbol{G} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 包含的信息有两方面：$\boldsymbol{X}$ 列向量的模、列向量两两夹角余弦值。

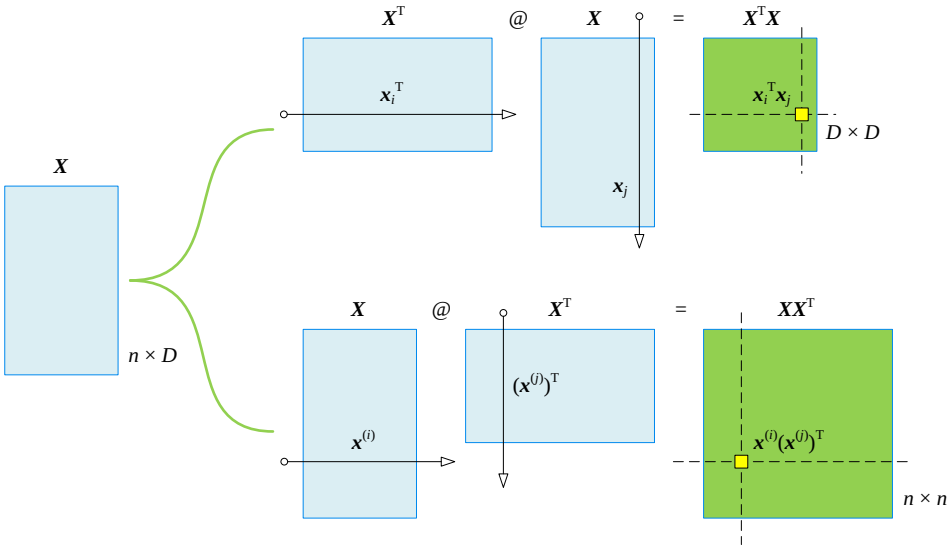


图 19. 两个格拉姆矩阵

而余弦相似度矩阵 $\boldsymbol{C}$ 则进一步减小信息量，只关注列向量夹角余弦值：

$$\boldsymbol{C}=\begin{bmatrix} \frac{\boldsymbol{x}_1\cdot\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|\|\boldsymbol{x}_1\|} & \frac{\boldsymbol{x}_1\cdot\boldsymbol{x}_2}{\|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\|} & \cdots & \frac{\boldsymbol{x}_1\cdot\boldsymbol{x}_D}{\|\boldsymbol{x}_1\|\|\boldsymbol{x}_D\|} \\ \frac{\boldsymbol{x}_2\cdot\boldsymbol{x}_1}{\|\boldsymbol{x}_2\|\|\boldsymbol{x}_1\|} & \frac{\boldsymbol{x}_2\cdot\boldsymbol{x}_2}{\|\boldsymbol{x}_2\|\|\boldsymbol{x}_2\|} & \cdots & \frac{\boldsymbol{x}_2\cdot\boldsymbol{x}_D}{\|\boldsymbol{x}_2\|\|\boldsymbol{x}_D\|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\boldsymbol{x}_D\cdot\boldsymbol{x}_1}{\|\boldsymbol{x}_D\|\|\boldsymbol{x}_1\|} & \frac{\boldsymbol{x}_D\cdot\boldsymbol{x}_2}{\|\boldsymbol{x}_D\|\|\boldsymbol{x}_2\|} & \cdots & \frac{\boldsymbol{x}_D\cdot\boldsymbol{x}_D}{\|\boldsymbol{x}_D\|\|\boldsymbol{x}_D\|} \end{bmatrix}=\begin{bmatrix} 1 & \cos\theta_{2,1} & \cdots & \cos\theta_{1,D} \\ \cos\theta_{1,2} & 1 & \cdots & \cos\theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \cos\theta_{1,D} & \cos\theta_{2,D} & \cdots & 1 \end{bmatrix} \tag{27}$$

计算 $\boldsymbol{X}^{\mathrm{T}}$ 的格拉姆矩阵，并定义其为 $\boldsymbol{H}$：

$$\boldsymbol{H}=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \tag{28}$$

如图 19 所示，$\boldsymbol{H}$ 为对称方阵，形状为 $n\times n$。

用向量内积来表达 $\boldsymbol{H}$：

$$\boldsymbol{H}=\begin{bmatrix} \left\langle \boldsymbol{x}^{(1)},\boldsymbol{x}^{(1)}\right\rangle & \left\langle \boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\right\rangle & \cdots & \left\langle \boldsymbol{x}^{(1)},\boldsymbol{x}^{(n)}\right\rangle \\ \left\langle \boldsymbol{x}^{(2)},\boldsymbol{x}^{(1)}\right\rangle & \left\langle \boldsymbol{x}^{(2)},\boldsymbol{x}^{(2)}\right\rangle & \cdots & \left\langle \boldsymbol{x}^{(2)},\boldsymbol{x}^{(n)}\right\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \left\langle \boldsymbol{x}^{(n)},\boldsymbol{x}^{(1)}\right\rangle & \left\langle \boldsymbol{x}^{(n)},\boldsymbol{x}^{(2)}\right\rangle & \cdots & \left\langle \boldsymbol{x}^{(n)},\boldsymbol{x}^{(n)}\right\rangle \end{bmatrix} \tag{29}$$

