10 Data Projection 数据投影

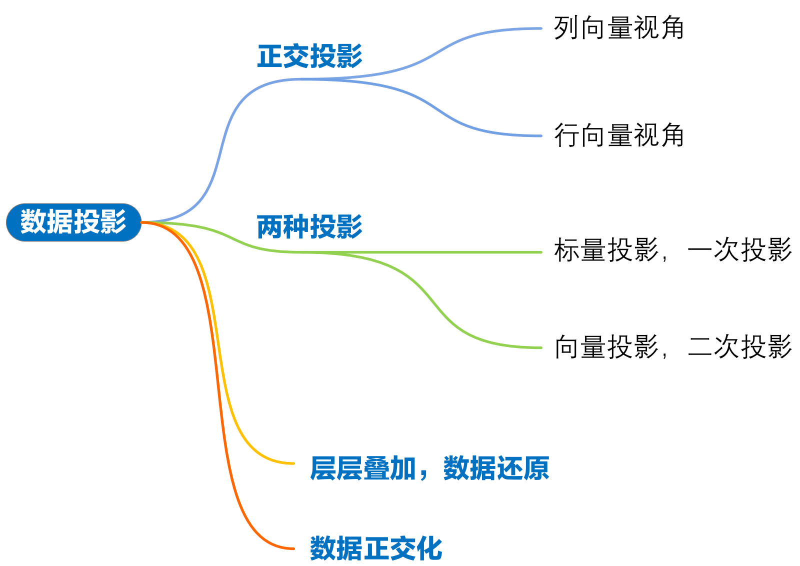
以鸢尾花数据集为例

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


数据投影
正交投影
列向量视角
行向量视角
两种投影
标量投影，一次投影
向量投影，二次投影
层层叠加，数据还原
数据正交化

10.1 从一个矩阵乘法运算说起

有数据的地方，就有向量！

有向量的地方，就有几何！

本章承前启后，结合数据、几何、向量三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：

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{V} \tag{1}$$

$\boldsymbol{X}$ 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，$\boldsymbol{X}$ 每一行代表一个数据点，每一列代表一个特征。

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

几何视角下，矩阵乘积 $\boldsymbol{X}\boldsymbol{V}$ 完成的是 $\boldsymbol{X}$ 向规范正交基 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_D]$ 投影，乘积 $\boldsymbol{X}\boldsymbol{V}$ 结果 $\boldsymbol{Z}$ 代表 $\boldsymbol{X}$ 在新的规范正交基下的坐标。

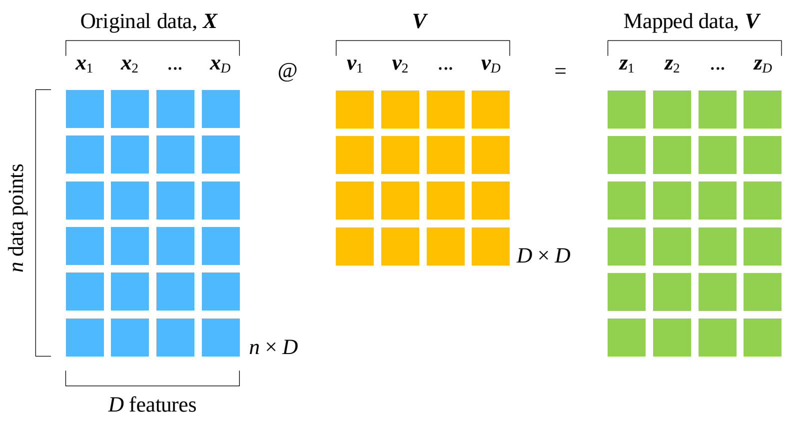


图 1. 数据矩阵 $\boldsymbol{X}$ 到 $\boldsymbol{Z}$ 线性变换

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

$$\begin{aligned}\begin{bmatrix}\boldsymbol{z}_1 & \boldsymbol{z}_2 & \cdots & \boldsymbol{z}_D\end{bmatrix} &= \boldsymbol{X}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D\end{bmatrix} \\ &= \begin{bmatrix}\boldsymbol{X}\boldsymbol{v}_1 & \boldsymbol{X}\boldsymbol{v}_2 & \cdots & \boldsymbol{X}\boldsymbol{v}_D\end{bmatrix}\end{aligned} \tag{2}$$

(2) 这个视角是数据列向量 (即特征) 之间的转换。

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

$$\boldsymbol{z}_j = \boldsymbol{X}\boldsymbol{v}_j \tag{3}$$

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

$$\boldsymbol{z}_j = \underbrace{\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D\end{bmatrix}}_{\boldsymbol{X}} \underbrace{\begin{bmatrix} v_{1,j} \\ v_{2,j} \\ \vdots \\ v_{D,j}\end{bmatrix}}_{\boldsymbol{v}_j} = v_{1,j}\boldsymbol{x}_1 + v_{2,j}\boldsymbol{x}_2 + \cdots v_{D,j}\boldsymbol{x}_D \tag{4}$$

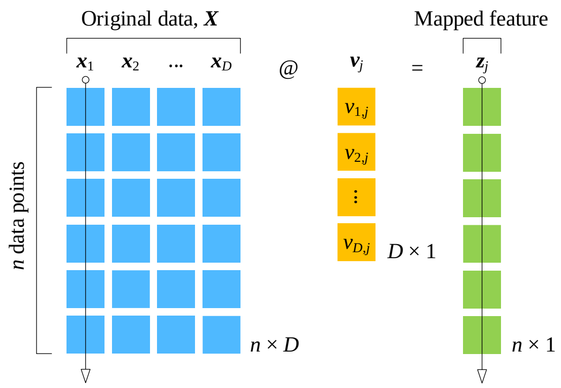


图 2. $\boldsymbol{Z}$ 第 j 列向量 $\boldsymbol{z}_j$ 的计算过程

行向量：点坐标

数据矩阵 $\boldsymbol{X}$ 的任意行向量 $\boldsymbol{x}^{(i)}$ 代表一个样本点在 $\mathbb{R}^D$ 的坐标。

注意 $\mathbb{R}^D$ 基底为标准正交基 $\boldsymbol{E} = [\boldsymbol{e}_1, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_D]$。

将 $\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 写成行向量形式，(1) 可以写作：

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

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

$$\boldsymbol{z}^{(i)} = \boldsymbol{x}^{(i)}\boldsymbol{V} \tag{6}$$

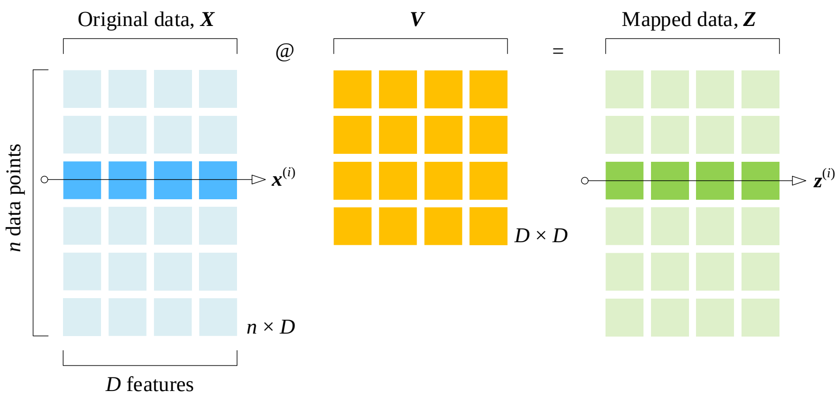


图 3. 每一行数据点之间的转换关系

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

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

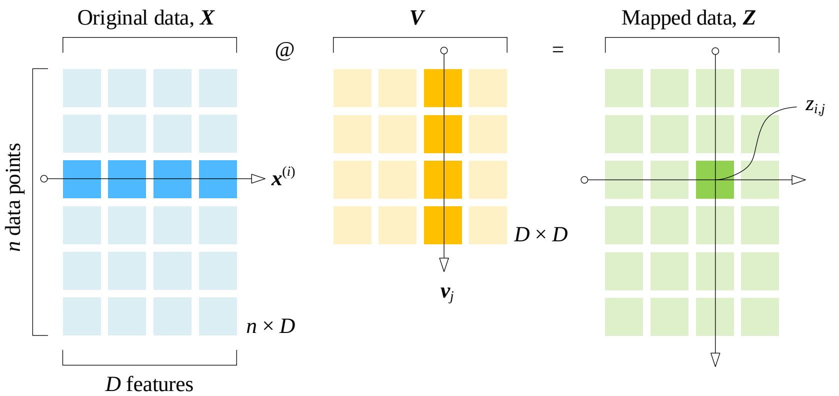


图 4. 每一行数据点向 $\boldsymbol{v}_j$ 投影

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

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

图 4 对应 (8) 运算。

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

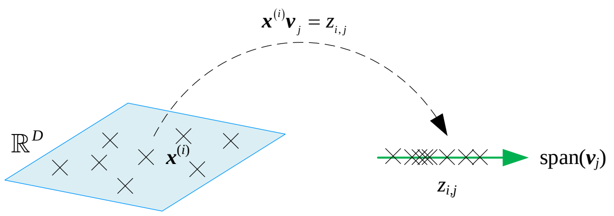


图 5. $\mathbb{R}^D$ 空间数据点投影到 span($\boldsymbol{v}_j$)

10.2 二次投影 + 层层叠加

本书上一章给出下面这个看似莫明其妙的矩阵乘法：

$$\boldsymbol{X} = \boldsymbol{X}\boldsymbol{I} = \boldsymbol{X}\underbrace{\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}}_{\boldsymbol{I}} = \boldsymbol{X} \tag{9}$$

数据矩阵 $\boldsymbol{X}$ 乘以单位阵 $\boldsymbol{I}$，结果为 $\boldsymbol{X}$ 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。

将 $\boldsymbol{V}$ 写成 [$\boldsymbol{v}_1$, $\boldsymbol{v}_1$, …, $\boldsymbol{v}_D$]，代入 (9) 得到：

$$\begin{aligned}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} &= \boldsymbol{X}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}}\end{bmatrix} \\ &= \underbrace{\boldsymbol{X}\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}}}_{\boldsymbol{X}_1} + \underbrace{\boldsymbol{X}\boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}}}_{\boldsymbol{X}_2} + \cdots + \underbrace{\boldsymbol{X}\boldsymbol{v}_D\boldsymbol{v}_D^{\mathrm{T}}}_{\boldsymbol{X}_D}\end{aligned} \tag{10}$$

令，

$$\boldsymbol{X}_j = \boldsymbol{X}\boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{T}} \tag{11}$$

图 6 所示为上述运算，$\boldsymbol{X}_j$ 的形状和原数据矩阵 $\boldsymbol{X}$ 完全相同。我们称图 6 为二次投影，一会解释原因。

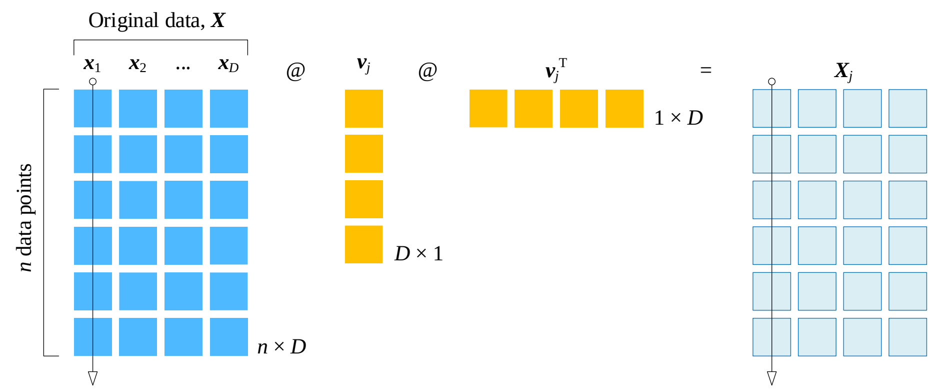


图 6. 二次投影

(10) 可以写成：

$$
\boldsymbol{X} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_D \tag{12}
$$

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 $\boldsymbol{X}$。

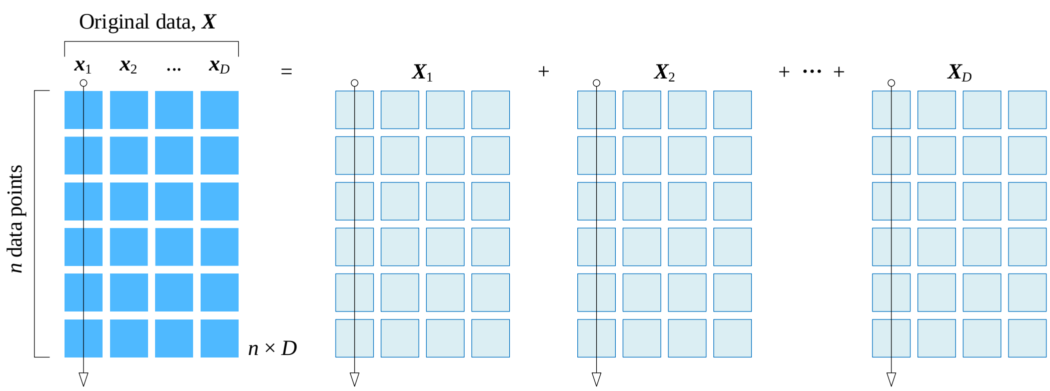


图 7. 层层叠加

二次投影

下面，我们聊聊"二次投影"。

取出 (7) 中向量 $\boldsymbol{X}_j$ 第 j 行元素，对应的运算为：

$$
\boldsymbol{x}_j^{(i)} = \underbrace{\boldsymbol{x}^{(i)} \boldsymbol{v}_j}_{z_{i,j}} \boldsymbol{v}_j^{\mathrm{T}} = z_{i,j} \boldsymbol{v}_j^{\mathrm{T}} \tag{13}
$$

