

9

Orthogonal Projection

正交投影

应用几乎无处不在



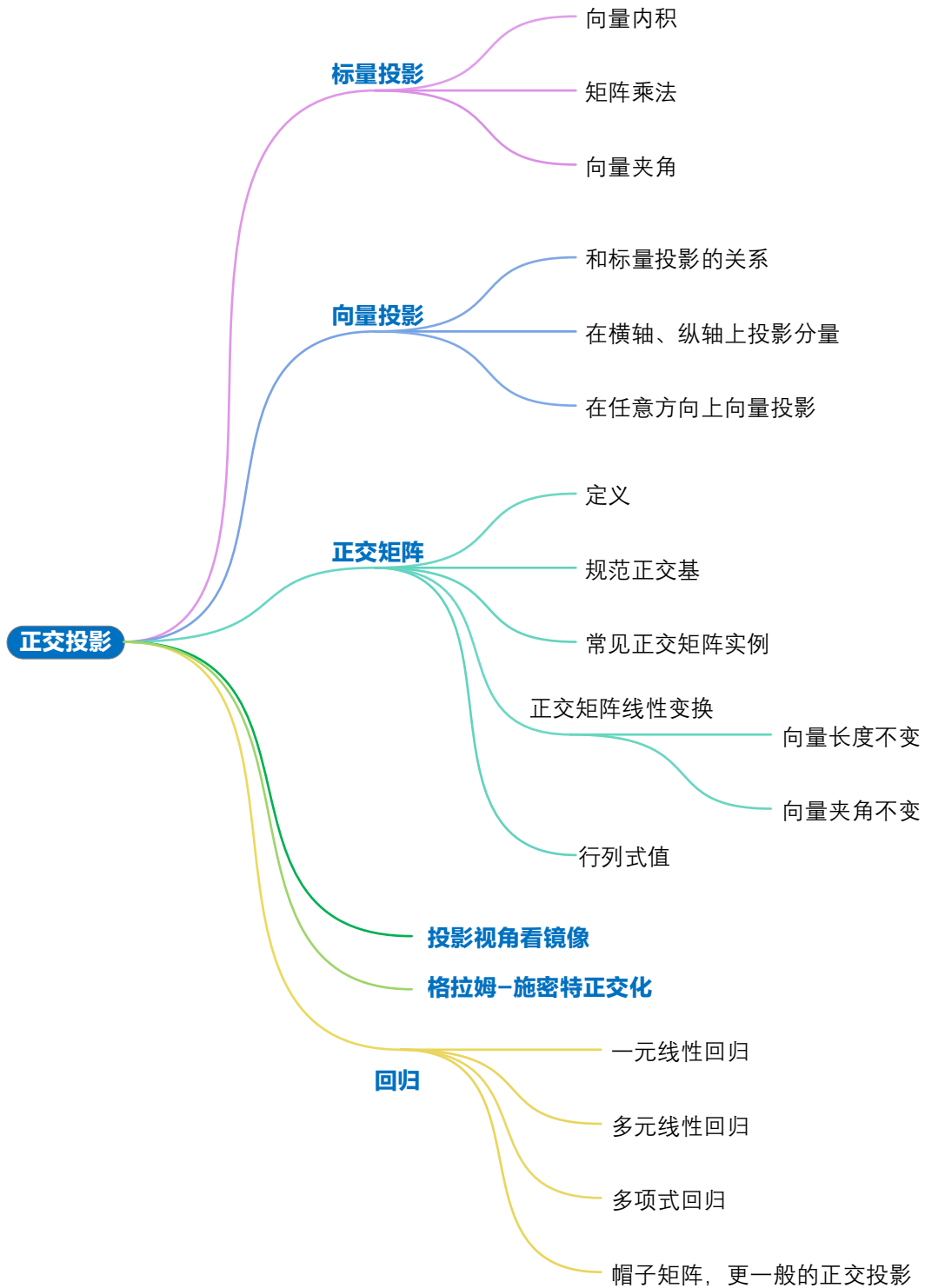
数学好比给了人类第六感。

Mathematics seems to endow one with something like a new sense.

—— 查尔斯·达尔文 (Charles Darwin) | 进化论之父 | 1809 ~ 1882



- ◀ `numpy.random.randn()` 生成满足正态分布的随机数
- ◀ `numpy.linalg.qr()` QR 分解
- ◀ `seaborn.heatmap()` 绘制热图



本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

9.1 标量投影：结果为标量

正交

打个比方，**正交投影** (orthogonal projection) 类似正午头顶阳光将物体投影到地面上，如图 1 所示。此时，假设光线之间相互平行和地面垂直。

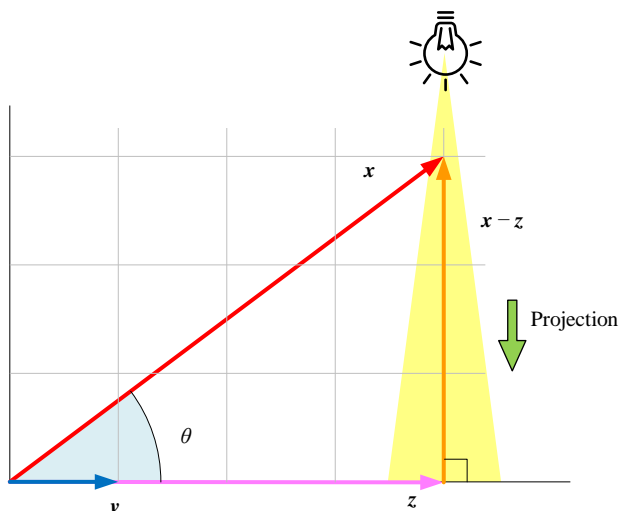


图 1. 正交投影的意义

把列向量 x 看成是一根木杆，而列向量 v 方向代表地面水平方向。 x 在 v 方向上的投影结果为 z 。向量 z 的长度 (向量模) 就是 x 在 v 方向上的**标量投影** (scalar projection)。

令，标量 s 为向量 z 的模。

由于 z 和非零向量 v 共线，因此 z 与 v 的单位向量共线，它们之间的关系为：

$$z = s \frac{v}{\|v\|} \quad (1)$$

很明显，如图 1 所示， $x - z$ 垂直于 v ，因此两者向量内积为 0：

$$(x - z) \cdot v = 0 \quad (2)$$

用矩阵乘法，(2) 可以写成，

$$(x - z)^T v = 0 \quad (3)$$

将 (1) 代入 (3) 得到：

$$\left(\mathbf{x} - s \frac{\mathbf{v}}{\|\mathbf{v}\|} \right)^T \mathbf{v} = 0 \quad (4)$$

(4) 经过整理，得到 s 的解析式，也就是 \mathbf{x} 在 \mathbf{v} 方向上的标量投影为：

$$s = \frac{\mathbf{x}^T \mathbf{v}}{\|\mathbf{v}\|} \quad (5)$$

上式可以写成如下几种形式：

$$s = \frac{\mathbf{x}^T \mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{v}^T \mathbf{x}}{\|\mathbf{v}\|} = \frac{\mathbf{x} \cdot \mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{v} \cdot \mathbf{x}}{\|\mathbf{v}\|} = \frac{\langle \mathbf{x}, \mathbf{v} \rangle}{\|\mathbf{v}\|} \quad (6)$$

⚠ 注意， \mathbf{x} 和 \mathbf{v} 为等行数列向量。

特别地，如果 \mathbf{v} 本身就是单位向量，(6) 可以写作：

$$s = \mathbf{x}^T \mathbf{v} = \mathbf{v}^T \mathbf{x} = \mathbf{x} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{x} = \langle \mathbf{x}, \mathbf{v} \rangle \quad (7)$$

本系列丛书，一般会用 \mathbf{e} 、 \mathbf{v} 、 \mathbf{u} 等代表单位向量。

向量夹角

下面介绍如何从向量夹角入手推导标量投影。

如图 1 所示，向量 \mathbf{x} 和 \mathbf{v} 的相对夹角为 θ ，这个夹角的余弦值 $\cos\theta$ 可以通过下式求解：

$$\cos\theta = \frac{\mathbf{x}^T \mathbf{v}}{\|\mathbf{x}\| \|\mathbf{v}\|} = \frac{\mathbf{v}^T \mathbf{x}}{\|\mathbf{x}\| \|\mathbf{v}\|} = \frac{\mathbf{x} \cdot \mathbf{v}}{\|\mathbf{x}\| \|\mathbf{v}\|} = \frac{\mathbf{v} \cdot \mathbf{x}}{\|\mathbf{x}\| \|\mathbf{v}\|} = \frac{\langle \mathbf{x}, \mathbf{v} \rangle}{\|\mathbf{x}\| \|\mathbf{v}\|} \quad (8)$$

而 \mathbf{x} 在 \mathbf{v} 方向上的标量投影 s 便是向量 \mathbf{x} 的模乘 $\cos\theta$ ：

$$s = \|\mathbf{x}\| \cos\theta = \frac{\mathbf{x}^T \mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{v}^T \mathbf{x}}{\|\mathbf{v}\|} = \frac{\mathbf{x} \cdot \mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{v} \cdot \mathbf{x}}{\|\mathbf{v}\|} = \frac{\langle \mathbf{x}, \mathbf{v} \rangle}{\|\mathbf{v}\|} \quad (9)$$