$\boldsymbol{H}=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 也包含两方面的信息：$\boldsymbol{X}$ 行向量的模、行向量之间两两夹角余弦值。

特征值分解

下面用特征值分解找到 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 和 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 之间联系。

先对 $\boldsymbol{G}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 进行特征值分解，得到：

$$\boldsymbol{G}=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} \tag{30}$$

假设 $\lambda_{\boldsymbol{G}}$ 为 $\boldsymbol{G}$ 的一个特征值，对应的特征向量为 $\boldsymbol{v}$，由此得到等式：

$$\boldsymbol{G}\boldsymbol{v}=\lambda_{\boldsymbol{G}}\boldsymbol{v} \tag{31}$$

即，

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}=\lambda_{\boldsymbol{G}}\boldsymbol{v} \tag{32}$$

然后对 $\boldsymbol{H}$ 特征值分解：

$$\boldsymbol{H}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{T}} \tag{33}$$

$\boldsymbol{U}$ 为特征向量矩阵，$\boldsymbol{D}$ 为特征值对角阵。

假设 $\lambda_{\boldsymbol{H}}$ 为 $\boldsymbol{H}$ 的一个特征值，对应特征向量为 $\boldsymbol{u}$，构造等式：

$$\boldsymbol{H}\boldsymbol{u}=\lambda_{\boldsymbol{H}}\boldsymbol{u} \tag{34}$$

即，

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{u}=\lambda_{\boldsymbol{H}}\boldsymbol{u} \tag{35}$$

(32) 左右乘以 $\boldsymbol{X}$，得到：

$$\underset{\boldsymbol{u}}{\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}}=\lambda_{\boldsymbol{G}}\underset{\boldsymbol{u}}{\boldsymbol{X}\boldsymbol{v}} \tag{36}$$

比较 (35) 和 (36)，可以发现 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 和 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 特征值分解得到的非零特征值存在等价关系。这就回答了为什么 (19) 和 (24) 都有 16 这个特征值这个问题。其实，我们在本书第 16 章也谈过这一现象。

23.6 标准差向量：以数据质心为起点

协方差矩阵可以看成是特殊的格拉姆矩阵，协方差矩阵也是一个“向量模”、“向量间夹角”信息的集合体。

对于形状为 $n\times D$ 的样本数据矩阵 $\boldsymbol{X}$，它的协方差矩阵 $\boldsymbol{\Sigma}$ 可以通过下式计算得到。

$$\boldsymbol{\Sigma}=\frac{\left(\underbrace{\boldsymbol{X}-\mathrm{E}(\boldsymbol{X})}_{\text{Centered}}\right)^{\mathrm{T}}\left(\underbrace{\boldsymbol{X}-\mathrm{E}(\boldsymbol{X})}_{\text{Centered}}\right)}{n-1}=\frac{\boldsymbol{X}_c^{\mathrm{T}}\boldsymbol{X}_c}{n-1} \tag{37}$$

分母上，$n-1$ 仅仅起到取平均作用。$\boldsymbol{X}_c$ 的格拉姆矩阵为：

$$\boldsymbol{X}_c^{\mathrm{T}}\boldsymbol{X}_c = (n-1)\boldsymbol{\Sigma} \tag{38}$$

图 20 所示，$\boldsymbol{X}$ 列向量的向量起点为 $\boldsymbol{0}$。而去均值获得 $\boldsymbol{X}_c$ 过程，相当于把列向量起点移动到质心 E($\boldsymbol{X}$)：

$$\boldsymbol{X}_c = \boldsymbol{X} - \mathrm{E}(\boldsymbol{X}) \tag{39}$$

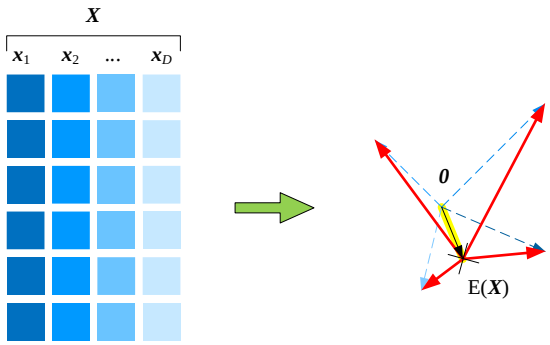


图 20. 数据质心为 $\boldsymbol{X}_c$ 列向量的起点

将 $\boldsymbol{X}_c$ 列向量的起点也平移到 $\boldsymbol{0}$，和 $\boldsymbol{X}$ 列向量起点对齐。图 21 比较 $\boldsymbol{X}$ 和 $\boldsymbol{X}_c$ 列向量，显然去均值之后，向量的长度和向量之间的夹角都发生了变化。有一种特例是，当质心 E($\boldsymbol{X}$) 本来就在 $\boldsymbol{0}$ 时，这样 $\boldsymbol{X} = \boldsymbol{X}_c$。

