

10

Data Projection

数据投影

以鸢尾花数据集为例，二次投影 + 层层叠加



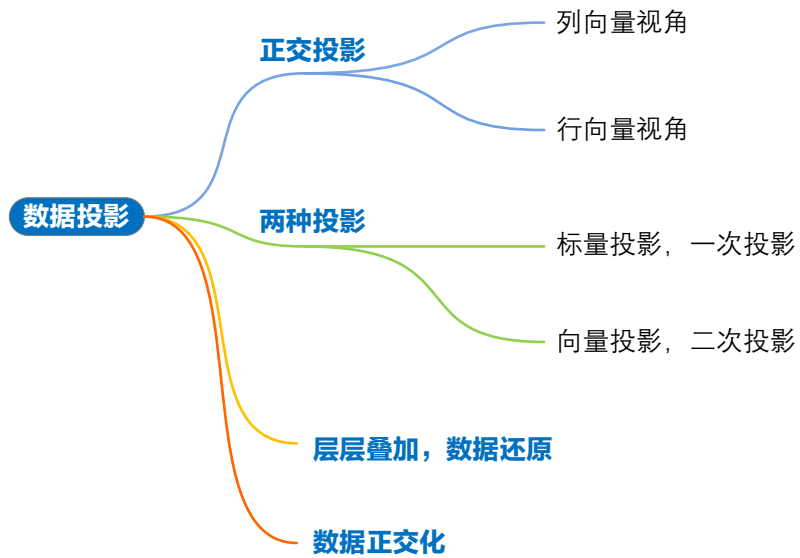
人生就像骑自行车。为了保持平衡，你必须不断移动。

Life is like riding a bicycle. To keep your balance, you must keep moving .

—— 阿尔伯特·爱因斯坦 (Albert Einstein) | 理论物理学家 | 1879 ~ 1955



- ◀ `numpy.linalg.eig()` 特征值分解
- ◀ `seaborn.heatmap()` 绘制热图



10.1 从一个矩阵乘法运算说起

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！

本章承前启后，结合数据、矩阵、向量、几何四个元素总结本书前九章主要内容，并开启本书下一个重要板块——矩阵分解。

本节和下一节内容会稍微枯燥，请大家耐心读完。之后，本章会用鸢尾花数据集作为例子，给大家展开讲解这两节内容。

正交投影

本章从一个矩阵乘法运算说起：

$$\mathbf{Z} = \mathbf{X}\mathbf{V} \quad (1)$$

\mathbf{X} 是数据矩阵，形状为 $n \times D$ ，即 n 行、 D 列。大家很清楚，以鸢尾花数据集为例， \mathbf{X} 每一行代表一个数据点，每一列代表一个特征。

\mathbf{V} 是正交矩阵，即满足 $\mathbf{V}^T \mathbf{V} = \mathbf{V} \mathbf{V}^T = \mathbf{I}$ 。这意味着 $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$ 是 \mathbb{R}^D 空间的一组规范正交基。

如图 1 所示，几何视角下，矩阵乘积 $\mathbf{X}\mathbf{V}$ 完成的是 \mathbf{X} 向规范正交基 $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$ 投影，乘积 $\mathbf{Z} = \mathbf{X}\mathbf{V}$ 代表 \mathbf{X} 在新的规范正交基下的坐标。矩阵乘法 $\mathbf{Z} = \mathbf{X}\mathbf{V}$ 也是一个线性映射过程。

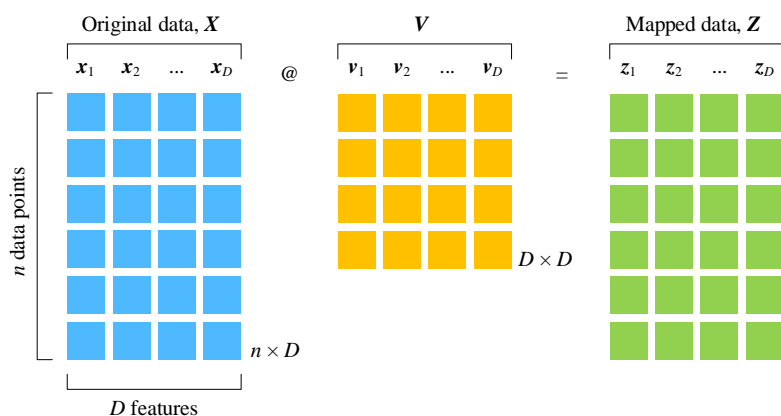


图 1. 数据矩阵 \mathbf{X} 到 \mathbf{Z} 线性变换

本书前文反复提到，一个矩阵可以看成由一系列行向量或列向量构造得到。下面，我们分别从这两个视角来分析 (1)。

列向量

将 \mathbf{Z} 和 \mathbf{V} 分别写成各自列向量，(1) 可以展开写成：

$$\begin{aligned} [\mathbf{z}_1 \quad \mathbf{z}_2 \quad \cdots \quad \mathbf{z}_D] &= \mathbf{X} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_D] \\ &= [\mathbf{X}\mathbf{v}_1 \quad \mathbf{X}\mathbf{v}_2 \quad \cdots \quad \mathbf{X}\mathbf{v}_D] \end{aligned} \quad (2)$$

(2) 这个视角是数据列向量 (即特征) 之间的转换。(2) 采用的工具是本书第 6 章介绍的分块矩阵乘法。

提取 (2) 等式左右第 j 列，得到 \mathbf{Z} 矩阵的第 j 列向量 \mathbf{z}_j 的计算式：

$$\mathbf{z}_j = \mathbf{X}\mathbf{v}_j \quad (3)$$

如图 2 所示，(3) 相当于 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D$ 通过线性组合得到 \mathbf{z}_j ，即：

$$\mathbf{z}_j = \underbrace{[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_D]}_{\mathbf{X}} \underbrace{\begin{bmatrix} v_{1,j} \\ v_{2,j} \\ \vdots \\ v_{D,j} \end{bmatrix}}_{\mathbf{v}_j} = v_{1,j}\mathbf{x}_1 + v_{2,j}\mathbf{x}_2 + \cdots + v_{D,j}\mathbf{x}_D \quad (4)$$

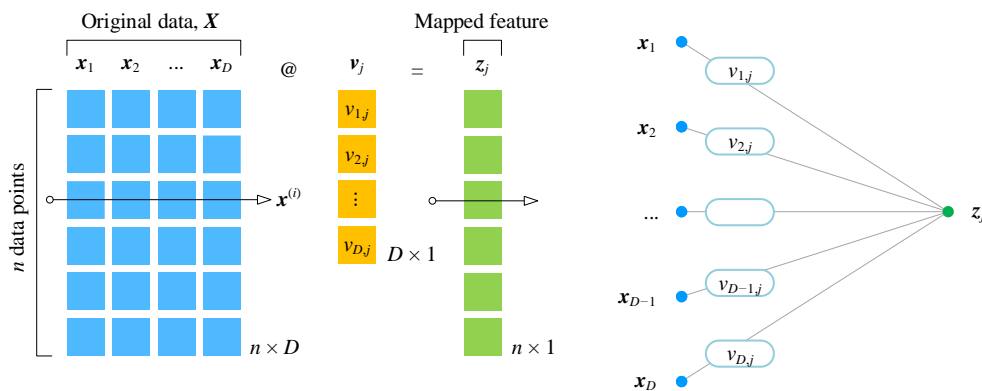


图 2. \mathbf{Z} 第 j 列向量 \mathbf{z}_j 的计算过程

行向量：点坐标

数据矩阵 \mathbf{X} 的任意行向量 $\mathbf{x}^{(i)}$ 代表一个样本点在 \mathbb{R}^D 标准正交基中坐标。将 \mathbf{X} 和 \mathbf{Z} 写成行向量形式，(1) 可以写作：

$$\begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \vdots \\ \mathbf{z}^{(n)} \end{bmatrix} = \begin{bmatrix} \mathbf{x}^{(1)}\mathbf{V} \\ \mathbf{x}^{(2)}\mathbf{V} \\ \vdots \\ \mathbf{x}^{(n)}\mathbf{V} \end{bmatrix} \quad (5)$$

如图 3 所示, (5) 代表每一行样本点之间的转换关系。即, $\mathbf{x}^{(i)}$ 投影得到 \mathbf{Z} 的第 i 行向量 $\mathbf{z}^{(i)}$:

$$\mathbf{z}^{(i)} = \mathbf{x}^{(i)}\mathbf{V} \quad (6)$$

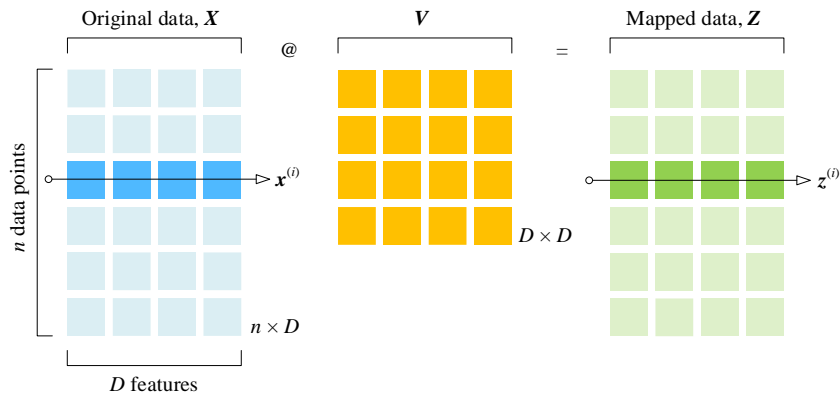


图 3. 每一行数据点之间的转换关系

进一步将 (6) 中 V 写成 $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$, (6) 可以展开得到:

$$\begin{bmatrix} z_{i,1} & z_{i,2} & \cdots & z_{i,D} \end{bmatrix} = \mathbf{x}^{(i)} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_D \end{bmatrix} \\ = \begin{bmatrix} \mathbf{x}^{(i)}\mathbf{v}_1 & \mathbf{x}^{(i)}\mathbf{v}_2 & \cdots & \mathbf{x}^{(i)}\mathbf{v}_D \end{bmatrix} \quad (7)$$

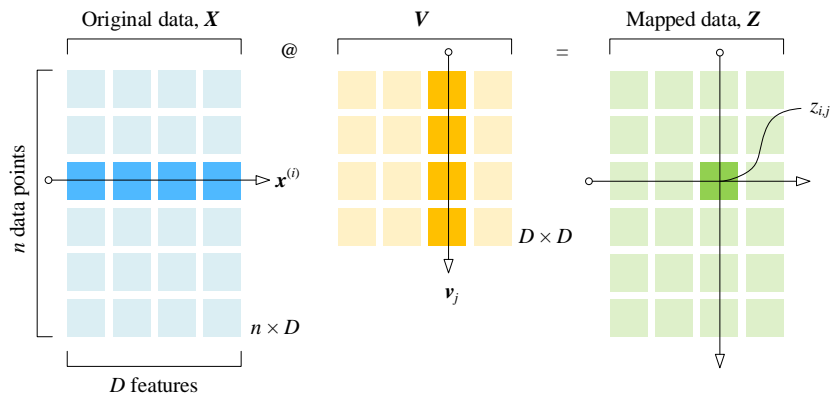


图 4. 每一行数据点向 \mathbf{v}_j 投影

取出 (7) 中向量 $\mathbf{z}^{(i)}$ 第 j 列元素 $z_{i,j}$ ，对应的运算为：

$$z_{i,j} = \mathbf{x}^{(i)} \mathbf{v}_j \quad (8)$$

图 4 对应 (8) 运算。

从空间视角来看，如图 5 所示，行向量 $\mathbf{x}^{(i)}$ 位于 \mathbb{R}^D 空间，而 $\mathbf{x}^{(i)}$ 正交投影到 \mathbb{R}^D 子空间 (subspace) $\text{span}(\mathbf{v}_j)$ 对应的坐标点就是 $z_{i,j}$ 。换句话说， $z_{i,j}$ 是 $\mathbf{x}^{(i)}$ 在 $\text{span}(\mathbf{v}_j)$ 的像 (image)。 $\mathbf{x}^{(i)}$ 在 \mathbb{R}^D 空间是 D 维，在 $\text{span}(\mathbf{v}_j)$ 仅是 1 维。图 5 中，从左边 \mathbb{R}^D 空间到右侧 $\text{span}(\mathbf{v}_j)$ 投影是个降维过程，数据发生压缩。

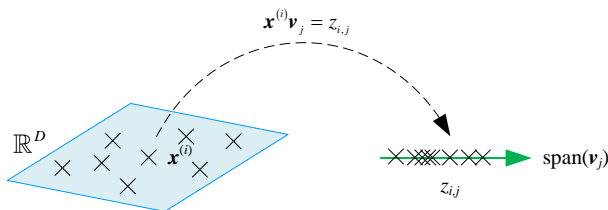


图 5. \mathbb{R}^D 空间数据点投影到 $\text{span}(\mathbf{v}_j)$

10.2 二次投影 + 层层叠加

本书上一章给出下面这个看似莫名其妙的矩阵乘法：

$$\mathbf{X} = \mathbf{X}\mathbf{I} = \mathbf{X}\mathbf{V}\mathbf{V}^T = \mathbf{X} \quad (9)$$

数据矩阵 \mathbf{X} 乘以单位阵 \mathbf{I} ，结果为 \mathbf{X} 其本身！这个显而易见的等式，有何意义？

其实，这个看似再简单不过的矩阵运算背后实际藏着“二次投影”和“层层叠加”这两重几何操作！下面，我们就解密这两个几何操作。

层层叠加

将 \mathbf{V} 写成 $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$ ，代入 (9) 得到：

$$\mathbf{X} = \mathbf{X}\mathbf{V}\mathbf{V}^T = \mathbf{X} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_D \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_D^T \end{bmatrix} = \underbrace{\mathbf{X}\mathbf{v}_1\mathbf{v}_1^T}_{\mathbf{x}_1} + \underbrace{\mathbf{X}\mathbf{v}_2\mathbf{v}_2^T}_{\mathbf{x}_2} + \dots + \underbrace{\mathbf{X}\mathbf{v}_D\mathbf{v}_D^T}_{\mathbf{x}_D} \quad (10)$$