这样，我们便得到和 (6) 一致的结果。

9.2 向量投影：结果为向量

相对标量投影，我们更经常使用**向量投影** (vector projection)。

顾名思义，向量投影就是标量投影结果再乘上 \mathbf{v} 的方向，即 s 乘以 \mathbf{v} 的单位向量。因此， \mathbf{x} 在 \mathbf{v} 方向上的向量投影实际上就是 (1)，即：

$$\text{proj}_v(\mathbf{x}) = s \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{x} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} \mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{x}}{\mathbf{v} \cdot \mathbf{v}} \mathbf{v} = \frac{\mathbf{x} \cdot \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v} = \frac{\mathbf{x}^T \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \mathbf{v} = \frac{\mathbf{v}^T \mathbf{x}}{\mathbf{v}^T \mathbf{v}} \mathbf{v} \quad (10)$$

用尖括号 $\langle \rangle$ 表达标量积， \mathbf{x} 在 \mathbf{v} 方向上的向量投影可以记做：

$$\text{proj}_v(\mathbf{x}) = \frac{\langle \mathbf{x}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v} \quad (11)$$

特别地，如果 \mathbf{v} 为单位向量， \mathbf{x} 在 \mathbf{v} 方向上的向量投影则可以写成：

$$\text{proj}_v(\mathbf{x}) = \langle \mathbf{x}, \mathbf{v} \rangle \mathbf{v} = (\mathbf{x} \cdot \mathbf{v}) \mathbf{v} = (\mathbf{v} \cdot \mathbf{x}) \mathbf{v} = (\mathbf{x}^T \mathbf{v}) \mathbf{v} = (\mathbf{v}^T \mathbf{x}) \mathbf{v} \quad (12)$$

举个例子

实际上，获得平面上某一个向量的横、纵轴坐标，或者计算横、纵轴的向量分量，也是一个投影过程。

下面看一个实例。给定如下列向量 \mathbf{x} ,

$$\mathbf{x} = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \quad (13)$$

如图 2 所示，列向量 \mathbf{x} 既可以代表平面直角坐标系上的一点，也可以代表一个起点为原点 (0, 0)、终点为 (4, 3) 的向量。

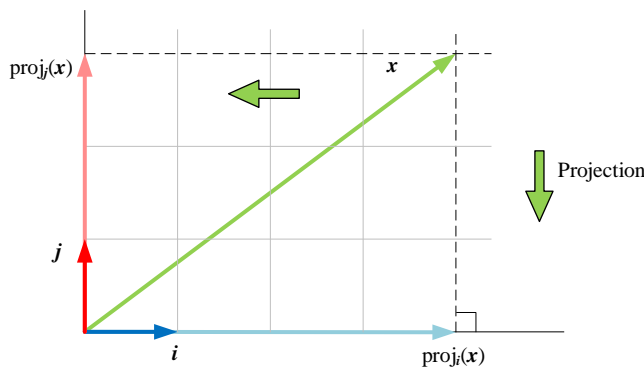


图 2. \mathbf{x} 向 \mathbf{i} 和 \mathbf{j} 投影

\mathbf{x} 向单位向量 $\mathbf{i} = [1, 0]^T$ 方向上投影得到的标量投影为 \mathbf{x} 横轴坐标：

$$\mathbf{i}^T \mathbf{x} = \mathbf{x}^T \mathbf{i} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 4 \quad (14)$$

\mathbf{x} 向单位向量 $\mathbf{j} = [0, 1]^T$ 方向上投影得到的标量投影就是 \mathbf{x} 纵轴坐标：

$$\mathbf{j}^T \mathbf{x} = \mathbf{x}^T \mathbf{j} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 3 \quad (15)$$

\mathbf{x} 在单位向量 $\mathbf{i} = [1, 0]^T$ 方向上向量投影就是 \mathbf{x} 在横轴上的分量：

$$\text{proj}_{\mathbf{i}}(\mathbf{x}) = (\mathbf{x}^T \mathbf{i}) \mathbf{i} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix} \mathbf{i} = 4\mathbf{i} \quad (16)$$

\mathbf{x} 在单位向量 $\mathbf{j} = [0, 1]^T$ 方向上向量投影就是 \mathbf{x} 在纵轴上的分量：

$$\text{proj}_{\mathbf{j}}(\mathbf{x}) = (\mathbf{x}^T \mathbf{j}) \mathbf{j} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathbf{j} = 3\mathbf{j} \quad (17)$$

如果单位向量 \mathbf{v} 为，

$$\mathbf{v} = \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix} \quad (18)$$

\mathbf{x} 在 \mathbf{v} 方向上投影得到的标量投影为：

$$\mathbf{x}^T \mathbf{v} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}^T \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix} = 5 = \|\mathbf{x}\| \quad (19)$$

如图 3 所示，可以发现， \mathbf{x} 和 \mathbf{v} 实际上共线，也就是夹角为 0° 。这显然是个特例。

从向量空间角度来看，向量 \mathbf{v} 张起的空间为 $\text{span}(\mathbf{v})$ ，这个向量空间维度为 1。由于 $\mathbf{x} = 5\mathbf{v}$ ， \mathbf{x} 在 $\text{span}(\mathbf{v})$ 坐标为 5。

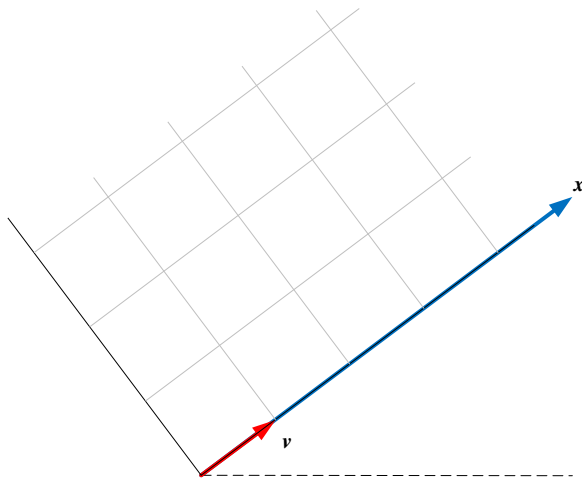


图 3. \mathbf{x} 向 \mathbf{v} 的投影

推导投影坐标

上一章在讲解线性变换时介绍过，点 (x_1, x_2) 在通过原点、切向量为 $\boldsymbol{\tau} [\tau_1, \tau_2]^T$ 直线方向上正交投影得到点的坐标 (z_1, z_2) 为：

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \frac{1}{\|\boldsymbol{\tau}\|^2} \begin{bmatrix} \tau_1^2 & \tau_1\tau_2 \\ \tau_1\tau_2 & \tau_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (20)$$

下面利用本节知识简单推导 (20)。

\mathbf{x} 在 $\boldsymbol{\tau}$ 方向上的向量投影为：

$$\begin{aligned} \mathbf{z} &= \frac{\mathbf{x} \cdot \boldsymbol{\tau}}{\|\boldsymbol{\tau}\|^2} \boldsymbol{\tau} = \frac{x_1\tau_1 + x_2\tau_2}{\|\boldsymbol{\tau}\|^2} \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \\ &= \frac{1}{\|\boldsymbol{\tau}\|^2} \begin{bmatrix} (x_1\tau_1 + x_2\tau_2)\tau_1 \\ (x_1\tau_1 + x_2\tau_2)\tau_2 \end{bmatrix} = \frac{1}{\|\boldsymbol{\tau}\|^2} \begin{bmatrix} \tau_1^2 x_1 + \tau_1\tau_2 x_2 \\ \tau_1\tau_2 x_1 + \tau_2^2 x_2 \end{bmatrix} = \frac{1}{\|\boldsymbol{\tau}\|^2} \begin{bmatrix} \tau_1^2 & \tau_1\tau_2 \\ \tau_1\tau_2 & \tau_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{aligned} \quad (21)$$

⚠ 注意，不做特殊说明的话，本书中“投影”都是正交投影。

图 4 所示为点 A 向一系列通过原点、方向不同直线的投影坐标。

➡ 本书第 7 章强调过，向量空间一定都通过原点。大家可能会问，空间某点朝任意直线或超平面投影时，如果直线或超平面不通过原点，该如何计算投影点的坐标？这个问题将在本书第 19 章揭晓答案。

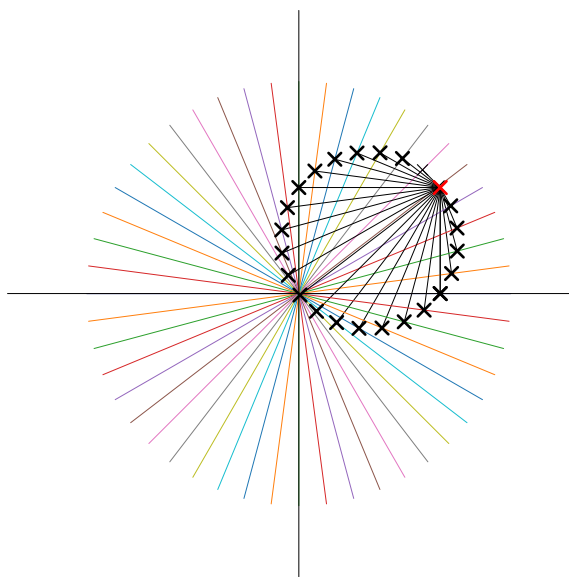


图 4. 点 A 向一系列通过原点的直线投影



Bk4_Ch9_01.py 绘制图 4。

向量张量积：无处不在

回过头再看 (12)，假设 \mathbf{v} 为单位列向量，(12) 可以写成如下含有向量张量积的形式：

$$\text{proj}_{\mathbf{v}}(\mathbf{x}) = \underbrace{(\mathbf{v}^T \mathbf{x})}_{\text{Scalar}} \mathbf{v} = \mathbf{v} \underbrace{(\mathbf{v}^T \mathbf{x})}_{\text{Scalar}} = \mathbf{v} \mathbf{v}^T \mathbf{x} = (\mathbf{v} \otimes \mathbf{v})_{2 \times 2} \mathbf{x} \quad (22)$$