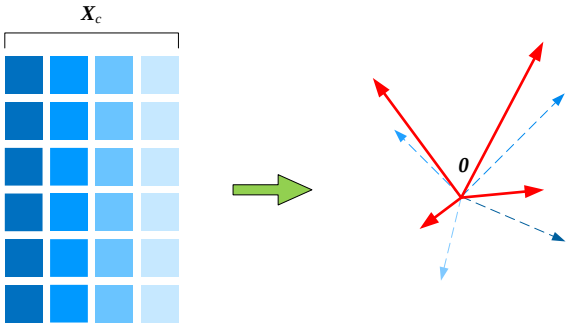


图 21. 比较 $\boldsymbol{X}$ 和 $\boldsymbol{X}_c$ 列向量

在数据科学和机器学习应用中，最常见的三大类数据矩阵就是：1) 原始数据矩阵 $\boldsymbol{X}$；2) 中心化数据矩阵 $\boldsymbol{X}_c$；3) 标准化数据矩阵 $\boldsymbol{Z}_X$ (z 分数)。

根据本章前文介绍数据矩阵 $\boldsymbol{X}$ 有四个空间；显然，中心化数据矩阵 $\boldsymbol{X}_c$ 也有自己的四个空间！那么大家立刻会想到，标准化数据矩阵 $\boldsymbol{Z}_X$，肯定也有对应的四个空间！

也就是说，如果用 SVD 分解 $\boldsymbol{X}$、$\boldsymbol{X}_c$、$\boldsymbol{Z}_X$ 这三个数据矩阵，会得到不同的结果。下一章则通过各种矩阵分解帮我们分析这三大类数据特点和区别。

标准差向量

整理 (37) 得到 $\boldsymbol{X}_c^{\mathrm{T}}\boldsymbol{X}_c$：

$$\boldsymbol{X}_c^{\mathrm{T}}\boldsymbol{X}_c=(n-1)\boldsymbol{\Sigma}=(n-1)\begin{bmatrix}\sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,D}\sigma_1\sigma_D\\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2,D}\sigma_2\sigma_D\\ \vdots & \vdots & \ddots & \vdots\\ \rho_{1,D}\sigma_1\sigma_D & \rho_{2,D}\sigma_2\sigma_D & \cdots & \sigma_D^2\end{bmatrix} \tag{40}$$

对比 (26) 和 (40)，我们可以把标准差 σ_j 也看做是向量 $\boldsymbol{\sigma}_j$，我们给它起个名字“标准差向量”。

标准差向量 $\boldsymbol{\sigma}_j$ 之间的夹角的余弦值便是相关性系数。这样 (40) 可以写成：

$$\boldsymbol{\Sigma}=\begin{bmatrix}\langle\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_1\rangle & \langle\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\rangle & \cdots & \langle\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_D\rangle\\ \langle\boldsymbol{\sigma}_2,\boldsymbol{\sigma}_1\rangle & \langle\boldsymbol{\sigma}_2,\boldsymbol{\sigma}_2\rangle & \cdots & \langle\boldsymbol{\sigma}_2,\boldsymbol{\sigma}_D\rangle\\ \vdots & \vdots & \ddots & \vdots\\ \langle\boldsymbol{\sigma}_D,\boldsymbol{\sigma}_1\rangle & \langle\boldsymbol{\sigma}_D,\boldsymbol{\sigma}_2\rangle & \cdots & \langle\boldsymbol{\sigma}_D,\boldsymbol{\sigma}_D\rangle\end{bmatrix}=\begin{bmatrix}\|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_1\|\cos\phi_{1,1} & \|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_2\|\cos\phi_{2,1} & \cdots & \|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_D\|\cos\phi_{1,D}\\ \|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_1\|\cos\phi_{1,2} & \|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_2\|\cos\phi_{2,2} & \cdots & \|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_D\|\cos\phi_{2,D}\\ \vdots & \vdots & \ddots & \vdots\\ \|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_1\|\cos\phi_{1,D} & \|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_2\|\cos\phi_{2,D} & \cdots & \|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_D\|\cos\phi_{D,D}\end{bmatrix} \tag{41}$$

如果两个随机变量线性相关，则对应标准差向量平行；如果两个随机变量线性无关，对应的标准差向量正交。

图 22 比较余弦相似度和相关性系数。

⚠ 注意，图 22 忽略了 $n-1$ 对缩放的影响。

相关性系数和余弦相似性都描述了两个“相似程度”，也就是靠近的程度；两者取值范围都是 [–1, 1]。越靠近 1，说明越相似，向量越贴近；越靠近–1，说明越不同，向量越背离。

不同的是，相关性系数量化“标准差向量”$\boldsymbol{\sigma}_j$ 之间相似，余弦相似性量化数据矩阵 $\boldsymbol{X}$ 列向量 $\boldsymbol{x}_j$ 之间相似。$\boldsymbol{x}_j$ 向量的始点为原点 $\boldsymbol{0}$，$\boldsymbol{\sigma}_j$ 向量始点为数据质心。

