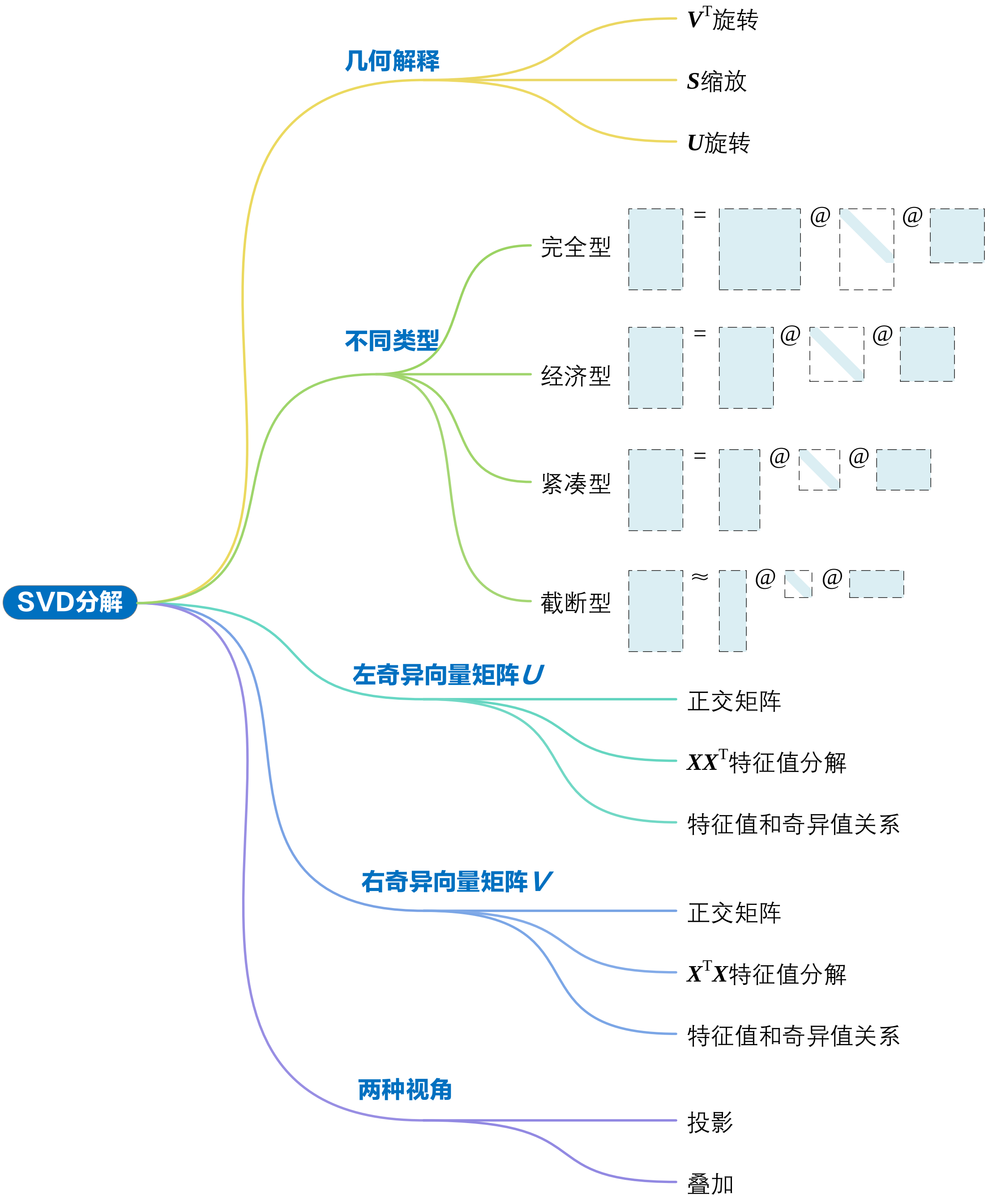
15 Singular Value Decomposition 奇异值分解

应该是最重要的矩阵分解，没有之一

> 人不过是一根芦苇，世界最脆弱的生灵；但是，人是会思考的芦苇。
>
> ***Man is but a reed, the most feeble thing in nature, but he is a thinking reed.***
>
> —— 布莱兹·帕斯卡 (Blaise Pascal) | 法国哲学家、科学家 | 1623 ~ 1662

