23 Data Space 数据空间

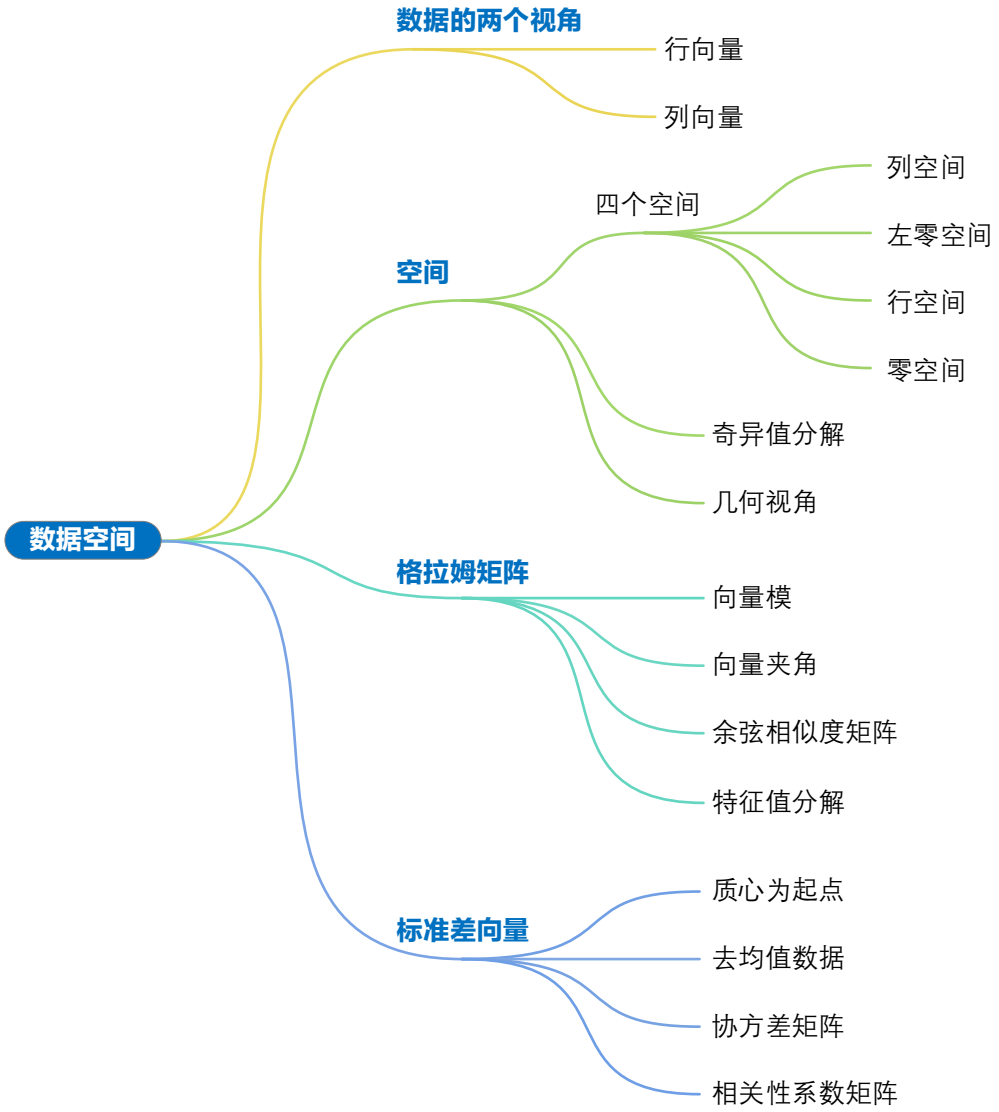
从向量、几何、空间视角看数据

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，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。

数据矩阵 (data matrix) 不过就是以表格形式整理的数据。

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，代表某个特征上的所有样本数据。

行向量

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：

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

其中，第 i 行行向量所有元素为：

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

图 1 所示为从行向量角度观察数据矩阵，每一个行向量 $\boldsymbol{x}^{(i)}$ 代表坐标系上一个点。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。这是本书后文要探讨的内容。

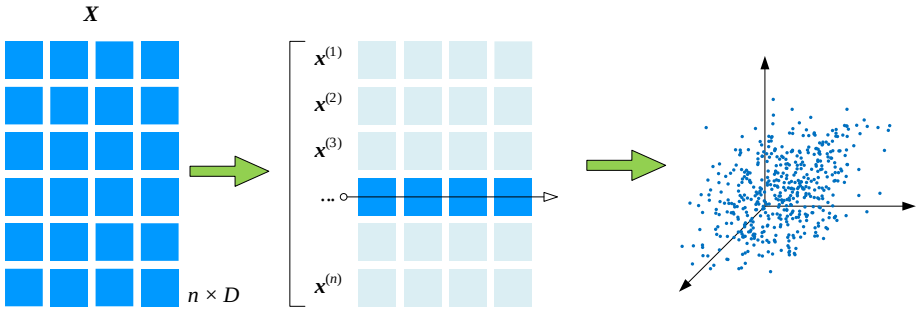


图 1. 从行向量角度观察数据矩阵

列向量

数据矩阵的列向量序号采用下标记法，比如：

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

其中，

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

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 $\boldsymbol{0}$。[$\boldsymbol{x}_1$, $\boldsymbol{x}_1$, … $\boldsymbol{x}_D$] 这些向量的方向和长度 (模、L^2 范数) 均包含在格拉姆矩阵 $\boldsymbol{G} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 之中。向量方向的表现形式是两两向量之间的夹角。更具体地说，是向量夹角余弦值。

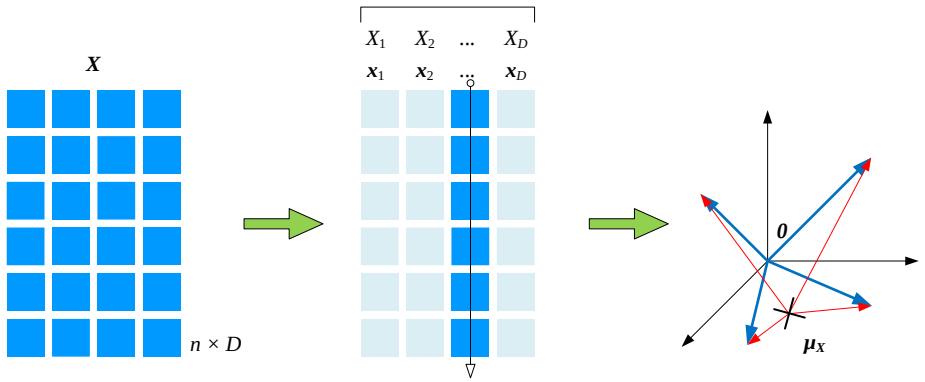


图 2. 从列向量角度观察数据矩阵

如果将图 2 向量起点放在数据质心 $\boldsymbol{\mu_X}$ (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，而向量之间夹角为随机变量之间的线性相关系数。与之对应的方阵为协方差矩阵 $\boldsymbol{\Sigma}$。

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。本章后文还将深入介绍这一重要视角。

区分相似符号

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 的采样得到的样本数据，也可以代表 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 $\boldsymbol{x}_1$ 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。

而从代数角度，斜体小写 x_1 代表变量，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 x_1 的一个取值，或代表随机变量 X_1 的一个取值，比如 $X_1 = \left\{x_1^{(1)}, x_1^{(2)}, ..., x_1^{(n)}\right\}$。