大家可能想要知道 $\boldsymbol{x}_j$ 向量和 $\boldsymbol{\sigma}_j$ 向量到底是什么？它们的具体坐标值又如何？我们下一章回答这个问题。

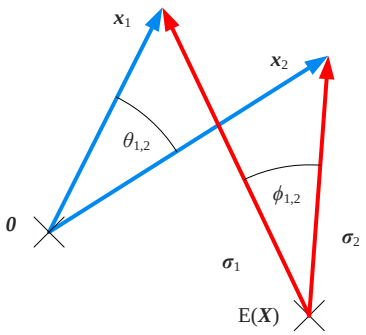


图 22. 余弦相似度和相关性系数的关系，图中忽略标准差向量的缩放系数

相关性系数

类似余弦相似度矩阵 $\boldsymbol{C}$，相关性系数矩阵 $\boldsymbol{P}$ 仅仅含有标准差向量夹角 (即相关性系数) 这一层信息：

$$\boldsymbol{P}=\begin{bmatrix} \frac{\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_1}{\|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_1\|} & \frac{\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2}{\|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_2\|} & \cdots & \frac{\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_D}{\|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_D\|} \\ \frac{\boldsymbol{\sigma}_2\cdot\boldsymbol{\sigma}_1}{\|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_1\|} & \frac{\boldsymbol{\sigma}_2\cdot\boldsymbol{\sigma}_2}{\|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_2\|} & \cdots & \frac{\boldsymbol{\sigma}_2\cdot\boldsymbol{\sigma}_D}{\|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_D\|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\boldsymbol{\sigma}_D\cdot\boldsymbol{\sigma}_1}{\|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_1\|} & \frac{\boldsymbol{\sigma}_D\cdot\boldsymbol{\sigma}_2}{\|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_2\|} & \cdots & \frac{\boldsymbol{\sigma}_D\cdot\boldsymbol{\sigma}_D}{\|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_D\|} \end{bmatrix}=\begin{bmatrix} 1 & \cos\phi_{2,1} & \cdots & \cos\phi_{1,D} \\ \cos\phi_{1,2} & 1 & \cdots & \cos\phi_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \cos\phi_{1,D} & \cos\phi_{2,D} & \cdots & 1 \end{bmatrix} \tag{42}$$

如图 23 所示，以二元随机数为例，相关性系数可以通过散点、二元高斯分布 PDF 曲面、PDF 等高线、椭圆表达。有了本节内容，在众多可视化方案基础上，相关性系数又多了一层几何表达。本系列丛书《概率统计》将讲解随机数、二元高斯分布、概率密度函数 PDF 等概念。此外，《概率统计》中大家会看到无处不在的椭圆。