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

而 $z_{i,j}\boldsymbol{v}_j^{\mathrm{T}}$ 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，$\boldsymbol{x}^{(i)} \to z_{i,j}$ 代表“标量投影”；$\boldsymbol{x}^{(i)} \to \boldsymbol{x}^{(i)}\boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{T}}$ 则是“向量投影”。

⚠ 注意，图 8 中 $\boldsymbol{x}^{(i)}$ 和 $z_{i,j}\boldsymbol{v}_j^{\mathrm{T}}$ 都用行向量表达坐标点。

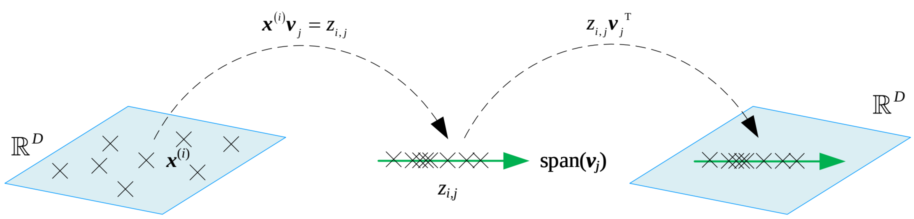


图 8. $\mathbb{R}^D$ 空间数据点先投影到 span($\boldsymbol{v}_j$)，再投影回到 $\mathbb{R}^D$

向量投影：张量积

将 (11) 写成张量积的形式：

$$\boldsymbol{X}_j = \boldsymbol{X}\boldsymbol{v}_j \otimes \boldsymbol{v}_j \tag{14}$$

$\boldsymbol{X}_j$ 就是 $\boldsymbol{X}$ 经过“降维”到 span($\boldsymbol{v}_j$) 后，再正交投影到 $\mathbb{R}^D$ 中得到的“像”。$\boldsymbol{X}_j$ 也是 $\boldsymbol{X}$ 在 $\boldsymbol{v}_j$ 上的向量投影。张量积 $\boldsymbol{v}_j \otimes \boldsymbol{v}_j$ 就是我们上一章提到的投影矩阵 (projection matrix)。

张量积 $\boldsymbol{v}_j \otimes \boldsymbol{v}_j$ 本身完成 “多维 → 一维” + “一维 → 多维” 这两步投影。很显然，

$$\operatorname{rank}\left(\boldsymbol{v}_j \otimes \boldsymbol{v}_j\right) = 1 \quad \Rightarrow \quad \operatorname{rank}\left(\boldsymbol{X}_j\right) = 1 \tag{15}$$

所以，在 $\mathbb{R}^D$ 空间中，$\boldsymbol{X}_j$ 所有数据点在一条直线上，和 $\boldsymbol{v}_j$ 平行 (同向或反向)。也就是说，虽然 $\boldsymbol{X}_j$ 在 D 维空间 $\mathbb{R}^D$ 中，$\boldsymbol{X}_j$ 实际上只有 1 个维度，即 dim($\boldsymbol{v}_j$) = dim($\boldsymbol{v}_j \otimes \boldsymbol{v}_j$) = 1，rank($\boldsymbol{X}_j$) = 1。

利用张量积，(10) 可以写成：

$$\boldsymbol{X} = \underbrace{\boldsymbol{X}\boldsymbol{v}_1 \otimes \boldsymbol{v}_1}_{\boldsymbol{X}_1} + \underbrace{\boldsymbol{X}\boldsymbol{v}_2 \otimes \boldsymbol{v}_2}_{\boldsymbol{X}_2} + \cdots + \underbrace{\boldsymbol{X}\boldsymbol{v}_D \otimes \boldsymbol{v}_D}_{\boldsymbol{X}_D} \tag{16}$$

可以这样理解上式，$\boldsymbol{X}$ 分别二次投影 (向量投影) 到规范正交基 [$\boldsymbol{v}_1$, $\boldsymbol{v}_1$, …, $\boldsymbol{v}_D$] 每个列向量 $\boldsymbol{v}_j$ 所代表的子空间 span($\boldsymbol{v}_j$) 中，获得 $\boldsymbol{X}_1$、$\boldsymbol{X}_2$ … $\boldsymbol{X}_D$。而 $\boldsymbol{X}_1$、$\boldsymbol{X}_2$ … $\boldsymbol{X}_D$ 层层叠加还原原始数据 $\boldsymbol{X}$。

再进一步，根据 $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{I}$，我们知道：

$$\boldsymbol{I} = \boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \cdots + \boldsymbol{v}_D \otimes \boldsymbol{v}_D \tag{17}$$

也就是说，$\boldsymbol{v}_j \otimes \boldsymbol{v}_j$ 层层叠加得到单位阵 $\boldsymbol{I}$。

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 $[\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_D]$ 替换 (16) 中的 $[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$，得到：

$$\boldsymbol{X} = \boldsymbol{X}\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{X}\boldsymbol{e}_2 \otimes \boldsymbol{e}_2 + \cdots + \boldsymbol{X}\boldsymbol{e}_D \otimes \boldsymbol{e}_D \tag{18}$$

展开 (18) 中等式右侧第一项得到：

$$\boldsymbol{X}_1 = \boldsymbol{X}\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 = \boldsymbol{X}\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} \otimes \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} = \underbrace{\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D\end{bmatrix}}_{X}\begin{bmatrix}1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0\end{bmatrix} = \begin{bmatrix}\boldsymbol{x}_1 & 0 & \cdots & 0\end{bmatrix} \tag{19}$$

$\boldsymbol{X}\boldsymbol{e}_1$ 得到的是 $\boldsymbol{X}$ 的每一行在 span($\boldsymbol{e}_1$) 这个子空间的坐标，即 $\boldsymbol{x}_1$。而 $\boldsymbol{X}\boldsymbol{e}_1 \otimes \boldsymbol{e}_1$ 告诉我们的是 $\boldsymbol{X}\boldsymbol{e}_1$ 在 D 维空间 $\mathbb{R}^D$ 中坐标值。

此后 (18) 右侧每一项 $\boldsymbol{X}_j$ 可以写成：

$$\begin{aligned}\boldsymbol{X}_2 &= \boldsymbol{X}\boldsymbol{e}_2 \otimes \boldsymbol{e}_2 = \begin{bmatrix}0 & \boldsymbol{x}_2 & \cdots & 0\end{bmatrix}\\ &\vdots\\ \boldsymbol{X}_D &= \boldsymbol{X}\boldsymbol{e}_D \otimes \boldsymbol{e}_D = \begin{bmatrix}0 & 0 & \cdots & \boldsymbol{x}_D\end{bmatrix}\end{aligned} \tag{20}$$

也就是说，这个每次计算 $\boldsymbol{X}\boldsymbol{e}_j \otimes \boldsymbol{e}_j$ 投影就是仅保留 $\boldsymbol{X}$ 的第 j 列 $\boldsymbol{x}_j$，其他位置元素置 0。

因此，(18) 可以写成：

$$\boldsymbol{X} = \underbrace{\begin{bmatrix}\boldsymbol{x}_1 & 0 & \cdots & 0\end{bmatrix}}_{X_1} + \underbrace{\begin{bmatrix}0 & \boldsymbol{x}_2 & \cdots & 0\end{bmatrix}}_{X_2} + \cdots + \underbrace{\begin{bmatrix}0 & 0 & \cdots & \boldsymbol{x}_D\end{bmatrix}}_{X_D} \tag{21}$$

图 9 所示为上式二次投影与叠加过程。

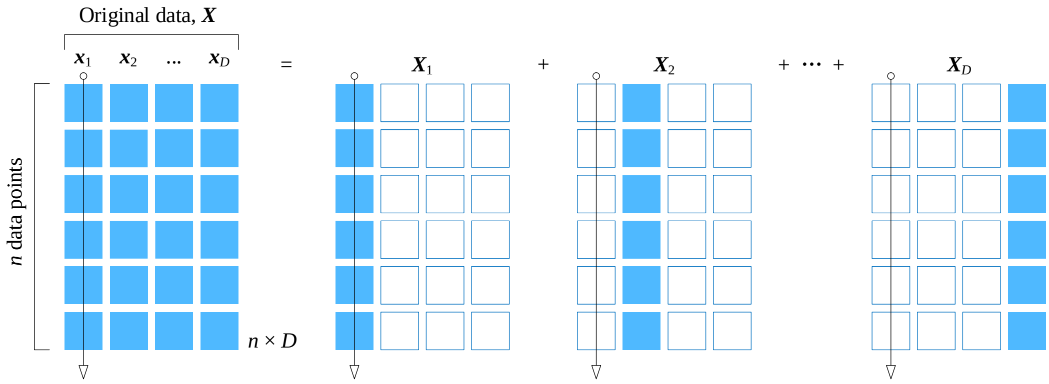


图 9. 标准正交基 $[\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_D]$ 中二次投影与叠加

回过头再看 (9)，我们知道这个运算过程代表先从标准正交基 $[\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_D]$ 到规范正交基 $[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$ 的投影，然后再投影回到标准正交基 $[\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_D]$：

$$\underset{}{\boldsymbol{X}} \xrightarrow[V]{} \underset{XV}{\boldsymbol{Z}} \xrightarrow[V^{\mathrm{T}}]{} \underset{XVV^{\mathrm{T}}}{\boldsymbol{X}} \tag{22}$$

看到这里，有些读者有可能已经晕头转向。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。

10.3 二特征数据投影：标准正交基

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 $\boldsymbol{X}_{150\times 2}$ 的形状为 150 × 2。

水平方向投影

如图 10 所示，$\boldsymbol{X}_{150\times 2}$ 向水平方向投影，即 $\boldsymbol{X}_{150\times 2}$ 向 $\boldsymbol{e}_1$ 投影。以图中红点 A 为例，A 的坐标为 (5, 2)，它在 $\boldsymbol{e}_1$ 方向上的投影坐标为：

$$\begin{bmatrix} 5 & 2 \end{bmatrix} \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\boldsymbol{e}_1} = 5 \tag{23}$$

注意，5 代表的是 A 在 $\mathrm{span}(\boldsymbol{e}_1)$ 空间中的坐标值，而 $\mathrm{span}(\boldsymbol{e}_1)$ 显然为一维空间。

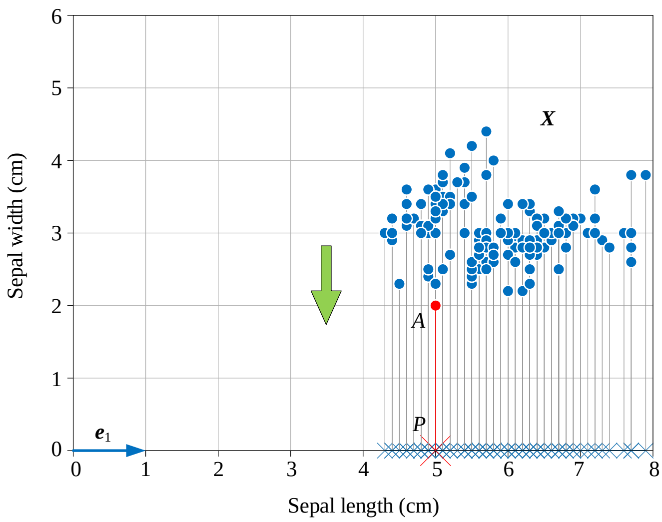


图 10. 二特征数据矩阵 $\boldsymbol{X}_{150\times 2}$ 向 $\boldsymbol{e}_1$ 投影，一次投影

如图 11 热图所示，$\boldsymbol{X}_{150\times 2}$ 向 $\boldsymbol{e}_1$ 投影结果相当于保留了 $\boldsymbol{X}_{150\times 2}$ 第一列数据：

$$\boldsymbol{z}_1 = \boldsymbol{X}\boldsymbol{e}_1 = \boldsymbol{x}_1 \tag{24}$$

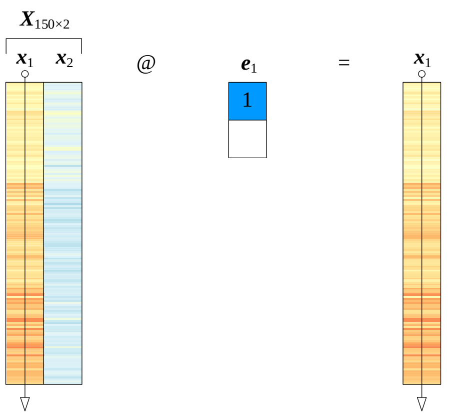


图 11. 数据热图，二特征数据矩阵 $\boldsymbol{X}_{150\times 2}$ 向 $\boldsymbol{e}_1$ 投影，一次投影

大家可能会好奇，既然图 10 中 $\boldsymbol{X}_{150\times 2}$ 向水平方向投影结果都可以画在图 10 直角坐标系中，也就是在二维空间 span($\boldsymbol{e}_1$, $\boldsymbol{e}_2$) 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 span($\boldsymbol{e}_1$, $\boldsymbol{e}_2$) 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 (23) 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：

$$\begin{bmatrix}5 & 2\end{bmatrix}\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 = \begin{bmatrix}5 & 2\end{bmatrix}\boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}} = \begin{bmatrix}5 & 2\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}5 & 0\end{bmatrix} \tag{25}$$