我们称 $\mathbf{v} \otimes \mathbf{v}$ 为**投影矩阵** (projection matrix)。

利用向量张量积，(21) 可以写成：

$$\mathbf{z} = \frac{1}{\|\boldsymbol{\tau}\|^2} (\boldsymbol{\tau} \otimes \boldsymbol{\tau})_{2 \times 2} \mathbf{x} = \left(\frac{\boldsymbol{\tau}}{\|\boldsymbol{\tau}\|} \otimes \frac{\boldsymbol{\tau}}{\|\boldsymbol{\tau}\|} \right)_{2 \times 2} \mathbf{x} = (\hat{\boldsymbol{\tau}} \otimes \hat{\boldsymbol{\tau}})_{2 \times 2} \mathbf{x} \quad (23)$$

其中， $\hat{\boldsymbol{\tau}}$ 代表 $\boldsymbol{\tau}$ 的单位向量。

一般情况，数据矩阵 \mathbf{X} 中样本点的坐标值以行向量表达， \mathbf{X} 向单位向量 \mathbf{v} 方向投影得到的标量投影，即 \mathbf{X} 在 $\text{span}(\mathbf{v})$ 的坐标：

$$\mathbf{Z} = \mathbf{X} \mathbf{v} \quad (24)$$

\mathbf{X} 向单位向量 \mathbf{v} 方向投影得到的向量投影坐标则为：

$$\mathbf{Z} = \mathbf{X} \mathbf{v} \mathbf{v}^T = \mathbf{X} (\mathbf{v} \otimes \mathbf{v})_{2 \times 2} \quad (25)$$

➡ 请大家格外注意 (25)，我们下一章还要继续这个话题。此外，(25) 也是下一章要讨论的核心运算。

9.3 正交矩阵：一个规范正交基

本章前文介绍的是朝一个向量方向投影，比如向量 \mathbf{x} 向 \mathbf{v} 方向投影，这可以视作 \mathbf{x} 向 \mathbf{v} 张起的向量空间 $\text{span}(\mathbf{v})$ 投影。同理，向量 \mathbf{x} 也可以向一个有序基构造的平面/超平面投影。这个有序基可以是正交基，可以是非正交基。

数据科学和机器学习实践中，最常用的基底是规范正交基。正交矩阵的本身就是规范正交基。本节主要介绍正交矩阵的性质。

正交矩阵

满足下式的方阵 \mathbf{V} 为**正交矩阵** (orthogonal matrix)：

$$\mathbf{V}^T \mathbf{V} = \mathbf{I} \quad (26)$$

强调一下， \mathbf{V} 为方阵是前提；否则即便满足上式也不能称之为正交矩阵。比如，如下长方形矩阵 \mathbf{A} 也满足上式，但 \mathbf{A} 不是正交矩阵：

$$\underbrace{\begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \\ 0 & 0 \end{bmatrix}}_A = \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_I \quad (27)$$

但是， A 的列向量为单位向量、且两两正交，所以 $A = [a_1, a_2]$ 是规范正交基。

正交矩阵基本性质：

$$\begin{aligned} \mathbf{V}\mathbf{V}^T &= \mathbf{V}^T\mathbf{V} = \mathbf{I} \\ \mathbf{V}^T &= \mathbf{V}^{-1} \end{aligned} \quad (28)$$

▲ (28) 中两式经常使用，必须烂熟于心。

举个实例，图 5 所示热图为一个 4×4 正交矩阵 V 和自己转置 V^T 乘积为单位阵 I 。

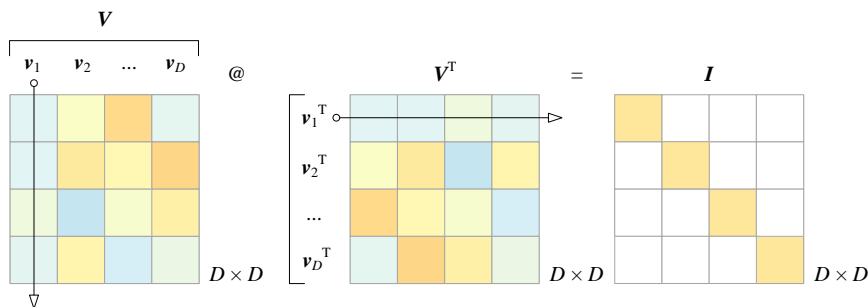


图 5. 正交阵 V 和自己转置 V^T 乘积为单位阵 I

前文的例子

其实我们已经接触过几种正交矩阵。本书前文提到的如下两个矩阵都是正交矩阵：

$$\mathbf{V} = [\mathbf{v}_1 \quad \mathbf{v}_2] = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}, \quad \mathbf{W} = [\mathbf{w}_1 \quad \mathbf{w}_2] = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \quad (29)$$

(29) 中 V 和 W 都满足方阵和自身转置乘积为单位阵，即：

$$\begin{aligned} \mathbf{V}^T\mathbf{V} &= \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \mathbf{W}^T\mathbf{W} &= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \quad (30)$$

本书上一章讲过的矩阵 \mathbf{R} 、 \mathbf{T} 和 \mathbf{P} 都是正交矩阵：

$$\mathbf{R} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} & & 1 \\ 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \quad (31)$$

其中， \mathbf{R} 代表旋转， \mathbf{T} 代表镜像， \mathbf{P} 是置换矩阵。也就是说，正交矩阵的几何操作可能对应“旋转”、“镜像”、“置换”，或者它们的组合，比如“旋转 + 镜像”。

矩阵乘法第一视角展开

将 (26) 中矩阵 \mathbf{V} 写成一排列向量：

$$\mathbf{V}_{D \times D} = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,D} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n,1} & v_{n,2} & \cdots & v_{n,D} \end{bmatrix} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_D] \quad (32)$$

(26) 左侧可以写成：

$$\mathbf{V}^T \mathbf{V} = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_D^T \end{bmatrix} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_D] \quad (33)$$

(33) 展开得到：

$$\mathbf{V}^T \mathbf{V} = \begin{bmatrix} \mathbf{v}_1^T \mathbf{v}_1 & \mathbf{v}_1^T \mathbf{v}_2 & \cdots & \mathbf{v}_1^T \mathbf{v}_D \\ \mathbf{v}_2^T \mathbf{v}_1 & \mathbf{v}_2^T \mathbf{v}_2 & \cdots & \mathbf{v}_2^T \mathbf{v}_D \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_D^T \mathbf{v}_1 & \mathbf{v}_D^T \mathbf{v}_2 & \cdots & \mathbf{v}_D^T \mathbf{v}_D \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (34)$$

大家应该已经意识到，(34) 就是 $\mathbf{V}^T \mathbf{V}$ 矩阵乘法的第一视角。

$\mathbf{V}^T \mathbf{V}$ 主对角线结果为 1，即，

$$\mathbf{v}_j^T \mathbf{v}_j = \mathbf{v}_j \cdot \mathbf{v}_j = \|\mathbf{v}_j\|^2 = 1 \quad j=1,2,\dots,D \quad (35)$$

也就是说，矩阵 \mathbf{V} 的每个列向量 \mathbf{v}_j 为单位向量。

(34) 主对角线以外元素均为 0：

$$\mathbf{v}_i^T \mathbf{v}_j = 0, \quad i \neq j \quad (36)$$

即 \mathbf{V} 中任意两个列向量两两正交，即垂直。

至此，可以判定 $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D\}$ 为规范正交基。写成有序基形式，就是矩阵 $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$ 。 \mathbf{V} 张起一个 D 维向量空间 $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D)$ ， $\mathbb{R}^D = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D)$ 。也就是说， $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$ 是张起 \mathbb{R}^D 无数规范正交基的一组。

顺便提一嘴，由于 $\mathbf{V}^T \mathbf{V} = \mathbf{V} \mathbf{V}^T = \mathbf{I}$ ， \mathbf{V}^T 本身也是一个规范正交基。 \mathbf{V}^T 可以展开写成 $\mathbf{V}^T = [\mathbf{v}^{(1)T}, \mathbf{v}^{(2)T}, \dots, \mathbf{v}^{(D)T}]$ 。

批量化计算向量模和夹角

此外，(34) 告诉我们“批量”计算一系列向量模和两两夹角的方式——**格拉姆矩阵** (Gram matrix)!