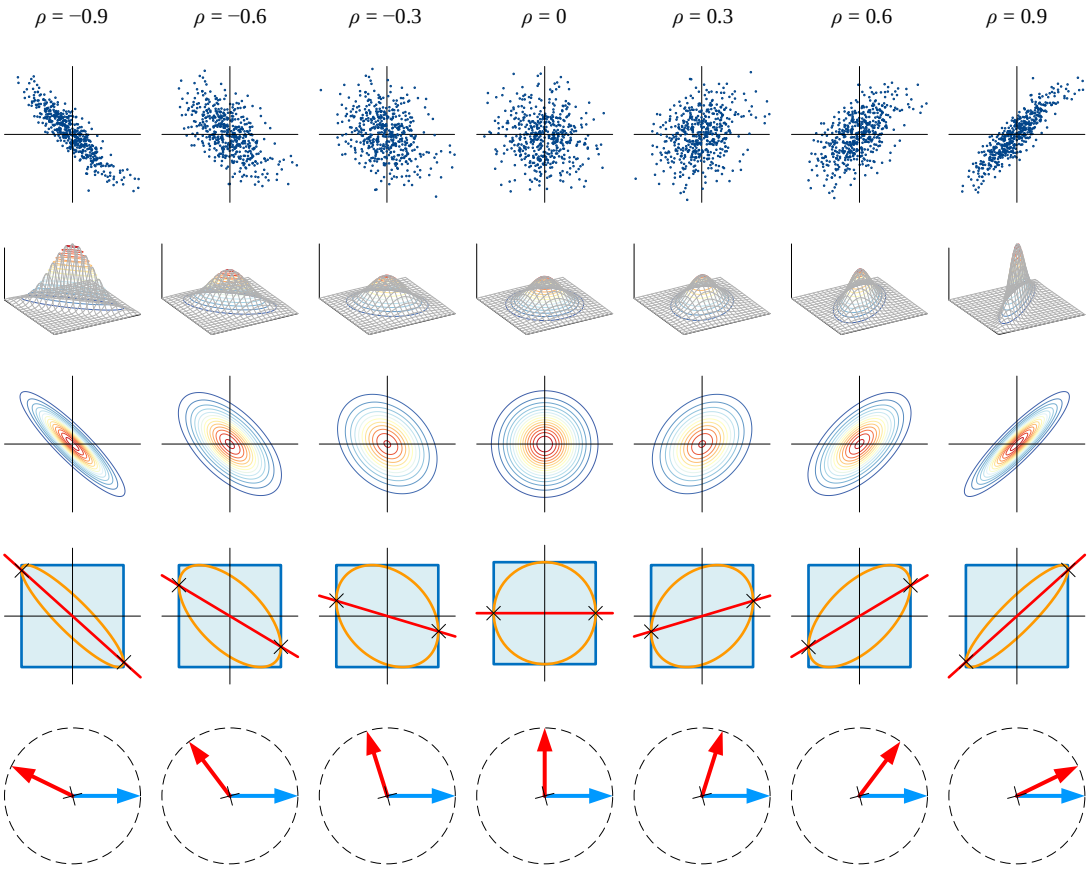


图 23. 相关性系数的几种表达，图中标准差相等，质心位于原点

23.7 白话说空间：以鸢尾花数据为例

本章最后一节，我们尝试尽量用大白话把本章之前讲解的四个空间说清楚。本节用的数据是鸢尾花数据前两列，即 $\boldsymbol{X}_{150\times 2} = [\boldsymbol{x}_1, \boldsymbol{x}_2]$。

标准正交基

矩阵 $\boldsymbol{X}$ 有 150 行、2 列，有 150 个行向量，它们就是图 24 中灰色带箭头的线段。为了装下这 150 个行向量，我们自然而然地想到了 $[\boldsymbol{e}_1, \boldsymbol{e}_2]$ —— 平面 $\mathbb{R}^2$ 的标准正交基。

图中散点横坐标就对应 $\boldsymbol{X}$ 的第一列向量 $\boldsymbol{x}_1$，纵坐标对应 $\boldsymbol{X}$ 的第二列向量 $\boldsymbol{x}_2$。

本书第 7 章讲过，$[\boldsymbol{e}_1, \boldsymbol{e}_2]$ 表示平面 $\mathbb{R}^2$ 最为自然，因此叫做“标准”正交基。

大家知道，1 维空间是相当于一条过原点的直线，显然图 24 的散点不在一条过原点的直线上。也就是说，要想装下 $\boldsymbol{X}$ 的行向量至少需要一个二维空间。因此 $[\boldsymbol{e}_1, \boldsymbol{e}_2]$ 对于图 24 向量来说，大小正好，没有任何富余。

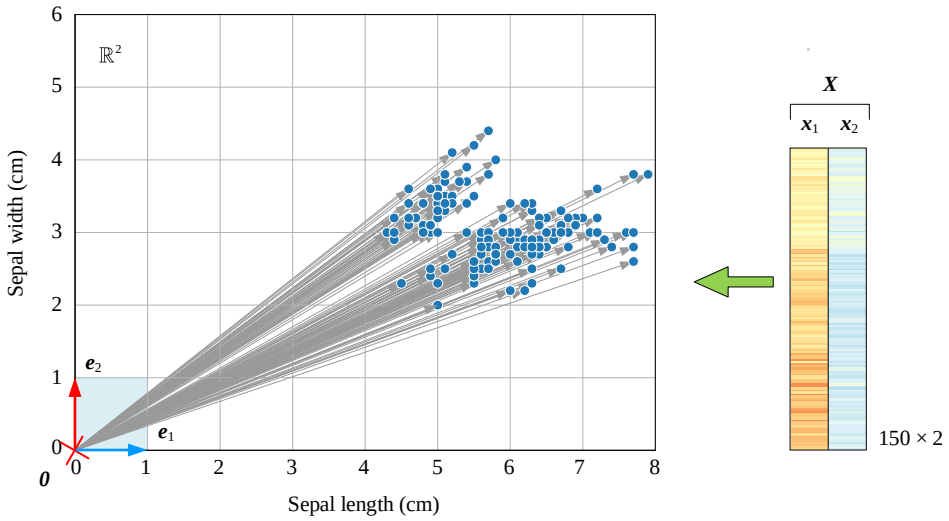


图 24. 找一个能够装下 $\boldsymbol{X}$ 行向量的空间

行空间、零空间

根据本章前文所学，为了计算 $\boldsymbol{X}$ 的**行空间、零空间**，我们可以首先计算格拉姆矩阵 (如所示)，然后对 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 特征值分解：

$$\boldsymbol{G} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \begin{bmatrix} 5223.85 & 2673.43 \\ 2673.43 & 1430.40 \end{bmatrix} = \underbrace{\begin{bmatrix} 0.888 & -0.459 \\ 0.459 & 0.888 \end{bmatrix}}_{\boldsymbol{V}} \underbrace{\begin{bmatrix} 6605.05 & \\ & 49.20 \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix} 0.888 & 0.459 \\ -0.459 & 0.888 \end{bmatrix}}_{\boldsymbol{V}^{\mathrm{T}}} \tag{43}$$