上式对应的计算如图 12 所示。

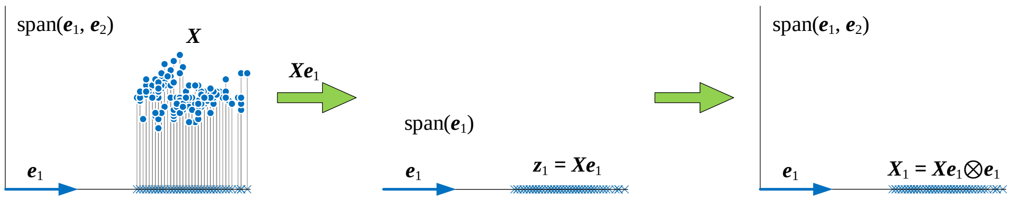


图 12. 二特征数据矩阵 $\boldsymbol{X}$ 向 $\boldsymbol{e}_1$ 投影，二次投影

$\boldsymbol{X}$ 在 $\boldsymbol{e}_1$ 二次投影对应 span($\boldsymbol{e}_1$, $\boldsymbol{e}_2$) 坐标值为 $\boldsymbol{X}_1$:

$$\boldsymbol{X}_1 = \boldsymbol{X}\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 = \boldsymbol{X}\boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}} = \boldsymbol{X}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{0} \end{bmatrix} \tag{26}$$

图 13 所示为上述运算对应热图。

很容易判断，(26) 上式中 $\boldsymbol{e}_1 \otimes \boldsymbol{e}_1$ 的行列式值为 0，即 $\det(\boldsymbol{e}_1 \otimes \boldsymbol{e}_1) = 0$。也就是说这个映射过程存在降维，映射矩阵不可逆，即几何操作不可逆。

⚠ 值得注意的是，从 $\boldsymbol{x}_1$ 到 $\boldsymbol{X}_1 = [\boldsymbol{x}_1, \boldsymbol{0}]$ 这种“升维”，不代表数据信息增多。显然，上式中 $\boldsymbol{X}_1$ 的秩仍为 1，即 rank($\boldsymbol{X}_1$)。

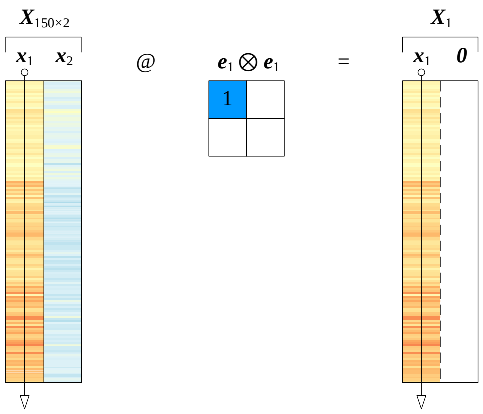


图 13. 数据热图，二特征数据矩阵 $\boldsymbol{X}_{150\times 2}$ 向 $\boldsymbol{e}_1$ 投影，二次投影

竖直方向投影

如图 14 所示，$\boldsymbol{X}_{150\times 2}$ 向竖直方向投影，即 $\boldsymbol{X}_{150\times 2}$ 向 $\boldsymbol{e}_2$ 投影。还是以 A 点为例，A (5, 2) 在 $\boldsymbol{e}_2$ 方向上的标量投影为:

$$\begin{bmatrix} 5 & 2 \end{bmatrix} \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\boldsymbol{e}_2} = 2 \tag{27}$$

2 代表的是 A 在 span($\boldsymbol{e}_2$) 空间中的坐标值，span($\boldsymbol{e}_2$) 同样为一维空间。图 15 为上述运算的热图。

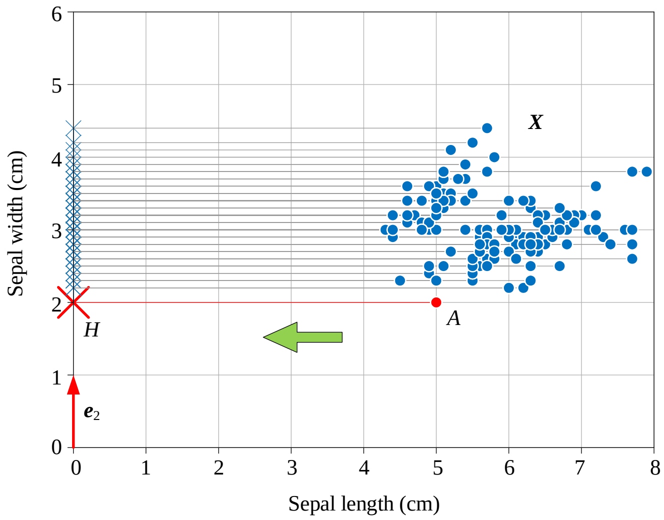


图 14. 二特征数据矩阵 $X_{150\times 2}$ 向 e_2 方向标量投影，一次投影

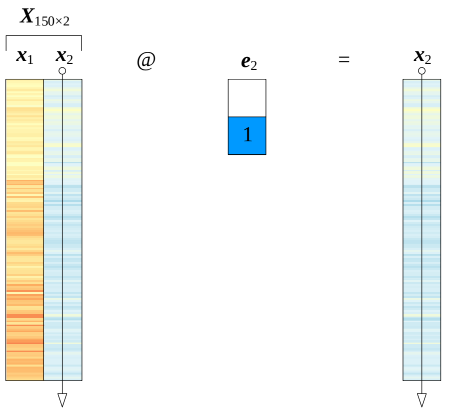


图 15. 数据热图，二特征数据矩阵 $X_{150\times 2}$ 向 e_2 投影，一次投影

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

$$
\begin{bmatrix} 5 & 2 \end{bmatrix} e_2 \otimes e_2 = \begin{bmatrix} 5 & 2 \end{bmatrix} e_2 e_2^{\mathrm{T}} = \begin{bmatrix} 5 & 2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 2 \end{bmatrix} \tag{28}
$$

上式对应的计算如图 16 所示。

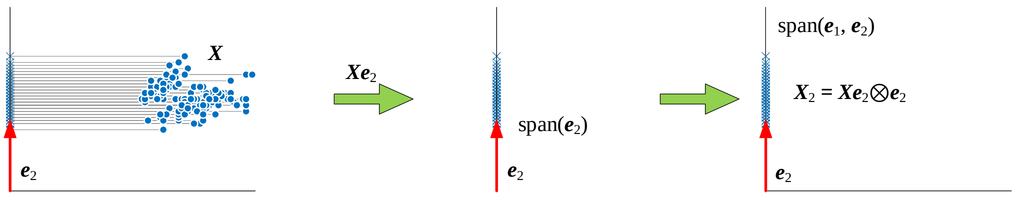


图 16. 二特征数据矩阵 $X_{150\times 2}$ 向 e_2 方向标量投影，二次投影

$X_{150\times2}$ 在 e_2 二次投影得到矩阵 X_2：

$$X_2 = Xe_2 \otimes e_2 = Xe_2e_2^{\mathrm{T}} = X\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{29}$$

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

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

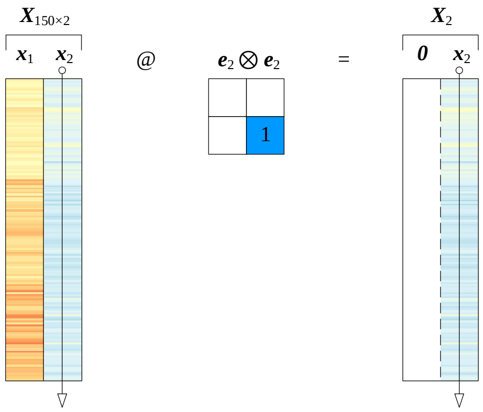


图 17. 数据热图，二特征数据矩阵 $X_{150\times2}$ 向 e_2 投影，二次投影

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)。这也相当于两个向量叠加得到一个向量，即：

$$\begin{bmatrix} 5 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix} \tag{30}$$

或，

$$\begin{bmatrix} 5 \\ 0 \end{bmatrix}^{\mathrm{T}} + \begin{bmatrix} 0 \\ 2 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}^{\mathrm{T}} \tag{31}$$

如图 19 所示，X_1 和 X_2 叠加还原 $X_{150\times2}$：

$$\begin{aligned} X_{150\times2} &= X_1 + X_2 \\ &= X(e_1 \otimes e_1 + Xe_2 \otimes e_2) \\ &= X(e_1e_1^{\mathrm{T}} + e_2e_2^{\mathrm{T}}) \\ &= X\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) = XI \end{aligned} \tag{32}$$

图 20 所示为上述运算对应的热图。

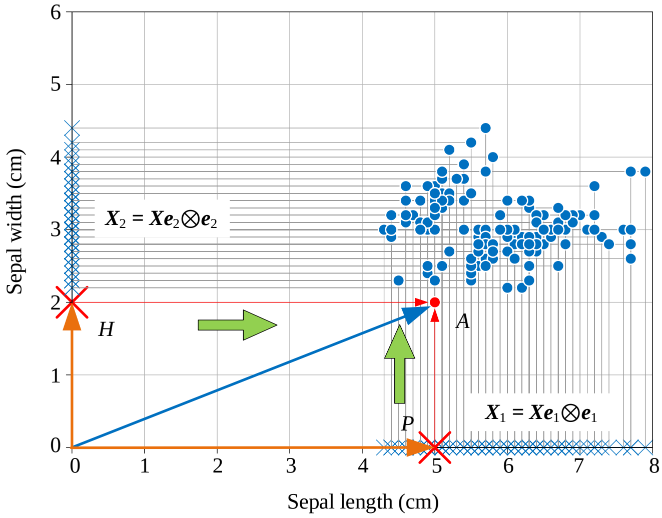


图 18. 数据叠加还原散点图

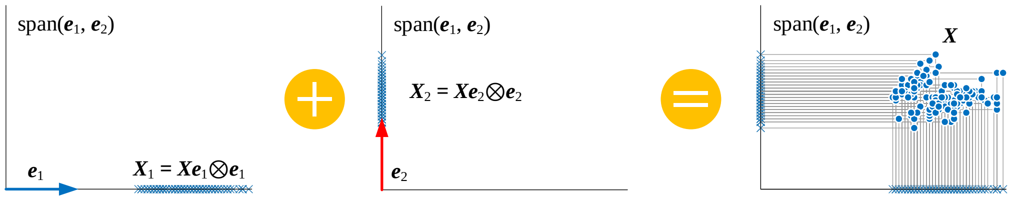


图 19. 数据叠加还原 $X_{150\times 2}$

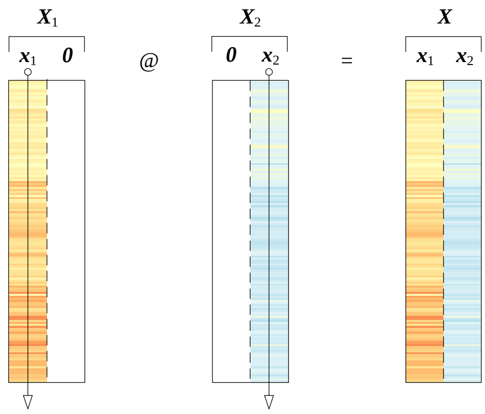


图 20. 数据热图，叠加还原 $X_{150\times 2}$

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 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2]$:

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \tag{33}$$

从几何变换角度来上，$\boldsymbol{V}$ 就是一个旋转矩阵。大家自己很容易计算得到 $\boldsymbol{V}$ 的行列式值为 1，即 $\det(\boldsymbol{V}) = 1$。

如图 21 所示，同样以点 A (5, 2) 为例，A 在 $\boldsymbol{v}_1$ 方向标量投影为:

$$\begin{bmatrix} 5 & 2 \end{bmatrix} \underbrace{\begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}}_{\boldsymbol{v}_1} \approx 5.33 \tag{34}$$

也就是说，A 在 span($\boldsymbol{v}_1$) 投影 H 的坐标值为 5.33，对应向量可以写成 $5.33\boldsymbol{v}_1$。

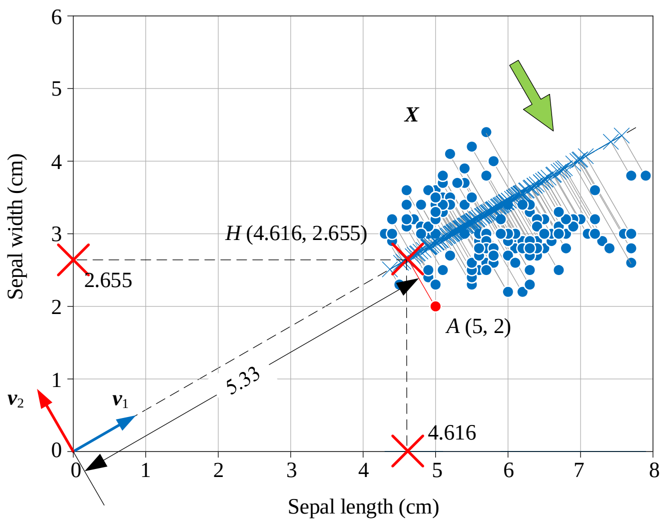


图 21. 二特征数据矩阵 $\boldsymbol{X}_{150\times2}$ 向 $\boldsymbol{v}_1$ 投影

通过二次投影获得 H 在 span($\boldsymbol{e}_1, \boldsymbol{e}_2$) 坐标值:

$$\begin{bmatrix} 5 & 2 \end{bmatrix} \boldsymbol{v}_1 \otimes \boldsymbol{v}_1 = \begin{bmatrix} 5 & 2 \end{bmatrix} \boldsymbol{v}_1 \boldsymbol{v}_1^{\mathrm{T}} = \begin{bmatrix} 5 & 2 \end{bmatrix} \begin{bmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{bmatrix} \approx \begin{bmatrix} 4.616 & 2.665 \end{bmatrix} \tag{35}$$