$\mathbf{V}^T \mathbf{V}$ 相当于 \mathbf{V} 的格拉姆矩阵，通过对 (34) 的分析，我们知道格拉姆矩阵包含原矩阵的所有向量模、向量两两夹角这两类信息。

再举个例子，给定矩阵 \mathbf{X} ，将其写成一组列向量 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D]$ 。 \mathbf{X} 的格拉姆矩阵为：

$$\mathbf{G} = \begin{bmatrix} \mathbf{x}_1 \cdot \mathbf{x}_1 & \mathbf{x}_1 \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_1 \cdot \mathbf{x}_D \\ \mathbf{x}_2 \cdot \mathbf{x}_1 & \mathbf{x}_2 \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_2 \cdot \mathbf{x}_D \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_D \cdot \mathbf{x}_1 & \mathbf{x}_D \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_D \cdot \mathbf{x}_D \end{bmatrix} = \begin{bmatrix} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle & \langle \mathbf{x}_1, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_1, \mathbf{x}_D \rangle \\ \langle \mathbf{x}_2, \mathbf{x}_1 \rangle & \langle \mathbf{x}_2, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_2, \mathbf{x}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{x}_D, \mathbf{x}_1 \rangle & \langle \mathbf{x}_D, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_D, \mathbf{x}_D \rangle \end{bmatrix} \quad (37)$$

借助向量夹角余弦展开 \mathbf{G} 中向量积：

$$\mathbf{G} = \begin{bmatrix} \|\mathbf{x}_1\| \|\mathbf{x}_1\| \cos \theta_{1,1} & \|\mathbf{x}_1\| \|\mathbf{x}_2\| \cos \theta_{1,2} & \cdots & \|\mathbf{x}_1\| \|\mathbf{x}_D\| \cos \theta_{1,D} \\ \|\mathbf{x}_2\| \|\mathbf{x}_1\| \cos \theta_{2,1} & \|\mathbf{x}_2\| \|\mathbf{x}_2\| \cos \theta_{2,2} & \cdots & \|\mathbf{x}_2\| \|\mathbf{x}_D\| \cos \theta_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \|\mathbf{x}_D\| \|\mathbf{x}_1\| \cos \theta_{D,1} & \|\mathbf{x}_D\| \|\mathbf{x}_2\| \cos \theta_{D,2} & \cdots & \|\mathbf{x}_D\| \|\mathbf{x}_D\| \cos \theta_{D,D} \end{bmatrix} \quad (38)$$

观察矩阵 \mathbf{G} ，它包含了数据矩阵 \mathbf{X} 中列向量的两个重要信息——模 $\|\mathbf{x}_i\|$ 、方向（向量两两夹角 $\cos \theta_{i,j}$ ）。再次强调， $\theta_{i,j}$ 为相对角度。



我们将会在本书第 12 章讲解 Cholesky 分解时继续深入探讨这一话题。

矩阵乘法第二视角展开

有了第一视角，大家自然会想到矩阵乘法的第二视角。

还是将 \mathbf{V} 写成 $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$ ， $\mathbf{V} \mathbf{V}^T$ 则可以按如下方式展开：

$$\mathbf{V} \mathbf{V}^T = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_D \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_D^T \end{bmatrix} = \mathbf{v}_1 \mathbf{v}_1^T + \mathbf{v}_2 \mathbf{v}_2^T + \cdots + \mathbf{v}_D \mathbf{v}_D^T = \mathbf{I}_{D \times D} \quad (39)$$

(39) 可以写成一系列张量积之和：

$$\mathbf{V}\mathbf{V}^T = \mathbf{v}_1 \otimes \mathbf{v}_1 + \mathbf{v}_2 \otimes \mathbf{v}_2 + \cdots + \mathbf{v}_D \otimes \mathbf{v}_D = \mathbf{I}_{D \times D} \quad (40)$$

上一节 (25) 对应数据矩阵 \mathbf{X} 向单位向量 \mathbf{v} 向量投影。如果 \mathbf{X} 向规范正交基 $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$ 张起的 D 维空间投影，得到的标量投影就是 $\mathbf{Z} = \mathbf{X}\mathbf{V}$ ，而向量投影结果为：

$$\begin{aligned} \mathbf{X}_{n \times D} \mathbf{V} \mathbf{V}^T &= \mathbf{X} (\mathbf{v}_1 \otimes \mathbf{v}_1 + \mathbf{v}_2 \otimes \mathbf{v}_2 + \cdots + \mathbf{v}_D \otimes \mathbf{v}_D) \\ &= \underbrace{\mathbf{X} \mathbf{v}_1}_{\mathbf{z}_1} \otimes \mathbf{v}_1 + \underbrace{\mathbf{X} \mathbf{v}_2}_{\mathbf{z}_2} \otimes \mathbf{v}_2 + \cdots + \underbrace{\mathbf{X} \mathbf{v}_D}_{\mathbf{z}_D} \otimes \mathbf{v}_D \\ &= \mathbf{X}_{n \times D} \mathbf{I}_{D \times D} \\ &= \mathbf{X}_{n \times D} \end{aligned} \quad (41)$$

大家可能已经糊涂了，上式折腾了半天，最后得到的还是原数据矩阵 \mathbf{X} 本身！

➡ (41) 已经非常接近本书第 15、16 章要讲解的奇异值分解的思路。下一章我们一起搞清楚 (41) 背后的数学思想。

再进一步，如图 6 所示，下式代表一个规范正交基对单位矩阵的分解：

$$\mathbf{I}_{D \times D} = \mathbf{v}_1 \otimes \mathbf{v}_1 + \mathbf{v}_2 \otimes \mathbf{v}_2 + \cdots + \mathbf{v}_D \otimes \mathbf{v}_D = \sum_{j=1}^D \mathbf{v}_j \otimes \mathbf{v}_j \quad (42)$$

其中，每个 $\mathbf{v}_j \otimes \mathbf{v}_j$ 都是一个特定方向的**投影矩阵** (projection matrix)。这个视角同样重要，本章和下一章还将继续深入讨论。

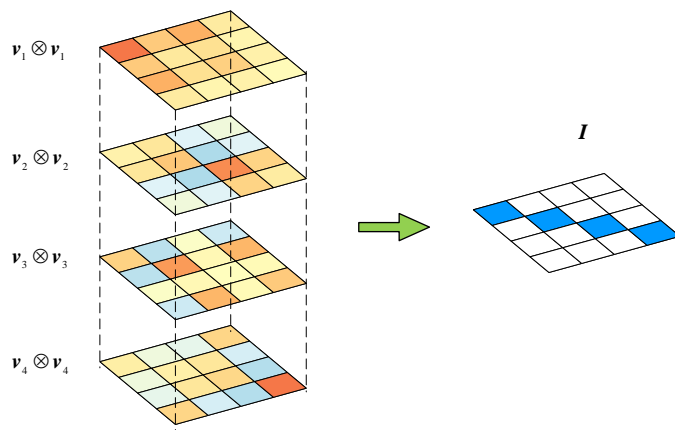


图 6. 对单位矩阵的分解

9.4 规范正交基性质

本节以 (29) 中矩阵 \mathbf{V} 为例介绍更多规范正交基的性质。

坐标

将 V 分解成两个列向量,

$$v_1 = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}, \quad v_2 = \begin{bmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \quad (43)$$

这两个向量长度为 1，都是单位向量。

显然， V 的转置和 V 本身乘积是一个 2×2 单位矩阵。用矩阵乘法第一视角展开 $V^T V$ 得到：

$$V^T V = \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} v_1^T v_1 & v_1^T v_2 \\ v_2^T v_1 & v_2^T v_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I \quad (44)$$

给定列向量 $x = [4, 3]^T$ 。如图 7 (a) 所示， x 在标准正交基 $[e_1, e_2]$ 中的坐标为 (4, 3)。

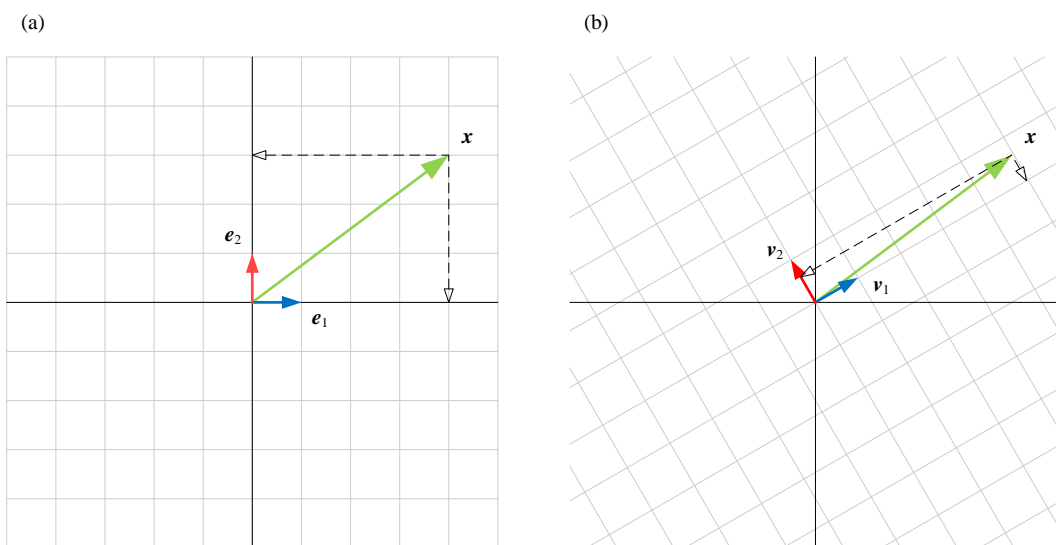


图 7. x 在不同规范正交系中的坐标

如图 7 (b) 所示，将 x 投影到 V 这个规范正交系中，得到的结果就是在 $[v_1, v_2]$ 这个规范正交系的坐标：

$$V^T x = \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix} x = \begin{bmatrix} v_1^T x \\ v_2^T x \end{bmatrix} = \begin{bmatrix} \text{proj}_{v_1}(x) \\ \text{proj}_{v_2}(x) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 4.964 \\ 0.598 \end{bmatrix} \quad (45)$$

这说明，向量 x 在规范正交系 $[v_1, v_2]$ 中的坐标为 (4.964, 0.598)。

向量长度不变

经过正交矩阵 V 线性变换后，向量 x 的 L^2 范数，即向量模，没有变化：

$$\begin{aligned}\|V^T x\|_2^2 &= V^T x \cdot V^T x = (V^T x)^T (V^T x) = x^T V^T V x \\ &= x^T I x = x^T x = x \cdot x = \|x\|_2^2\end{aligned}\quad (46)$$