令，

$$\mathbf{X}_j = \mathbf{X}\mathbf{v}_j\mathbf{v}_j^T \quad (11)$$

图 6 所示为上述运算， X_j 的形状和原数据矩阵 X 完全相同。我们称图 6 为二次投影，一会儿解释原因。

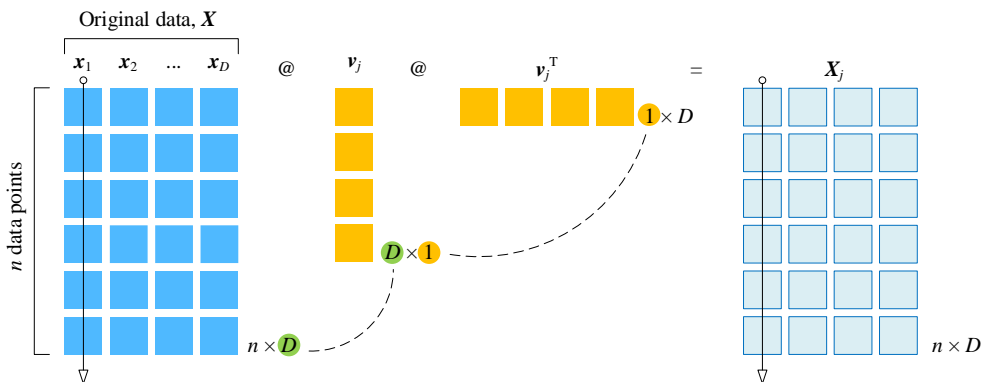


图 6. 二次投影

(10) 可以写成：

$$X = X_1 + X_2 + \dots + X_D \quad (12)$$

上式就是“层层叠加”。如图 7 所示， D 个形状完全相同的数据，层层叠加还原原始数据 X 。这本质上是矩阵乘法的第二视角。

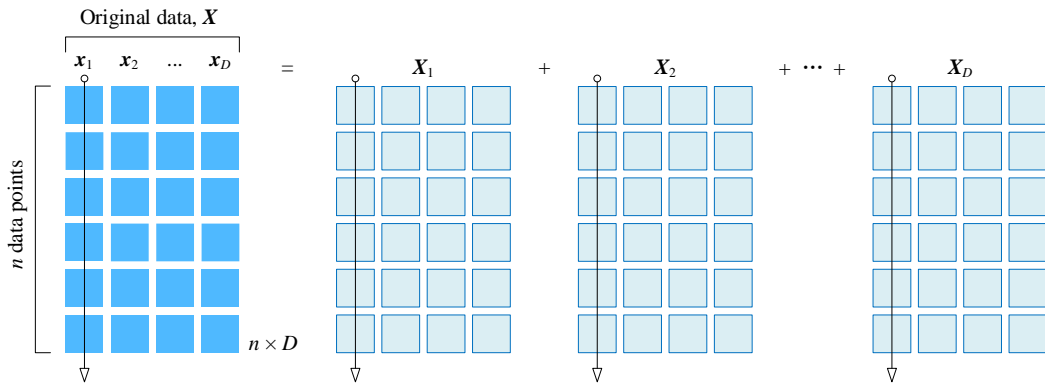


图 7. 层层叠加

二次投影

下面，我们专门聊聊“二次投影”。

取出 (11) X_j 中第 i 行行向量 $x_j^{(i)}$ ， $x_j^{(i)}$ 对应的运算为：

$$\mathbf{x}_j^{(i)} = \mathbf{x}^{(i)} \mathbf{v}_j \mathbf{v}_j^T = z_{i,j} \mathbf{v}_j^T \quad (13)$$

如 (8) 所示，上式中 $z_{i,j}$ 就是 $\mathbf{x}^{(i)}$ 正交投影到子空间 $\text{span}(\mathbf{v}_j)$ 对应的坐标点，这是第一次投影，具体过程如图 5 所示。

而 $z_{i,j} \mathbf{v}_j^T$ 得到的是 $z_{i,j}$ 在 \mathbb{R}^D 的坐标点，这便是第二次投影。

上述两次投影合并，得到所谓“二次投影”。整个二次投影的过程如图 8 所示。可以这样理解， $\mathbf{x}^{(i)} \rightarrow z_{i,j}$ 代表“标量投影”； $\mathbf{x}^{(i)} \rightarrow \mathbf{x}^{(i)} \mathbf{v}_j \mathbf{v}_j^T$ 则是“向量投影”。图 8 这个过程显然不可逆，方阵 $\mathbf{v}_j \mathbf{v}_j^T$ 的秩为 1，因此不可逆。

▲ 注意，图 8 中 $\mathbf{x}^{(i)}$ 和 $z_{i,j} \mathbf{v}_j^T$ 都用行向量表达坐标点。这和本书第 23 章要介绍的行空间有直接联系。

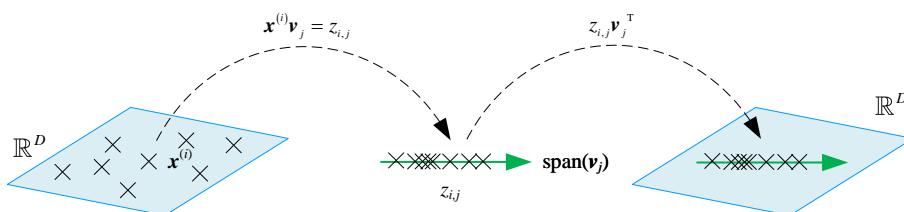


图 8. \mathbb{R}^D 空间数据点先投影到 $\text{span}(\mathbf{v}_j)$ ，再投影回到 \mathbb{R}^D

向量投影：张量积

将 (11) 写成张量积的形式：

$$\mathbf{X}_j = \mathbf{X} \mathbf{v}_j \otimes \mathbf{v}_j \quad (14)$$

\mathbf{X}_j 就是 \mathbf{X} 经过“降维”到子空间 $\text{span}(\mathbf{v}_j)$ 后，再投影到 \mathbb{R}^D 中得到的“像”。 \mathbf{X}_j 也是 \mathbf{X} 在 \mathbf{v}_j 上的向量投影。张量积 $\mathbf{v}_j \otimes \mathbf{v}_j$ 就是我们上一章提到的**投影矩阵** (projection matrix)。

张量积 $\mathbf{v}_j \otimes \mathbf{v}_j$ 本身完成“多维 \rightarrow 一维” + “一维 \rightarrow 多维”这两步映射。很显然，对于非 \mathbf{O} 矩阵 \mathbf{X} 来说，

$$\text{rank}(\mathbf{v}_j \otimes \mathbf{v}_j) = 1 \Rightarrow \text{rank}(\mathbf{X}_j) = 1 \quad (15)$$

所以，在 \mathbb{R}^D 空间中， \mathbf{X}_j 所有数据点在一条通过原点的直线上，直线和 \mathbf{v}_j 平行。也就是说，虽然 \mathbf{X}_j 表面上来看在 D 维空间 \mathbb{R}^D 中， \mathbf{X}_j 实际上只有 1 个维度， $\text{rank}(\mathbf{X}_j) = 1$ 。

利用张量积，(10) 可以写成：

$$\mathbf{X} = \underbrace{\mathbf{X} \mathbf{v}_1 \otimes \mathbf{v}_1}_{\mathbf{X}_1} + \underbrace{\mathbf{X} \mathbf{v}_2 \otimes \mathbf{v}_2}_{\mathbf{X}_2} + \cdots + \underbrace{\mathbf{X} \mathbf{v}_D \otimes \mathbf{v}_D}_{\mathbf{X}_D} \quad (16)$$

可以这样理解上式， X 分别二次投影（向量投影）到规范正交基 $[v_1, v_2, \dots, v_D]$ 每个列向量 v_j 所代表的子空间 $\text{span}(v_j)$ 中，获得 $X_1, X_2 \dots X_D$ 。而 $X_1, X_2 \dots X_D$ 层层叠加还原原始数据 X 。

再进一步，根据 $V^T V = I$ ，我们知道：

$$I = v_1 \otimes v_1 + v_2 \otimes v_2 + \dots + v_D \otimes v_D \quad (17)$$

也就是说， $v_j \otimes v_j$ 层层叠加得到单位阵 I 。

此外， $i \neq j$ 时， $v_i \otimes v_i$ 和 $v_j \otimes v_j$ 这两个张量积的矩阵乘积为零矩阵 O ：

$$(v_i \otimes v_i) @ (v_j \otimes v_j) = v_i v_i^T v_j v_j^T = 0 v_i v_j^T = O \quad (18)$$

标准正交基：便于理解

标准正交基是特殊的规范正交基。为了方便理解，我们用标准正交基 $[e_1, e_2, \dots, e_D]$ 替换 (16) 中的 $[v_1, v_2, \dots, v_D]$ ，得到：

$$X = X e_1 \otimes e_1 + X e_2 \otimes e_2 + \dots + X e_D \otimes e_D \quad (19)$$

展开 (19) 中等式右侧第一项得到：

$$X_1 = X e_1 \otimes e_1 = X \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} x_1 & x_2 & \dots & x_D \end{bmatrix}}_X \begin{bmatrix} 1 & 0 & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{bmatrix} = \begin{bmatrix} x_1 & 0 & \dots & 0 \end{bmatrix} \quad (20)$$

$X e_1$ 得到的是 X 的每一行在 $\text{span}(e_1)$ 这个子空间的坐标，即 x_1 。而 $X e_1 \otimes e_1$ 告诉我们的是 $X e_1$ 在 D 维空间 \mathbb{R}^D 中坐标值。

因此 (19) 右侧每一项 X_j 可以写成：

$$\begin{aligned} X_1 &= X e_1 \otimes e_1 = \begin{bmatrix} x_1 & 0 & \dots & 0 \end{bmatrix} \\ X_2 &= X e_2 \otimes e_2 = \begin{bmatrix} 0 & x_2 & \dots & 0 \end{bmatrix} \\ &\vdots \\ X_D &= X e_D \otimes e_D = \begin{bmatrix} 0 & 0 & \dots & x_D \end{bmatrix} \end{aligned} \quad (21)$$

也就是说， $X e_j \otimes e_j$ 仅保留 X 的第 j 列 x_j ，其他位置元素置 0。

因此，(19) 可以写成：

$$X = \underbrace{\begin{bmatrix} x_1 & 0 & \dots & 0 \end{bmatrix}}_{x_1} + \underbrace{\begin{bmatrix} 0 & x_2 & \dots & 0 \end{bmatrix}}_{x_2} + \dots + \underbrace{\begin{bmatrix} 0 & 0 & \dots & x_D \end{bmatrix}}_{x_D} \quad (22)$$

图 9 所示为上式“二次投影”与“层层叠加”过程。

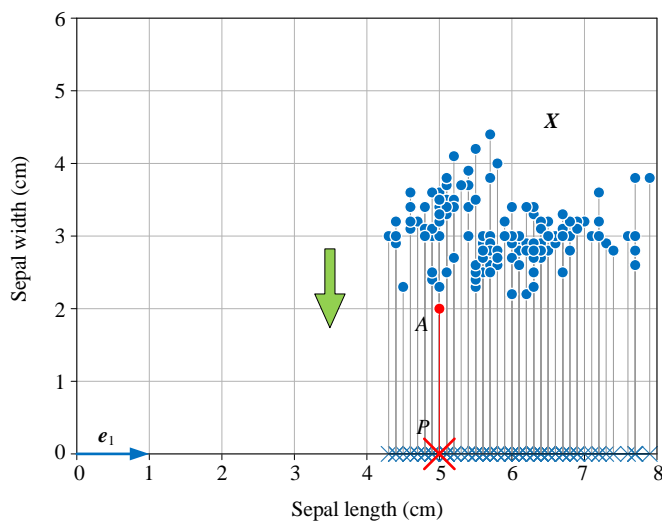
看到这里，有些读者怕是已经晕头转向。下面利用鸢尾花数据集做例子，帮大家更直观理解本节内容。

本节以二特征矩阵为例讲解何谓“二次投影”和“层层叠加”。数据矩阵 $X_{150 \times 2}$ 选取鸢尾花数据集前两列——花萼长度、花萼宽度，这样数据矩阵 $X_{150 \times 2}$ 的形状为 150×2 。投影的方向为标准正交基 $[e_1, e_2]$ 。

如图 10 所示, $\mathbf{X}_{150 \times 2}$ 向水平方向标量投影, 即 $\mathbf{X}_{150 \times 2}$ 向 \mathbf{e}_1 投影。以图中红点 A 为例, A 的坐标为 $(5, 2)$, 它在 \mathbf{e}_1 方向上的标量投影对应 A 在横轴坐标:

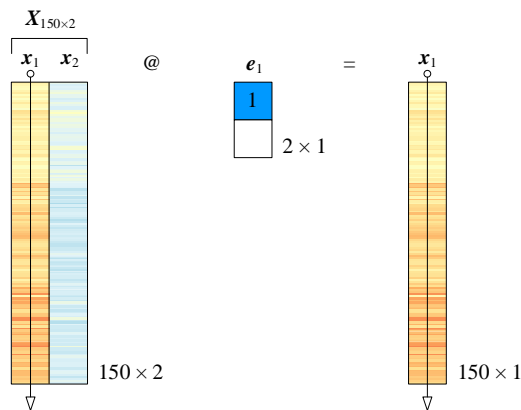
$$[5 \quad 2] \underset{e_1}{\begin{bmatrix} 1 \\ 0 \end{bmatrix}} = 5 \quad (24)$$

⚠ 注意，5 代表的是 A 在 $\text{span}(e_1)$ 空间中的坐标值，而 $\text{span}(e_1)$ 显然为一维空间。

图 10. 二特征数据矩阵 $X_{150 \times 2}$ 向 e_1 投影，一次投影

如图 11 热图所示， $X_{150 \times 2}$ 向 e_1 投影结果为列向量 x_1 ，相当于保留了 $X_{150 \times 2}$ 第一列数据：

$$z_1 = X e_1 = x_1 \quad (25)$$

图 11. 数据热图，二特征数据矩阵 $X_{150 \times 2}$ 向 e_1 投影，一次投影 (标量投影)

大家可能会好奇，既然图 10 中 $X_{150 \times 2}$ 向水平方向投影结果都可以画在图 10 直角坐标系中，在二维空间 $\mathbb{R}^2 = \text{span}(e_1, e_2)$ 中，这些投影点一定有其二维坐标值。

很明显，以 A 为例， A 在横轴投影点 P 在 $\mathbb{R}^2 = \text{span}(e_1, e_2)$ 的坐标值为 $(5, 0)$ 。这个结果是怎么得到的？

这就用到了本章前文讲到的“二次投影”，相当于在 (24) 基础上再次投影。第二次投影相当于“升维”，从一维升到二维。

以点 A 为例，“二次投影”对应的计算为：

$$[5 \ 2]e_1 \otimes e_1 = [5 \ 2]e_1 e_1^T = [5 \ 2] \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = [5 \ 0] \quad (26)$$