这就是 H 在图 21 中坐标值。很容易计算，(35) 中 $\boldsymbol{v}_1 \otimes \boldsymbol{v}_1$ 的行列式值为 0，即 $\det(\boldsymbol{v}_1 \otimes \boldsymbol{v}_1) = 0$。

$\boldsymbol{X}_{150\times2}$ 在 $\boldsymbol{v}_1$ 投影 $\boldsymbol{z}_1$ 为:

$$z_1 = Xv_1 = \underbrace{\begin{bmatrix} x_1 & x_2 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix}}_{v_1} \approx 0.866 x_1 + 0.5 x_2 \tag{36}$$

即，z_1 相当于 x_1 和 x_2 的线性组合。

$X_{150\times 2}$ 在 v_1 二次投影结果 X_1 为：

$$X_1 = Xv_1 \otimes v_1 = Xv_1 v_1^{\mathrm{T}} \approx \underbrace{\begin{bmatrix} x_1 & x_2 \end{bmatrix}}_{X} \begin{bmatrix} 0.750 & 0.433 \\ 0.433 & 0.250 \end{bmatrix} = \begin{bmatrix} 0.750 x_1 + 0.433 x_2 & 0.433 x_1 + 0.250 x_2 \end{bmatrix} \tag{37}$$

如图 22 所示，同样以点 A (5, 2) 为例，A 在 v_2 方向标量为：

$$\begin{bmatrix} 5 & 2 \end{bmatrix} \underbrace{\begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix}}_{v_2} \approx -0.7679 \tag{38}$$

即 A 在 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 ×)：

$$\begin{bmatrix} 5 & 2 \end{bmatrix} v_2 \otimes v_2 = \begin{bmatrix} 5 & 2 \end{bmatrix} v_2 v_2^{\mathrm{T}} = \begin{bmatrix} 5 & 2 \end{bmatrix} \begin{bmatrix} 1/4 & -\sqrt{3}/4 \\ -\sqrt{3}/4 & 3/4 \end{bmatrix} \approx \begin{bmatrix} 0.384 & -0.665 \end{bmatrix} \tag{39}$$

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

(35) 和 (39) 之和还原 A 坐标值 (5, 2)：

$$\begin{bmatrix} 5 & 2 \end{bmatrix} (v_1 \otimes v_1 + v_2 \otimes v_2) = \begin{bmatrix} 5 & 2 \end{bmatrix} \left\{ \begin{bmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{bmatrix} + \begin{bmatrix} 1/4 & -\sqrt{3}/4 \\ -\sqrt{3}/4 & 3/4 \end{bmatrix} \right\} = \begin{bmatrix} 5 & 2 \end{bmatrix} \tag{40}$$

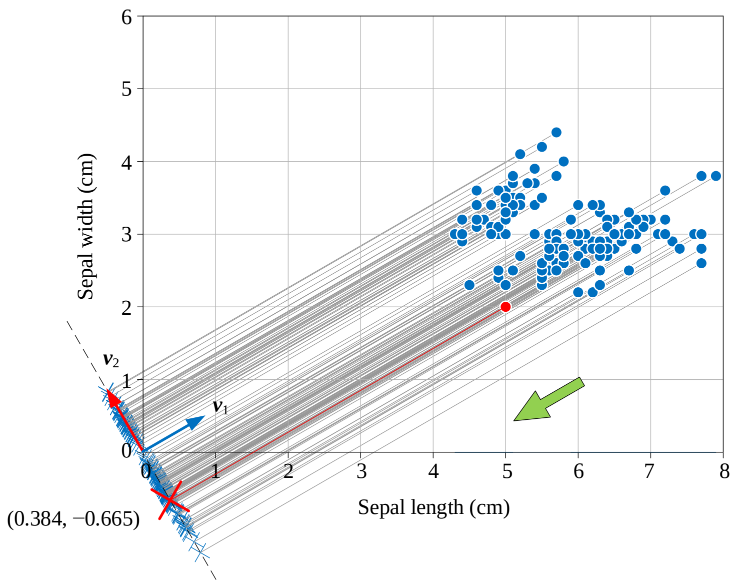


图 22. 二特征数据矩阵 $X_{150\times 2}$ 向 v_2 投影

$\boldsymbol{X}_{150\times 2}$ 在 $\boldsymbol{v}_2$ 投影 $\boldsymbol{z}_2$ 为：

$$\boldsymbol{z}_2 = \boldsymbol{X}\boldsymbol{v}_2 = \underbrace{\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix}}_{\boldsymbol{X}} \underbrace{\begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix}}_{\boldsymbol{v}_2} \approx -0.5\boldsymbol{x}_1 + 0.866\boldsymbol{x}_2 \tag{41}$$

$\boldsymbol{z}_2$ 也是 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的线性组合。

$\boldsymbol{X}_{150\times 2}$ 在 $\boldsymbol{v}_2$ 二次投影 $\boldsymbol{X}_2$ 为：

$$\boldsymbol{X}_2 = \boldsymbol{X}\boldsymbol{v}_2 \otimes \boldsymbol{v}_2 = \boldsymbol{X}\boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}} \approx \underbrace{\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix}}_{\boldsymbol{X}} \begin{bmatrix} 0.250 & -0.433 \\ -0.433 & 0.750 \end{bmatrix} = \begin{bmatrix} 0.250\boldsymbol{x}_1 - 0.433\boldsymbol{x}_2 & -0.433\boldsymbol{x}_1 + 0.750\boldsymbol{x}_2 \end{bmatrix} \tag{42}$$

(37) 和 (42) 叠加还原 $\boldsymbol{X}$：

$$\boldsymbol{X}_1 + \boldsymbol{X}_2 = \boldsymbol{X}\boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \boldsymbol{X}\boldsymbol{v}_2 \otimes \boldsymbol{v}_2 = \boldsymbol{X}\left\{\begin{bmatrix} 0.750 & 0.433 \\ 0.433 & 0.250 \end{bmatrix} + \begin{bmatrix} 0.250 & -0.433 \\ -0.433 & 0.750 \end{bmatrix}\right\} = \boldsymbol{X} \tag{43}$$

第二个规范正交基

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

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \tag{44}$$

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

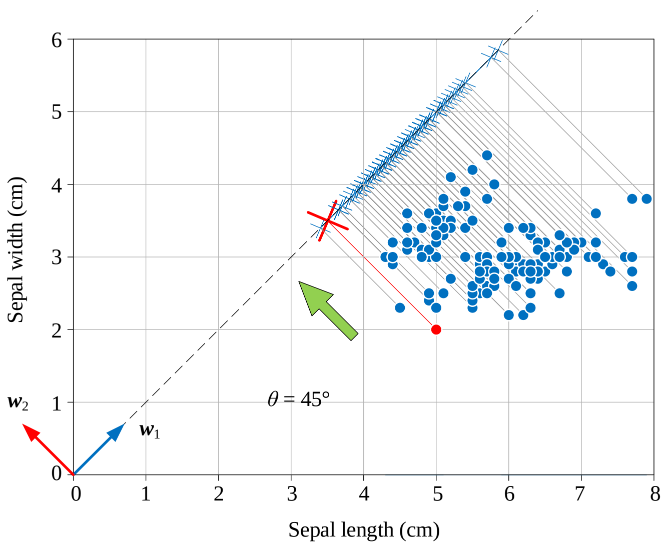


图 23. 二特征数据矩阵 $\boldsymbol{X}_{150\times 2}$ 向 $\boldsymbol{w}_1$ 投影

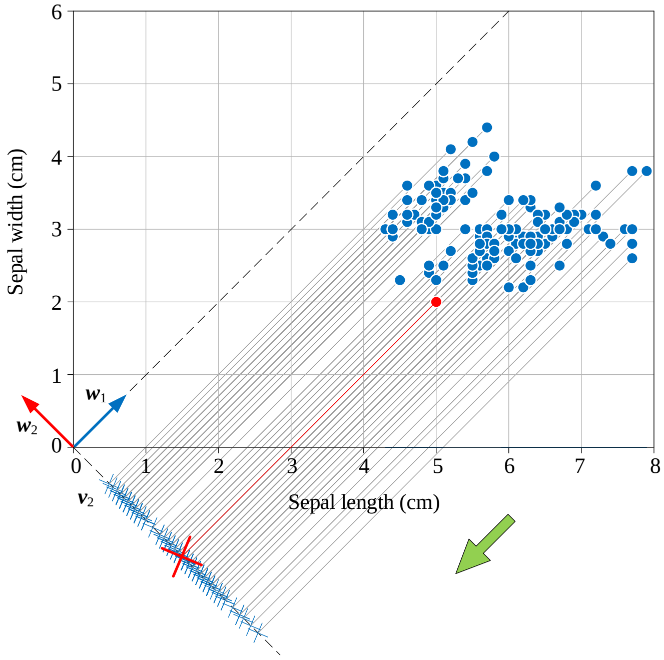


图 24. 二特征数据矩阵 $X_{150\times2}$ 向 w_2 投影

第三个规范正交基

给定如下规范正交基 $U = [u_1, u_2]$:

$$
U = \begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \tag{45}
$$

图 25 和图 26 所示为二特征数据矩阵 $X_{150\times2}$ 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。

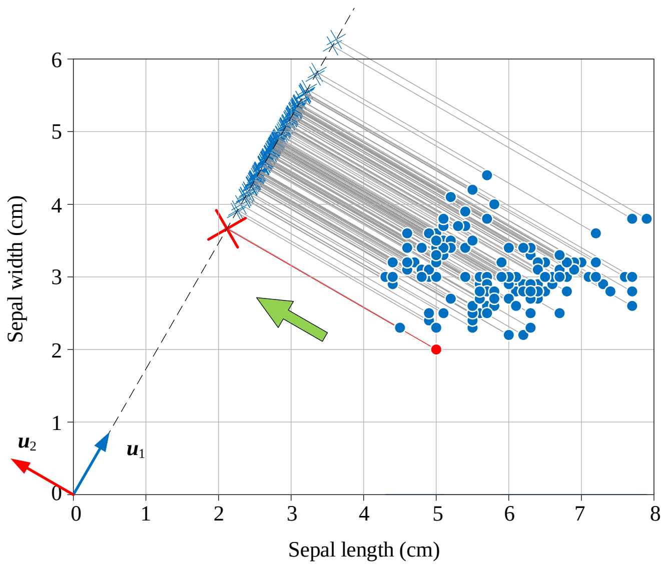


图 25. 二特征数据矩阵 $X_{150\times2}$ 向 u_1 投影

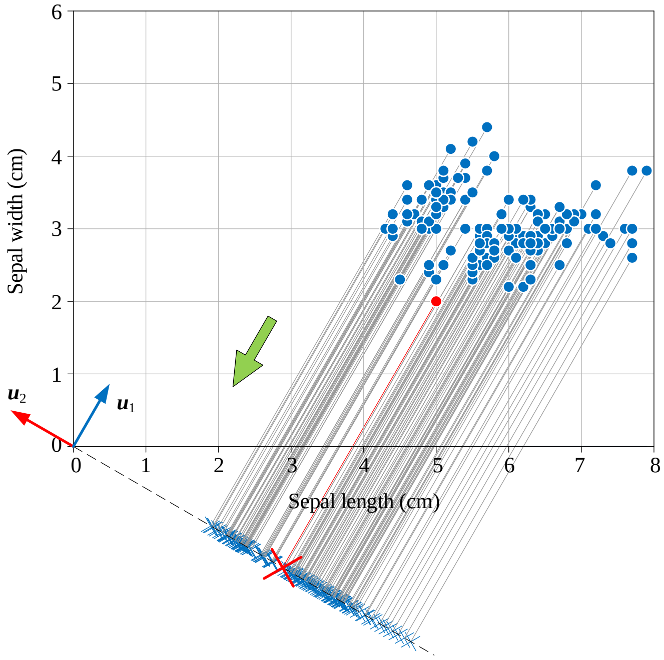


图 26. 二特征数据矩阵 $\boldsymbol{X}_{150\times2}$ 向 $\boldsymbol{u}_2$ 投影

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，扩展前文分析思路。

本节先从最简单的标准正交基 $[\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_D]$ 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 $\boldsymbol{X}$ 在 $\boldsymbol{e}_1$ 方向上标量投影的运算热图。

从行向量角度来看，$\boldsymbol{x}^{(i)}\boldsymbol{e}_1 \rightarrow x_{i,1}$ 代表 $\mathbb{R}^D$ 空间坐标值 $\boldsymbol{x}^{(i)}$投影到 span($\boldsymbol{e}_1$) 这个子空间后，结果坐标值为 $x_{i,1}$。

⚠ 再次强调，$x_{i,1}$ 是 $\boldsymbol{x}^{(i)}$ 在 span($\boldsymbol{e}_1$) 的坐标值。

从列向量角度来看，$[\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4]\boldsymbol{e}_1 \rightarrow \boldsymbol{x}_1$，是一个线性组合过程。而 $\boldsymbol{e}_1 = [1, 0, 0, 0]^T$，所以组合的结果只保留了鸢尾花数据集第一列 $\boldsymbol{x}_1$，即花萼长度，这个特征的所有样本数据。