比较图 7 (a) 和 (b) 可以发现，不同规范正交系中 x 的长度确实没有变化。向量 x 在 $[v_1, v_2]$ 中坐标为 (4.964, 0.598)，计算其向量模：

$$\sqrt{4.964^2 + 0.598^2} = \sqrt{4^2 + 3^2} = 5 \quad (47)$$

图 8 所示为平面上给定向量在不同规范正交基中的投影结果。

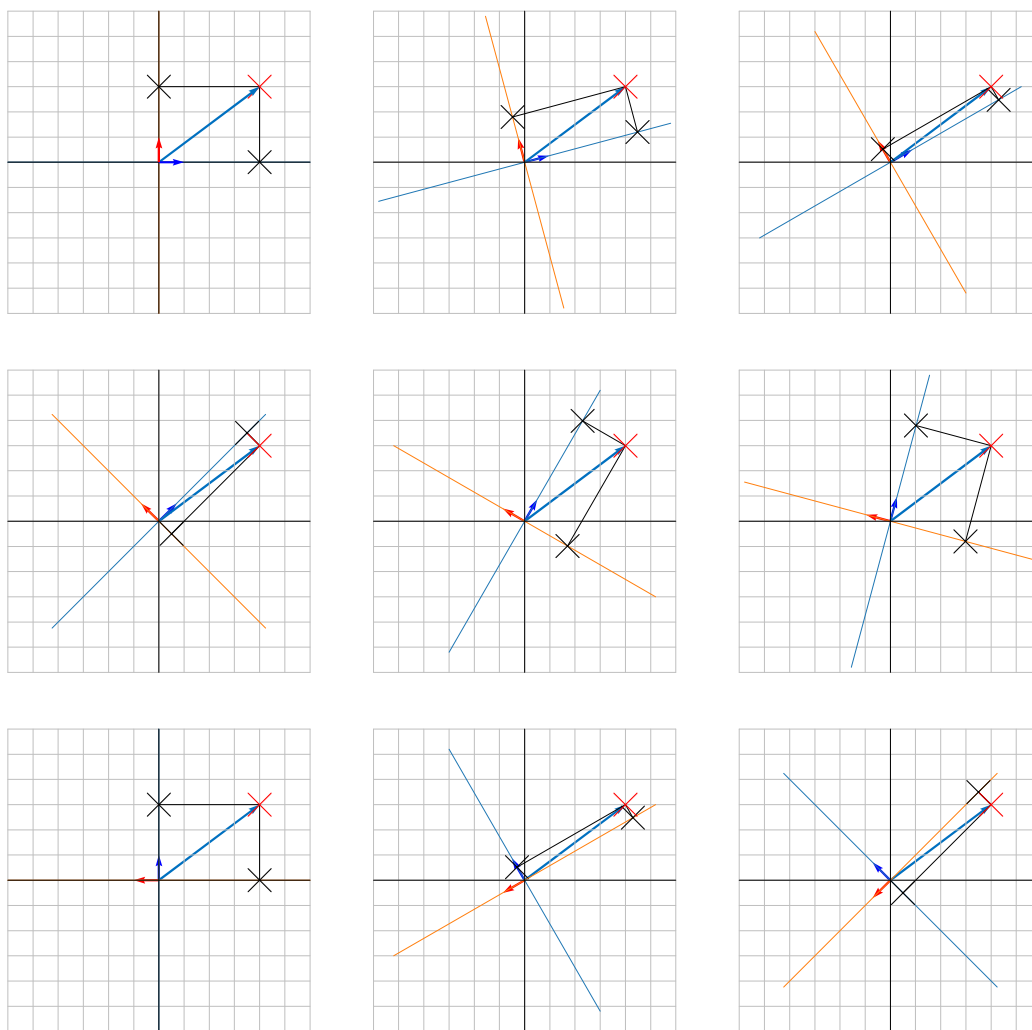


图 8. 平面中向量在不同坐标系的投影



Bk4_Ch9_02.py 绘制图 8。

夹角不变

\mathbf{x}_i 和 \mathbf{x}_j 经过正交矩阵 \mathbf{V} 线性转化得到 \mathbf{z}_i 和 \mathbf{z}_j 。 \mathbf{z}_i 和 \mathbf{z}_j 夹角等同于 \mathbf{x}_i 和 \mathbf{x}_j 夹角：

$$\frac{\mathbf{z}_i \cdot \mathbf{z}_j}{\|\mathbf{z}_i\| \|\mathbf{z}_j\|} = \frac{\mathbf{z}_i \cdot \mathbf{z}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|} = \frac{\mathbf{V}^T \mathbf{x}_i \cdot \mathbf{V}^T \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|} = \frac{(\mathbf{V}^T \mathbf{x}_i)^T \mathbf{V}^T \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|} = \frac{\mathbf{x}_i^T \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|} = \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|} \quad (48)$$

如图 9 所示，经过正交矩阵 \mathbf{V} 线性变换后， \mathbf{x}_i 和 \mathbf{x}_j 两者相对角度等同于 \mathbf{z}_i 和 \mathbf{z}_j 相对角度。这也不难理解，变化前后，向量都还在 \mathbb{R}^2 中，只不过是坐标参考系发生了旋转，而 \mathbf{x}_i 和 \mathbf{x}_j 之间的“相对角度”完全没有发生改变。

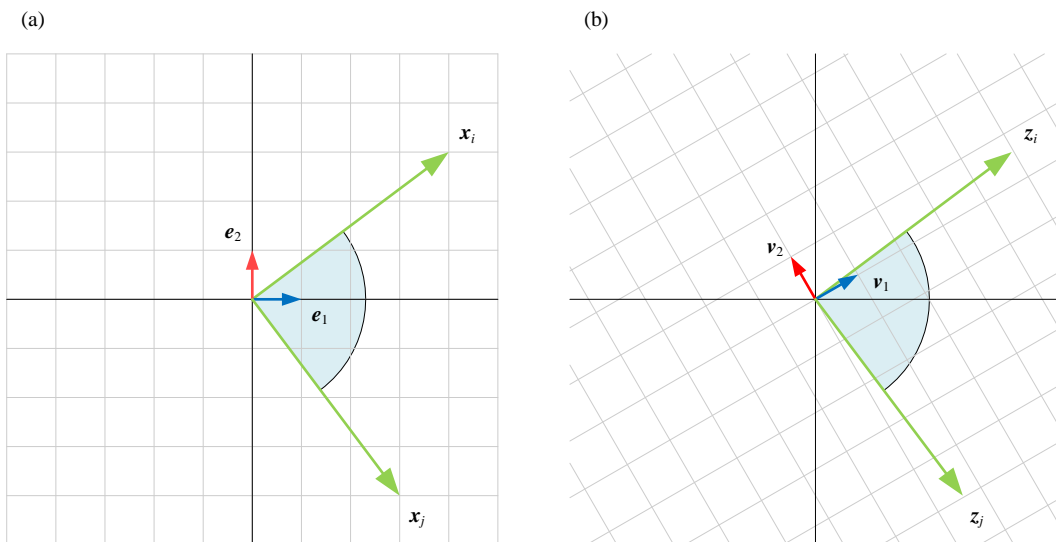


图 9. 不同规范正交系中， \mathbf{x}_i 和 \mathbf{x}_j 的夹角不变

行列式值

正交矩阵 \mathbf{V} 还有一个有趣性质， \mathbf{V} 行列式值为 1 或 -1：

$$(\det(\mathbf{V}))^2 = \det(\mathbf{V}^T) \det(\mathbf{V}) = \det(\mathbf{V}^T \mathbf{V}) = \det(\mathbf{I}) = 1 \quad (49)$$

也就是说，经过 2×2 方阵 \mathbf{V} 线性变换后，图形面积不变。当 $\det(\mathbf{V}) = -1$ 时，图形会发生翻转。

9.5 再谈镜像：从投影视角

上一章聊几何变换时，我们介绍了镜像，并且直接给出完成镜像操作转换矩阵 T 的一种形式，具体如下：

$$T = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \quad (50)$$

本节用正交投影推导 (50)。

如图 10 所示，镜像对称轴 l 这条直线通过原点，直线切向量 τ 为：

$$\tau = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \quad (51)$$

向量 x 关于对称轴 l 镜像得到 z 。

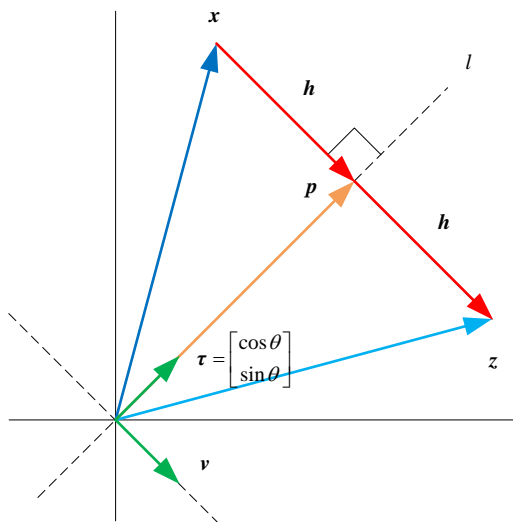


图 10. 投影视角看镜像

从投影角度来看，向量 x 在 τ 方向投影为向量 p 。利用张量积（投影矩阵）形式，向量 p 可以写成：

$$p = (\tau \otimes \tau) x \quad (52)$$

将 (51) 代入 (52)，整理得到：

$$p = (\tau \otimes \tau) x = \begin{bmatrix} \cos \theta \cos \theta & \cos \theta \sin \theta \\ \cos \theta \sin \theta & \sin \theta \sin \theta \end{bmatrix} x \quad (53)$$

利用三角恒等式，上式可以整理为：

$$\mathbf{p} = \begin{bmatrix} (\cos 2\theta + 1)/2 & \sin 2\theta/2 \\ \sin 2\theta/2 & (1 - \cos 2\theta)/2 \end{bmatrix} \mathbf{x} \quad (54)$$