上式对应的计算如图 12 所示。

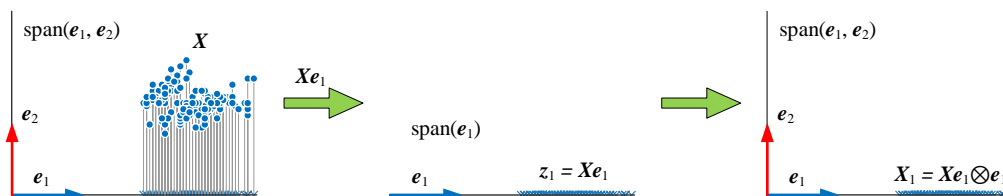


图 12. 二特征数据矩阵 X 向 e_1 投影，二次投影

X 在 e_1 二次投影对应 $\mathbb{R}^2 = \text{span}(e_1, e_2)$ 坐标值为 X_1 ：

$$X_1 = Xe_1 \otimes e_1 = Xe_1 e_1^T = X \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = [x_1 \ 0] \quad (27)$$

图 13 所示为上述运算对应热图。

很容易判断，(27) 上式中 $e_1 \otimes e_1$ 的行列式值为 0，即 $\det(e_1 \otimes e_1) = 0$ 。也就是说这个映射过程存在降维，映射矩阵 $e_1 \otimes e_1$ 不存在逆，即几何操作不可逆。

⚠ 值得注意的是，从 x_1 到 $X_1 = [x_1, 0]$ 这种“升维”只是名义上的维度提高，不代表数据信息增多。显然，上式中 X_1 的秩仍为 1，即 $\text{rank}(X_1)$ 。举个形象点的例子，我们给桌面上马克杯拍了张照片，再把照片平放在桌面上。马克杯本身就是 X ，桌面上的照片就是 X_1 。

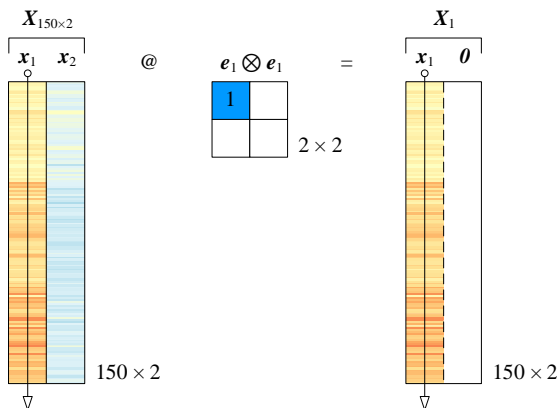


图 13. 数据热图，二特征数据矩阵 $X_{150 \times 2}$ 向 e_1 投影，二次投影

竖直方向投影

如图 14 所示, $X_{150 \times 2}$ 向竖直方向投影, 即 $X_{150 \times 2}$ 向 e_2 投影。还是以 A 点为例, $A(5, 2)$ 在 e_2 方向上的标量投影为:

$$\begin{bmatrix} 5 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2 \quad (28)$$

e_2

2 代表的是 A 在 $\text{span}(e_2)$ 空间中的坐标值, $\text{span}(e_2)$ 同样为一维空间。图 15 为上述运算的热图。

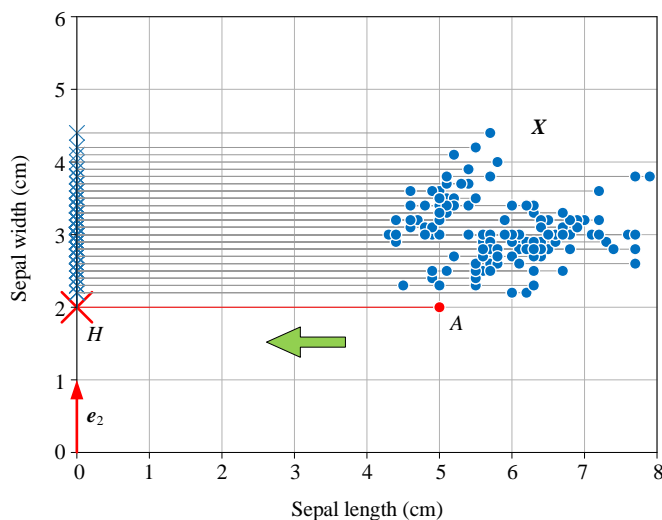


图 14. 二特征数据矩阵 $X_{150 \times 2}$ 向 e_2 方向标量投影, 一次投影

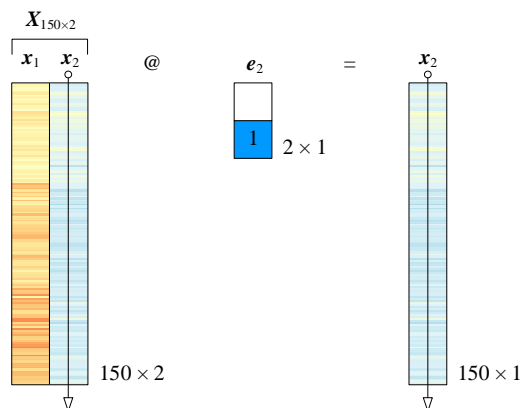


图 15. 数据热图, 二特征数据矩阵 $X_{150 \times 2}$ 向 e_2 投影, 一次投影

同样利用“二次投影”, 得到 A 在竖直方向投影点 H 在 $\text{span}(e_1, e_2)$ 的坐标值为 $(0, 2)$:

$$[5 \ 2]e_2 \otimes e_2 = [5 \ 2]e_2 e_2^T = [5 \ 2] \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = [0 \ 2] \quad (29)$$

上式对应的计算如图 16 所示。

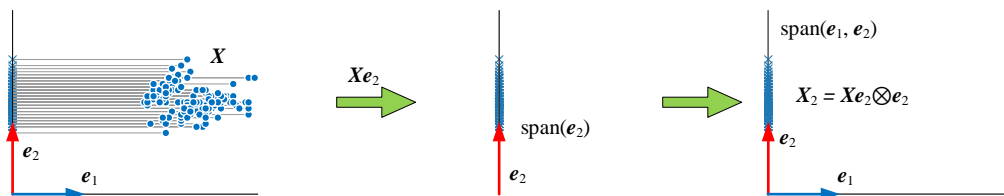


图 16. 二特征数据矩阵 $X_{150 \times 2}$ 向 e_2 方向标量投影，二次投影

$X_{150 \times 2}$ 在 e_2 二次投影得到矩阵 X_2 :

$$X_2 = Xe_2 \otimes e_2 = Xe_2 e_2^T = X \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (30)$$

上式对应的热图运算为图 17。 X_2 第一列向量为 $\mathbf{0}$ ，第二列向量为 x_2 。

(30) 中 $e_2 \otimes e_2$ 的行列式值为 0，即 $\det(e_2 \otimes e_2) = 0$ 。

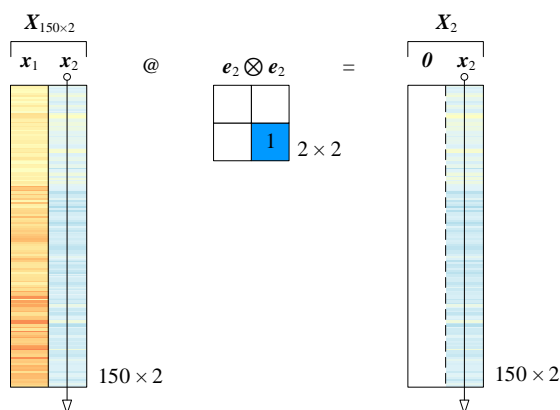


图 17. 数据热图，二特征数据矩阵 $X_{150 \times 2}$ 向 e_2 投影，二次投影

叠加

如图 18 所示，以 A 为例， $P(5, 0)$ 和 $H(0, 2)$ 叠加得到点 A 坐标 $(5, 2)$ 。这相当于两个向量合成，即：

$$\begin{bmatrix} 5 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix} \quad (31)$$

或者以行向量来表示，

$$\begin{bmatrix} 5 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 2 \end{bmatrix} \quad (32)$$

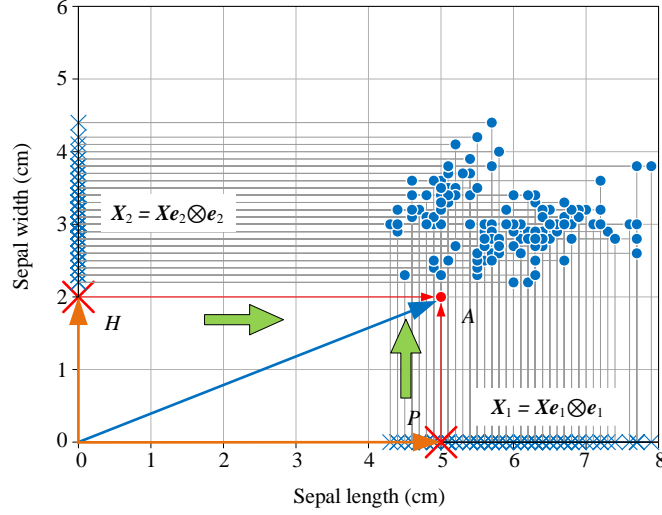


图 18. 数据叠加还原散点图

如图 19 所示， X_1 和 X_2 叠加还原 $X_{150 \times 2}$ ：

$$\begin{aligned} X_{150 \times 2} &= X_1 + X_2 \\ &= X(e_1 \otimes e_1 + X e_2 \otimes e_2) \\ &= X(e_1 e_1^T + e_2 e_2^T) \\ &= X \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = XI \end{aligned} \quad (33)$$

图 20 所示为上述运算对应的热图。

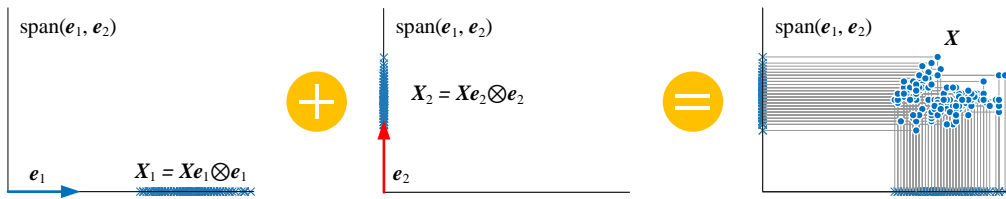
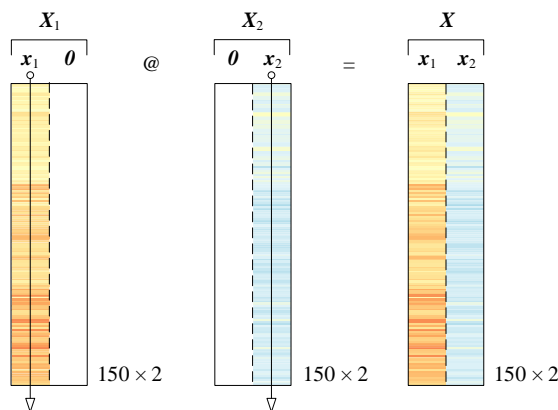


图 19. 数据叠加还原 $X_{150 \times 2}$

图 20. 数据热图，叠加还原 $X_{150 \times 2}$

10.4 二特征数据投影：规范正交基

本节分析 $X_{150 \times 2}$ 在三个不同规范正交基投影情况。

第一个规范正交基

给定如下规范正交基 $V = [v_1, v_2]$:

$$V = [v_1 \ v_2] = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \quad (34)$$

从几何变换角度来看， V 就是一个旋转矩阵。请大家自行验证 $V^T V = I$ 。此外，很容易计算得到 V 的行列式值为 1，即 $\det(V) = 1$ 。也就是说，旋转不改变面积。

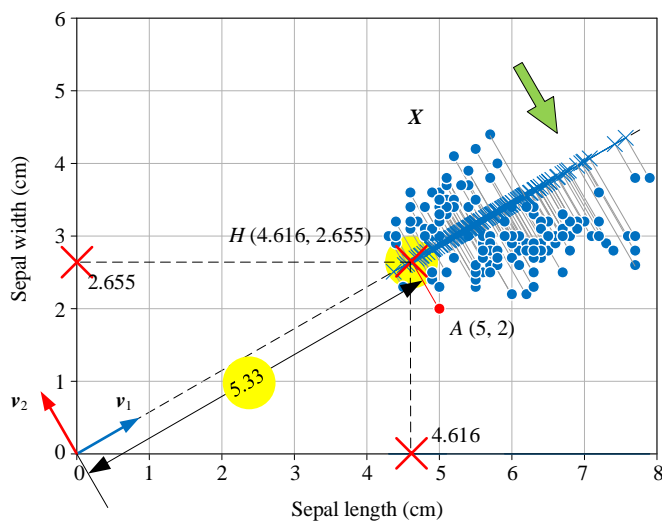
v_1 和 v_2 也相当于 e_1 和 e_2 的线性组合，即：

$$\begin{aligned} v_1 &= \sqrt{3}/2 e_1 + 1/2 e_2 \\ v_2 &= -1/2 e_1 + \sqrt{3}/2 e_2 \end{aligned} \quad (35)$$

如图 21 所示，同样以点 $A(5, 2)$ 为例， A 在 v_1 方向标量投影为：

$$\begin{bmatrix} 5 & 2 \end{bmatrix} \underbrace{\begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}}_{v_1} \approx 5.33 \quad (36)$$

也就是说， A 在 $\text{span}(v_1)$ 投影点 H 的坐标值为 5.33，对应向量可以写成 $5.33v_1$ 。

图 21. 二特征数据矩阵 $X_{150 \times 2}$ 向 v_1 投影

通过二次投影获得 H 在 $\text{span}(e_1, e_2)$ 坐标值：

$$[5 \ 2]v_1 \otimes v_1 = [5 \ 2]v_1 v_1^T = [5 \ 2] \begin{bmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{bmatrix} \approx [4.616 \ 2.665] \quad (37)$$

这就是 H 在图 21 中坐标值。很容易计算，(37) 中 $v_1 \otimes v_1$ 的行列式值为 0，即 $\det(v_1 \otimes v_1) = 0$ 。

数据矩阵 $X_{150 \times 2}$ 在 v_1 投影 z_1 为：

$$z_1 = Xv_1 = \underbrace{\begin{bmatrix} x_1 & x_2 \end{bmatrix}}_X \underbrace{\begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}}_{v_1} \approx 0.866x_1 + 0.5x_2 \quad (38)$$

观察上式发现， z_1 相当于 x_1 和 x_2 的线性组合。请大家关注一下单位， x_1 和 x_2 的单位均为厘米，因此上式线性组合结果的单位还是厘米。

如果， x_1 和 x_2 的分别代表身高、体重数据，单位为米、公斤。这种情况， x_1 和 x_2 线性组合结果的单位就“尴尬”。因此，对于单位不统一的矩阵，可以考虑先通过标准化“去单位”。