请大家按照这个思路分析图 27 (b)、(c)、(d)三幅热图运算。

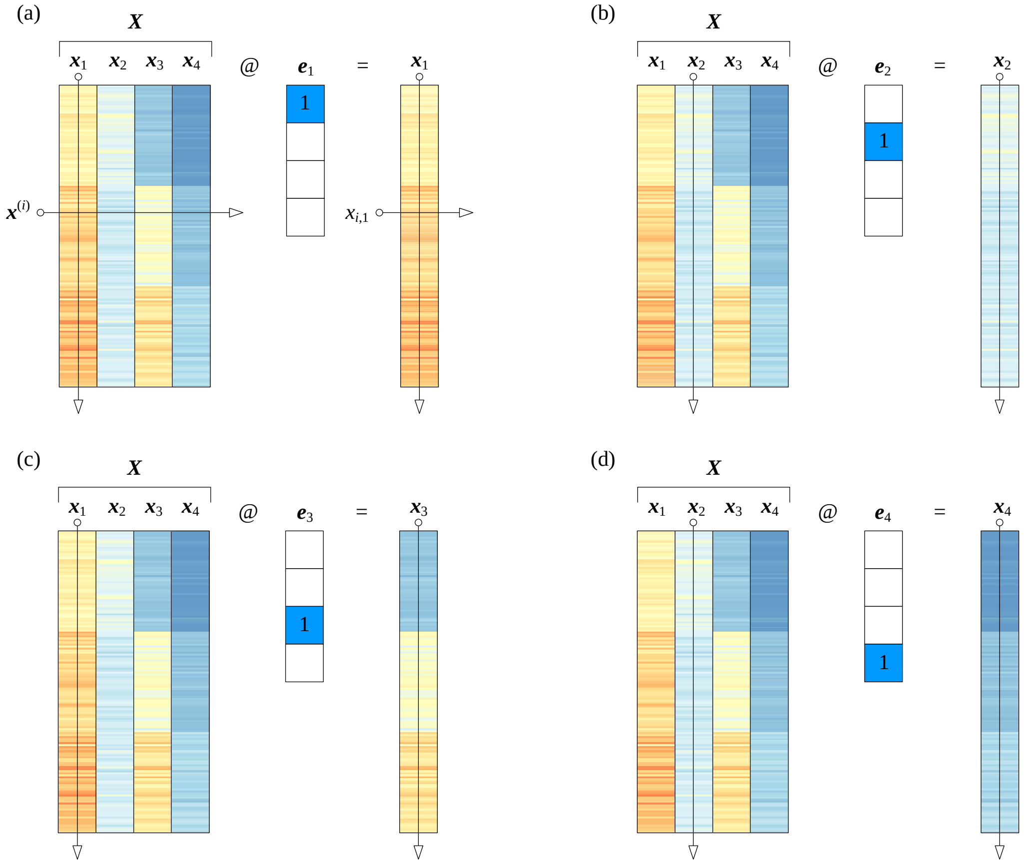


图 27. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 分别向 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$、$\boldsymbol{e}_3$、$\boldsymbol{e}_4$ 投影，一次投影

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，$\boldsymbol{X}$ 向 $\boldsymbol{e}_1$ 方向向量投影结果就是 $\boldsymbol{X}$ 和 $\boldsymbol{e}_1 \otimes \boldsymbol{e}_1$ 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 $\boldsymbol{e}_1 \otimes \boldsymbol{e}_1$、$\boldsymbol{e}_2 \otimes \boldsymbol{e}_2$、$\boldsymbol{e}_3 \otimes \boldsymbol{e}_3$、$\boldsymbol{e}_4 \otimes \boldsymbol{e}_4$ 的行列式值都为 0。

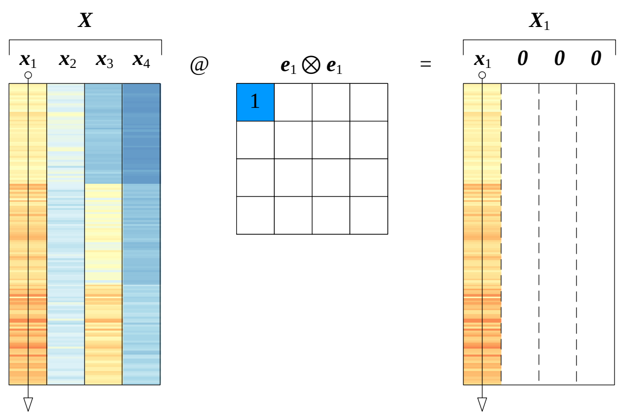


图 28. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{e}_1$ 方向向量投影，二次投影

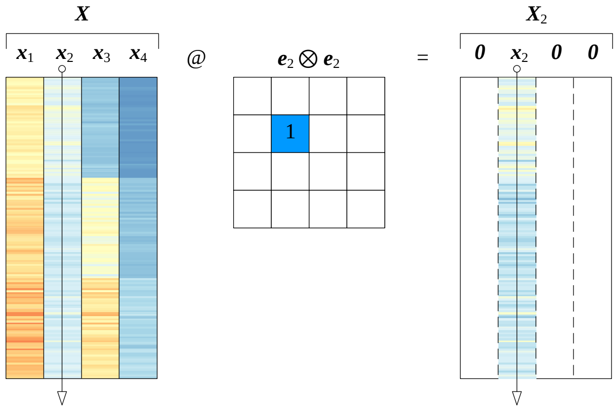


图 29. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{e}_2$ 方向向量投影，二次投影

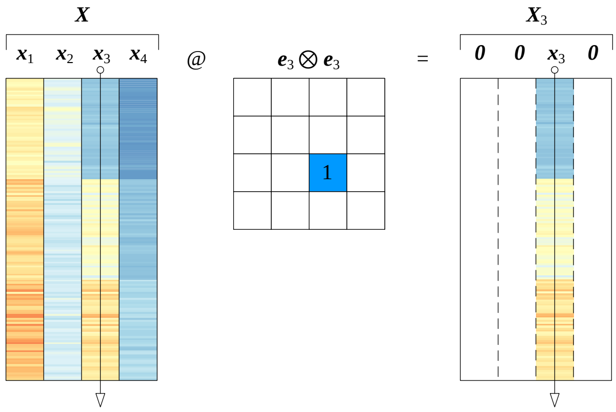


图 30. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{e}_3$ 方向向量投影，二次投影

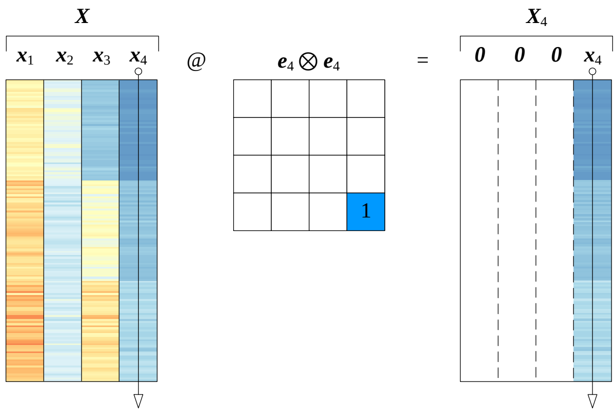


图 31. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{e}_4$ 方向向量投影，二次投影

两个方向投影

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两个方向投影。

如图 32 所示，$\boldsymbol{X}$ 向 $[\boldsymbol{e}_1, \boldsymbol{e}_2]$ 方向标量投影，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。

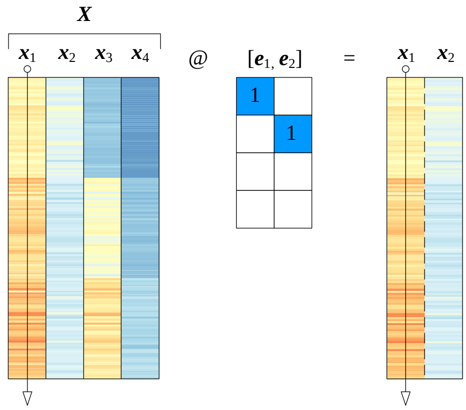


图 32. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $[\boldsymbol{e}_1, \boldsymbol{e}_2]$ 方向标量投影

图 33 所示为 $\boldsymbol{X}$ 向 $[\boldsymbol{e}_1, \boldsymbol{e}_2]$ 方向向量投影，结果相当于图 28 和图 29 结果“叠加”，即 $\boldsymbol{X}_1 + \boldsymbol{X}_2$。很明显，$\boldsymbol{X}_1 + \boldsymbol{X}_2$ 并没有还原 $\boldsymbol{X}$。

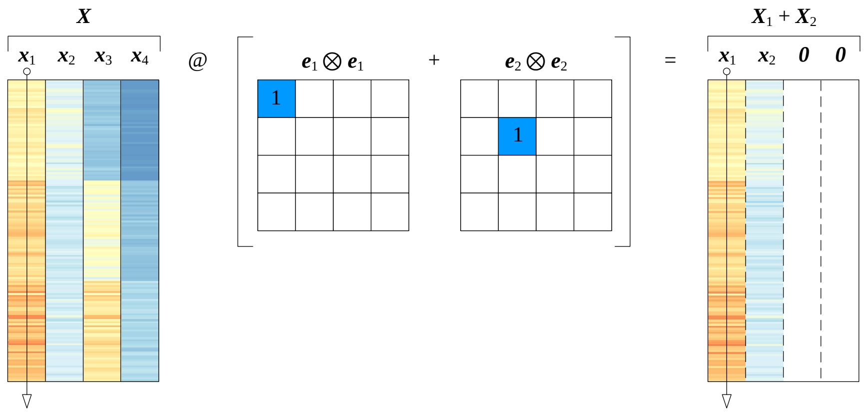


图 33. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $[\boldsymbol{e}_1, \boldsymbol{e}_2]$ 方向向量投影

层层叠加：还原原始矩阵

本章前文 (12) 告诉我们，数据矩阵 $\boldsymbol{X}$ 在规范正交基 $[\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_D]$ 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 $[\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_D]$ 可以视作特殊的规范正交基。

观察图 34 得知，要想完整还原 $\boldsymbol{X}$，需要图 28、图 29、图 30、图 31 四副热图叠加，即 $\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3 + \boldsymbol{X}_4$。显然，$\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$、$\boldsymbol{X}_4$ 这四个矩阵的秩都是 1。

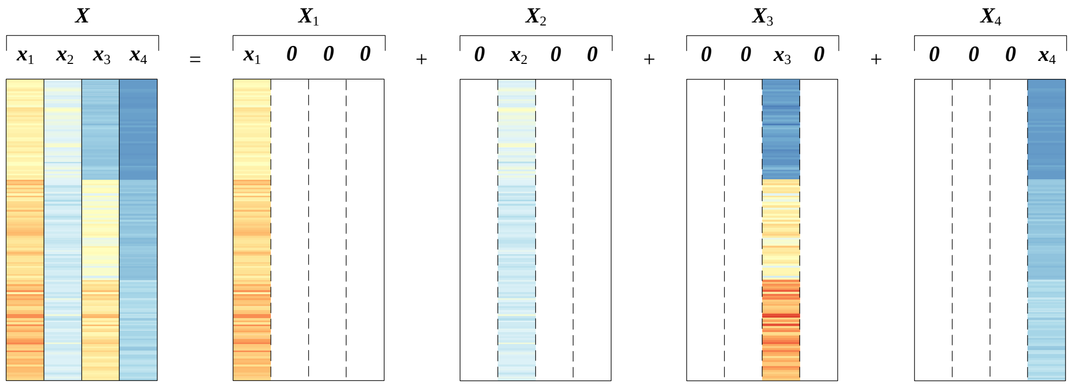


图 34. 投影数据矩阵的层层叠加还原数据矩阵 $\boldsymbol{X}_{150\times4}$

图 35 是张量积层层叠加得到单位矩阵 $\boldsymbol{I}$，它是数据还原的另外一个侧面：

$$\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_2 + \boldsymbol{e}_3 \otimes \boldsymbol{e}_3 + \boldsymbol{e}_4 \otimes \boldsymbol{e}_4 = \boldsymbol{I} \tag{46}$$

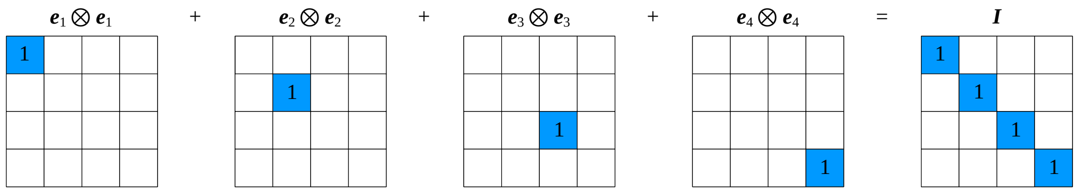


图 35. 张量积的层层叠加还原 4 × 4 单位矩阵

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。

一个“无数里挑一”的规范正交基

先别问怎么办到的。假设我们恰好找到了一个 4 × 4 规范正交基 $\boldsymbol{V}$，具体如下：

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 \end{bmatrix} = \begin{bmatrix} 0.751 & 0.284 & 0.502 & 0.321 \\ 0.380 & 0.547 & -0.675 & -0.317 \\ 0.513 & -0.709 & -0.059 & -0.481 \\ 0.168 & -0.344 & -0.537 & 0.752 \end{bmatrix} \tag{47}$$