令，向量 \mathbf{h} 为 \mathbf{p} 、 \mathbf{x} 之差，即：

$$\mathbf{h} = \mathbf{p} - \mathbf{x} \quad (55)$$

根据正交投影， \mathbf{h} 显然垂直 \mathbf{p} 。观察图 10，由于 \mathbf{z} 和 \mathbf{x} 为镜像关系，因此两者之差为 $2\mathbf{h}$ ，也就是下式成立：

$$\mathbf{z} = \mathbf{x} + 2\mathbf{h} \quad (56)$$

将 (55) 代入 (56) 整理得到：

$$\mathbf{z} = 2\mathbf{p} - \mathbf{x} \quad (57)$$

从另外一个角度来看， $\mathbf{x} + \mathbf{z} = 2\mathbf{p}$ 。

将 (54) 代入 (57) 得到：

$$\begin{aligned} \mathbf{z} &= 2 \begin{bmatrix} (\cos 2\theta + 1)/2 & \sin 2\theta/2 \\ \sin 2\theta/2 & (1 - \cos 2\theta)/2 \end{bmatrix} \mathbf{x} - \mathbf{x} \\ &= \begin{bmatrix} 2 \times (\cos 2\theta + 1)/2 - 1 & 2 \times \sin 2\theta/2 \\ 2 \times \sin 2\theta/2 & 2 \times (1 - \cos 2\theta)/2 - 1 \end{bmatrix} \mathbf{x} \\ &= \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \mathbf{x} \end{aligned} \quad (58)$$

这样，我们使用投影视角推导得到 (50) 结果。

豪斯霍尔德矩阵

此外，将 (52) 代入 (57)，整理得到：

$$\mathbf{z} = 2(\boldsymbol{\tau} \otimes \boldsymbol{\tau})\mathbf{x} - \mathbf{x} = (2\boldsymbol{\tau} \otimes \boldsymbol{\tau} - \mathbf{I})\mathbf{x} \quad (59)$$

在图 10 中，定义单位向量 \mathbf{v} 垂直于切向量 $\boldsymbol{\tau}$ ， $[\boldsymbol{\tau}, \mathbf{v}]$ 为规范正交基，满足：

$$\boldsymbol{\tau} \otimes \boldsymbol{\tau} + \mathbf{v} \otimes \mathbf{v} = \mathbf{I} \quad (60)$$

$\boldsymbol{\tau} \otimes \boldsymbol{\tau}$ 可以写成：

$$\boldsymbol{\tau} \otimes \boldsymbol{\tau} = \mathbf{I} - \mathbf{v} \otimes \mathbf{v} \quad (61)$$

将 (61) 代入 (59) 得到：

$$\mathbf{z} = \underbrace{(\mathbf{I} - 2\mathbf{v} \otimes \mathbf{v})}_{\mathbf{H}} \mathbf{x} \quad (62)$$

令 \mathbf{H} 为：

$$H = I - 2v \otimes v \quad (63)$$

矩阵 H 有自己的名字——**豪斯霍尔德矩阵** (Householder matrix)。矩阵 H 完成的转换叫做**豪斯霍尔德反射** (Householder reflection)，也叫初等反射。图 10 中向量 v 的方向就是反射面所在方向。

9.6 格拉姆-施密特正交化

格拉姆-施密特正交化 (Gram-Schmidt orthogonalization) 是求解规范正交基的一种方法。整个过程用到核心数学工具就是正交投影。

给定非正交 D 个线性不相关的向量 $[x_1, x_2, x_3, \dots, x_D]$ ，通过格拉姆-施密特正交化，可以得到 D 个单位正交向量 $\{q_1, q_2, q_3, \dots, q_D\}$ ，它们可以构造一个规范正交基 $[q_1, q_2, q_3, \dots, q_D]$ 。

正交化过程

格拉姆-施密特正交化过程如下所示：

$$\begin{aligned} \eta_1 &= x_1 \\ \eta_2 &= x_2 - \text{proj}_{\eta_1}(x_2) \\ \eta_3 &= x_3 - \text{proj}_{\eta_1}(x_3) - \text{proj}_{\eta_2}(x_3) \\ &\vdots \\ \eta_D &= x_D - \sum_{j=1}^{D-1} \text{proj}_{\eta_j}(x_D) \end{aligned} \quad (64)$$

前两步

图 11 所示为格拉姆-施密特正交化前两步。

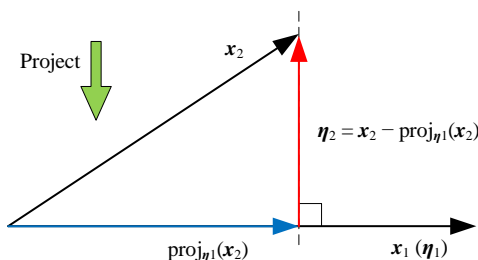


图 11. 格拉姆-施密特正交化前两步

获得 η_1 很容易，只需要 $\eta_1 = x_1$ 。

求解 η_2 需要利用 η_2 垂直于 η_1 这一条件，即：

$$(\eta_1)^T \eta_2 = 0 \quad (65)$$

如图 11 所示， x_2 在 η_1 方向上向量投影为 $\text{proj}_{\eta_1}(x_2)$ ，剩余的向量分量垂直于 x_1 (η_1)，这个分量就是 η_2 ：

$$\eta_2 = x_2 - \text{proj}_{\eta_1}(x_2) = x_2 - \frac{x_2^T \eta_1}{\eta_1^T \eta_1} \eta_1 \quad (66)$$

η_2 也有自己的名字，叫 η_1 的**正交补** (orthogonal complement)。也可以说， η_1 和 η_2 互为正交补。下面验证 η_1 和 η_2 相互垂直：

$$\begin{aligned} (\eta_1)^T \eta_2 &= (x_1)^T \left(x_2 - \frac{x_2^T \eta_1}{\eta_1^T \eta_1} \eta_1 \right) \\ &= x_1^T x_2 - \frac{x_1^T x_1 x_2^T x_1}{x_1^T x_1} = x_1^T x_2 - x_2^T x_1 = 0 \end{aligned} \quad (67)$$

第三步

如图 12 所示，第三步是 x_3 向 $[\eta_1, \eta_2]$ 张成的平面投影。令 η_3 为 x_3 中不在 $[\eta_1, \eta_2]$ 平面上的向量分量，即：

$$\eta_3 = x_3 - \text{proj}_{\eta_1}(x_3) - \text{proj}_{\eta_2}(x_3) \quad (68)$$

显然， η_3 垂直 $\text{span}(\eta_1, \eta_2)$ ，也就是说 η_3 分别垂直 η_1 和 η_2 。 η_3 和 $\text{span}(\eta_1, \eta_2)$ 互为正交补。

按此思路，不断反复投影直至得到所有正交向量 $\{\eta_1, \eta_2, \eta_3, \dots, \eta_D\}$ 。

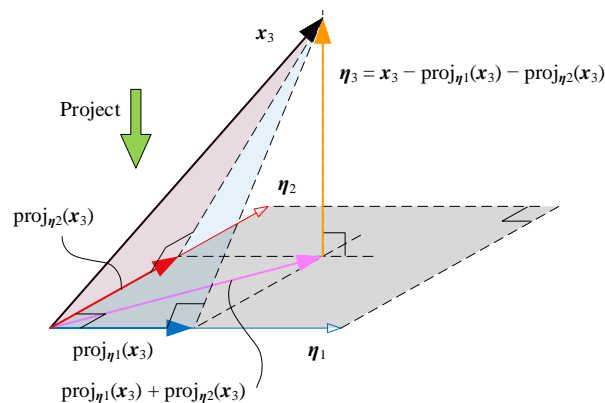


图 12. 格拉姆-施密特正交化第三步

单位化

最后单位化，获得单位正交向量 $\{q_1, q_2, q_3, \dots, q_D\}$ ：

$$q_1 = \frac{\eta_1}{\|\eta_1\|}, \quad q_2 = \frac{\eta_2}{\|\eta_2\|}, \quad q_3 = \frac{\eta_3}{\|\eta_3\|}, \quad \dots, \quad q_D = \frac{\eta_D}{\|\eta_D\|} \quad (69)$$

值得强调的是，规范正交基 $[q_1, q_2, q_3, \dots, q_D]$ 的特别之处在于 q_1 平行 x_1 。本书后续还会介绍其他获得规范正交基的算法，请大家注意比对。

举个实例

给定 x_1 和 x_2 两个向量，利用格拉姆-施密特正交化获得两个正交向量：

$$x_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad (70)$$

η_1 就是 x_1 ，即，

$$\eta_1 = x_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \quad (71)$$

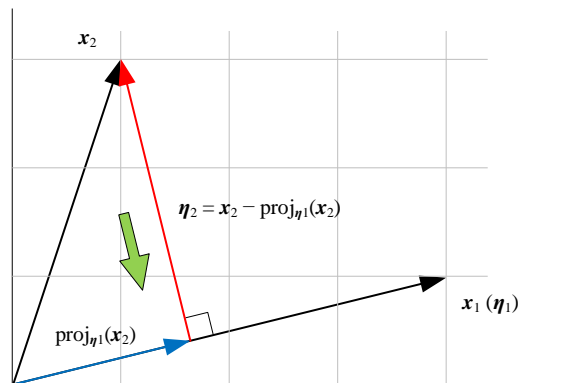


图 13. 格拉姆-施密特正交化第三步

x_2 在 η_1 (x_1) 方向上投影，得到向量投影：

$$\text{proj}_{\eta_1}(x_2) = \frac{x_2 \cdot \eta_1}{\eta_1 \cdot \eta_1} \eta_1 = \frac{4 \times 1 + 1 \times 3}{4 \times 4 + 1 \times 1} \times \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \frac{7}{17} \times \begin{bmatrix} 4 \\ 1 \end{bmatrix} \quad (72)$$