$\boldsymbol{X}$ 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 $\underline{x_{i,j}}$。

我们还会用粗体斜体小写希腊字母 $\boldsymbol{\chi}$ 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 $\boldsymbol{\chi}$ 主要用在多元统计计算式中。

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。

四个空间

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 $\mathrm{span}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D)$ 为 $\boldsymbol{X}$ 的列空间 (column space)，记做 $C(\boldsymbol{X})$。

与 $C(\boldsymbol{X})$ 相对应的是左零空间 (left null space) $\mathrm{Null}(\boldsymbol{X}^{\mathrm{T}})$。$C(\boldsymbol{X})$ 和 $\mathrm{Null}(\boldsymbol{X}^{\mathrm{T}})$ 构成了 $\mathbb{R}^n$。而 $C(\boldsymbol{X})$ 和 $\mathrm{Null}(\boldsymbol{X}^{\mathrm{T}})$ 分别都是 $\mathbb{R}^n$ 的子空间。

由 $\boldsymbol{X}$ 的行向量 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(j)}, \ldots, \boldsymbol{x}^{(D)}$ 张成的子空间 $\mathrm{span}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(D)})$ 为 $\boldsymbol{X}$ 的行空间 (row space)，记做 $R(\boldsymbol{X})$。

与 $R(\boldsymbol{X})$ 相对应的是零空间 (null space 或 right null space)，记做 $\mathrm{Null}(\boldsymbol{X})$。$R(\boldsymbol{X})$ 和 $\mathrm{Null}(\boldsymbol{X})$ 构成了 $\mathbb{R}^D$。而 $R(\boldsymbol{X})$ 和 $\mathrm{Null}(\boldsymbol{X})$ 分别都是 $\mathbb{R}^D$ 的子空间。$R(\boldsymbol{X})$的维度为 $\dim(R(\boldsymbol{X})) = \mathrm{rank}(\boldsymbol{X})$。

相信大家读完以上两段话已经晕头转向，云里雾里不知所云。

的确，这四个空间的定义恐怕让很多人望而却步。很多线性代数教材多是从线性方程角度讲解这四个空间，而线性方程角度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：

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}} \tag{5}$$

图 3 所示为 $\boldsymbol{X}$ 完全型 SVD 分解示意图，请大家注意矩阵形状。

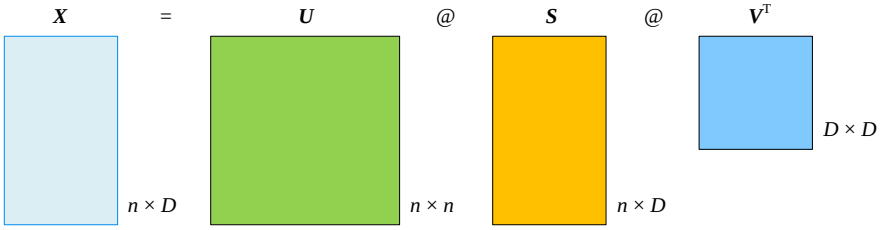


图 3. $\boldsymbol{X}$ 进行完全型 SVD 分解

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，且向量模均为 1。$[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 为张成 $\mathbb{R}^n$ 空间规范正交基。

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

如图 4 所示，$\mathbb{R}^n$ 空间和 $\mathbb{R}^D$ 空间之间的联系为 $\boldsymbol{X}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S}$。

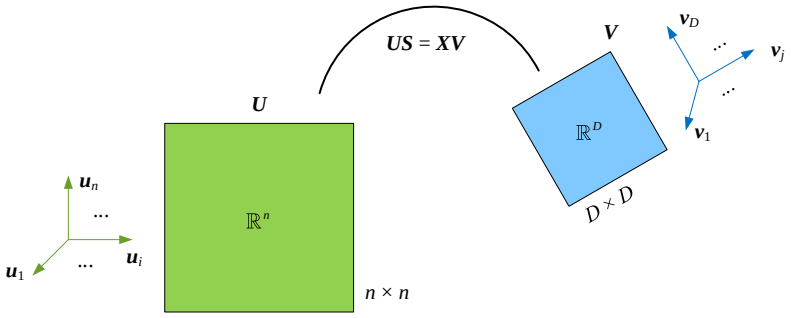


图 4. 对于矩阵 $\boldsymbol{X}$ 来说，$\mathbb{R}^n$ 空间和 $\mathbb{R}^D$ 空间关系

另外，对“粗短”$\boldsymbol{X}^{\mathrm{T}}$矩阵进行完全型 SVD 分解，就是对 (5) 转置。

$$\boldsymbol{X}^{\mathrm{T}} = \left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\right)^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} \tag{6}$$

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 $\boldsymbol{S}^{\mathrm{T}} \neq \boldsymbol{S}$。

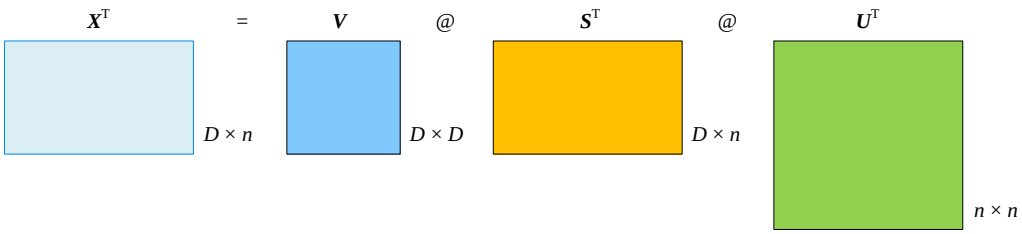


图 5. $\boldsymbol{X}^{\mathrm{T}}$进行完全型 SVD 分解

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：

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{r\times r} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} \tag{7}$$

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，本书前文介绍过图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 $[\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n]$。

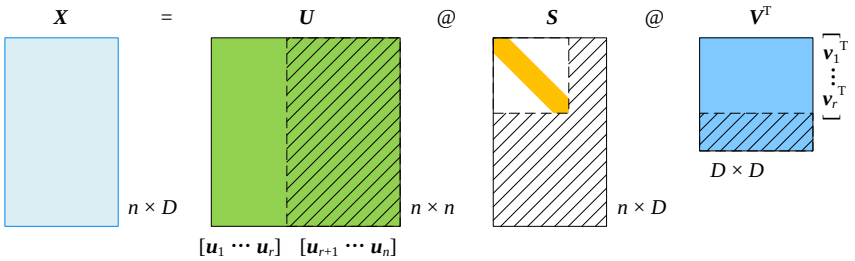


图 6. $\boldsymbol{X}$ 进行紧凑型 SVD 分解

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。

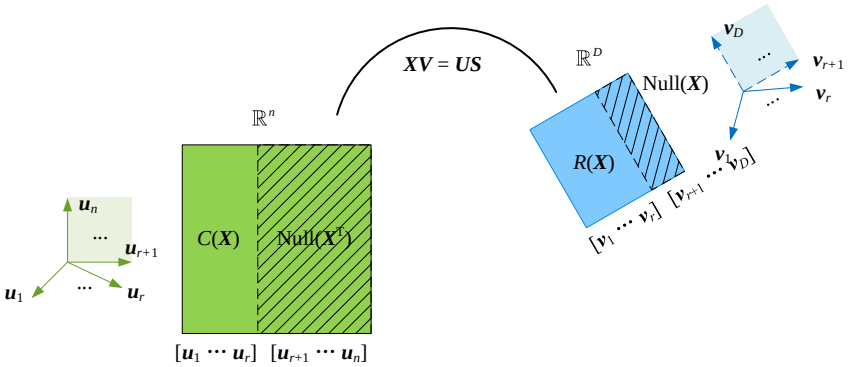


图 7. $\mathbb{R}^n$ 空间和 $\mathbb{R}^D$ 空间关系，考虑列空间、行空间和零空间

列空间，左零空间

$[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r]$ 是代表 $\boldsymbol{X}$ 的列空间 $C(\boldsymbol{X})$ 基底。而 $[\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n]$ 是左零空间 Null($\boldsymbol{X}^T$) 基底。