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

从几何角度来看，规范正交基 $\boldsymbol{V}$ 对应的几何操作是四维空间旋转。

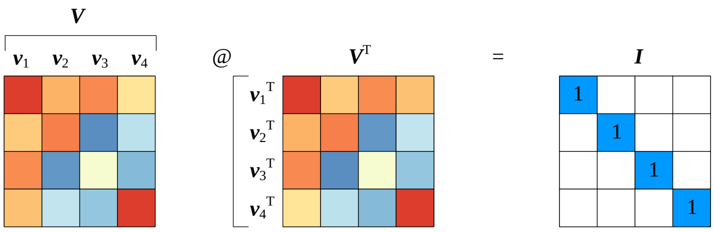


图 36. 规范正交基 $\boldsymbol{V}$ 乘其转置得到 4×4 单位矩阵

$\boldsymbol{V}$ 中的像

如图 37 所示，直接将 $\boldsymbol{X}$ 投影到规范正交基 $\boldsymbol{V}$，得到 $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{V}$。$\boldsymbol{Z}$ 就是 $\boldsymbol{X}$ 在 $\boldsymbol{V}$ 中的像，根据 $\boldsymbol{X}\boldsymbol{v}_j = \boldsymbol{z}_j$，下面我们逐一分析矩阵 $\boldsymbol{Z}$ 的列向量。

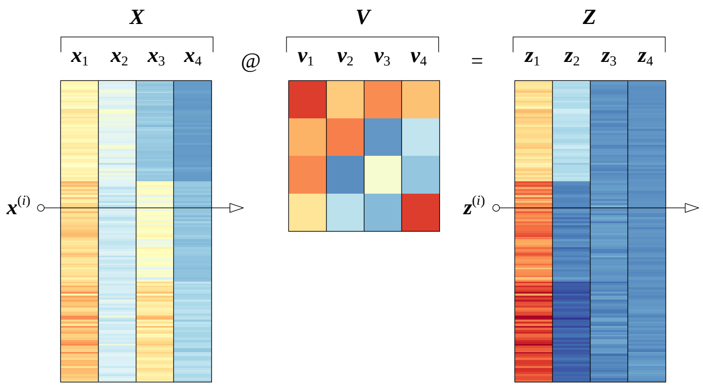


图 37. 四特征数据矩阵 $\boldsymbol{X}_{150\times 4}\boldsymbol{X}$ 投影到规范正交基 $\boldsymbol{V}$ 得到 $\boldsymbol{Z}$

第 1 列向量 $\boldsymbol{v}_1$

图 38 所示为鸢尾花数据集矩阵 $\boldsymbol{X}$ 在 $\boldsymbol{v}_1$ 方向上标量投影的运算热图。

从行向量角度来看，$\boldsymbol{x}^{(i)}\boldsymbol{v}_1 \to z_{i,1}$ 代表 $\mathbb{R}^D$ 空间坐标值 $\boldsymbol{x}^{(i)}$ 投影到 $\mathrm{span}(\boldsymbol{v}_1)$ 这个子空间后坐标值变成 $z_{i,1}$。

从列向量角度来看，$[\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4]\boldsymbol{v}_1 \to \boldsymbol{z}_1$，是一个线性组合过程，即：

$$z_1 = Xv_1 = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} \begin{bmatrix} 0.751 \\ 0.380 \\ 0.513 \\ 0.168 \end{bmatrix} = 0.751x_1 + 0.380x_2 + 0.513x_3 + 0.168x_4 \tag{48}$$

上式说明，0.7512 倍 $\boldsymbol{x}_1$、0.380 倍 $\boldsymbol{x}_2$、0.513 倍 $\boldsymbol{x}_3$、0.168 倍 $\boldsymbol{x}_4$ 合成得到了向量 $\boldsymbol{z}_1$。

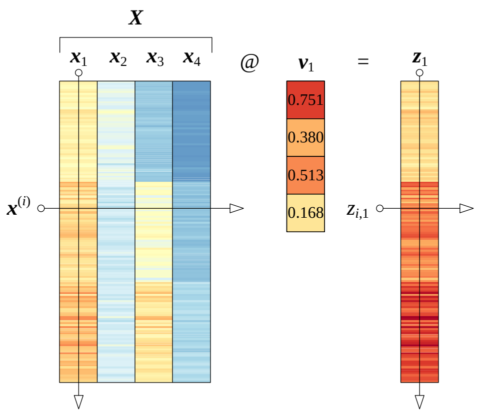


图 38. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}\boldsymbol{X}$ 向 $\boldsymbol{v}_1$ 方向标量投影，一次投影

如图 39 所示，$\boldsymbol{z}_1$ 再乘 $\boldsymbol{v}_1^{\mathrm{T}}$，便得到 $\boldsymbol{X}_1$。不难理解，$\boldsymbol{X}_1$ 的每一列都是 $\boldsymbol{z}_1$ 乘一个标量系数。显然，$\boldsymbol{X}_1$ 的秩为 1，即 $\mathrm{rank}(\boldsymbol{X}_1) = 1$。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。

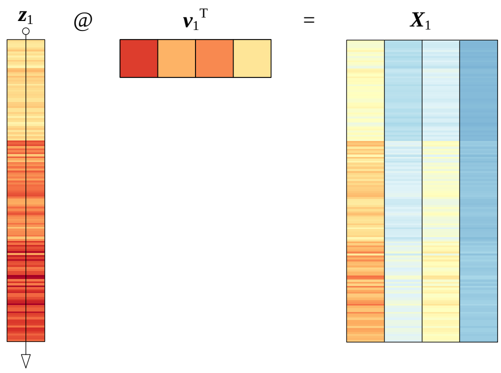


图 39. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}\boldsymbol{z}_1$ 乘 $\boldsymbol{v}_1^{\mathrm{T}}$ 得到 $\boldsymbol{X}_1$

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 $\boldsymbol{v}_1 \otimes \boldsymbol{v}_1$：

$$v_1 \otimes v_1 = v_1 v_1^{\mathrm{T}} = \begin{bmatrix} 0.751 \\ 0.380 \\ 0.513 \\ 0.168 \end{bmatrix} @ \begin{bmatrix} 0.751 \\ 0.380 \\ 0.513 \\ 0.168 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 0.564 & 0.285 & 0.385 & 0.126 \\ 0.285 & 0.144 & 0.194 & 0.063 \\ 0.385 & 0.194 & 0.263 & 0.086 \\ 0.126 & 0.063 & 0.086 & 0.028 \end{bmatrix} \tag{49}$$

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，(49) 上式仅仅保留小数点后 3 位数值。

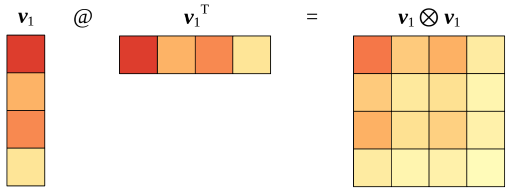


图 40. 计算张量积 $\boldsymbol{v}_1 \otimes \boldsymbol{v}_1$

图 41 所示为 $\boldsymbol{X}$ 和张量积 $\boldsymbol{v}_1 \otimes \boldsymbol{v}_1$ 乘积。几何视角，即 $\boldsymbol{X}$ 向 $\boldsymbol{v}_1$ 方向向量投影得到 $\boldsymbol{X}_1$，即所谓"二次投影"。

➔ 请大家特别注意一点，$\boldsymbol{X}$ 和 $\boldsymbol{X}_1$ 在热图上已经非常接近。这是因为我们在设定 $\boldsymbol{v}_1$ 时，有特殊的"讲究"。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 $\boldsymbol{v}_1$。

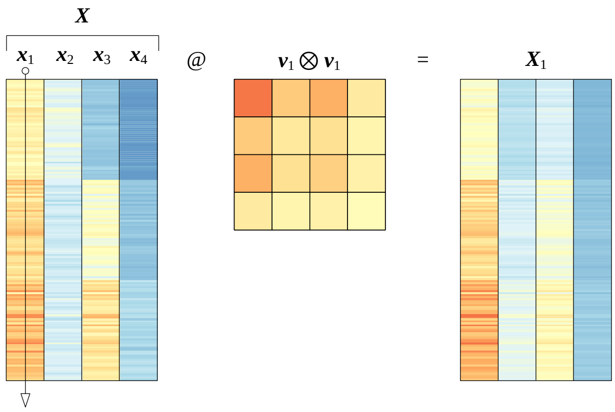


图 41. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{v}_1$ 方向向量投影，二次投影

第 2 列向量 $\boldsymbol{v}_2$

图 42 展示获得 $\boldsymbol{z}_2$ 和 $\boldsymbol{X}_2$ 的过程。请大家根据之前分析 $\boldsymbol{v}_1$ 的思路自行分析这两图。

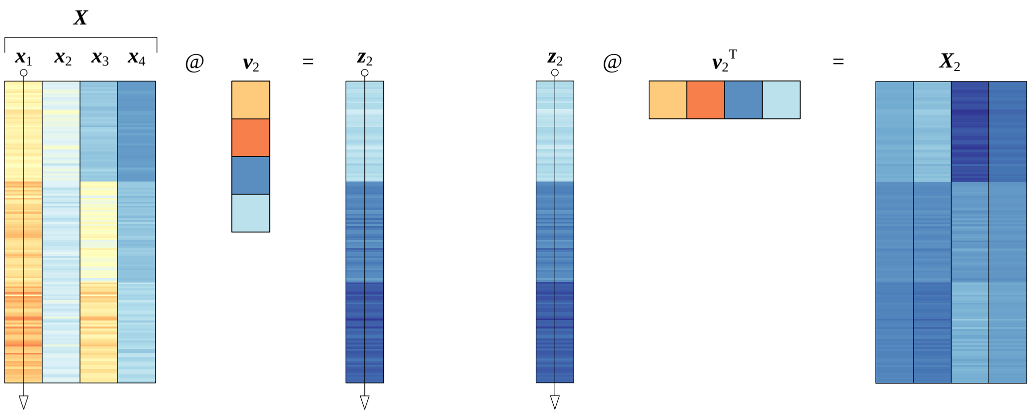


图 42. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}\boldsymbol{X}$ 向 $\boldsymbol{v}_2$ 投影，一次投影，二次投影

同样，利用张量积完成 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{v}_2$ 二次投影。大家自行计算张量积 $\boldsymbol{v}_2 \otimes \boldsymbol{v}_2$ 具体值，按照前文思路分析图 43。有必要指出一点，对比 $\boldsymbol{X}_1$，$\boldsymbol{X}_2$ 热图和 $\boldsymbol{X}$ 相差很大。