计算 η_2 ：

$$\begin{aligned}\eta_2 &= \mathbf{x}_2 - \text{proj}_{\eta_1}(\mathbf{x}_2) \\ &= \begin{bmatrix} 1 \\ 3 \end{bmatrix} - \frac{7}{17} \times \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \frac{1}{17} \times \begin{bmatrix} -11 \\ 44 \end{bmatrix}\end{aligned}\quad (73)$$

最后对 η_1 和 η_2 单位化，得到 q_1 和 q_2 ：

$$\begin{cases} q_1 = \frac{\eta_1}{\|\eta_1\|} = \frac{1}{\sqrt{17}} \begin{bmatrix} 4 \\ 1 \end{bmatrix} \\ q_2 = \frac{\eta_2}{\|\eta_2\|} = \frac{1}{\sqrt{17}} \begin{bmatrix} -1 \\ 4 \end{bmatrix} \end{cases}\quad (74)$$

➡ 格拉姆-施密特正交化可以通过 QR 分解完成，这是第 11 章矩阵分解要讲解的内容之一。

9.7 投影视角看回归

本系列丛书《数学要素》鸡兔同笼三部曲中简单介绍过如何通过投影视角理解线性回归。本节在此基础上展开讲解。

一元线性回归

列向量 y 在 x 方向上正交投影得到向量 \hat{y} 。向量差 $y - \hat{y}$ 垂直于 x 。据此构造如下等式：

$$\mathbf{x}^T (\mathbf{y} - \hat{\mathbf{y}}) = 0 \quad (75)$$

显然 \hat{y} 和 x 共线，因此下式成立：

$$\hat{\mathbf{y}} = b\mathbf{x} \quad (76)$$

其中， b 为实数系数。大家在上式中是否已经看到线性回归的影子？

从向量空间角度来看， $\text{span}(\mathbf{x})$ 张起的向量空间维度为 1。 \hat{y} 在 $\text{span}(\mathbf{x})$ 中， \hat{y} 和 x 线性相关。

从数据角度思考， x 为自变量， y 为因变量。数据 x 方向能够解释 y 的一部分，即 \hat{y} 。不能解释的部分就是残差 (residuals)，即 $\varepsilon = y - \hat{y}$ 。 ε 和 x 互为正交补。

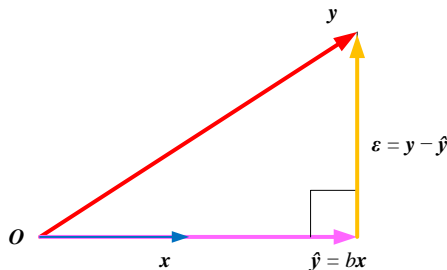


图 14. 向量 y 向 x 正交投影得到向量投影 \hat{y}

将 (76) 代入 (75)，得到：

$$\mathbf{x}^T (\mathbf{y} - b\mathbf{x}) = 0 \quad (77)$$

容易求得系数 b 为：

$$b = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y} \quad (78)$$

从而， \hat{y} 为：

$$\hat{\mathbf{y}} = \mathbf{x} (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y} \quad (79)$$

这样，利用向量投影这个数学工具，我们解释了一元线性回归。

⚠ 注意，在上述分析中，我们没有考虑常数项。也就是说，上述线性回归模型为比例函数，截距为 0。从图像上来看，比例函数过原点。

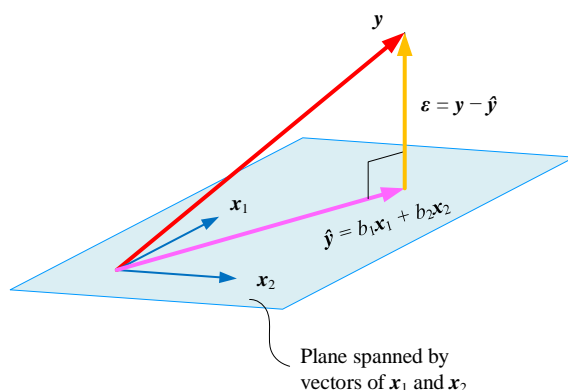
二元线性回归

下面我们聊一下二元线性回归。

如图 15 所示，两个线性无关向量 \mathbf{x}_1 和 \mathbf{x}_2 张成一个平面 $\text{span}(\mathbf{x}_1, \mathbf{x}_2)$ 。向量 \mathbf{y} 向该平面投影得到向量 $\hat{\mathbf{y}}$ 。向量 $\hat{\mathbf{y}}$ 是 \mathbf{x}_1 和 \mathbf{x}_2 线性组合：

$$\hat{\mathbf{y}} = b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2 = \underbrace{\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}_{\mathbf{b}} = \mathbf{X}\mathbf{b} \quad (80)$$

其中， b_1 和 b_2 为系数。 $\text{span}(\mathbf{x}_1, \mathbf{x}_2)$ 和 $\boldsymbol{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}}$ 互为正交补。

图 15. 向量 y 向平面 $\text{span}(\mathbf{x}_1, \mathbf{x}_2)$ 投影

$y - \hat{y}$ 垂直于 $X = [\mathbf{x}_1, \mathbf{x}_2]$, 也就是说 $y - \hat{y}$ 分别垂直 \mathbf{x}_1 和 \mathbf{x}_2 , 据此构造如下两个等式:

$$\begin{cases} \mathbf{x}_1^T (y - \hat{y}) = 0 \\ \mathbf{x}_2^T (y - \hat{y}) = 0 \end{cases} \Rightarrow \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \end{bmatrix} (y - \hat{y}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (81)$$

注意, 并不要求 \mathbf{x}_1 和 \mathbf{x}_2 相互正交。

整理 (81) 得到:

$$X^T (y - \hat{y}) = 0 \quad (82)$$

将 (80) 代入 (82) 得到:

$$X^T (y - Xb) = 0 \quad (83)$$

从而推导得到 b 的解析式:

$$b = (X^T X)^{-1} X^T y \quad (84)$$

(84) 代入 (80), 可以得到:

$$\hat{y} = X (X^T X)^{-1} X^T y \quad (85)$$

上式中, $X(X^T X)^{-1} X^T$ 常被称作**帽子矩阵** (hat matrix)。必须强调一点, 只有 X 为列满秩时, $X^T X$ 才存在逆。

$X(X^T X)^{-1} X^T$ 是我们在本书第 5 章提到的**幂等矩阵** (idempotent matrix), 即下式成立:

$$\left(X (X^T X)^{-1} X^T \right)^2 = X \underbrace{(X^T X)^{-1} X^T X}_{I} (X^T X)^{-1} X^T = X (X^T X)^{-1} X^T \quad (86)$$

多元线性回归

以上结论也可以推广到如图 16 所示多元线性回归情形。 D 个向量 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D$ 张成超平面 $H = \text{span}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D)$, 向量 y 在超平面 H 上投影结果为 \hat{y} , 即,

$$\hat{y} = b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2 + \dots + b_D \mathbf{x}_D \quad (87)$$

误差 $y - \hat{y}$ 垂直垂直于 $H = \text{span}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D)$, 也就是说 $y - \hat{y}$ 分别垂直于 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D$, 即:

$$\begin{cases} \mathbf{x}_1^T (y - \hat{y}) = 0 \\ \mathbf{x}_2^T (y - \hat{y}) = 0 \\ \vdots \\ \mathbf{x}_D^T (y - \hat{y}) = 0 \end{cases} \Rightarrow \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_D^T \end{bmatrix}_{X^T} (y - \hat{y}) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (88)$$

$\text{span}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D)$ 和 $\varepsilon = y - \hat{y}$ 互为正交补。

用之前的推导思路，我们也可以得到 (85)。

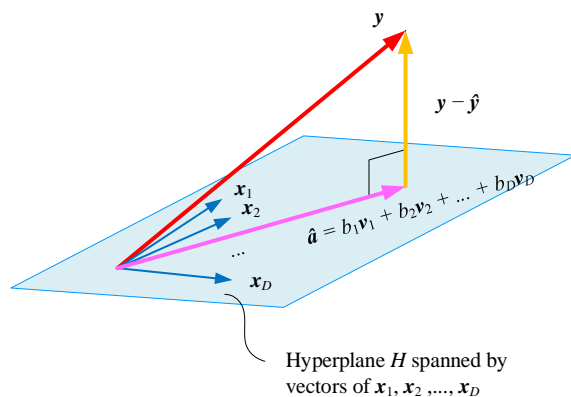


图 16. 向量 y 向超平面 $\text{span}(x_1, x_2, \dots, x_D)$ 投影

考虑常数项

而考虑常数项 b_0 ，无非就是在 (87) 中加入一个全 1 列向量 $\mathbf{1}$ ，即，

$$\hat{y} = b_0 \mathbf{1} + b_1 x_1 + b_2 x_2 + \dots + b_D x_D \quad (89)$$

而 $D + 1$ 个向量 $\mathbf{1}$ 、 x_1 、 x_2 、 \dots 、 x_D 张成一个全新超平面 $\text{span}(\mathbf{1}, x_1, x_2, \dots, x_D)$ 。而 $\mathbf{1}$ 经常写成 x_0 ，新的 X 则为 $[x_0, x_1, x_2, \dots, x_D]$ 。按照本节前文思路，我们同样可以得到 (85)。