$X_{150 \times 2}$ 在 v_1 二次投影结果 X_1 为：

$$X_1 = Xv_1 \otimes v_1 = Xv_1 v_1^T \approx \underbrace{\begin{bmatrix} x_1 & x_2 \end{bmatrix}}_X \begin{bmatrix} 0.750 & 0.433 \\ 0.433 & 0.250 \end{bmatrix} = [0.750x_1 + 0.433x_2 \quad 0.433x_1 + 0.250x_2] \quad (39)$$

而 X_1 的两个列向量都存在如下倍数关系，因此 X_1 的秩为 1：

$$X_1 \approx [0.866 \times (0.866x_1 + 0.5x_2) \quad 0.5 \times (0.866x_1 + 0.5x_2)] \quad (40)$$

如图 21 所示， X_1 所有点在一条通过原点的直线上。这条直线等价于 $\text{span}(v_1)$ 。

如图 22 所示，同样以点 $A(5, 2)$ 为例， A 在 v_2 方向标量投影结果为：

$$\begin{bmatrix} 5 & 2 \end{bmatrix} \underbrace{\begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix}}_{v_2} \approx -0.7679 \quad (41)$$

即 A 在 $\text{span}(v_2)$ 投影点的坐标值为 -0.7679 ，对应向量可以写成 $-0.7679v_2$ 。通过二次投影获得投影点坐标值 (图 22 中 \times)：

$$\begin{bmatrix} 5 & 2 \end{bmatrix} v_2 \otimes v_2 = \begin{bmatrix} 5 & 2 \end{bmatrix} v_2 v_2^T = \begin{bmatrix} 5 & 2 \end{bmatrix} \begin{bmatrix} 1/4 & -\sqrt{3}/4 \\ -\sqrt{3}/4 & 3/4 \end{bmatrix} \approx \begin{bmatrix} 0.384 & -0.665 \end{bmatrix} \quad (42)$$

(42) 中 $v_2 \otimes v_2$ 的行列式值为 0，即 $\det(v_2 \otimes v_2) = 0$ 。

(37) 和 (42) 之和还原 A 坐标值 $(5, 2)$ ：

$$\begin{bmatrix} 5 & 2 \end{bmatrix} (v_1 \otimes v_1 + v_2 \otimes v_2) = \begin{bmatrix} 5 & 2 \end{bmatrix} \left\{ \begin{bmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{bmatrix} + \begin{bmatrix} 1/4 & -\sqrt{3}/4 \\ -\sqrt{3}/4 & 3/4 \end{bmatrix} \right\} = \begin{bmatrix} 5 & 2 \end{bmatrix} \quad (43)$$

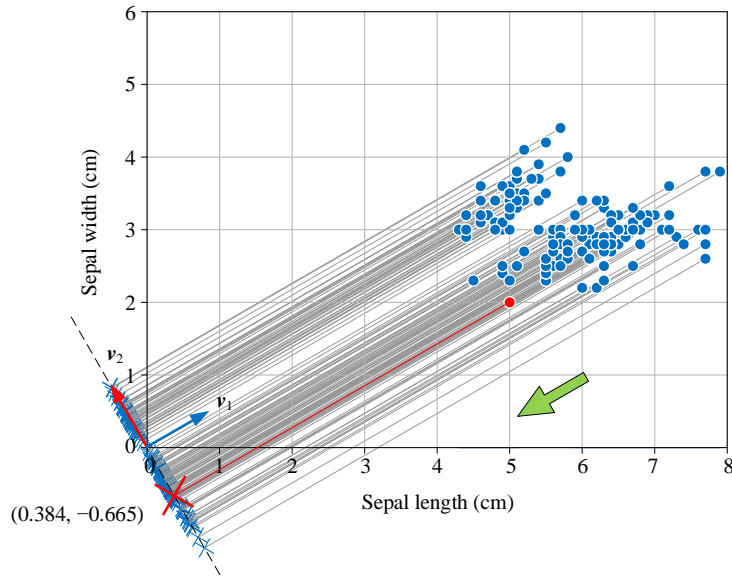


图 22. 二特征数据矩阵 $X_{150 \times 2}$ 向 v_2 投影

$X_{150 \times 2}$ 在 v_2 投影 z_2 为：

$$z_2 = Xv_2 = \underbrace{\begin{bmatrix} x_1 & x_2 \end{bmatrix}}_X \underbrace{\begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix}}_{v_2} \approx -0.5x_1 + 0.866x_2 \quad (44)$$

z_2 也是 x_1 和 x_2 的线性组合。

$X_{150 \times 2}$ 在 v_2 二次投影 X_2 为：

$$\mathbf{X}_2 = \mathbf{X}\mathbf{v}_2 \otimes \mathbf{v}_2 = \mathbf{X}\mathbf{v}_2\mathbf{v}_2^T \approx \underbrace{\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix}}_{\mathbf{X}} \begin{bmatrix} 0.250 & -0.433 \\ -0.433 & 0.750 \end{bmatrix} = \begin{bmatrix} 0.250\mathbf{x}_1 - 0.433\mathbf{x}_2 & -0.433\mathbf{x}_1 + 0.750\mathbf{x}_2 \end{bmatrix} \quad (45)$$

\mathbf{X}_2 的秩也为 1。如图 22 所示， \mathbf{X}_2 对应的坐标也在一条通过原点的直线上。

(39) 和 (45) 叠加还原 \mathbf{X} ：

$$\mathbf{X}_1 + \mathbf{X}_2 = \mathbf{X}\mathbf{v}_1 \otimes \mathbf{v}_1 + \mathbf{X}\mathbf{v}_2 \otimes \mathbf{v}_2 = \mathbf{X} \left\{ \begin{bmatrix} 0.750 & 0.433 \\ 0.433 & 0.250 \end{bmatrix} + \begin{bmatrix} 0.250 & -0.433 \\ -0.433 & 0.750 \end{bmatrix} \right\} = \mathbf{X}\mathbf{I} = \mathbf{X} \quad (46)$$

顺便提一嘴，对于 2×2 方阵 \mathbf{A} 和 \mathbf{B} ， $\mathbf{A} + \mathbf{B}$ 行列式值存在如下关系：

$$\det(\mathbf{A} + \mathbf{B}) = \det(\mathbf{A}) + \det(\mathbf{B}) + \text{tr}(\mathbf{A})\text{tr}(\mathbf{B}) - \text{tr}(\mathbf{AB}) \quad (47)$$

请大家将 $\mathbf{v}_1 \otimes \mathbf{v}_1$ 和 $\mathbf{v}_2 \otimes \mathbf{v}_2$ 代入上式验证。

第二个规范正交基

给定如下规范正交基 $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2]$ ：

$$\mathbf{W} = [\mathbf{w}_1 \quad \mathbf{w}_2] = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \quad (48)$$

图 23 和图 24 所示为二特征数据矩阵 $\mathbf{X}_{150 \times 2}$ 向 \mathbf{w}_1 和 \mathbf{w}_2 投影。请按照本节之前分析 \mathbf{V} 的逻辑，自行分析数据在 \mathbf{W} 中的投影，并计算 \mathbf{W} 的行列式值。

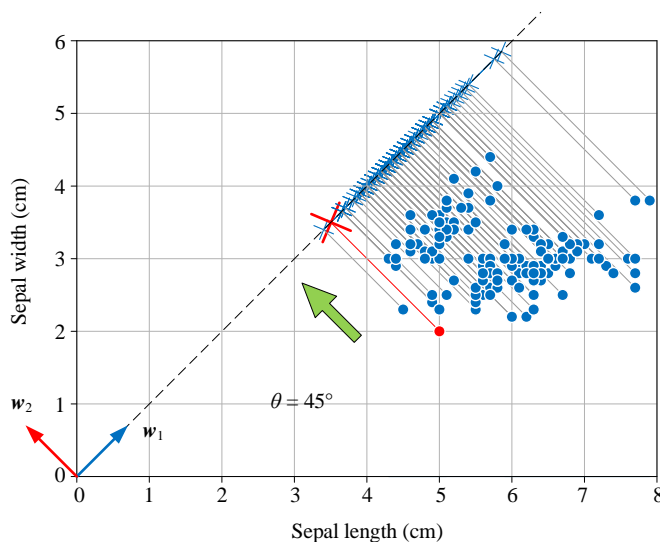
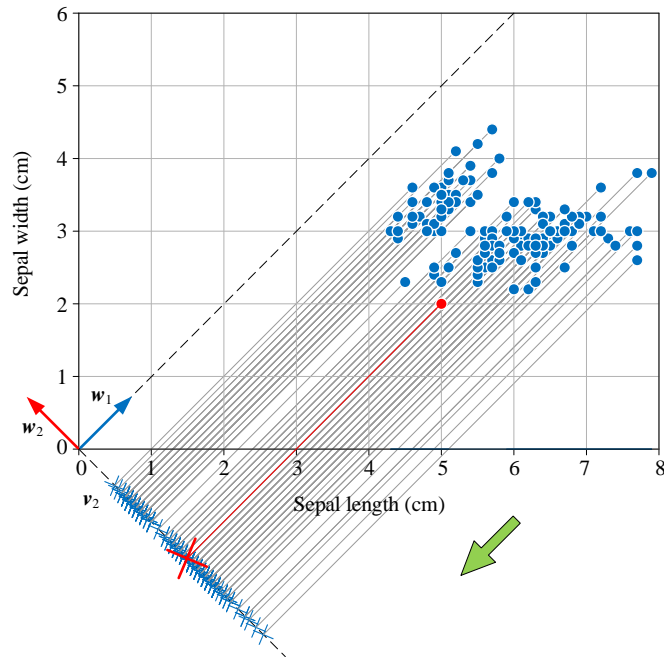


图 23. 二特征数据矩阵 $\mathbf{X}_{150 \times 2}$ 向 \mathbf{w}_1 投影

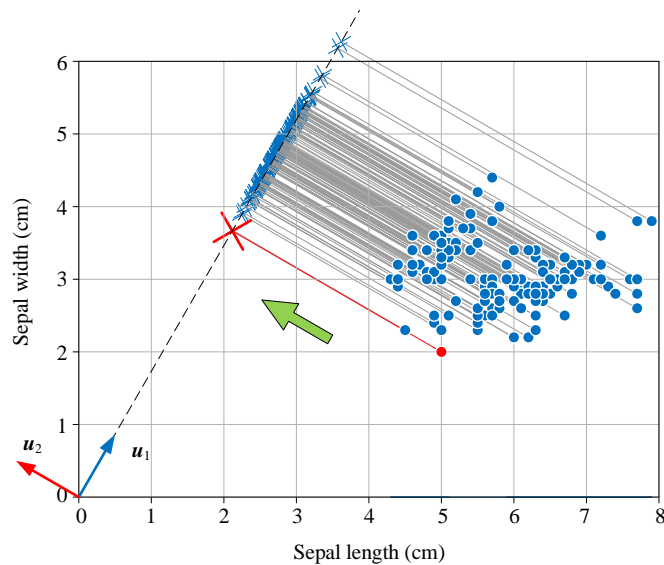
图 24. 二特征数据矩阵 $X_{150 \times 2}$ 向 w_2 投影

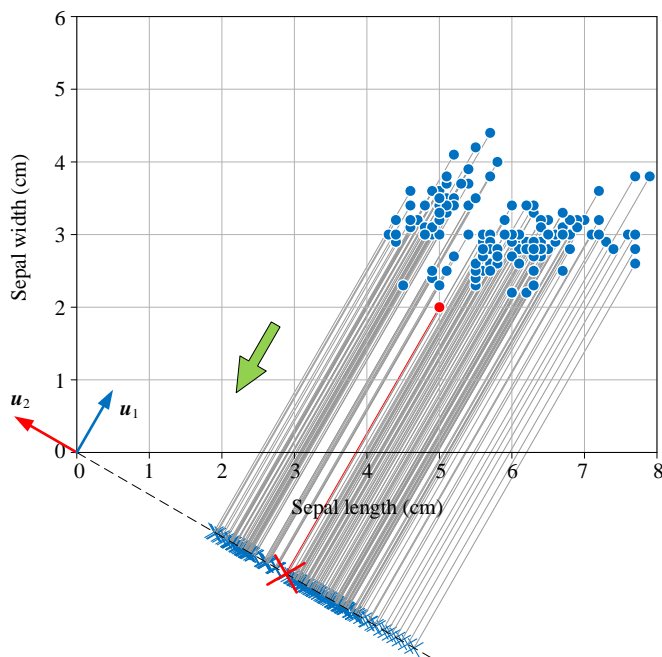
第三个规范正交基

给定如下规范正交基 $U = [u_1, u_2]$:

$$U = [u_1 \quad u_2] = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \quad (49)$$

图 25 和图 26 所示为二特征数据矩阵 $X_{150 \times 2}$ 向 u_1 和 u_2 投影。请大家分析数据在 U 中的投影，并计算 U 的行列式值。

图 25. 二特征数据矩阵 $X_{150 \times 2}$ 向 u_1 投影

图 26. 二特征数据矩阵 $X_{150 \times 2}$ 向 u_2 投影

旋转角度连续变化

前文提过，在 \mathbb{R}^2 中不同规范正交基之间仅差在旋转角度上。比较图 21 ~ 图 26 这六幅图，当旋转角度连续变化时，投影结果 z_1 和 z_2 也会连续变化。给出如下更具一般性的矩阵 V ：

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

其中， θ 代表逆时针旋转角度。 $Z = XV$ 可以展开写成：

$$\begin{bmatrix} z_1 & z_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} = \begin{bmatrix} \cos \theta x_1 - \sin \theta x_2 & \sin \theta x_1 + \cos \theta x_2 \end{bmatrix} \quad (51)$$

对于上式中 z_1 和 z_2 ，我们可以分析它们各自的向量模，也可以计算 z_1 和 z_2 之间的向量夹角余弦值、夹角弧度、角度等。

从统计视角来看， z_1 和 z_2 代表两列数值，我们可以分析它们各自的均值、方差、标准差，也可以计算 z_1 和 z_2 的协方差、相关性系数。

而上述这些量值都随着 θ 变化而连续变化。有变化就有最大值、最小值，就有优化问题。本书后续介绍的特征值分解和奇异值分解背后都离不开优化视角。这是本书第 18 章要讨论的话题。