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

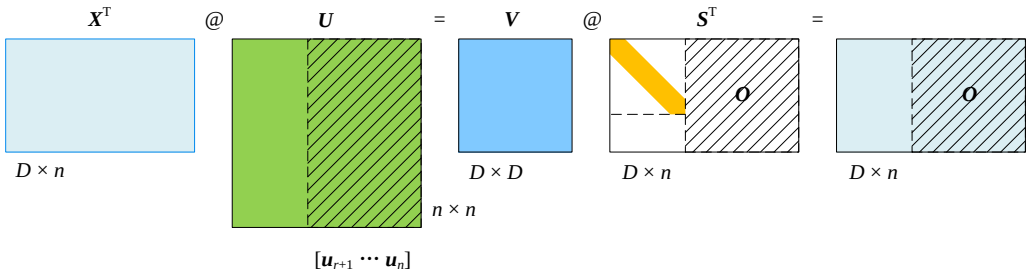


图 8. $\boldsymbol{X}^T$ 向 Null($\boldsymbol{X}^T$) $[\boldsymbol{u}_{r+1}, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 投影的结果为 $\boldsymbol{O}$

这就是为什么 Null($\boldsymbol{X}^T$) 被称作左“零”空间的原因。而且，我们也同时在图 8 中 $\boldsymbol{X}^T$ 上看到了“转置”，这就解释了为什么列空间 $C(\boldsymbol{X})$ 对应 Null($\boldsymbol{X}^T$)。

多说一句，(6) 可以写成：

$$\boldsymbol{X}^T\boldsymbol{U} = \boldsymbol{V}\boldsymbol{S}^T \tag{8}$$

矩阵“$\boldsymbol{X}^T$”对应的投影矩阵是 $\boldsymbol{U}$，$\boldsymbol{X}^T$ 的每一行是 $\boldsymbol{X}$ 的“列”向量。大家在这句话中看到列空间 $C(\boldsymbol{X})$ 和 Null($\boldsymbol{X}^T$) 中“列”和“$\boldsymbol{X}^T$”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。

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($\boldsymbol{X}$) $[\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_D]$ 投影的结果为全 0 矩阵 $\boldsymbol{O}$。

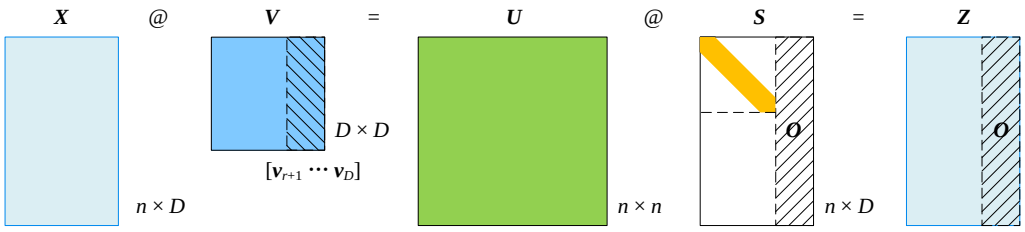


图 9. $\boldsymbol{X}$ 向 Null($\boldsymbol{X}$) [$\boldsymbol{v}_{r+1}$, $\boldsymbol{v}_2$, …, $\boldsymbol{v}_D$] 投影的结果为 $\boldsymbol{O}$

同理，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($\boldsymbol{X}$)。

正交关系

列空间 $C(\boldsymbol{X})$ 和 左零空间 Null($\boldsymbol{X}^{\mathrm{T}}$) 之间正交，而行空间 $R(\boldsymbol{X})$ 和零空间 Null($\boldsymbol{X}$) 正交。

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。

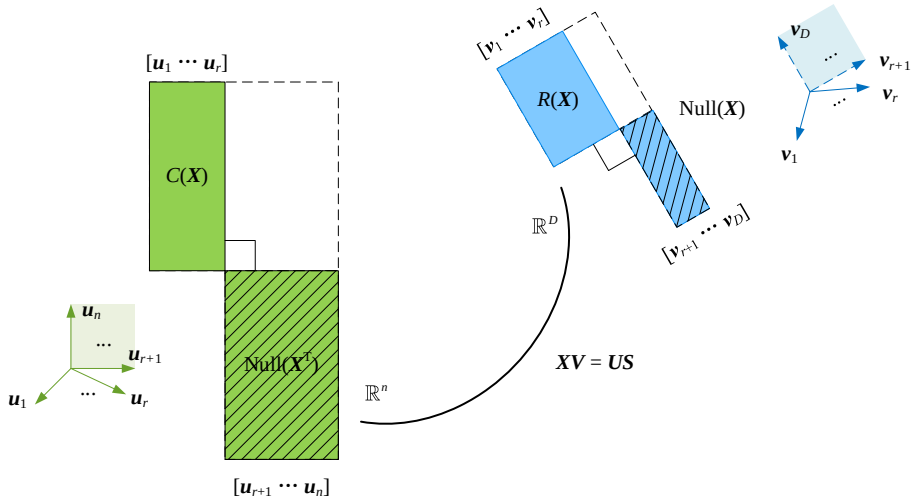


图 10. $\mathbb{R}^n$ 空间和 $\mathbb{R}^D$ 空间关系，考虑列空间、行空间和零空间的正交关系

格外强调，$\mathbb{R}^n$ 空间和 $\mathbb{R}^D$ 空间是“永恒”存在的，能张成这两个空间的规范正交基有无数组。

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

[$\boldsymbol{v}_1$, …, $\boldsymbol{v}_D$]，即 $R(\boldsymbol{X})$ + 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，$C(\boldsymbol{X})$、Null($\boldsymbol{X}^{\mathrm{T}}$)、$R(\boldsymbol{X})$ 和 Null($\boldsymbol{X}$) 是为矩阵 $\boldsymbol{X}$ 而生。

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，即列向量长度相等，因此一定包含列空间。

$\boldsymbol{U}$ 在矩阵乘积左边，因此包含“左”零空间。

$\boldsymbol{V}$ 和 $\boldsymbol{X}$ 等宽，即行向量长度相等，且 $\boldsymbol{XV}$ 中的 $\boldsymbol{V}$ 完成行向量投影，因此 $\boldsymbol{V}$ 包含行空间。

$\boldsymbol{V}$ 在矩阵乘积右边，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，怎么记 $\boldsymbol{XV} = \boldsymbol{US}$?