- ◀ `matplotlib.pyplot.quiver()` 绘制箭头图
- ◀ `numpy.linspace()` 在指定的间隔内，返回固定步长的数据
- ◀ `numpy.linalg.svd()` 进行 SVD 分解
- ◀ `numpy.diag()` 以一维数组的形式返回方阵的对角线元素，或将一维数组转换成对角阵



SVD分解
几何解释
V^T旋转
S缩放
U旋转
不同类型
完全型
经济型
紧凑型
截断型
=
@
≈
左奇异向量矩阵U
正交矩阵
XX^T特征值分解
特征值和奇异值关系
右奇异向量矩阵V
正交矩阵
X^TX特征值分解
特征值和奇异值关系
两种视角
投影
叠加

15.1 几何视角：旋转 → 缩放 → 旋转

本书第 11 章简要介绍过**奇异值分解** (Singular Value Decomposition, SVD)，这个宇宙中最重要的矩阵分解。本节将从几何视角解剖奇异值分解。

对数据矩阵 $\boldsymbol{X}_{n\times D}$ 奇异值分解得到：

$$\boldsymbol{X}_{n\times D} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}} \tag{1}$$

其中，$\boldsymbol{S}$ 为对角阵，其主对角线元素 s_j ($j = 1, 2, \ldots, D$) 为**奇异值** (singular value)。

⚠ 注意，SVD 分解得到的奇异值非负，即 $s_j \geqslant 0$。

$\boldsymbol{U}$ 的列向量称作**左奇异向量** (left singular vector)。

$\boldsymbol{V}$ 的列向量称作**右奇异向量** (left singular vector)。

⚠ 特别注意，(1) 中矩阵 $\boldsymbol{V}$ 的转置运算。

$\boldsymbol{U}$ 和 $\boldsymbol{V}$ 为正交矩阵，即 $\boldsymbol{U}$ 和自己转置 $\boldsymbol{U}^{\mathrm{T}}$ 的乘积为单位矩阵；$\boldsymbol{V}$ 和自己转置 $\boldsymbol{V}^{\mathrm{T}}$ 的乘积也是单位矩阵。

从向量空间角度来看，$\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ 为 $\mathbb{R}^n$ 的规范正交基，$\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$ 为 $\mathbb{R}^D$ 的规范正交基。

根据这三个矩阵的形态，我们知道，从几何视角来看，正交矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 矩阵作用是旋转，而对角矩阵 $\boldsymbol{S}$ 的作用是缩放。

几何视角

大家可能会问这和特征值分解对应的“旋转 → 缩放 →旋转”有何不同？

特征值分解中，三步几何变换是旋转 ($\boldsymbol{V}^{-1}$) → 缩放 ($\boldsymbol{\Lambda}$) →旋转 ($\boldsymbol{V}$)。

奇异值分解中，三步几何变换是旋转 ($\boldsymbol{V}^{\mathrm{T}}$) → 缩放 ($\boldsymbol{S}$) →旋转 ($\boldsymbol{U}$)。一个明显的区别是，$\boldsymbol{V}^{\mathrm{T}}$ 的旋转发生在 $\mathbb{R}^D$ 空间，$\boldsymbol{U}$ 的旋转则发生在 $\mathbb{R}^n$ 空间。