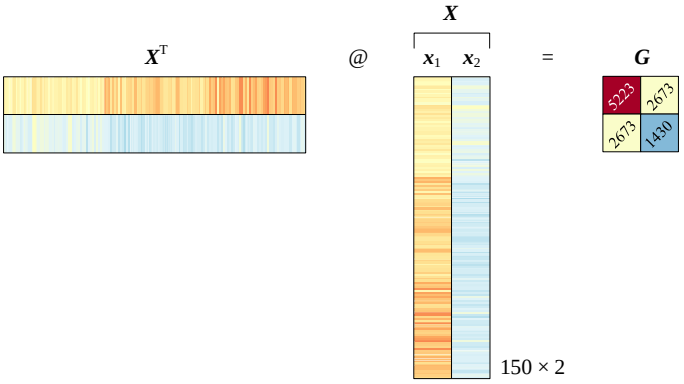


图 25. 计算格拉姆矩阵 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$

可以张起 $\mathbb{R}^2$ 的规范正交基有无数个，(43) 中的 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2]$ 只是其中一个。$[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 在平面上的网格如图 26 所示。$\boldsymbol{X}$ 在这个 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 坐标系中有全新的坐标点。请大家自己回忆怎么计算新的坐标点。

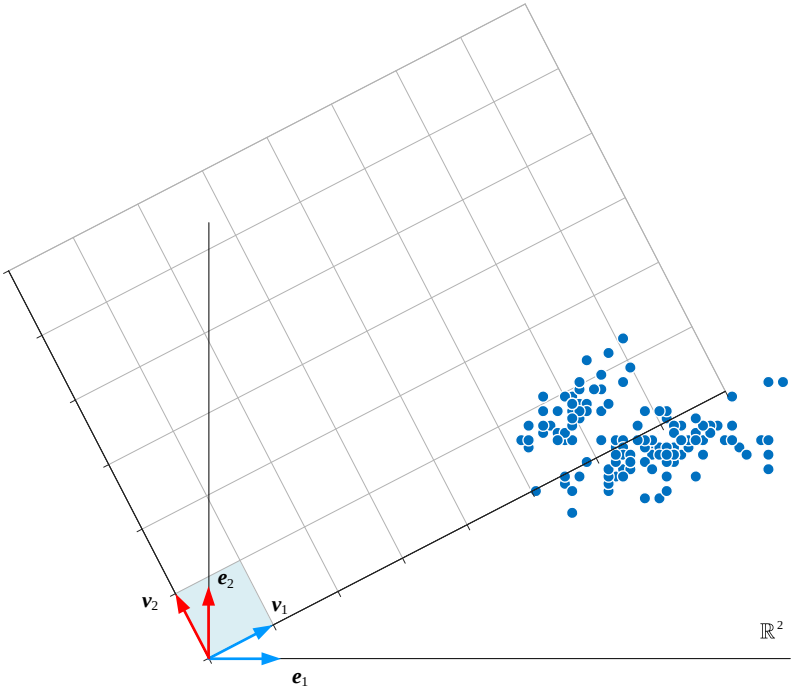


图 26. 规范正交基 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$

大家可能会问，之前我们已经在 $[\boldsymbol{e}_1, \boldsymbol{e}_2]$ 这个坐标系中“自然地”描绘了数据矩阵 $\boldsymbol{X}$，为何还要劳神费力地寻找 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$。

这是因为对 $\boldsymbol{X}$ 来说，$[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 可谓“量身打造”！下面，我们看看 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 有何特殊之处。

如图 27 所示，$\boldsymbol{X}$ 向 $\boldsymbol{v}_1$ 投影结果为 $\boldsymbol{y}_1 = \boldsymbol{X}\boldsymbol{v}_1$。$\boldsymbol{X}$ 是图中的蓝色点 ●，$\boldsymbol{y}_1$ 为图中的蓝色叉 × 在 span($\boldsymbol{v}_1$) 上的坐标值。蓝色叉 × 距离原点欧氏距离的平方对应 (43) 中特征值 $\lambda_1 = 6605.05$。

利用本书第 18 章介绍的优化视角来观察，给定平面内任意单位向量 $\boldsymbol{v}$，$\|\boldsymbol{X}\boldsymbol{v}\|_2^2$ 的最大值就是 λ_1。而 $\|\boldsymbol{X}\boldsymbol{v}\|_2$ 能取得的最大值就是 $\sqrt{\lambda_1}$，对应 $\boldsymbol{X}$ 的最大奇异值，即 $s_1 = \sqrt{\lambda_1}$。

反之，如图 28 所示，$\boldsymbol{X}$ 向 $\boldsymbol{v}_2$ 投影结果为 $\boldsymbol{y}_2 = \boldsymbol{X}\boldsymbol{v}_2$。给定平面内任意单位向量 $\boldsymbol{v}$，$\|\boldsymbol{X}\boldsymbol{v}\|_2^2$ 的最小值就是 λ_2。

基底 $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ 对于 $\boldsymbol{X}$ 来说，也显得“捉襟见肘”，维度不能再进一步减小。