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聊聊上节介绍的几个空间。

举个例子

给定矩阵 $\boldsymbol{X}$ 如下:

$$\boldsymbol{X} = \begin{bmatrix} 1 & -1 \\ -\sqrt{3} & 3 \\ 2 & -2 \end{bmatrix} \tag{9}$$

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

$$\boldsymbol{x}_1 + \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ -\sqrt{3} \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ \sqrt{3} \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{10}$$

也就是说 $\boldsymbol{X}$ 的 $\text{rank}(\boldsymbol{X}) = r = 1$。

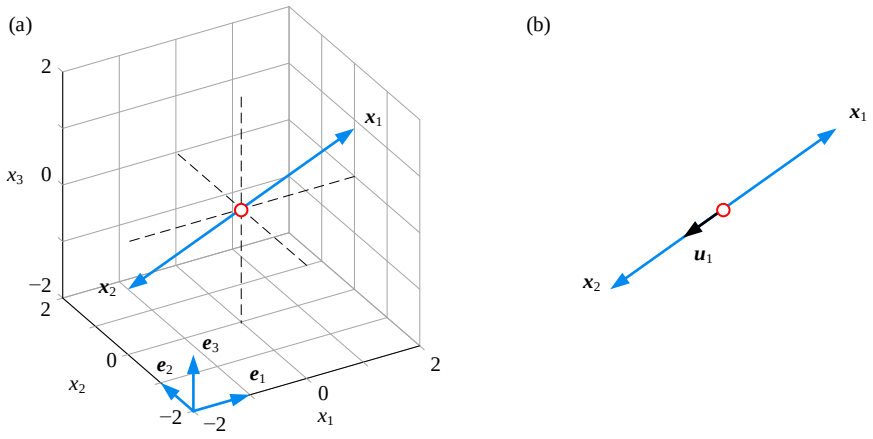


图 11. 从三维空间到一维空间

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。

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：

$$\boldsymbol{x}_1 = a\boldsymbol{u}_1, \quad \boldsymbol{x}_2 = b\boldsymbol{u}_1 \tag{11}$$

也就是一维空间就足够描述 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$，这就是为什么 $\text{rank}(\boldsymbol{X}) = 1$。

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。

$$\boldsymbol{X} = \begin{bmatrix} 1 & -1 \\ -\sqrt{3} & 3 \\ 2 & -2 \end{bmatrix} = \underbrace{\begin{bmatrix} -0.3536 & -0.9297 & 0.1034 \\ 0.6124 & -0.1465 & 0.7769 \\ -0.7071 & 0.3380 & 0.6211 \end{bmatrix}}_{\boldsymbol{U}} \underbrace{\begin{bmatrix} 4 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\boldsymbol{S}} \left(\underbrace{\begin{bmatrix} -0.7071 & -0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}}_{\boldsymbol{V}} \right)^{\mathrm{T}} \tag{12}$$

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 SVD 分解中，我们顺路还得到了 $\boldsymbol{u}_2$ 和 $\boldsymbol{u}_3$，它俩张起了左零空间 $\text{Null}(\boldsymbol{X}^{\mathrm{T}})$。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。

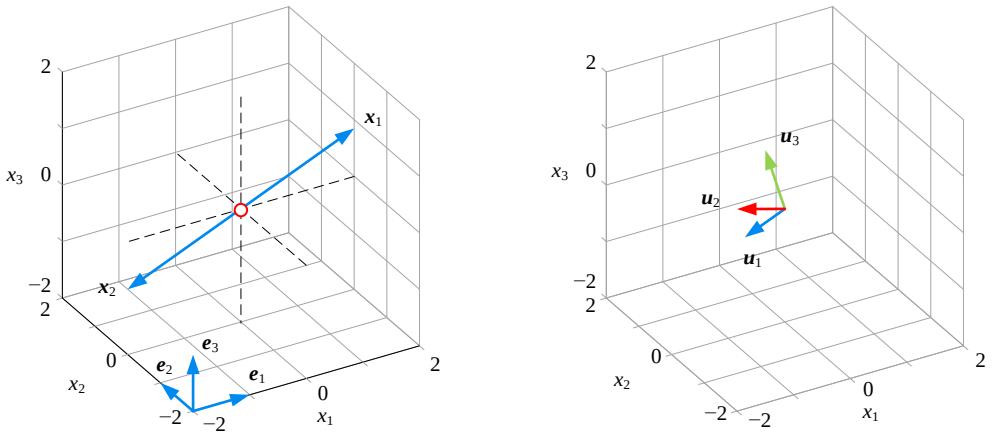


图 12. 矩阵 $\boldsymbol{X}$ 的列空间 $C(\boldsymbol{X})$ 和左零空间 $\text{Null}(\boldsymbol{X}^{\mathrm{T}})$

投影

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

$$\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{U}=\begin{bmatrix}1 & -\sqrt{3} & 2\end{bmatrix}\left[\underbrace{\begin{bmatrix}-0.3536\\0.6124\\-0.7071\end{bmatrix}}_{\boldsymbol{u}_1}\underbrace{\begin{bmatrix}-0.9297\\-0.1465\\0.3380\end{bmatrix}}_{\boldsymbol{u}_2}\underbrace{\begin{bmatrix}0.1034\\0.7769\\0.6211\end{bmatrix}}_{\boldsymbol{u}_3}\right]=\begin{bmatrix}-2.8284 & 0 & 0\end{bmatrix} \tag{13}$$

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 $\boldsymbol{u}_3$ 上投影均为 0，这就是为什么它俩构成 $\text{Null}(\boldsymbol{X}^{\mathrm{T}})$。(13) 中的 $\boldsymbol{x}_1$ 转置运算也解释了为什么，$\text{Null}(\boldsymbol{X}^{\mathrm{T}})$ 括号里面为 $\boldsymbol{X}^{\mathrm{T}}$。

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：

$$\boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{U}=\begin{bmatrix}-1 & \sqrt{3} & -2\end{bmatrix}\left[\underbrace{\begin{bmatrix}-0.3536\\0.6124\\-0.7071\end{bmatrix}}_{\boldsymbol{u}_1}\underbrace{\begin{bmatrix}-0.9297\\-0.1465\\0.3380\end{bmatrix}}_{\boldsymbol{u}_2}\underbrace{\begin{bmatrix}0.1034\\0.7769\\0.6211\end{bmatrix}}_{\boldsymbol{u}_3}\right]=\begin{bmatrix}2.8284 & 0 & 0\end{bmatrix} \tag{14}$$

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 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$：

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = \begin{bmatrix} 1 & -1 \\ -\sqrt{3} & 3 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -\sqrt{3} & 3 \\ 2 & -2 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 2 & -3.4641 & 4 \\ -3.4641 & 6 & -6.9282 \\ 4 & -6.9282 & 8 \end{bmatrix} \tag{15}$$

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

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

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

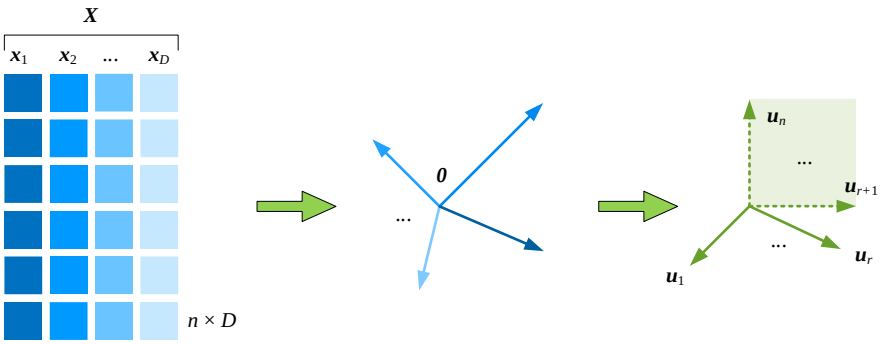


图 13. 矩阵 $\boldsymbol{X}$ 的列空间 $C(\boldsymbol{X})$ 和左零空间 Null($\boldsymbol{X}^{\mathrm{T}}$)