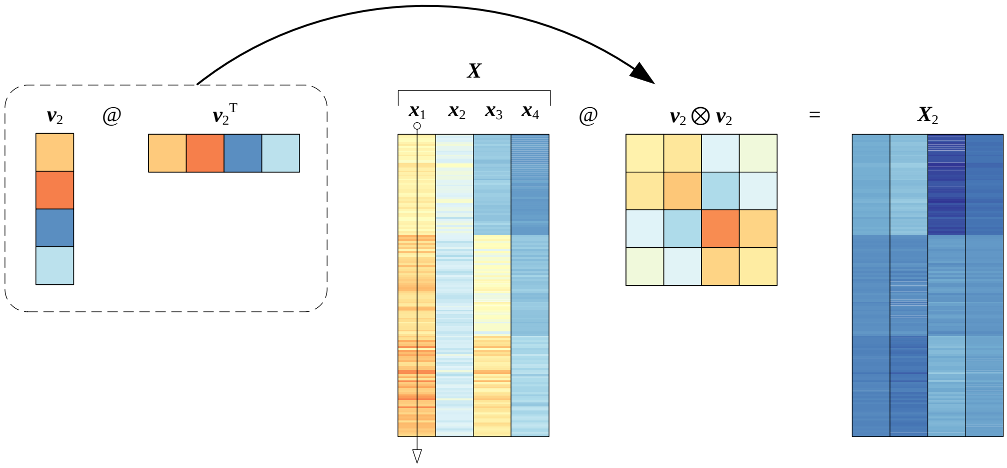


图 43. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{v}_2$ 投影，二次投影

第 3 列向量 $\boldsymbol{v}_3$

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。

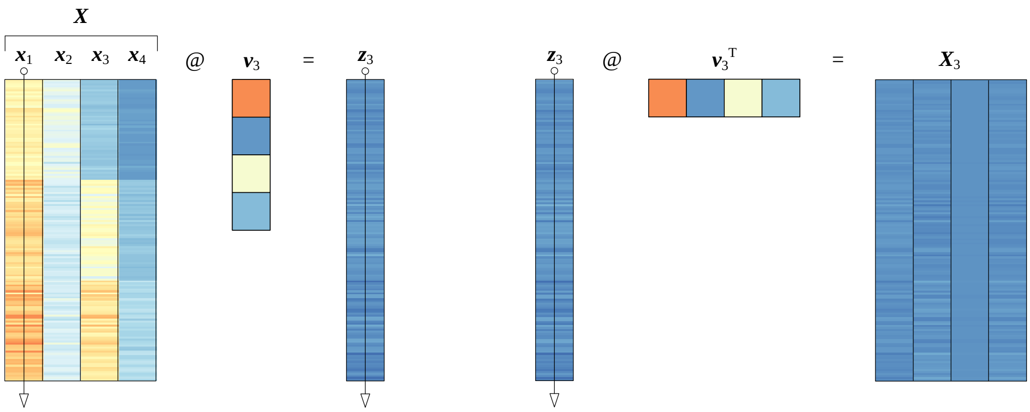


图 44. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{v}_3$ 投影，一次投影，二次投影

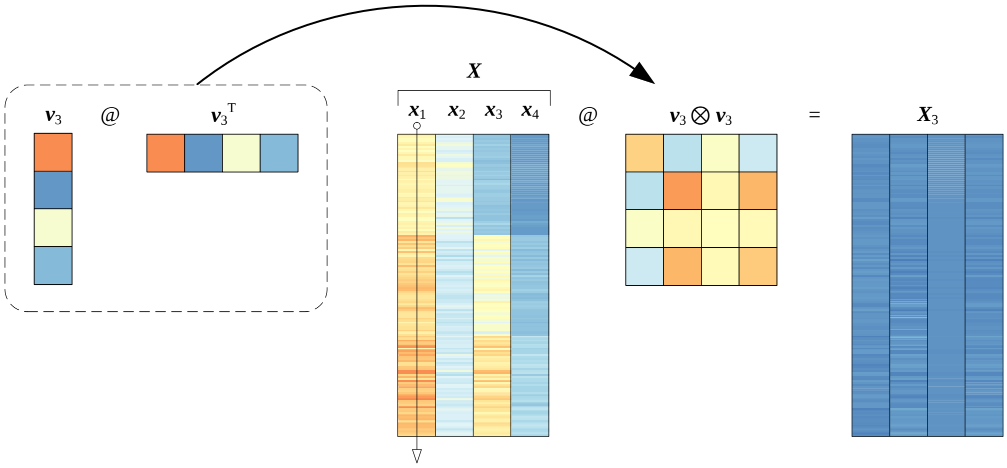


图 45. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{v}_3$ 投影，二次投影

第 4 列向量 $\boldsymbol{v}_4$

大家自行分析图 46、图 47。特别注意比较 $\boldsymbol{X}$ 和 $\boldsymbol{X}_4$ 的热图差异。

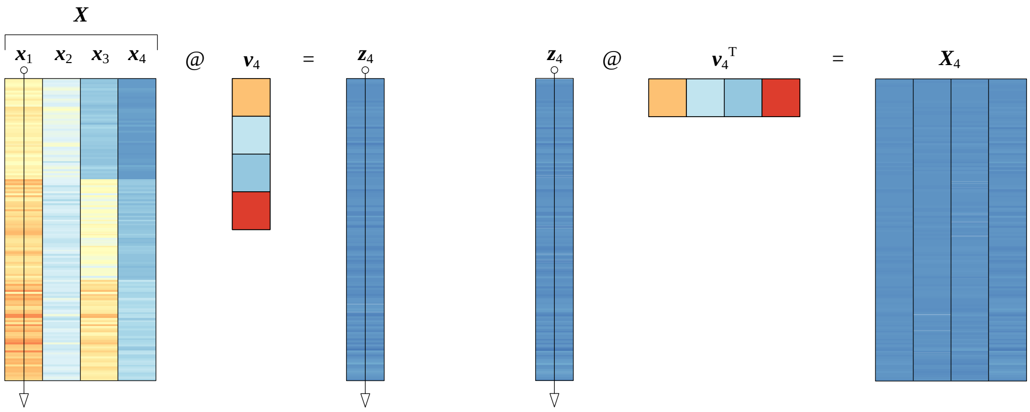


图 46. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{v}_4$ 投影，一次投影和二次投影

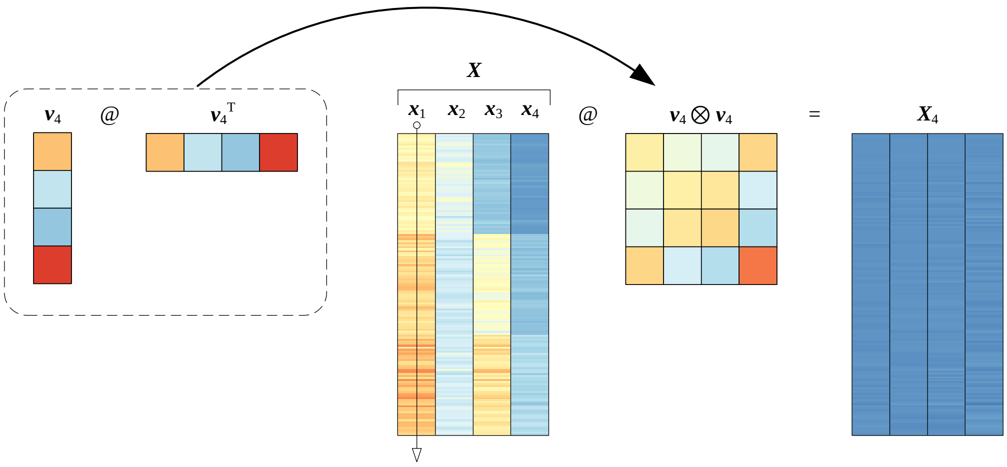


图 47. 四特征数据矩阵 $\boldsymbol{X}_{150\times4}$ 向 $\boldsymbol{v}_4$ 投影，二次投影

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 $\boldsymbol{X}$ 在规范正交基 $[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$ 中每个方向上向量投影层层叠加完全还原原始数据。

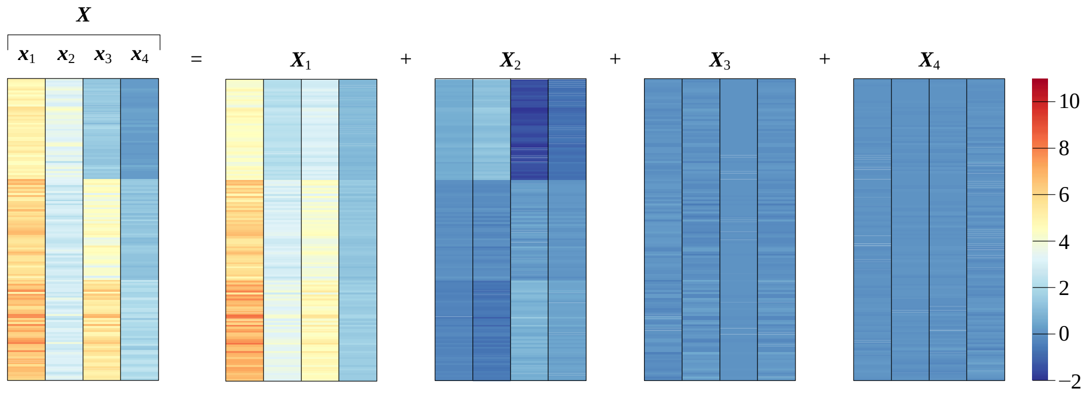


图 48. 层层叠加还原四特征数据矩阵 $\boldsymbol{X}_{150\times4}$

图 48 告诉我们，要想完整还原 $\boldsymbol{X}$，需要四副热图叠加，即 $\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3 + \boldsymbol{X}_4$。我们已经很清楚 $\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$、$\boldsymbol{X}_4$ 这四个矩阵的秩都是 1，而 $\boldsymbol{X}$ 就是这四个秩为 1 的不同矩阵层层叠加之和。

前文已经提到 $\boldsymbol{X}_1$ 已经非常接近 $\boldsymbol{X}$。也就是说，我们可以用 $\boldsymbol{X}_1$ 近似 $\boldsymbol{X}$。特别考虑到 $\boldsymbol{X}_1$ 的秩为 1，即 $\text{rank}(\boldsymbol{X}_1) = 1$。也就是说，$\boldsymbol{X}_1$ 的四个列向量之间存在倍数关系，即，

$$\boldsymbol{X}_1 = \boldsymbol{z}_1\boldsymbol{v}_1^{\mathrm{T}} = \boldsymbol{z}_1\begin{bmatrix} 0.751 & 0.380 & 0.513 & 0.168 \end{bmatrix} = \begin{bmatrix} 0.751\boldsymbol{z}_1 & 0.380\boldsymbol{z}_1 & 0.513\boldsymbol{z}_1 & 0.168\boldsymbol{z}_1 \end{bmatrix} \tag{50}$$

建议大家仔细对比图 48 中 $\boldsymbol{X}$、$\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$、$\boldsymbol{X}_4$ 这五幅热图色差，它们采用完全相同的色谱。

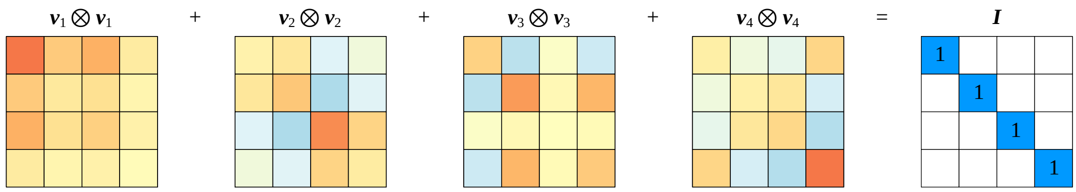


图 49. 张量积层层累加获得 4 × 4 单位矩阵

如图 49 所示，这四个张量积层层叠加得到单位矩阵，即：

$$\boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \boldsymbol{v}_3 \otimes \boldsymbol{v}_3 + \boldsymbol{v}_4 \otimes \boldsymbol{v}_4 = \boldsymbol{I} \tag{51}$$

如前文所述，(51) 是数据还原的另外一个侧面。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 (51) 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。

容易计算 $v_1 \otimes v_1$、$v_2 \otimes v_2$、$v_3 \otimes v_3$、$v_4 \otimes v_4$ 的行列式值都为 0。

Bk4_Ch10_01.py 绘制本章前文大部分热图。

10.7 数据正交化

两个格拉姆矩阵

本节再回过头来分析图 37 中数据矩阵 Z。如图 50 所示，矩阵 Z 转置乘自身得到 Z 的格拉姆矩阵：

$$Z^{\mathrm{T}} Z = \begin{bmatrix} z_1^{\mathrm{T}} \\ z_2^{\mathrm{T}} \\ \vdots \\ z_D^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} z_1 & z_2 & \cdots & z_D \end{bmatrix} = \begin{bmatrix} z_1^{\mathrm{T}} z_1 & z_1^{\mathrm{T}} z_2 & \cdots & z_1^{\mathrm{T}} z_D \\ z_2^{\mathrm{T}} z_1 & z_2^{\mathrm{T}} z_2 & \cdots & z_2^{\mathrm{T}} z_D \\ \vdots & \vdots & \ddots & \vdots \\ z_D^{\mathrm{T}} z_1 & z_D^{\mathrm{T}} z_2 & \cdots & z_D^{\mathrm{T}} z_D \end{bmatrix} \tag{52}$$

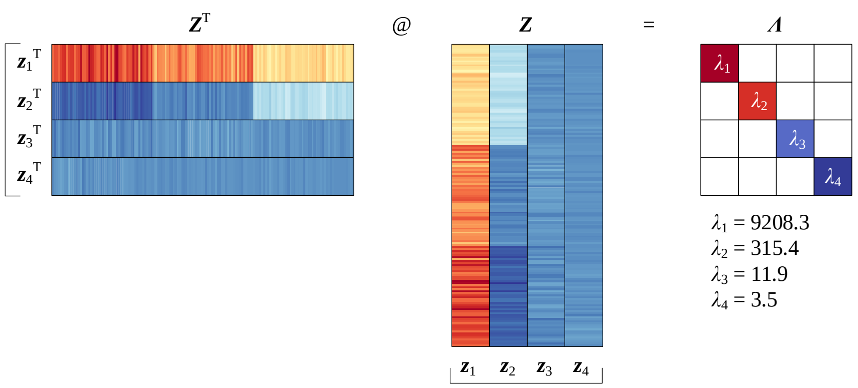


图 50. 矩阵 Z 的格拉姆矩阵

(52) 写成向量内积形式：

$$
Z^{\mathrm{T}}Z=\begin{bmatrix} z_1\cdot z_1 & z_1\cdot z_2 & \cdots & z_1\cdot z_D \\ z_2\cdot z_1 & z_2\cdot z_2 & \cdots & z_2\cdot z_D \\ \vdots & \vdots & \ddots & \vdots \\ z_D\cdot z_1 & z_D\cdot z_2 & \cdots & z_D\cdot z_D \end{bmatrix}=\begin{bmatrix} \langle z_1,z_1\rangle & \langle z_1,z_2\rangle & \cdots & \langle z_1,z_D\rangle \\ \langle z_2,z_1\rangle & \langle z_2,z_2\rangle & \cdots & \langle z_2,z_D\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle z_D,z_1\rangle & \langle z_D,z_2\rangle & \cdots & \langle z_D,z_D\rangle \end{bmatrix} \tag{53}
$$

观察图 50，发现 $Z^{\mathrm{T}}Z$ 恰好是对角方阵，即：

$$
Z^{\mathrm{T}}Z=\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_D \end{bmatrix}=\Lambda \tag{54}
$$

这说明，Z 的列向量两两正交，即：

$$
z_i^{\mathrm{T}}z_j=z_j^{\mathrm{T}}z_i=z_i\cdot z_j=z_j\cdot z_i=\langle z_i,z_j\rangle=\langle z_j,z_i\rangle=0,\quad i\neq j \tag{55}
$$