如果 $\boldsymbol{X}$ 非列满秩，$\boldsymbol{V}$ 就会出现“余富”，这个余富就是零空间。

比如，$\boldsymbol{X}_1 = \boldsymbol{X}\boldsymbol{v}_1 \otimes \boldsymbol{v}_1$ 就是图 27 中蓝色叉 × 在 $\mathbb{R}^2$ 中坐标。蓝色叉 × 显然都在一条过原点的直线上。$\boldsymbol{X}_1$ 的秩为 1 也印证了这一点。对于来说，span($\boldsymbol{v}_1$) 足够装下 $\boldsymbol{X}_1$，余富的 span($\boldsymbol{v}_2$) 就是 $\boldsymbol{X}_1$ 的零空间。很明显，$\boldsymbol{X}_1$ 在 span($\boldsymbol{v}_2$) 投影为零向量 $\boldsymbol{0}$。感兴趣的话，大家可以自己计算 $\boldsymbol{X}_1$ 的特征值，它的一个特征值是 $\lambda_1 = 6605.05$，另一个特征值为 0。

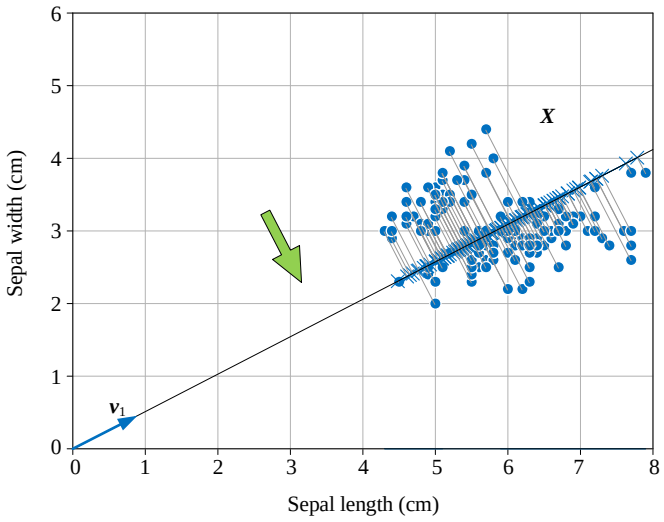


图 27. $\boldsymbol{X}$ 向 $\boldsymbol{v}_1$ 投影

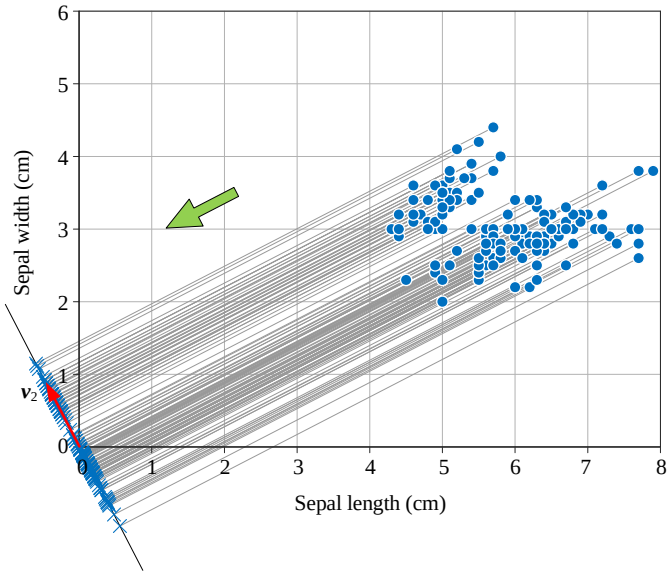


图 28. $\boldsymbol{X}$ 向 $\boldsymbol{v}_2$ 投影

列空间、左零空间

矩阵 $\boldsymbol{X}$ 有两个列向量 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$，$\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的行数都是 150。为了装下 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$，我们自然想到 $\mathbb{R}^{150}$。但是 $\mathbb{R}^{150}$ 对于矩阵 $\boldsymbol{X}$ 来说简直就是“高射炮打蚊子”，小题大做！

下面解释为什么。

为了计算矩阵 $\boldsymbol{X}$ 的**列空间、左零空间**，我们首先计算格拉姆矩阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$，计算过程如图 29 所示。格拉姆矩阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 形状为 150 × 150。格拉姆矩阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 看着很大，实际上它的秩只有 2。也就是说，$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 所有 150 个列向量都可以用两个列向量线性组合来表达。