10.5 四特征数据投影：标准正交基

本章最后两节以四特征数据矩阵为例，扩展前文分析案例。本节先从最简单的标准正交基 $[e_1, e_2, \dots, e_D]$ 入手。

一次投影：标量投影

前文提到过，一次投影实际上就是“标量投影”。图 27 (a) 所示为鸢尾花数据集矩阵 X 在 e_1 方向上标量投影的运算热图。

从行向量角度来看， $x^{(i)}e_1 \rightarrow x_{i,1}$ 代表 \mathbb{R}^D 空间坐标值 $x^{(i)}$ 投影到 $\text{span}(e_1)$ 这个子空间后，坐标值为 $x_{i,1}$ 。

▲ 再次强调，向量空间 $\text{span}(e_1)$ 维度为 1。 $x_{i,1}$ 是 $x^{(i)}$ 在 $\text{span}(e_1)$ 的坐标值。

从列向量角度来看， $[x_1, x_2, x_3, x_4]e_1 \rightarrow x_1$ ，是一个线性组合过程。而 $e_1 = [1, 0, 0, 0]^T$ ，线性组合结果只保留了鸢尾花数据集第一列 x_1 ，即花萼长度。

请大家按照这个思路分析图 27 (b)、(c)、(d) 三幅热图运算。请大家思考，要是想计算鸢尾花花萼长度和花萼宽度之和，用矩阵乘法怎样完成？

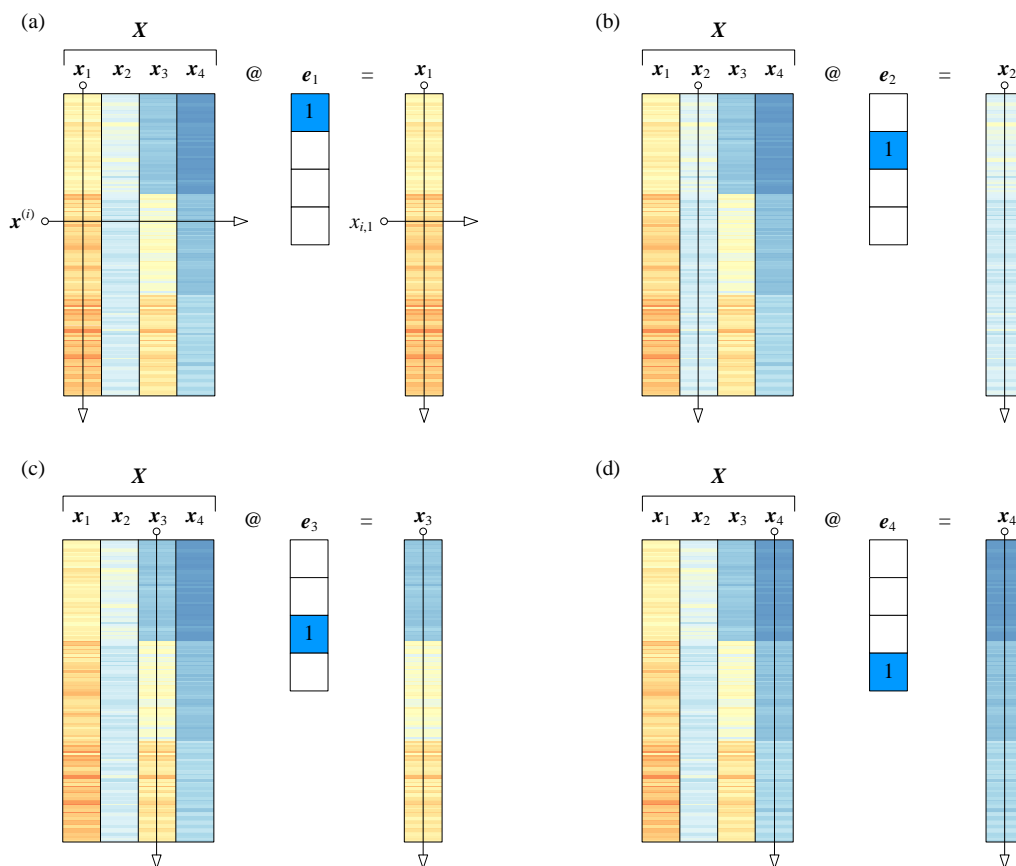


图 27. 四特征数据矩阵 $X_{150 \times 4}$ 分别向 e_1 、 e_2 、 e_3 、 e_4 投影，一次投影

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

二次投影

如前文所述，本章所谓的“二次投影”实际上就是向量投影。如图 28 所示， X 向 e_1 方向向量投影结果就是 X 和 $e_1 \otimes e_1$ 的矩阵乘积。乘积结果是，只保留鸢尾花数据集第一列——花萼长度，其他数据均置 0。请大家按照这个思路自行分析图 29、图 30、图 31。此外，容易计算 $e_1 \otimes e_1$ 、 $e_2 \otimes e_2$ 、 $e_3 \otimes e_3$ 、 $e_4 \otimes e_4$ 的行列式值都为 0。

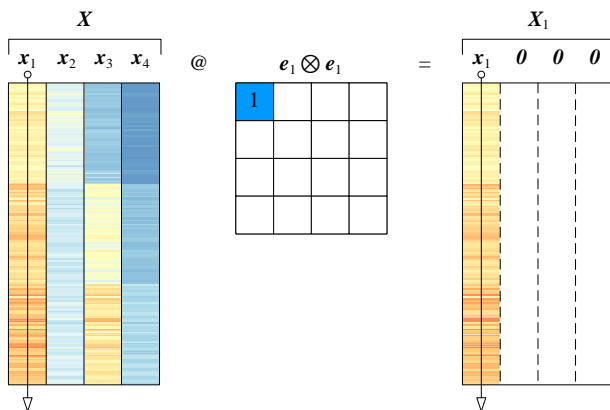


图 28. 四特征数据矩阵 $X_{150 \times 4}$ 向 e_1 方向向量投影，二次投影

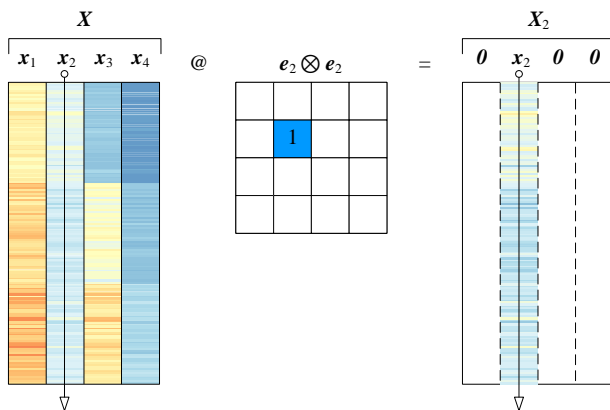


图 29. 四特征数据矩阵 $X_{150 \times 4}$ 向 e_2 方向向量投影，二次投影

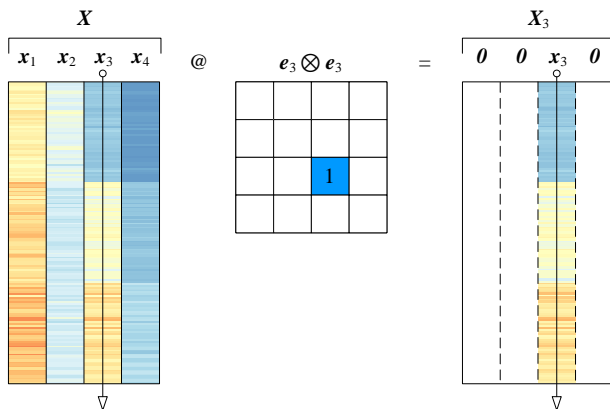
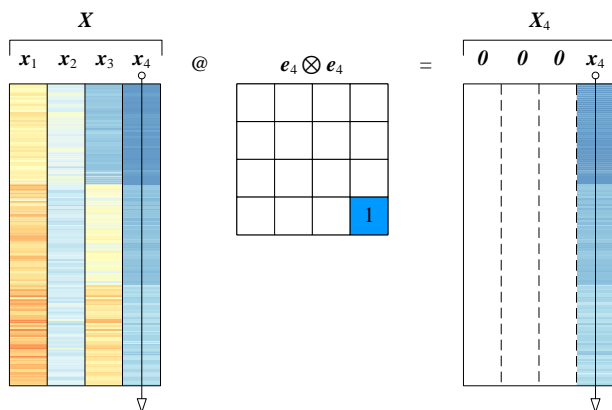


图 30. 四特征数据矩阵 $X_{150 \times 4}$ 向 e_3 方向向量投影，二次投影图 31. 四特征数据矩阵 $X_{150 \times 4}$ 向 e_4 方向向量投影，二次投影

向平面投影

本节之前提到的都是向单一方向投影。下面，我们用一个例子说明向某个二维向量空间投影。

如图 32 所示， X 向 $[e_1, e_2]$ 基底张成的向量空间标量投影，这个过程也相当于降维，从 4 维降到 2 维，只保留了鸢尾花花萼长度、花萼宽度两个特征。

本书第 1 章介绍过成对特征散点图，请大家思考如何用矩阵乘法运算获得每幅散点图数据矩阵。

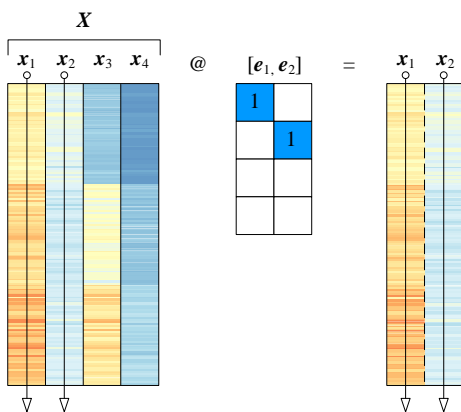
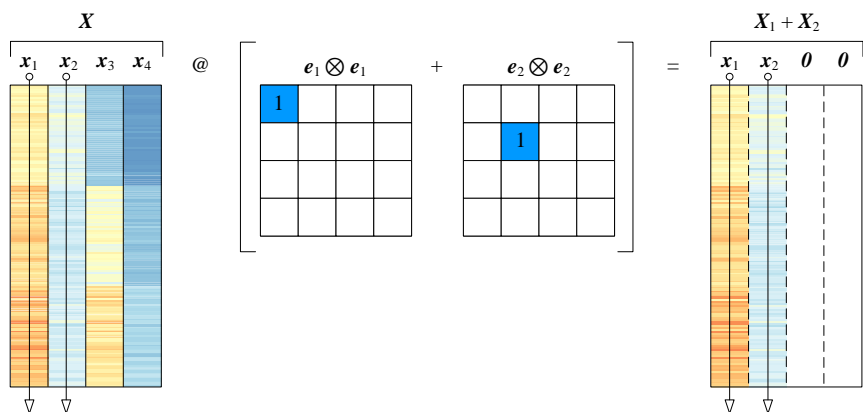
图 32. 四特征数据矩阵 $X_{150 \times 4}$ 向 $[e_1, e_2]$ 方向标量投影

图 33 所示为 X 向 $[e_1, e_2]$ 基底张成的向量空间向量投影，结果相当于图 28 和图 29 结果“叠加”，即 $X_1 + X_2$ 。很明显， $X_1 + X_2$ 并没有还原 X 。

图 33. 四特征数据矩阵 $X_{150 \times 4}$ 向 $[e_1, e_2]$ 方向向量投影

层层叠加：还原原始矩阵

本章前文 (12) 告诉我们，数据矩阵 X 在规范正交基 $[v_1, v_2, \dots, v_D]$ 中每个方向上向量投影层层叠加可以完全还原原始数据。而标准正交基 $[e_1, e_2, \dots, e_D]$ 可以视作特殊的规范正交基。

观察图 34 得知，要想完整还原 X ，需要图 28、图 29、图 30、图 31 四幅热图叠加，即 $X = X_1 + X_2 + X_3 + X_4$ 。显然， X_1, X_2, X_3, X_4 这四个矩阵的秩都是 1。

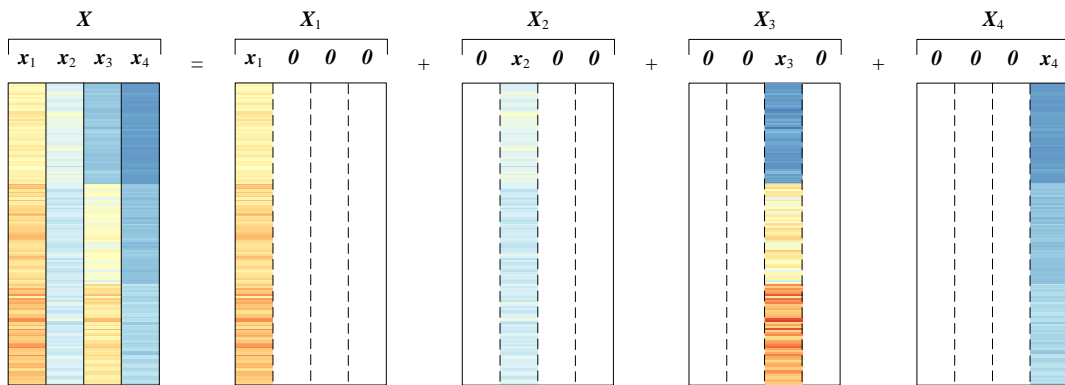
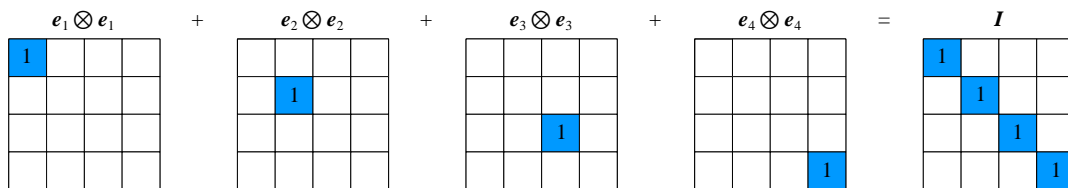
图 34. 投影数据矩阵的层层叠加还原数据矩阵 $X_{150 \times 4}$

图 35 是张量积层层叠加得到单位矩阵 I ，它是数据还原的另外一个视角：

$$e_1 \otimes e_1 + e_2 \otimes e_2 + e_3 \otimes e_3 + e_4 \otimes e_4 = I \quad (52)$$