在多元线性回归中， X 也叫**设计矩阵** (design matrix)。

➡ 数据角度来看， $x_0, x_1, x_2, \dots, x_D$ 是一列列数值，但是几何视角下它们又是什么？本书第 12 章就试图回答这个问题。

多项式回归

有些应用场合，自变量和因变量之间存在明显的非线性关系，线性回归不足以描述这种关系。这种情况，我们需要借助非线性回归模型，比如**多项式回归** (polynomial regression)。

举个例子，一元三次多项式回归模型可以写成：

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + b_3 x^3 \quad (90)$$

这时，设计矩阵 X 为：

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}_{n \times 4} \quad (91)$$

举个例子， x 和 y 取值图 17 所示。

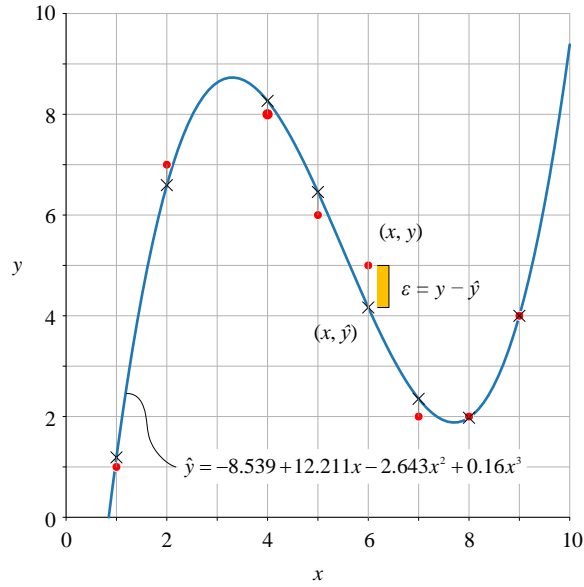


图 17. 一元三次多项式回归

一元三次多项式回归模型的自变量 x 、因变量 y 、设计矩阵 X 分别为：

$$x = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{bmatrix}_{8 \times 1}, \quad y = \begin{bmatrix} 7 \\ 8 \\ 6 \\ 5 \\ 2 \\ 2 \\ 4 \end{bmatrix}_{8 \times 1}, \quad X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 4 & 16 & 64 \\ 1 & 5 & 25 & 125 \\ 1 & 6 & 36 & 216 \\ 1 & 7 & 49 & 343 \\ 1 & 8 & 64 & 512 \\ 1 & 9 & 81 & 729 \end{bmatrix}_{8 \times 4} \quad (92)$$

利用 (84) 计算得到系数向量 b ：

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \left(\begin{bmatrix} 8 & 42 & 276 & 1998 \\ 42 & 276 & 1998 & 15252 \\ 276 & 1998 & 15252 & 120582 \\ 1998 & 15252 & 120582 & 977676 \end{bmatrix} \right)^{-1} X^T y \approx \begin{bmatrix} -8.539 \\ 12.211 \\ -2.643 \\ 0.160 \end{bmatrix} \quad (93)$$

三次一元多项式回归模型可以写成：

$$\hat{y} = -8.539 + 12.211x - 2.643x^2 + 0.16x^3 \quad (94)$$

对于给定的因变量 y ，因变量预测值为 \hat{y} ，误差为 ε ，它们的具体值如下：

$$\mathbf{y} = \begin{bmatrix} 1 \\ 7 \\ 8 \\ 6 \\ 5 \\ 2 \\ 2 \\ 4 \end{bmatrix}_{8 \times 1}, \quad \hat{\mathbf{y}} = \begin{bmatrix} 1.189 \\ 6.592 \\ 8.266 \\ 6.457 \\ 4.165 \\ 2.351 \\ 1.976 \\ 4.001 \end{bmatrix}_{8 \times 1}, \quad \boldsymbol{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -0.189 \\ 0.408 \\ -0.266 \\ -0.457 \\ 0.835 \\ -0.351 \\ 0.024 \\ -0.001 \end{bmatrix}_{8 \times 1} \quad (95)$$

更具一般性的正交投影

最后再回过头来看 (85)，我们可以发现这个式子实际上代表了更具一般性的正交投影。

数据矩阵 $\mathbf{X}_{n \times D}$ 的列向量 $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D]$ 张成超平面 $H = \text{span}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D)$ 。即便 $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D]$ 之间并非两两正交，向量 \mathbf{y} 依然可以在超平面 H 上正交投影，得到 $\hat{\mathbf{y}}$ 。

特殊地，如果假设 \mathbf{X} 的列向量 $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D]$ 两两正交，且列向量本身都是单位向量，可以得到：

$$\underbrace{\begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_D^T \end{bmatrix}}_{\mathbf{X}^T} \underbrace{\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_D \end{bmatrix}}_{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_1^T \mathbf{x}_1 & \mathbf{x}_1^T \mathbf{x}_2 & \cdots & \mathbf{x}_1^T \mathbf{x}_D \\ \mathbf{x}_2^T \mathbf{x}_1 & \mathbf{x}_2^T \mathbf{x}_2 & \cdots & \mathbf{x}_2^T \mathbf{x}_D \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_D^T \mathbf{x}_1 & \mathbf{x}_D^T \mathbf{x}_2 & \cdots & \mathbf{x}_D^T \mathbf{x}_D \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (96)$$

即：

$$\mathbf{X}^T \mathbf{X} = \mathbf{I} \quad (97)$$

显然， $\mathbf{X}_{n \times D}$ 不能叫做正交矩阵，这是因为 $\mathbf{X}_{n \times D}$ 的形状为 $n \times D$ ，不是方阵。

将 (97) 代入 (85) 得到：

$$\hat{\mathbf{y}} = \mathbf{X} \mathbf{X}^T \mathbf{y} \quad (98)$$

将 \mathbf{X} 写成 $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D]$ ，并展开上式得到：

$$\hat{\mathbf{y}} = \underbrace{\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_D \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_D^T \end{bmatrix}}_{\mathbf{X}^T} \mathbf{y} = (\mathbf{x}_1 \mathbf{x}_1^T + \mathbf{x}_2 \mathbf{x}_2^T + \cdots + \mathbf{x}_D \mathbf{x}_D^T) \mathbf{y} \quad (99)$$

进一步，使用向量张量积将上式写成：

$$\hat{\mathbf{y}} = (\mathbf{x}_1 \otimes \mathbf{x}_1 + \mathbf{x}_2 \otimes \mathbf{x}_2 + \cdots + \mathbf{x}_D \otimes \mathbf{x}_D) \mathbf{y} \quad (100)$$

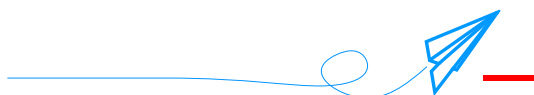
▲再次强调，上式成立的前提是—— X 的列向量 $[x_1, x_2, \dots, x_D]$ 两两正交，且列向量本身都是单位向量。

这从另外一个侧面解释了我们为什么需要格拉姆-施密特正交化！也就是说，通过格拉姆-施密特正交化， $X = [x_1, x_2, \dots, x_D]$ 变成 $Q = [q_1, q_2, \dots, q_D]$ 。而 $[q_1, q_2, \dots, q_D]$ 两两正交，且列向量都是单位向量，即满足下式：

$$Q^T Q = \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_D^T \end{bmatrix} \underbrace{\begin{bmatrix} q_1 & q_2 & \cdots & q_D \end{bmatrix}}_Q = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (101)$$

➡从 X 到 Q ，本章利用的是格拉姆-施密特正交化，而本书第 11 章将用 QR 分解。此外，本书最后一章将介绍如何用矩阵分解结果计算线性回归系数。

到目前为止，相信大家已经领略到了矩阵乘法的伟力所在！本章前前后后用的无非就是矩阵乘法的各种变形、各种视角。强烈建议大家回过头来再读一遍本书第 5 章，相信你会有一番新的收获。



本章从几何角度讲解正交投影及其应用，以下四幅图总结本书重要内容。

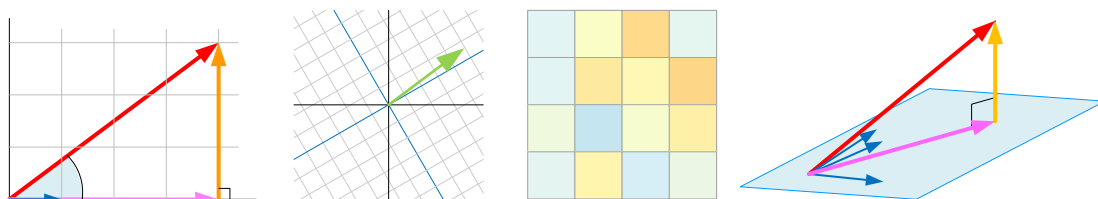


图 18. 总结本章重要内容的四幅图

本书后续内容离不开投影这个线性代数工具！大家务必熟练掌握标量/向量投影，不管是用向量内积、矩阵乘法，还是张量积。

正交矩阵本身就是规范正交基。我们将会在数据投影、矩阵分解、数据空间等一系列话题中，反复用到正交矩阵。请大家务必注意正交矩阵的性质，以及两个展开视角。

手算格拉姆-施密特正交化没有意义，大家理解这个正交化思想就好。本书后续还会介绍其他正交化方法，重要的是大家能从几何、空间、数据视角区分不同正交化方法得到结果差异。

重要的事情，强调多少遍都不为过。有向量的地方，就有几何！几何视角是理解线性回归的最佳途径，本系列丛书《概率统计》、《数据科学》还会从不同角度展开讲解线性回归。

下一章以数据为视角，和大家聊聊正交投影如何帮助我们解密数据。