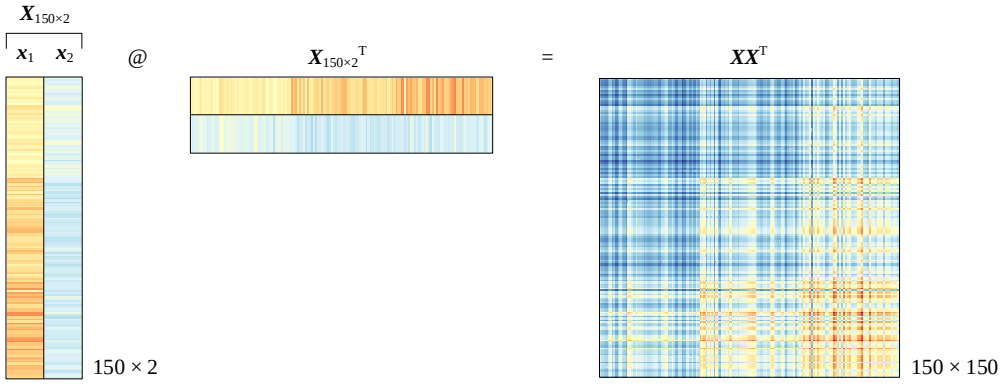


图 29. 计算格拉姆矩阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$

对 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 特征值分解得到特征向量构成的矩阵 $\boldsymbol{U}$ 如图 30 所示。$\boldsymbol{U}$ 的形状也是 150×150。$\boldsymbol{U}$ 是 $\mathbb{R}^{150}$ 中无数个规范正交阵中的一个。

$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 的非零特征值就是 (43) 中的两个特征值，剩余的特征值都为 0。也就是说，$\boldsymbol{U}$ 的前两列 $[\boldsymbol{u}_1, \boldsymbol{u}_2]$ 就是我们要找的列空间，$[\boldsymbol{u}_1, \boldsymbol{u}_2]$ 正好可以装下 $\boldsymbol{X}$。剩余的 148 列构成左零空间 $\mathrm{Null}(\boldsymbol{X}^{\mathrm{T}})$。也就是说，想要装下 $\boldsymbol{X}$ 的列向量，$\mathbb{R}^{150}$ 富富有余。

请大家用同样的思路分析 $\boldsymbol{X}_c$、$\boldsymbol{Z}_{\boldsymbol{X}}$。

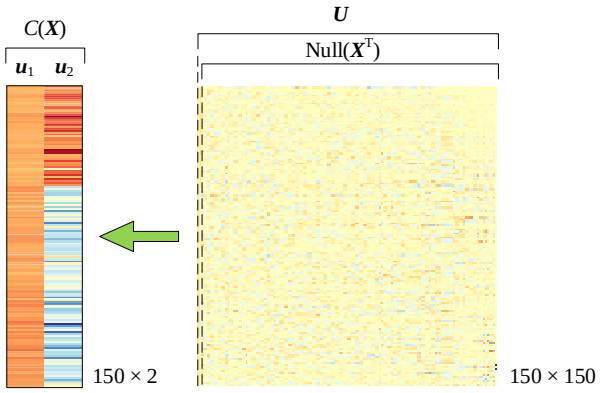


图 30. 格拉姆矩阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 的特征向量矩阵 $\boldsymbol{U}$

有数据的地方，就有矩阵！

有矩阵的地方，就有向量！

有向量的地方，就有几何！

有向量的地方，就有空间！

本书最后三章开启了一场特殊的旅行——“数据三部曲”。这三章梳理总结本书前文核心内容，同时展望这些数学工具的应用。本章作为“数据三部曲”的第一部，主要通过数据矩阵奇异值分解介绍了四个空间。

下图虽然是一幅图，但是其中有四幅子图，它们最能总结本章的核心内容——四个空间。强烈建议大家自行脑补图中缺失的各种符号，以及它们的意义。

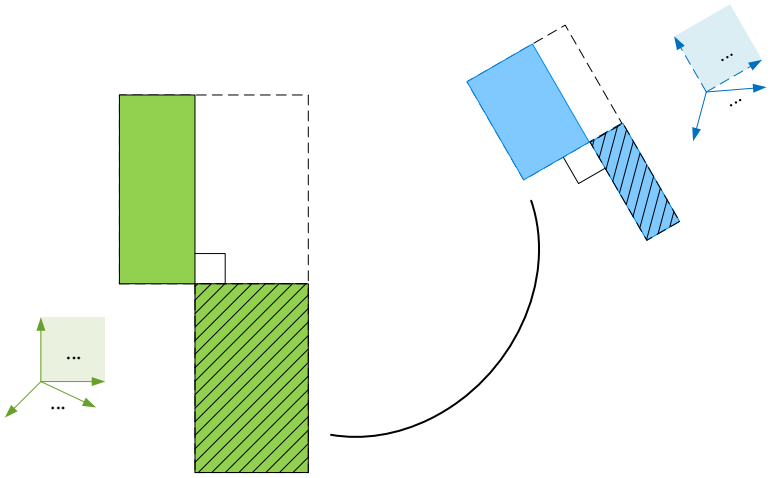


图 31. 总结本章重要内容的四幅图

此外，本书还引出了中心化数据、标准化数据，并创造了“标准差向量”这个概念。格拉姆矩阵是原始数据矩阵列向量长度和两两角度信息的集合体，协方差矩阵则是标准差向量长度和两两角度的结合体。这种类比有助于我们理解线性代数工具在多元统计领域的应用。

推荐大家阅读 MIT 数学教授 Gilbert Strang 的 *Linear Algebra and Learning from Data*。这本书可谓线性代数工具的弹药库，从知识体系上给了本书作者很多启发。图书目前没有免费电子版图书，该书的专属网站提供样张和勘误等资源：

https://math.mit.edu/~gs/learningfromdata/