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：

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} 1 & -1 \end{bmatrix}, \quad \boldsymbol{x}^{(2)} = \begin{bmatrix} -\sqrt{3} & 3 \end{bmatrix}, \quad \boldsymbol{x}^{(3)} = \begin{bmatrix} 2 & -2 \end{bmatrix} \tag{17}$$

如图 14 (a) 所示，可视化 $\boldsymbol{x}^{(1)}$、$\boldsymbol{x}^{(2)}$ 和 $\boldsymbol{x}^{(3)}$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：

$$\boldsymbol{x}^{(1)} = a\boldsymbol{v}_1^{\mathrm{T}}, \quad \boldsymbol{x}^{(2)} = b\boldsymbol{v}_1^{\mathrm{T}}, \quad \boldsymbol{x}^{(3)} = c\boldsymbol{v}_1^{\mathrm{T}} \tag{18}$$

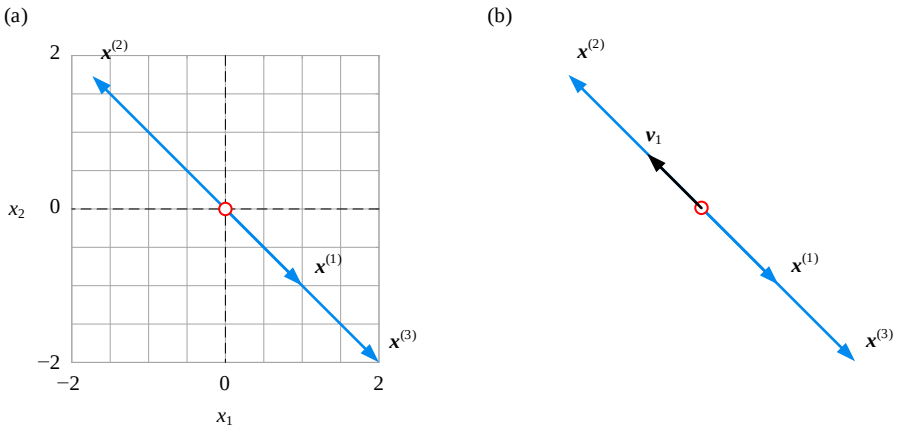


图 14. 从二维空间到一维空间

(12) 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 $R(\boldsymbol{X})$，$\boldsymbol{v}_2$ 张成零空间 Null($\boldsymbol{X}$)。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。

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，一般情况都是不重合的，本例中两者分别重合的原因是 rank($\boldsymbol{X}$) = 1。

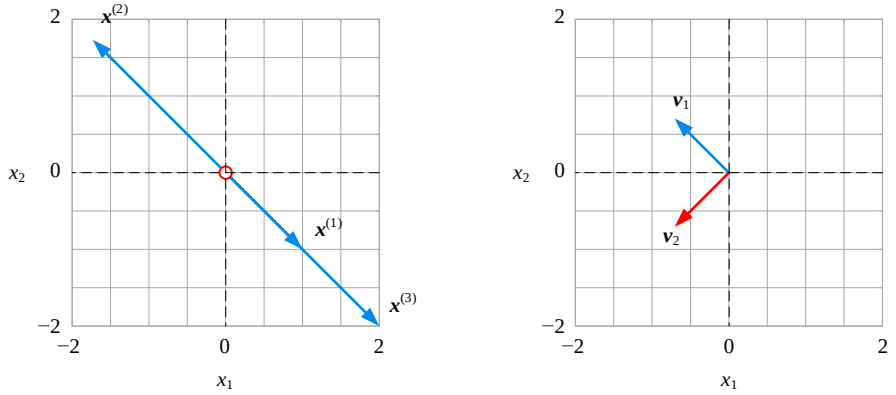


图 15. 矩阵 $\boldsymbol{X}$ 的行空间 $R(\boldsymbol{X})$ 和零空间 Null($\boldsymbol{X}$)

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

$$\boldsymbol{x}^{(1)}\boldsymbol{V} = \begin{bmatrix} 1 & -1 \end{bmatrix} \left[\underbrace{\begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix}}_{\boldsymbol{v}_1} \underbrace{\begin{bmatrix} -0.7071 \\ -0.7071 \end{bmatrix}}_{\boldsymbol{v}_2} \right] = \begin{bmatrix} -1.4142 & 0 \end{bmatrix} \tag{19}$$

也就是说，$\boldsymbol{x}^{(1)}$ 在 $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$这个直角坐标系中的坐标为 $(-1.4142, 0)$。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 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$:

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \begin{bmatrix} 1 & -1 \\ -\sqrt{3} & 3 \\ 2 & -2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 1 & -1 \\ -\sqrt{3} & 3 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 8 & -8 \\ -8 & 8 \end{bmatrix} \tag{20}$$

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

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \begin{bmatrix} 8 & -8 \\ -8 & 8 \end{bmatrix} = \underbrace{\begin{bmatrix} -0.7071 & -0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}}_{\boldsymbol{V}} \begin{bmatrix} 16 & 0 \\ 0 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} -0.7071 & 0.7071 \\ -0.7071 & -0.7071 \end{bmatrix}}_{\boldsymbol{V}^{\mathrm{T}}} \tag{21}$$

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。

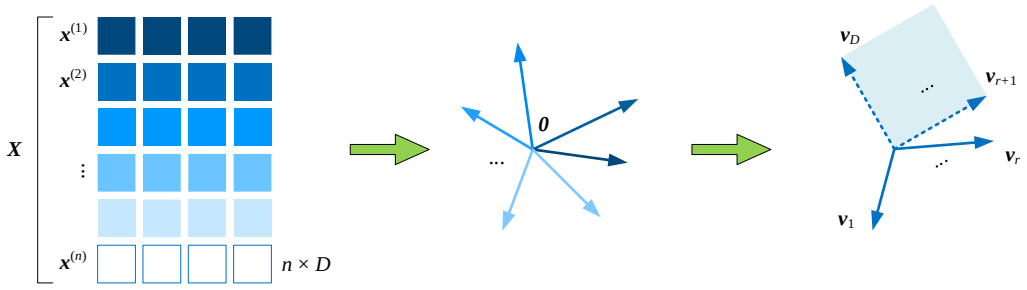


图 16. 矩阵 $\boldsymbol{X}$ 的行空间 $R(\boldsymbol{X})$ 和零空间 Null($\boldsymbol{X}$)

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。对于一个向量而言，最重要的元素有两个——长度和方向。

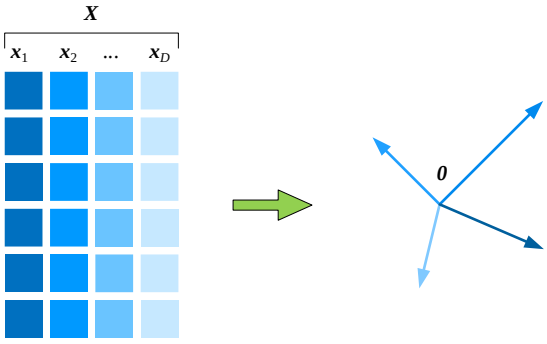


图 17. 矩阵 $\boldsymbol{X}$ 列向量几何化为空间向量

向量的长度不难确定，向量的模 (L^2 范数) 就是向量的长度。

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。

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，极坐标系中需要量化向量和坐标系极轴的夹角，即绝对角度值。而一般情况，我们分析的是向量两两夹角，即相对角度值。