图 35. 张量积的层层叠加还原 4×4 单位矩阵

10.6 四维数据投影：规范正交基

有了上一节内容作为基础，这一节提高难度，我们用一个规范正交基重复上一节所有计算。大家阅读这一节时，请对比上一节内容。

某个“无数里挑一”的规范正交基

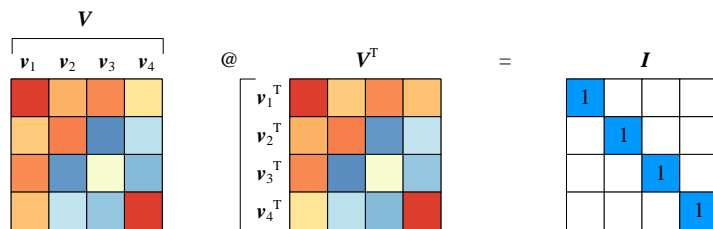
我们恰好找到了一个 4×4 规范正交基 V ，具体如下：

$$V = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3 \quad \mathbf{v}_4] = \begin{bmatrix} 0.751 & 0.284 & 0.502 & 0.321 \\ 0.380 & 0.547 & -0.675 & -0.317 \\ 0.513 & -0.709 & -0.059 & -0.481 \\ 0.168 & -0.344 & -0.537 & 0.752 \end{bmatrix} \quad (53)$$

大家可能好奇我们怎么找到这个 V ，本章后面会揭晓答案。

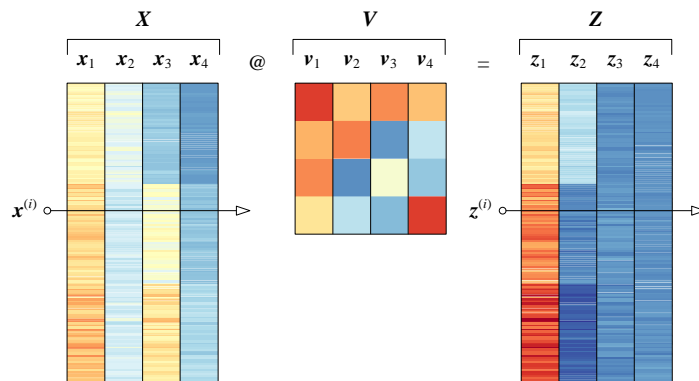
图 36 所示为规范正交基 V 乘其转置 V^T 得到单位矩阵。大家可以自己试着验算上式是否满足 $VV^T = I$ ，即方阵 V 每一列列向量都是单位向量，且 V 的列向量两两正交。上式， V 仅保留小数点后 3 位， VV^T 结果非常接近单位矩阵 I 。

从几何角度来看，规范正交基 V 对应的几何操作是四维空间旋转。

图 36. 规范正交基 V 乘其转置得到 4×4 单位矩阵

V 中的像

如图 37 所示，以为规范正交基 V 桥梁，矩阵乘法 $Z = XV$ 完成 X 到 Z 的映射。 Z 就是 X 在 V 中的像，根据 $X\mathbf{v}_j = \mathbf{z}_j$ ，下面逐一分析矩阵 Z 的列向量。

图 37. 四特征数据矩阵 $X_{150 \times 4}$ 投影到规范正交基 V 得到 Z

第 1 列向量 v_1

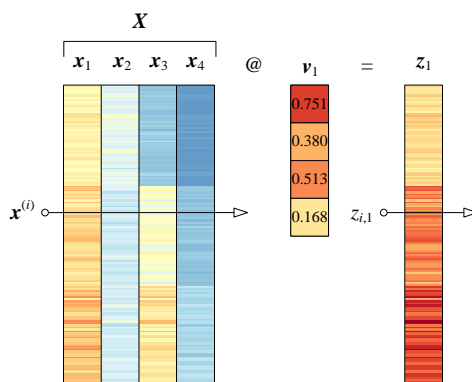
图 38 所示为鸢尾花数据集矩阵 X 在 v_1 方向上标量投影的运算热图。

从行向量角度来看, $x^{(i)}v_1 \rightarrow z_{i,1}$ 代表 \mathbb{R}^D 空间坐标值 $x^{(i)}$ 投影到 $\text{span}(v_1)$ 这个子空间后坐标值变成 $z_{i,1}$ 。

从列向量角度来看, $[x_1, x_2, x_3, x_4]v_1 \rightarrow z_1$, 是一个线性组合过程, 即:

$$z_1 = Xv_1 = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} \begin{bmatrix} 0.751 \\ 0.380 \\ 0.513 \\ 0.168 \end{bmatrix} = 0.751x_1 + 0.380x_2 + 0.513x_3 + 0.168x_4 \quad (54)$$

上式说明, 0.7512 倍 x_1 、0.380 倍 x_2 、0.513 倍 x_3 、0.168 倍 x_4 合成得到了向量 z_1 。

图 38. 四特征数据矩阵 $X_{150 \times 4}$ 向 v_1 方向标量投影, 一次投影

如图 39 所示, z_1 再乘 v_1^T , 便得到 X_1 。不难理解, X_1 的每一列都是 z_1 乘一个标量系数。也就是说, X_1 的四个列向量之间存在倍数关系, 即,

$$X_1 = z_1 v_1^T = z_1 \begin{bmatrix} 0.751 & 0.380 & 0.513 & 0.168 \end{bmatrix} = \begin{bmatrix} 0.751z_1 & 0.380z_1 & 0.513z_1 & 0.168z_1 \end{bmatrix} \quad (55)$$

显然, X_1 的秩为 1, 即 $\text{rank}(X_1) = 1$ 。

总结来说, 图 38 和图 39 用了两步完成了“二次投影”, 即向量投影。

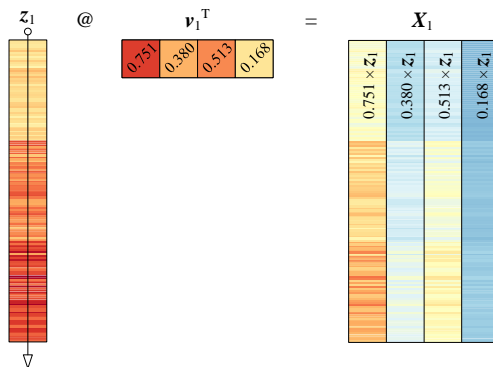


图 39. 四特征数据矩阵 $X_{150 \times 4} Z_1$ 乘 v_1^T 得到 X_1

下面, 我们用向量张量积方法完成同样的计算。

首先计算张量积 $v_1 \otimes v_1$:

$$v_1 \otimes v_1 = v_1 v_1^T = \begin{bmatrix} 0.751 \\ 0.380 \\ 0.513 \\ 0.168 \end{bmatrix} @ \begin{bmatrix} 0.751 & 0.380 & 0.513 & 0.168 \end{bmatrix}^T = \begin{bmatrix} 0.564 & 0.285 & 0.385 & 0.126 \\ 0.285 & 0.144 & 0.194 & 0.063 \\ 0.385 & 0.194 & 0.263 & 0.086 \\ 0.126 & 0.063 & 0.086 & 0.028 \end{bmatrix} \quad (56)$$

图 40 所示为上述运算热图。很容易发现, 张量积为对称矩阵。请大家自行计算张量积的秩是否为 1。

⚠ 注意, (56) 上式仅仅保留小数点后 3 位数值。

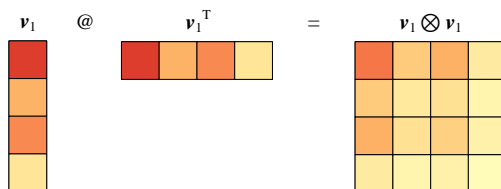
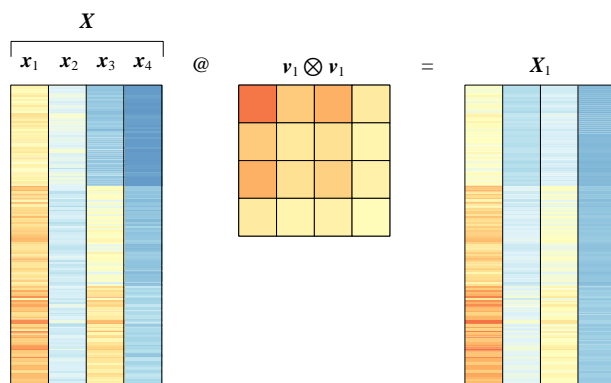


图 40. 计算张量积 $v_1 \otimes v_1$

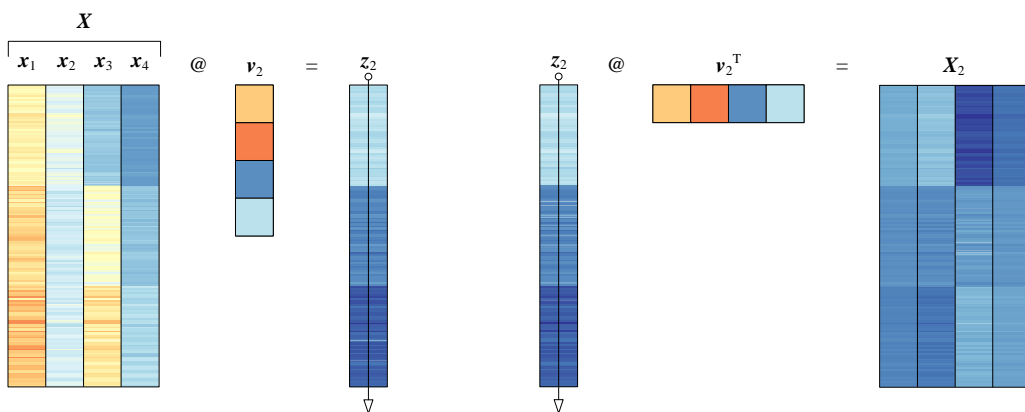
图 41 所示为 X 和张量积 $v_1 \otimes v_1$ 乘积。几何视角来看, X 向 v_1 向量投影得到 X_1 , 即所谓“二次投影”。

请大家特别注意一点, X 和 X_1 在热图上已经非常接近。我们在选取 v_1 时, 有特殊的“讲究”, 这就是为什么在本节开头说 V 是“无数里挑一”的原因。我们将会在本书下一个板块——矩阵分解, 和大家深入探讨如何获得这个特殊的 V 。

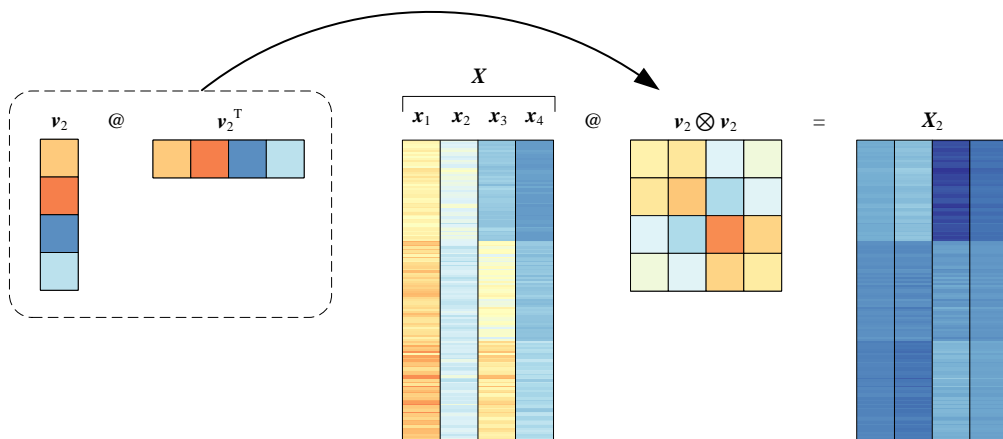
图 41. 四特征数据矩阵 $X_{150 \times 4}$ 向 v_1 方向向量投影，二次投影

第 2 列向量 v_2

图 42 展示获得 z_2 和 X_2 的过程。请大家根据之前分析 v_1 的思路自行分析这两图。

图 42. 四特征数据矩阵 $X_{150 \times 4}$ 向 v_2 投影，一次投影，二次投影

同样，利用张量积完成 $X_{150 \times 4}$ 向 v_2 二次投影。大家自行计算张量积 $v_2 \otimes v_2$ 具体值，按照前文思路分析图 43。有必要指出一点，对比 X_1 ， X_2 热图和 X 相差很大。