对比 X 的格拉姆矩阵：

$$
\begin{aligned}
G=X^{\mathrm{T}}X&=\begin{bmatrix} x_1^{\mathrm{T}} \\ x_2^{\mathrm{T}} \\ \vdots \\ x_D^{\mathrm{T}} \end{bmatrix}\begin{bmatrix} x_1 & x_2 & \cdots & x_D \end{bmatrix}=\begin{bmatrix} x_1^{\mathrm{T}}x_1 & x_1^{\mathrm{T}}x_2 & \cdots & x_1^{\mathrm{T}}x_D \\ x_2^{\mathrm{T}}x_1 & x_2^{\mathrm{T}}x_2 & \cdots & x_2^{\mathrm{T}}x_D \\ \vdots & \vdots & \ddots & \vdots \\ x_D^{\mathrm{T}}x_1 & x_D^{\mathrm{T}}x_2 & \cdots & x_D^{\mathrm{T}}x_D \end{bmatrix} \\
&=\begin{bmatrix} x_1\cdot x_1 & x_1\cdot x_2 & \cdots & x_1\cdot x_D \\ x_2\cdot x_1 & x_2\cdot x_2 & \cdots & x_2\cdot x_D \\ \vdots & \vdots & \ddots & \vdots \\ x_D\cdot x_1 & x_D\cdot x_2 & \cdots & x_D\cdot x_D \end{bmatrix}=\begin{bmatrix} \langle x_1,x_1\rangle & \langle x_1,x_2\rangle & \cdots & \langle x_1,x_D\rangle \\ \langle x_2,x_1\rangle & \langle x_2,x_2\rangle & \cdots & \langle x_2,x_D\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_D,x_1\rangle & \langle x_D,x_2\rangle & \cdots & \langle x_D,x_D\rangle \end{bmatrix}
\end{aligned} \tag{56}
$$

图 51 所示为计算矩阵 X 的格拉姆矩阵的热图。此外，请大家注意一点，图 51 中矩阵 G 的迹，即对角线元素之和，$\mathrm{tr}(G)$ = 9539.29。而图 50 中矩阵 Λ 的迹和 G 的迹相同，$\mathrm{tr}(G)$ = $\mathrm{tr}(\Lambda)$ = 9539.29。本书后面还会反复提到这一点。

V 因 X 而生

细细想来，本章介绍的 $Z = XV$ 的数据转换很神奇！

还是以鸢尾花数据为例，如图 51 所示，矩阵 X 的格拉姆矩阵 G 主对角线以外元素代表任意两个向量的内积。G 中没有一个元素为 0。

但是经过数据转换 $Z = XV$，如图 50 所示，矩阵 Z 的格拉姆矩阵 Λ，主对角线以外元素都为 0。z_i 和 z_j 都是长度为 150 的列向量，两者的向量内积竟然为 0，也就是说 150 个成对元素乘积之和为 0！

对于鸢尾花数据矩阵 X 来说，(47) 中给出的这个 V 真可谓"无数里挑一"，！

换句话说，V 因 X 而生！

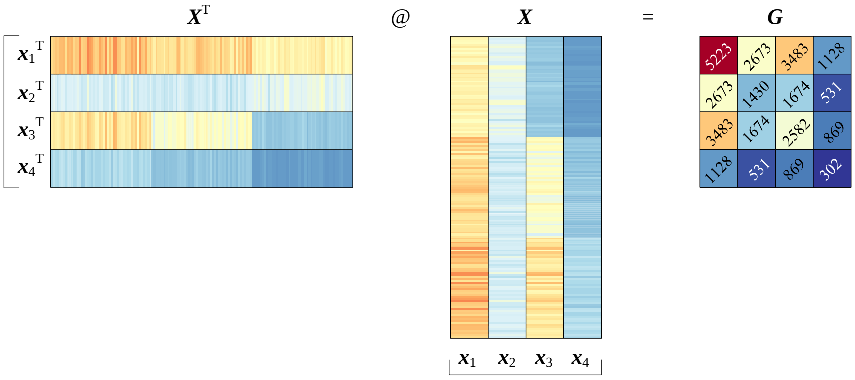


图 51. 矩阵 X 的格拉姆矩阵

对角化

将 $Z = XV$ 其代入 (54) 得到：

$$Z^TZ = (XV)^T XV = V^T \underbrace{X^TX}_{G} V = V^TGV = \Lambda \tag{57}$$

再进一步，由于 V 为规范正交基，因此 $V^TV = I$，根据 (57) 等式关系，G 可以写成：

$$G = V\Lambda V^T \tag{58}$$

这就是说，如图 52 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，Λ 为对角方阵。从 G 到 Λ 也是一个方阵**对角化** (diagonalization) 的过程。

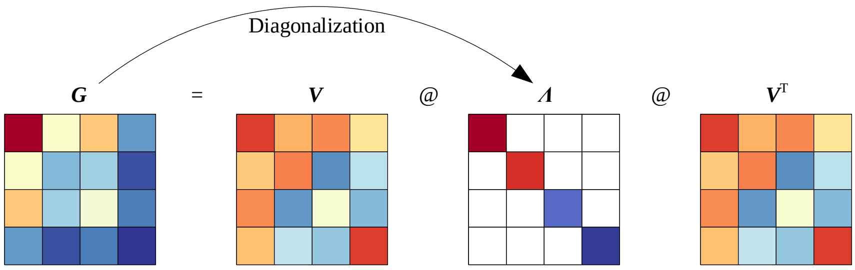


图 52. 对 G 矩阵分解

为了获得 (58) 等式，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 (48)，我们可以把 z_1、z_2、z_3、z_4 写成如下线性组合：

$$
\begin{aligned}
z_1 &= Xv_1 = 0.751x_1 + 0.380x_2 + 0.513x_3 + 0.168x_4 \\
z_2 &= Xv_2 = 0.284x_1 + 0.547x_2 - 0.709x_3 - 0.344x_4 \\
z_3 &= Xv_3 = 0.502x_1 - 0.675x_2 - 0.059x_3 - 0.537x_4 \\
z_4 &= Xv_4 = 0.321x_1 - 0.317x_2 - 0.481x_3 + 0.752x_4 \\
V &= \begin{bmatrix} 0.751 & 0.284 & 0.502 & 0.321 \\ 0.380 & 0.547 & -0.675 & -0.317 \\ 0.513 & -0.709 & -0.059 & -0.481 \\ 0.168 & -0.344 & -0.537 & 0.752 \end{bmatrix}
\end{aligned}
\tag{59}
$$

请大家格外注意 (59) 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 53 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 53 中阴影背景子图为例介绍如何理解双标图。

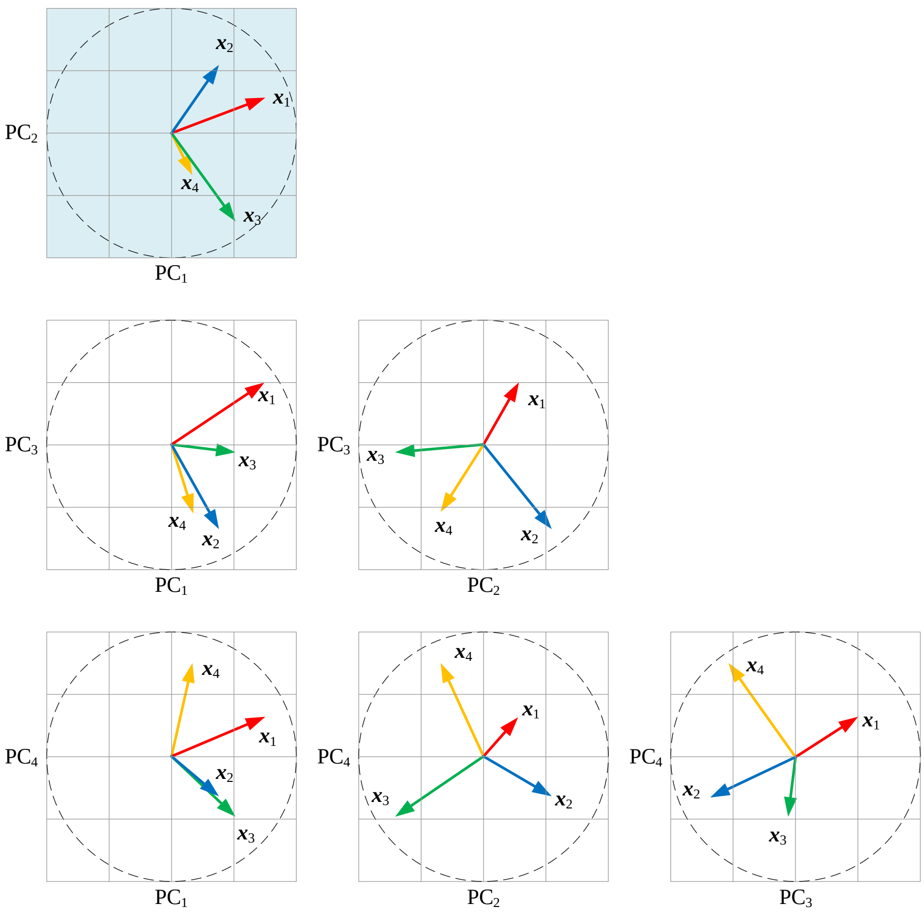


图 53. 分解主成分

在 PC_1-PC_2 平面上，$\boldsymbol{x}_1$ 对应坐标点为 (0.751, 0.284)，这意味着 $\boldsymbol{x}_1$ 分别给 $\boldsymbol{z}_1$ 和 $\boldsymbol{z}_2$ 贡献 $0.751\boldsymbol{x}_1$ 和 $0.284\boldsymbol{x}_1$。同理，我们可以发现 $\boldsymbol{x}_2$ 分别给 $\boldsymbol{z}_1$ 和 $\boldsymbol{z}_2$ 贡献 $0.380\boldsymbol{x}_2$ 和 $0.547\boldsymbol{x}_2$。以此类推。

反向来看，$\boldsymbol{x}_1$ 在 PC_1、PC_2、PC_3、PC_4 方向上的分量分别为 $0.751\boldsymbol{x}_1$、$0.284\boldsymbol{x}_1$、$0.502\boldsymbol{x}_1$、$0.321\boldsymbol{x}_1$，这四个成分满足：

$$0.751^2 + 0.284^2 + 0.502^2 + 0.321^2 = 1 \tag{60}$$

反向正交投影

由于 $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{V}$，$\boldsymbol{X}$ 则可以通过 $\boldsymbol{Z}$ 反推得到，即：

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{V}^{-1} = \boldsymbol{Z}\boldsymbol{V}^{\mathrm{T}} \tag{61}$$

图 54 所示为 $\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 相互转化的关系。这幅图告诉我们 $\boldsymbol{V}$ 和 $\boldsymbol{V}^{\mathrm{T}}$ 都是规范正交基。

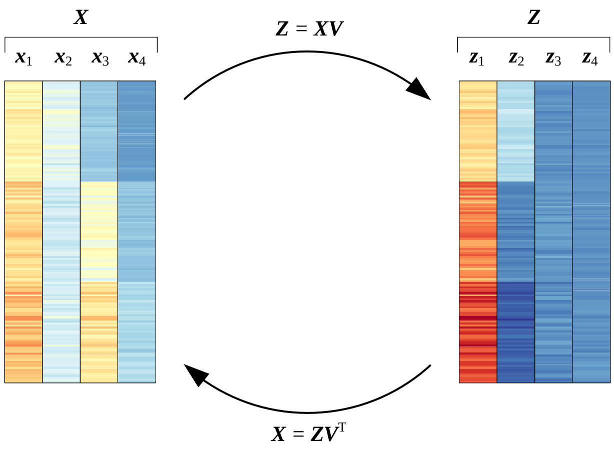


图 54. $\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 之间关系

将 (61) 展开写：

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{Z}\begin{bmatrix} \boldsymbol{v}^{(1)} \\ \boldsymbol{v}^{(2)} \\ \vdots \\ \boldsymbol{v}^{(D)} \end{bmatrix}^{\mathrm{T}} = \boldsymbol{Z}\begin{bmatrix} \boldsymbol{v}^{(1)\mathrm{T}} & \boldsymbol{v}^{(2)\mathrm{T}} & \cdots & \boldsymbol{v}^{(D)\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z}\boldsymbol{v}^{(1)\mathrm{T}} & \boldsymbol{Z}\boldsymbol{v}^{(2)\mathrm{T}} & \cdots & \boldsymbol{Z}\boldsymbol{v}^{(D)\mathrm{T}} \end{bmatrix} \tag{62}$$

取出 (62) 矩阵 $\boldsymbol{X}$ 第 j 列对应的等式：

$$
\boldsymbol{x}_j = \boldsymbol{Z}\boldsymbol{v}^{(j)\mathrm{T}} = \begin{bmatrix} \boldsymbol{z}_1 & \boldsymbol{z}_2 & \cdots & \boldsymbol{z}_D \end{bmatrix} \begin{bmatrix} v_{j,1} \\ v_{j,2} \\ \vdots \\ v_{j,D} \end{bmatrix} = v_{j,1}\boldsymbol{z}_1 + v_{j,2}\boldsymbol{z}_2 + \cdots + v_{j,D}\boldsymbol{z}_D \tag{63}
$$

$\boldsymbol{v}^{(j)}$ 代表 $\boldsymbol{V}$ 的第 j 行行向量，也是一个单位向量。(63) 是一个“反向”正交投影的过程。

(63) 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。

本章内容是个分水岭。如果本章内容，特别是前两节内容，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，如果你感觉很吃力，请回头重读前 9 章内容。

大家可能会好奇，本章中神奇的 $\boldsymbol{V}$ 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！

请大家带着这三句话，进入本书下一阶段的学习。