换句话说，格拉姆矩阵中包含向量两两之间的相对夹角，而非绝对角度。

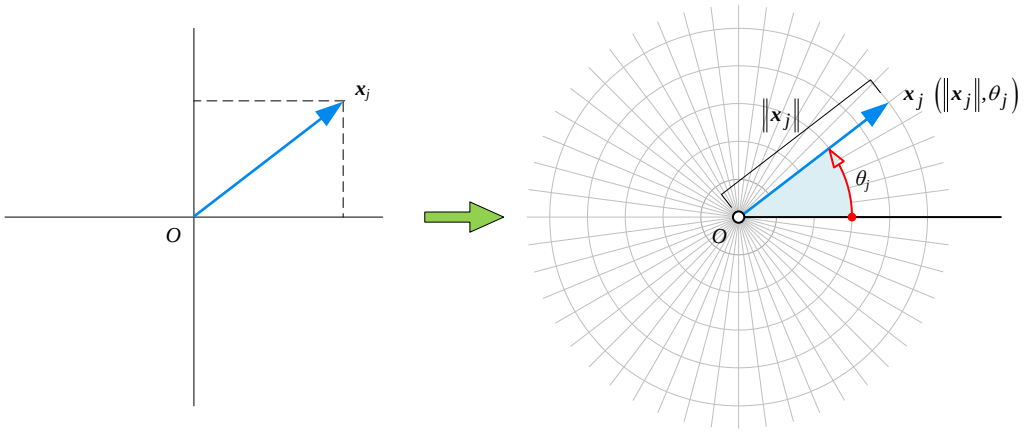


图 18. 从平面直角坐标系到极坐标系

长度，相对夹角

任意两个向量的夹角很容易算，我们可以通过向量向量内积计算夹角的余弦值。

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) 个余弦值。该怎么有序保存这些结果？

实际上，我们反复提到的格拉姆矩阵就是解决方案。

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：

$$\boldsymbol{G} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \tag{22}$$

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 $\boldsymbol{G}$：

$$\boldsymbol{G} = \begin{bmatrix} \langle \boldsymbol{x}_1, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_1, \boldsymbol{x}_D \rangle \\ \langle \boldsymbol{x}_2, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_2, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_2, \boldsymbol{x}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{x}_D, \boldsymbol{x}_1 \rangle & \langle \boldsymbol{x}_D, \boldsymbol{x}_2 \rangle & \cdots & \langle \boldsymbol{x}_D, \boldsymbol{x}_D \rangle \end{bmatrix} = \begin{bmatrix} \|\boldsymbol{x}_1\|\|\boldsymbol{x}_1\|\cos\theta_{1,1} & \|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\|\cos\theta_{2,1} & \cdots & \|\boldsymbol{x}_1\|\|\boldsymbol{x}_D\|\cos\theta_{1,D} \\ \|\boldsymbol{x}_2\|\|\boldsymbol{x}_1\|\cos\theta_{1,2} & \|\boldsymbol{x}_2\|\|\boldsymbol{x}_2\|\cos\theta_{2,2} & \cdots & \|\boldsymbol{x}_2\|\|\boldsymbol{x}_D\|\cos\theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \|\boldsymbol{x}_D\|\|\boldsymbol{x}_1\|\cos\theta_{1,D} & \|\boldsymbol{x}_D\|\|\boldsymbol{x}_2\|\cos\theta_{2,D} & \cdots & \|\boldsymbol{x}_D\|\|\boldsymbol{x}_D\|\cos\theta_{D,D} \end{bmatrix} \tag{23}$$

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，和列向量两两夹角余弦值。

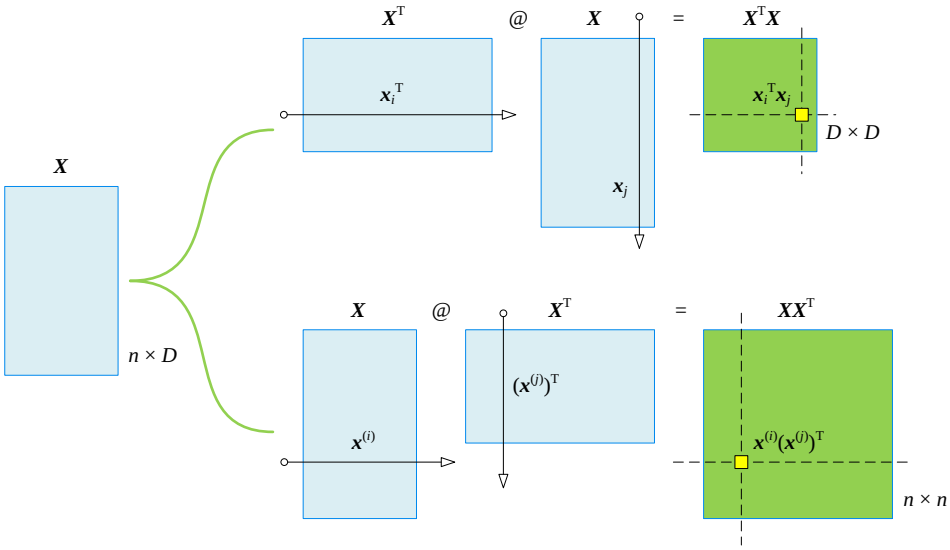


图 19. 两个格拉姆矩阵

而余弦相似度矩阵 $\boldsymbol{C}$ 则更进一步，只关注列向量夹角余弦值：

$$\boldsymbol{C}=\begin{bmatrix} \frac{\boldsymbol{x}_1\cdot\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|\|\boldsymbol{x}_1\|} & \frac{\boldsymbol{x}_1\cdot\boldsymbol{x}_2}{\|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\|} & \cdots & \frac{\boldsymbol{x}_1\cdot\boldsymbol{x}_D}{\|\boldsymbol{x}_1\|\|\boldsymbol{x}_D\|} \\ \frac{\boldsymbol{x}_2\cdot\boldsymbol{x}_1}{\|\boldsymbol{x}_2\|\|\boldsymbol{x}_1\|} & \frac{\boldsymbol{x}_2\cdot\boldsymbol{x}_2}{\|\boldsymbol{x}_2\|\|\boldsymbol{x}_2\|} & \cdots & \frac{\boldsymbol{x}_2\cdot\boldsymbol{x}_D}{\|\boldsymbol{x}_2\|\|\boldsymbol{x}_D\|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\boldsymbol{x}_D\cdot\boldsymbol{x}_1}{\|\boldsymbol{x}_D\|\|\boldsymbol{x}_1\|} & \frac{\boldsymbol{x}_D\cdot\boldsymbol{x}_2}{\|\boldsymbol{x}_D\|\|\boldsymbol{x}_2\|} & \cdots & \frac{\boldsymbol{x}_D\cdot\boldsymbol{x}_D}{\|\boldsymbol{x}_D\|\|\boldsymbol{x}_D\|} \end{bmatrix}=\begin{bmatrix} 1 & \cos\theta_{2,1} & \cdots & \cos\theta_{1,D} \\ \cos\theta_{1,2} & 1 & \cdots & \cos\theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \cos\theta_{1,D} & \cos\theta_{2,D} & \cdots & 1 \end{bmatrix} \tag{24}$$