图 43. 四特征数据矩阵 $X_{150 \times 4}$ 向 v_2 投影，二次投影

第 3 列向量 v_3

大家自行分析图 44、图 45。再次强调，一次投影就是标量投影；二次投影相当于向量投影。

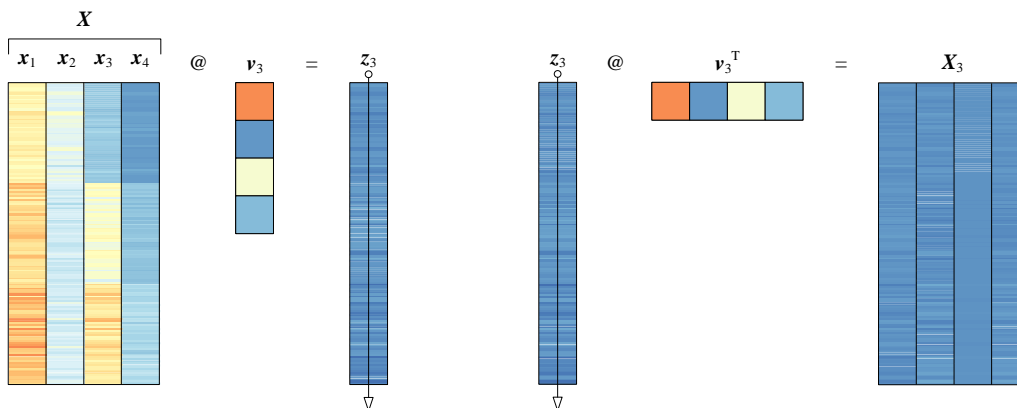


图 44. 四特征数据矩阵 $X_{150 \times 4}$ 向 v_3 投影，一次投影，二次投影

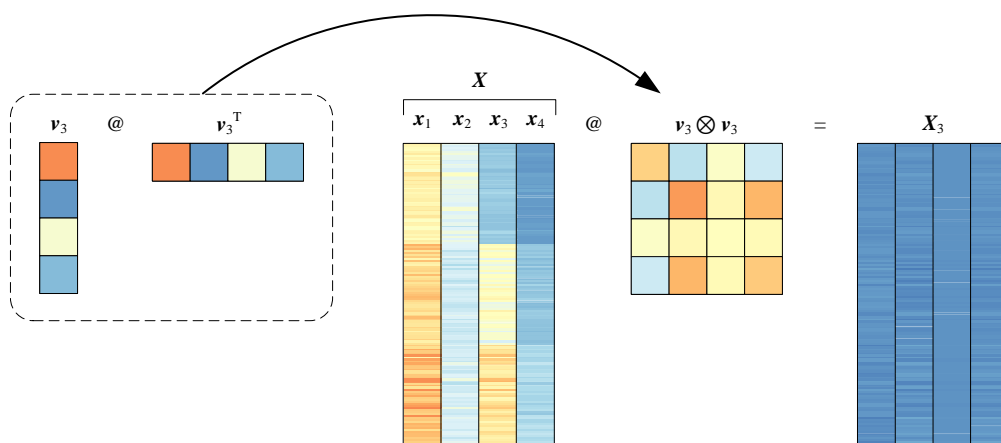


图 45. 四特征数据矩阵 $X_{150 \times 4}$ 向 v_3 投影，二次投影

第 4 列向量 v_4

大家自行分析图 46、图 47。特别注意比较 X_1 、 X_2 、 X_3 、 X_4 的四幅热图差异。

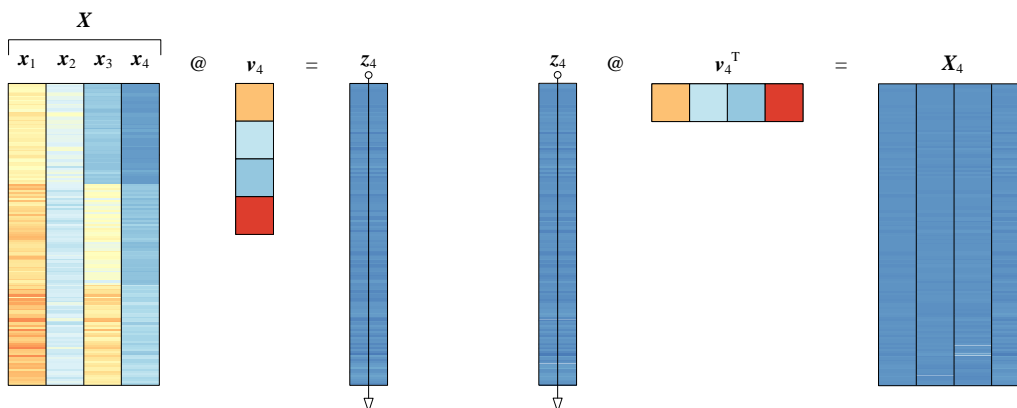
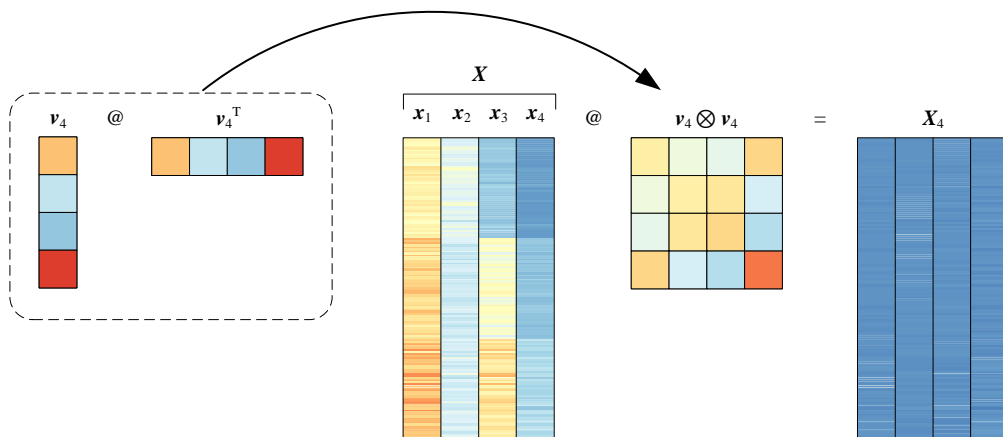


图 46. 四特征数据矩阵 $X_{150 \times 4}$ 向 v_4 投影，一次投影和二次投影

图 47. 四特征数据矩阵 $X_{150 \times 4}$ 向 v_4 投影，二次投影

层层叠加

类似前文，我们也从两个视角讨论层层叠加还原原矩阵。

如图 48 所示，数据矩阵 X 在规范正交基 $[v_1, v_2, \dots, v_D]$ 中每个方向上向量投影层层叠加完全还原原始数据。

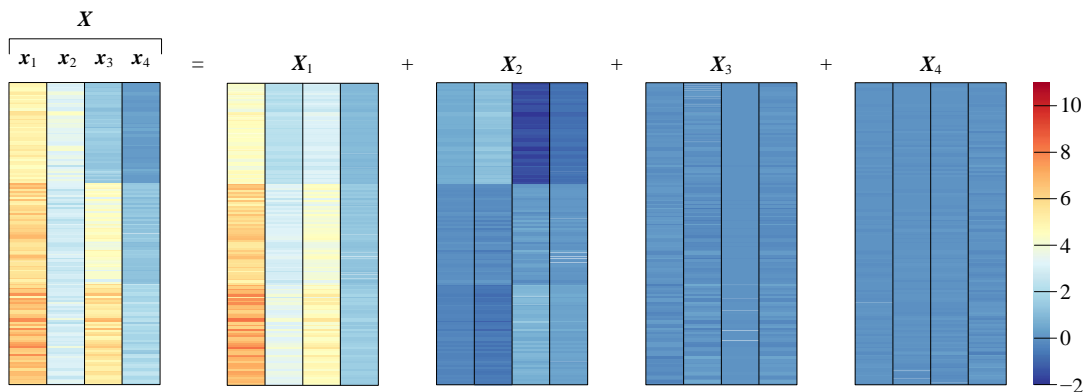
图 48. 层层叠加还原四特征数据矩阵 $X_{150 \times 4}$

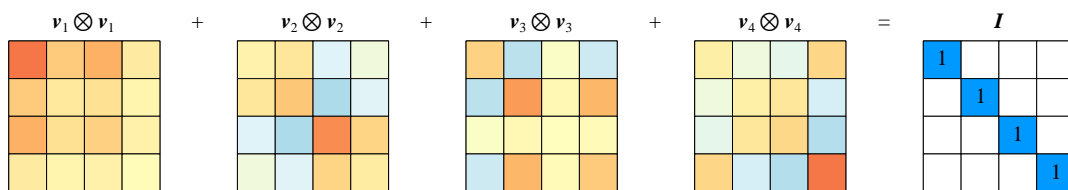
图 48 告诉我们，要想完整还原 X ，需要四幅热图叠加，即 $X = X_1 + X_2 + X_3 + X_4$ 。我们已经很清楚 X_1 、 X_2 、 X_3 、 X_4 这四个矩阵的秩都是 1。而 X 本身的秩为 4，即 $\text{rank}(X) = 4$ 。

建议大家仔细对比图 48 中 X 、 X_1 、 X_2 、 X_3 、 X_4 这五幅热图色差，它们采用完全相同的色谱。前文已经提到 X_1 已经非常接近 X 。也就是说，我们可以用秩为 1 的 X_1 近似秩为 4 的 X 。

如图 49 所示，这四个张量积层层叠加得到单位矩阵，即：

$$v_1 \otimes v_1 + v_2 \otimes v_2 + v_3 \otimes v_3 + v_4 \otimes v_4 = I \quad (57)$$

如前文所述，(57) 是数据还原的另外一个视角。本章前文提到 (9)，矩阵乘单位矩阵结果为其本身，即 $XI = X$ 。而单位矩阵 I 可以按 (57) 分解。这也就是说，张量积层层叠加得到了单位矩阵 I ，等价于还原原始数据。

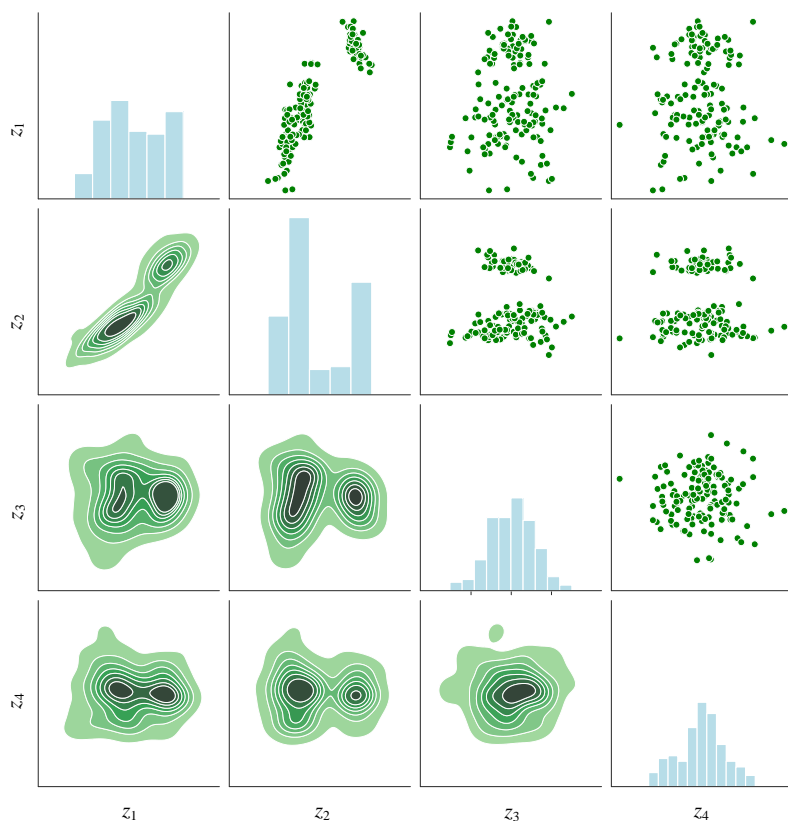
图 49. 张量积层层累加获得 4×4 单位矩阵

Bk4_Ch10_01.py 绘制本章前文大部分热图。

10.7 数据正交化

成对特征散点图

本节再回过头来分析图 37 中数据矩阵 \mathbf{Z} 。本书第 1 章提到，对于多特征 ($D > 3$) 数据矩阵，成对特征散点图可以帮助我们可视化数据分布。图 50 所示为矩阵 \mathbf{Z} 的成对特征散点图。这幅图中，对角线上的六幅图是每个特征数据分布的直方图，左下角六幅图是二元概率密度估计等高线图。

图 50. \mathbf{Z} 成对特征分析图

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

两个格拉姆矩阵

如图 51 所示， \mathbf{Z}^T 乘 \mathbf{Z} 得到 \mathbf{Z} 的格拉姆矩阵：

$$\mathbf{Z}^T \mathbf{Z} = \begin{bmatrix} \mathbf{z}_1^T \\ \mathbf{z}_2^T \\ \vdots \\ \mathbf{z}_D^T \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_D \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1^T \mathbf{z}_1 & \mathbf{z}_1^T \mathbf{z}_2 & \cdots & \mathbf{z}_1^T \mathbf{z}_D \\ \mathbf{z}_2^T \mathbf{z}_1 & \mathbf{z}_2^T \mathbf{z}_2 & \cdots & \mathbf{z}_2^T \mathbf{z}_D \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{z}_D^T \mathbf{z}_1 & \mathbf{z}_D^T \mathbf{z}_2 & \cdots & \mathbf{z}_D^T \mathbf{z}_D \end{bmatrix} \quad (58)$$

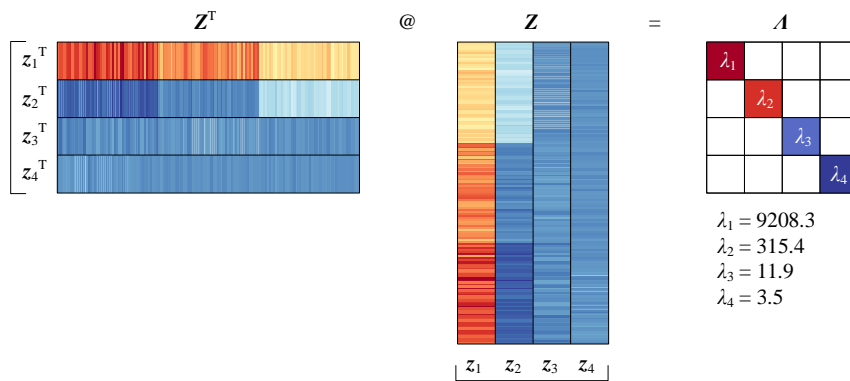


图 51. 矩阵 \mathbf{Z} 的格拉姆矩阵

(58) 写成向量内积形式：

$$\mathbf{Z}^T \mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 \cdot \mathbf{z}_1 & \mathbf{z}_1 \cdot \mathbf{z}_2 & \cdots & \mathbf{z}_1 \cdot \mathbf{z}_D \\ \mathbf{z}_2 \cdot \mathbf{z}_1 & \mathbf{z}_2 \cdot \mathbf{z}_2 & \cdots & \mathbf{z}_2 \cdot \mathbf{z}_D \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{z}_D \cdot \mathbf{z}_1 & \mathbf{z}_D \cdot \mathbf{z}_2 & \cdots & \mathbf{z}_D \cdot \mathbf{z}_D \end{bmatrix} = \begin{bmatrix} \langle \mathbf{z}_1, \mathbf{z}_1 \rangle & \langle \mathbf{z}_1, \mathbf{z}_2 \rangle & \cdots & \langle \mathbf{z}_1, \mathbf{z}_D \rangle \\ \langle \mathbf{z}_2, \mathbf{z}_1 \rangle & \langle \mathbf{z}_2, \mathbf{z}_2 \rangle & \cdots & \langle \mathbf{z}_2, \mathbf{z}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{z}_D, \mathbf{z}_1 \rangle & \langle \mathbf{z}_D, \mathbf{z}_2 \rangle & \cdots & \langle \mathbf{z}_D, \mathbf{z}_D \rangle \end{bmatrix} \quad (59)$$

观察图 51，发现 $\mathbf{Z}^T \mathbf{Z}$ 恰好是对角方阵，即：

$$\mathbf{Z}^T \mathbf{Z} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_D \end{bmatrix} = \mathbf{A} \quad (60)$$

这说明， \mathbf{Z} 的列向量两两正交，即：

$$\mathbf{z}_i^T \mathbf{z}_j = \mathbf{z}_j^T \mathbf{z}_i = \mathbf{z}_i \cdot \mathbf{z}_j = \mathbf{z}_j \cdot \mathbf{z}_i = \langle \mathbf{z}_i, \mathbf{z}_j \rangle = \langle \mathbf{z}_j, \mathbf{z}_i \rangle = 0, \quad i \neq j \quad (61)$$