两个格拉姆矩阵

本章前文提到了两个不同的格拉姆矩阵——$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 和 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$，有必要对两者性质进行对比介绍。

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

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

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 $\boldsymbol{H}$：

$$\boldsymbol{H}=\begin{bmatrix} \left\langle \boldsymbol{x}^{(1)},\boldsymbol{x}^{(1)}\right\rangle & \left\langle \boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\right\rangle & \cdots & \left\langle \boldsymbol{x}^{(1)},\boldsymbol{x}^{(n)}\right\rangle \\ \left\langle \boldsymbol{x}^{(2)},\boldsymbol{x}^{(1)}\right\rangle & \left\langle \boldsymbol{x}^{(2)},\boldsymbol{x}^{(2)}\right\rangle & \cdots & \left\langle \boldsymbol{x}^{(2)},\boldsymbol{x}^{(n)}\right\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \left\langle \boldsymbol{x}^{(n)},\boldsymbol{x}^{(1)}\right\rangle & \left\langle \boldsymbol{x}^{(n)},\boldsymbol{x}^{(2)}\right\rangle & \cdots & \left\langle \boldsymbol{x}^{(n)},\boldsymbol{x}^{(n)}\right\rangle \end{bmatrix} \tag{26}$$

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

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

$$\boldsymbol{G}=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} \tag{27}$$

假设 $\lambda_{\boldsymbol{G}}$ 为 $\boldsymbol{G}$ 的一个特征值：

$$\boldsymbol{G}\boldsymbol{v}=\lambda_{\boldsymbol{G}}\boldsymbol{v} \tag{28}$$

即，

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}=\lambda_{\boldsymbol{G}}\boldsymbol{v} \tag{29}$$

然后对 $\boldsymbol{H}$ 特征值分解：

$$\boldsymbol{H}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{T}} \tag{30}$$

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

假设 λ_H 为 $\boldsymbol{H}$ 的一个特征值：

$$\boldsymbol{H}\boldsymbol{u} = \lambda_H \boldsymbol{u} \tag{31}$$

即，

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{u} = \lambda_H \boldsymbol{u} \tag{32}$$

(29) 左右乘以 $\boldsymbol{X}$，得到：

$$\underset{u}{\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}}\ \underset{u}{\boldsymbol{X}\boldsymbol{v}} = \lambda_G\ \underset{u}{\boldsymbol{X}\boldsymbol{v}} \tag{33}$$

比较 (32) 和 (33)，可以发现 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 和 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 具有对应关系。

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，因此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 $\boldsymbol{X}$，$\boldsymbol{X}$ 的协方差矩阵 $\boldsymbol{\Sigma}$ 可以通过下式计算得到。

$$\boldsymbol{\Sigma} = \frac{\underbrace{\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)}_{\text{Centered}}^{\mathrm{T}} \underbrace{\left(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\right)}_{\text{Centered}}}{n-1} = \frac{\boldsymbol{X}_c^{\mathrm{T}}\boldsymbol{X}_c}{n-1} \tag{34}$$

分母上，$n-1$ 仅仅起到取平均作用。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 $\mathrm{E}(\boldsymbol{X})$：

$$\boldsymbol{X}_c = \boldsymbol{X} - \mathrm{E}(\boldsymbol{X}) \tag{35}$$

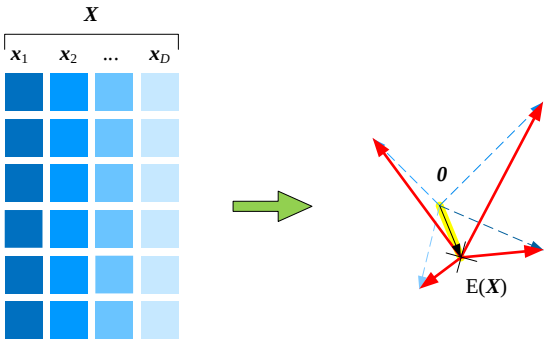


图 20. 数据质心为 $\boldsymbol{X}_c$ 列向量的起点

将 $\boldsymbol{X}_c$ 列向量的起点也移动到 $\boldsymbol{0}$，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 $\boldsymbol{X} = \boldsymbol{X}_c$。

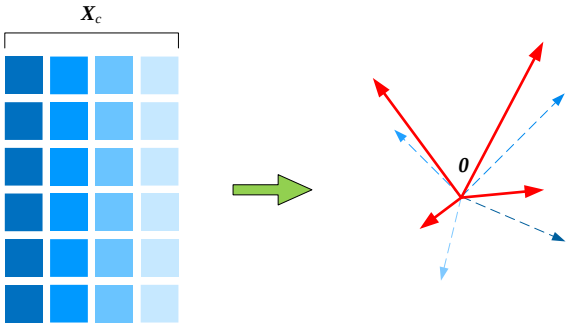


图 21. 比较 $\boldsymbol{X}$ 和 $\boldsymbol{X}_c$ 列向量

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 z 分数数据矩阵 $\boldsymbol{Z}$。

根据本章前文定义的四个空间，中心化数据矩阵 $\boldsymbol{X}_c$ 也应该有自己的四个空间！那么大家立刻会想到，标准化后的 z 分数矩阵 $\boldsymbol{Z}$，不也会有自己的四个空间吗？

答案是肯定的！

也就是说，如果用 SVD 分解 $\boldsymbol{X}$、$\boldsymbol{X}_c$、$\boldsymbol{Z}$ 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构造的空间特点和区别。

标准差向量

整理 (34) 得到 $\boldsymbol{X}_c^{\mathrm{T}}\boldsymbol{X}_c$：

$$\boldsymbol{X}_c^{\mathrm{T}}\boldsymbol{X}_c = (n-1)\boldsymbol{\Sigma} = (n-1)\begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,D}\sigma_1\sigma_D \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2,D}\sigma_2\sigma_D \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,D}\sigma_1\sigma_D & \rho_{2,D}\sigma_2\sigma_D & \cdots & \sigma_D^2 \end{bmatrix} \tag{36}$$

对比 (23) 和 (36)，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”。

向量 $\boldsymbol{\sigma}_j$ 之间的夹角的余弦值便是相关性系数。这样 (36) 可以写成：

$$\boldsymbol{\Sigma} = \begin{bmatrix} \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1 \rangle & \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \rangle & \cdots & \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_D \rangle \\ \langle \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_1 \rangle & \langle \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_2 \rangle & \cdots & \langle \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{\sigma}_D, \boldsymbol{\sigma}_1 \rangle & \langle \boldsymbol{\sigma}_D, \boldsymbol{\sigma}_2 \rangle & \cdots & \langle \boldsymbol{\sigma}_D, \boldsymbol{\sigma}_D \rangle \end{bmatrix} = \begin{bmatrix} \|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_1\|\cos\phi_{1,1} & \|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_2\|\cos\phi_{2,1} & \cdots & \|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_D\|\cos\phi_{1,D} \\ \|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_1\|\cos\phi_{1,2} & \|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_2\|\cos\phi_{2,2} & \cdots & \|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_D\|\cos\phi_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_1\|\cos\phi_{1,D} & \|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_2\|\cos\phi_{2,D} & \cdots & \|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_D\|\cos\phi_{D,D} \end{bmatrix} \tag{37}$$

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，图中忽略了 $n-1$ 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；取值范围都是 [–1, 1]。越靠近 1，说明越相似，越贴近；越靠近–1，说明越相反，越背离。