对比 \mathbf{X} 的格拉姆矩阵：

$$\begin{aligned}
 G = X^T X &= \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_D^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_D \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T \mathbf{x}_1 & \mathbf{x}_1^T \mathbf{x}_2 & \cdots & \mathbf{x}_1^T \mathbf{x}_D \\ \mathbf{x}_2^T \mathbf{x}_1 & \mathbf{x}_2^T \mathbf{x}_2 & \cdots & \mathbf{x}_2^T \mathbf{x}_D \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_D^T \mathbf{x}_1 & \mathbf{x}_D^T \mathbf{x}_2 & \cdots & \mathbf{x}_D^T \mathbf{x}_D \end{bmatrix} \\
 &= \begin{bmatrix} \mathbf{x}_1 \cdot \mathbf{x}_1 & \mathbf{x}_1 \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_1 \cdot \mathbf{x}_D \\ \mathbf{x}_2 \cdot \mathbf{x}_1 & \mathbf{x}_2 \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_2 \cdot \mathbf{x}_D \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_D \cdot \mathbf{x}_1 & \mathbf{x}_D \cdot \mathbf{x}_2 & \cdots & \mathbf{x}_D \cdot \mathbf{x}_D \end{bmatrix} = \begin{bmatrix} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle & \langle \mathbf{x}_1, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_1, \mathbf{x}_D \rangle \\ \langle \mathbf{x}_2, \mathbf{x}_1 \rangle & \langle \mathbf{x}_2, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_2, \mathbf{x}_D \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{x}_D, \mathbf{x}_1 \rangle & \langle \mathbf{x}_D, \mathbf{x}_2 \rangle & \cdots & \langle \mathbf{x}_D, \mathbf{x}_D \rangle \end{bmatrix} \quad (62)
 \end{aligned}$$

图 52 所示为计算矩阵 X 的格拉姆矩阵的热图。请大家格外注意一点，图 52 中矩阵 G 的迹，即对角线元素之和， $\text{tr}(G) = 9539.29$ 。而图 51 中矩阵 A 的迹和 G 的迹相同， $\text{tr}(G) = \text{tr}(A) = 9539.29$ 。本书后面还会反复提到这一点。

V 因 X 而生

细细想来，上一节介绍的 $Z = XV$ 的数据转换很神奇！

还是以鸢尾花数据为例，如图 52 所示， G 中没有一个元素为 0！ G 主对角线元素代表 X 的列向量模的平方， G 主对角线以外元素代表 X 两个特定列向量的内积。

如图 51 所示，经过数据转换 $Z = XV$ ，矩阵 Z 的格拉姆矩阵为对角方阵 A 。 A 的主对角线以外元素都为 0。也就是说， $i \neq j$ 时， z_i 和 z_j 都是行数为 150 的列向量， z_i 和 z_j 的向量内积竟然为 0。也就是说 150 个成对元素乘积之和为 0！这种情况在图 51 中竟然发生了 12 次，本质上发生了 6 次。

对于鸢尾花数据矩阵 X 来说，(53) 中给出的这个 V 真可谓“无数里挑一”！

换句话说， V 因 X 而生！

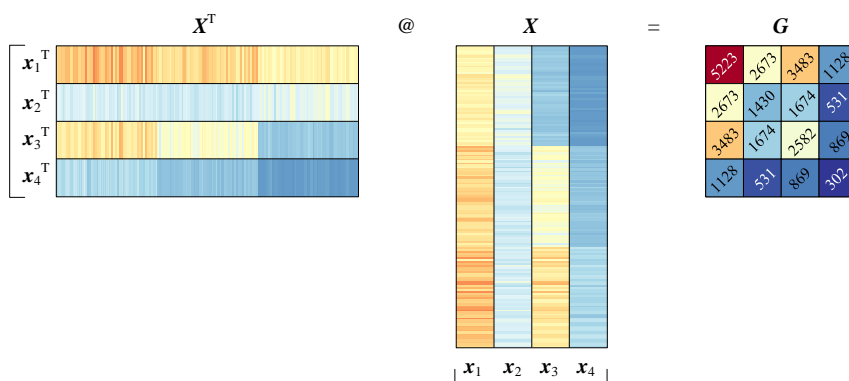


图 52. 矩阵 X 的格拉姆矩阵

⚠ 注意，统计视角下，矩阵 Z 的列向量两两内积为 0，不代表两两相关性系数为 0。本系列丛书《概率统计》将介绍如何通过正交投影获得两两相关性系数为 0 的数据矩阵。

对角化

将 $Z = XV$ 其代入 (60) 得到：

$$Z^T Z = (XV)^T XV = V^T \underset{G}{X^T X} V = V^T G V = A \quad (63)$$

再进一步，由于 V 为规范正交基，因此 $V^T V = I$ ，根据 (63) 等式关系， G 可以写成：

$$G = V A V^T \quad (64)$$

这就是说，如图 53 所示， X 的格拉姆矩阵 G 可以通过某种矩阵分解得到三个矩阵的乘积。其中， V 为正交矩阵， A 为对角方阵。从 G 到 A 也是一个方阵**对角化** (diagonalization) 的过程。

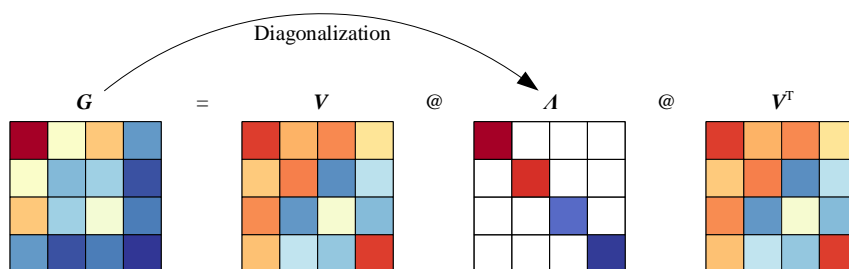


图 53. 对 G 矩阵分解

➡ 为了获得 (64)，就需要本书下一个板块要介绍的重要线性代数工具——特征值分解 (eigen decomposition)。

回看规范正交基 V ：双标图

像 Z 这样具有这种**正交性** (orthogonality) 的数据应用场合很多，因此我们再深究一步。

类似 (54)，我们可以把 z_1 、 z_2 、 z_3 、 z_4 写成如下线性组合：

$$\begin{aligned} z_1 &= Xv_1 = 0.751x_1 + 0.380x_2 + 0.513x_3 + 0.168x_4 \\ z_2 &= Xv_2 = 0.284x_1 + 0.547x_2 - 0.709x_3 - 0.344x_4 \\ z_3 &= Xv_3 = 0.502x_1 - 0.675x_2 - 0.059x_3 - 0.537x_4 \\ z_4 &= Xv_4 = 0.321x_1 - 0.317x_2 - 0.481x_3 + 0.752x_4 \end{aligned}$$

$$V = \begin{bmatrix} 0.751 & 0.284 & 0.502 & 0.321 \\ 0.380 & 0.547 & -0.675 & -0.317 \\ 0.513 & -0.709 & -0.059 & -0.481 \\ 0.168 & -0.344 & -0.537 & 0.752 \end{bmatrix} \quad (65)$$

请大家格外注意 (65) 各个元素颜色对应关系。

我们给 z_1 、 z_2 、 z_3 、 z_4 取一个新的名字——**主成分** (Principal Component, PC)。 z_1 、 z_2 、 z_3 、 z_4 分别对应 PC_1 、 PC_2 、 PC_3 、 PC_4 。显然 PC_1 、 PC_2 、 PC_3 、 PC_4 相互垂直。

有了 PC_1 、 PC_2 、 PC_3 、 PC_4 ，我们可以绘制图 54 这幅图，图中有 6 幅子图，每幅子图都是一个双标图 (biplot)。

我们以图 54 中浅蓝色阴影背景子图为例介绍如何理解双标图。

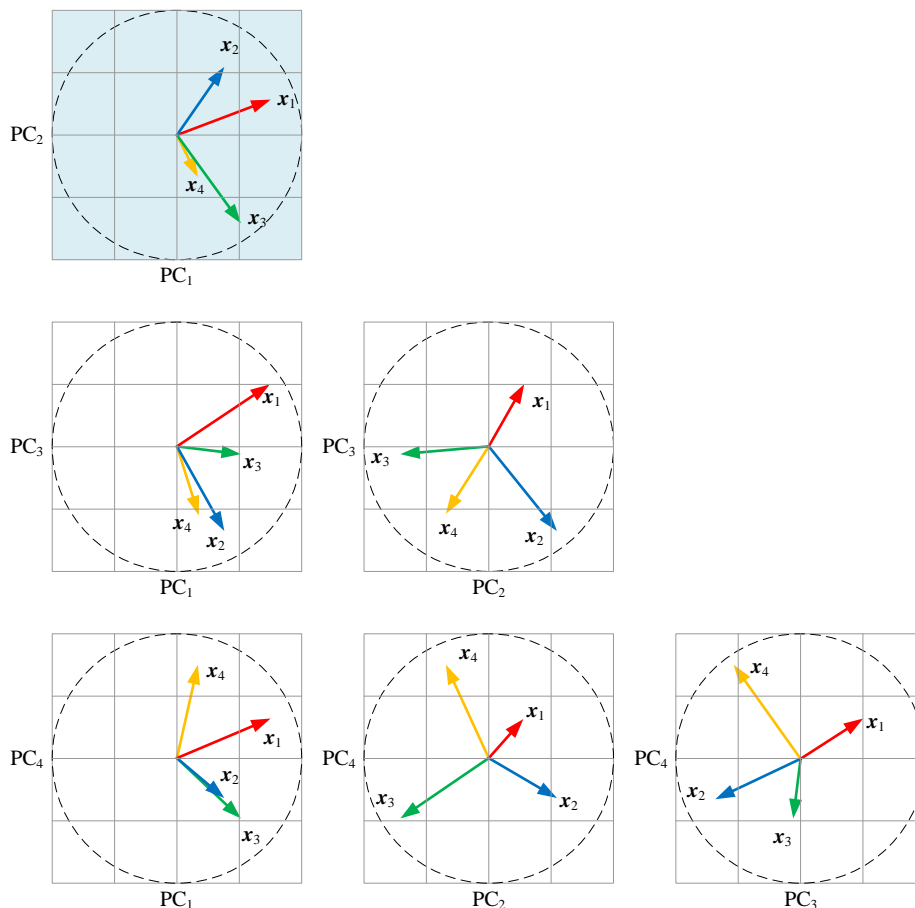


图 54. 分解主成分

在 PC_1 - PC_2 平面上， x_1 对应坐标点为 $(0.751, 0.284)$ ，这意味着 x_1 分别给 z_1 和 z_2 贡献 $0.751x_1$ 和 $0.284x_1$ 。同理，我们可以发现 x_2 分别给 z_1 和 z_2 贡献 $0.380x_2$ 和 $0.547x_2$ 。以此类推。

反向来看， x_1 在 PC_1 、 PC_2 、 PC_3 、 PC_4 方向上的分量分别为 $0.751x_1$ 、 $0.284x_1$ 、 $0.502x_1$ 、 $0.321x_1$ ，这四个成分满足：

$$0.751^2 + 0.284^2 + 0.502^2 + 0.321^2 = 1 \quad (66)$$

反向正交投影

由于 $Z = XV$ ，且正交矩阵 V 可逆， X 则可以通过 Z 反推得到，即：

$$X = ZV^{-1} = ZV^T \quad (67)$$

图 55 所示为 X 和 Z 相互转化的关系。这幅图告诉我们另外一个重要性质—— V 和 V^T 都是规范正交基！

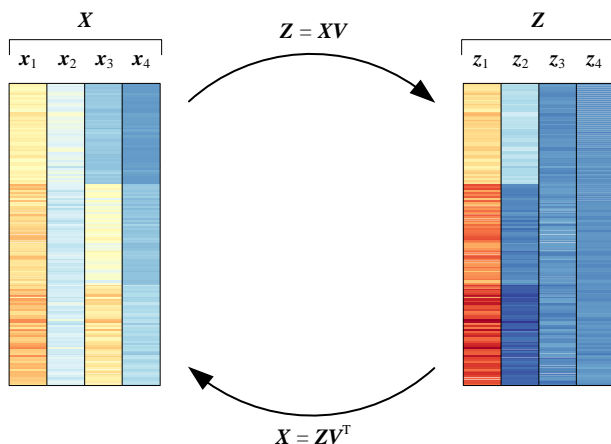


图 55. X 和 Z 之间关系

将 (67) 展开写：

$$X = ZV^T = Z \begin{bmatrix} \mathbf{v}^{(1)T} \\ \mathbf{v}^{(2)T} \\ \vdots \\ \mathbf{v}^{(D)T} \end{bmatrix} = Z \begin{bmatrix} \mathbf{v}^{(1)T} & \mathbf{v}^{(2)T} & \cdots & \mathbf{v}^{(D)T} \end{bmatrix} = \begin{bmatrix} Z\mathbf{v}^{(1)T} & Z\mathbf{v}^{(2)T} & \cdots & Z\mathbf{v}^{(D)T} \end{bmatrix} \quad (68)$$

$V^T = [\mathbf{v}^{(1)T}, \mathbf{v}^{(2)T}, \dots, \mathbf{v}^{(D)T}]$ 也是一个规范正交基。上式代表“反向”正交投影的过程。

取出 (68) 矩阵 X 第 j 列对应的等式：

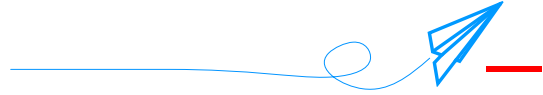
$$\mathbf{x}_j = Z\mathbf{v}^{(j)T} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_D \end{bmatrix} \begin{bmatrix} v_{j,1} \\ v_{j,2} \\ \vdots \\ v_{j,D} \end{bmatrix} = v_{j,1}\mathbf{z}_1 + v_{j,2}\mathbf{z}_2 + \cdots + v_{j,D}\mathbf{z}_D \quad (69)$$



(69) 这一视角在主成分分析中非常重要，我们将会在《数据科学》一书中深入探讨。



本书第 1 章用 Streamlit 制作了一个 App，我们利用 Plotly 可视化鸢尾花数据集的热图、平面散点图、三维散点图、成对特征散点图。本章“照葫芦画瓢”照搬这个 App，采用完全一致的图像可视化转换得到的数据矩阵 Z 。请大家参考 Streamlit_Bk4_Ch10_01.py。



本章是个分水岭。如果本章前两节内容，你读起来毫无压力，恭喜你，你可以顺利进入本书下一个板块——矩阵分解——的学习。阅读本章时，如果感觉很吃力，请回头重读前 9 章内容。

大家可能会好奇，本章中神奇的 V 是怎么算出来的？其实本章代码文件已经给出了答案——特征值分解。这是本书下一个板块要讲的重要内容之一。

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！

再加一句，有向量的地方，肯定有空间！

请大家带着这四句话，进入本书下一阶段的学习。