唯一不同点，相关性系数量化“标准差向量”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；相关性系数 ρ 标准差向量 $\boldsymbol{\sigma}_j$ 之间的相似。$\boldsymbol{x}_j$ 向量的始点为 $\boldsymbol{0}$，$\boldsymbol{\sigma}_j$ 向量始点为质心。

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是如何？我们下一章回答这个问题。

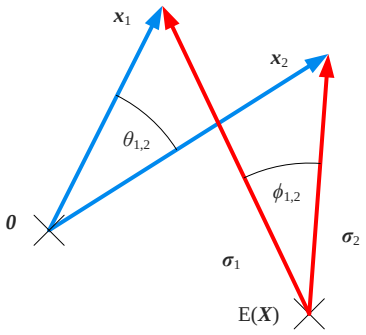


图 22. 余弦相似度和相关性系数的关系

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。

$$\boldsymbol{P}=\begin{bmatrix} \frac{\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_1}{\|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_1\|} & \frac{\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2}{\|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_2\|} & \cdots & \frac{\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_D}{\|\boldsymbol{\sigma}_1\|\|\boldsymbol{\sigma}_D\|} \\ \frac{\boldsymbol{\sigma}_2\cdot\boldsymbol{\sigma}_1}{\|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_1\|} & \frac{\boldsymbol{\sigma}_2\cdot\boldsymbol{\sigma}_2}{\|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_2\|} & \cdots & \frac{\boldsymbol{\sigma}_2\cdot\boldsymbol{\sigma}_D}{\|\boldsymbol{\sigma}_2\|\|\boldsymbol{\sigma}_D\|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\boldsymbol{\sigma}_D\cdot\boldsymbol{\sigma}_1}{\|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_1\|} & \frac{\boldsymbol{\sigma}_D\cdot\boldsymbol{\sigma}_2}{\|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_2\|} & \cdots & \frac{\boldsymbol{\sigma}_D\cdot\boldsymbol{\sigma}_D}{\|\boldsymbol{\sigma}_D\|\|\boldsymbol{\sigma}_D\|} \end{bmatrix}=\begin{bmatrix} 1 & \cos\phi_{2,1} & \cdots & \cos\phi_{1,D} \\ \cos\phi_{1,2} & 1 & \cdots & \cos\phi_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \cos\phi_{1,D} & \cos\phi_{2,D} & \cdots & 1 \end{bmatrix} \tag{38}$$

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。

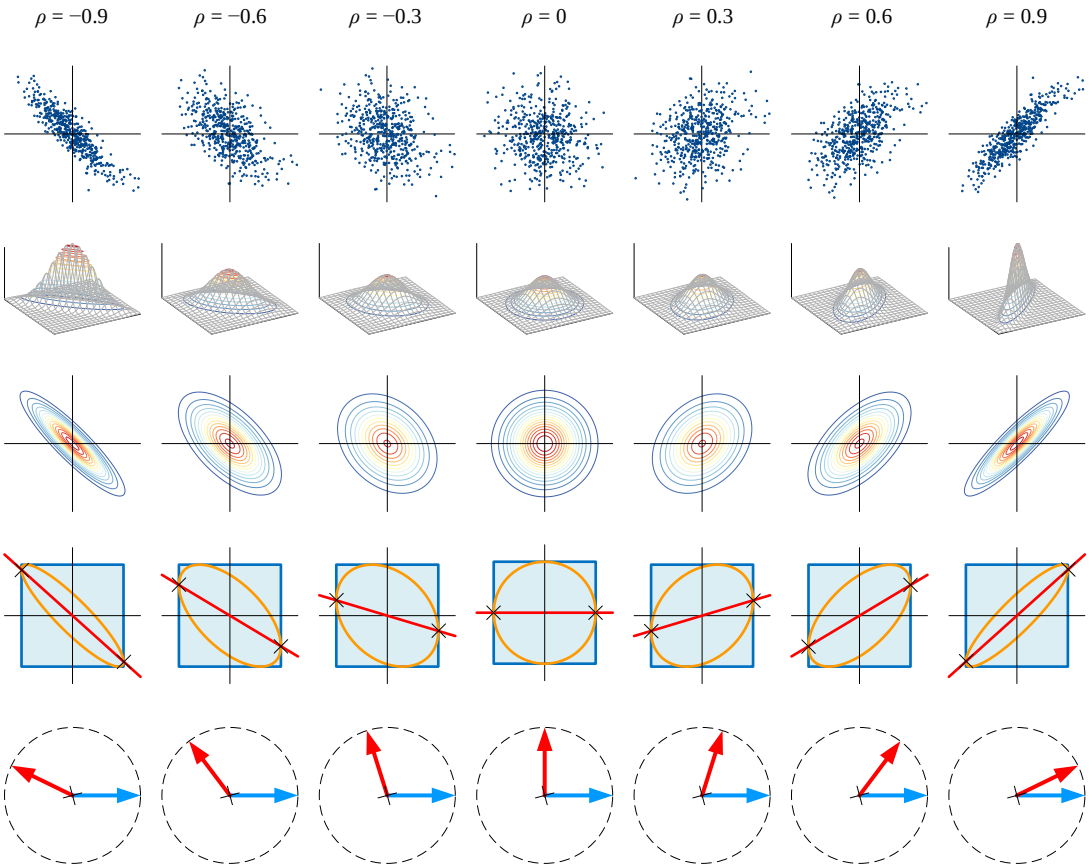
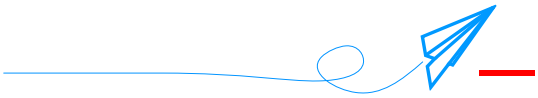


图 23. 相关性系数的几种表达，图中标准差相等，质心位于原点



有数据的地方，就有向量！

有向量的地方，就有几何！

有向量的地方，肯定有空间！

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，首先讲解了四个空间。下图虽然是一幅图，但是其中有四副子图，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。

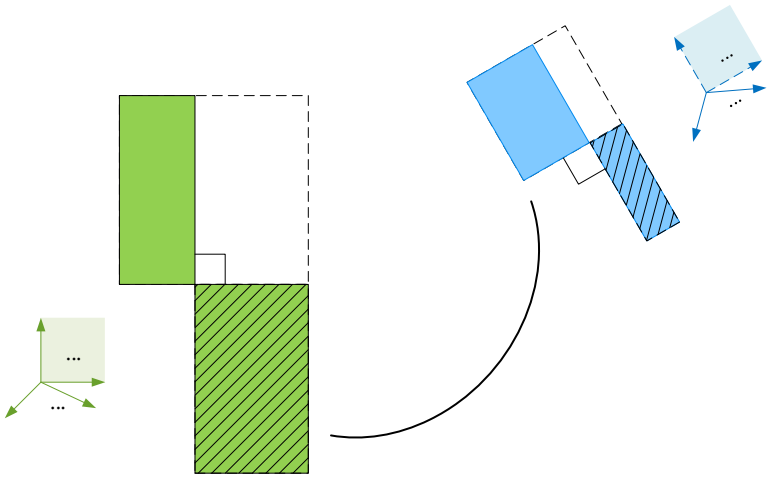


图 24. 总结本章重要内容的四副图

本章特别强化了四个空间和 SVD 分解以及特征值分解之间的关系。

然后，我们又给均值、均方差、协方差、相关性系数这些统计工具赋予了向量、几何、空间层面的意义。