为了方便解释，我们用 2 × 2 矩阵 $\boldsymbol{A}$ 做例子。

矩阵 $\boldsymbol{A}$ 坐标向量 $\boldsymbol{x}$ 得到 $\boldsymbol{z}$。对 $\boldsymbol{A}$ 奇异值分解，得到 $\boldsymbol{U}$、$\boldsymbol{S}$ 和 $\boldsymbol{V}$ 三个矩阵：

$$\boldsymbol{A}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underset{\text{Rotate}}{\boldsymbol{U}}\ \underset{\text{Scale}}{\boldsymbol{S}}\ \underset{\text{Rotate}}{\boldsymbol{V}^{\mathrm{T}}}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{2}$$

图 1 所示为几何角度解释奇异值分解，$\boldsymbol{A}$ 乘 $[x_1, x_2]^{\mathrm{T}}$，相当于先用 $\boldsymbol{V}^{\mathrm{T}}$ 旋转，再用 $\boldsymbol{S}$ 缩放，最后用 $\boldsymbol{U}$ 旋转。

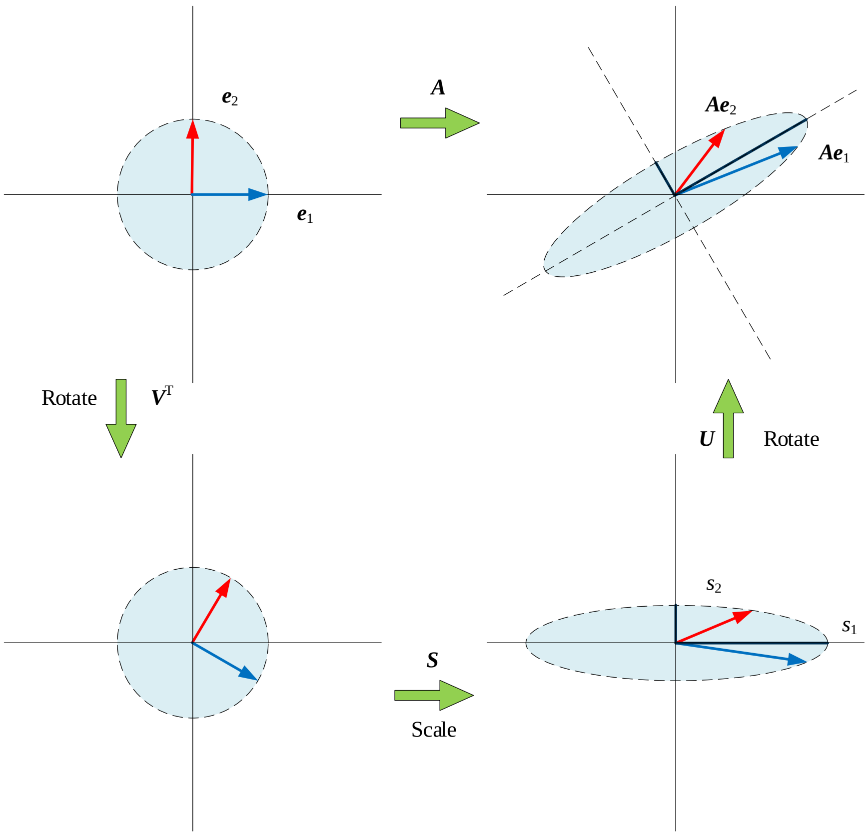


图 1. 几何角度解释奇异值分解

举个实例

下面用具体实例解释图 1。

给定 2 × 2 矩阵 $\boldsymbol{A}$:

$$\boldsymbol{A}=\begin{bmatrix}1.625 & 0.6495\\ 0.6495 & 0.875\end{bmatrix} \tag{3}$$

对矩阵 $\boldsymbol{A}$ 进行 SVD 分解:

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}=\underbrace{\begin{bmatrix}0.866 & -0.5\\ 0.5 & 0.866\end{bmatrix}}_{U}\underbrace{\begin{bmatrix}2 & 0\\ 0 & 0.5\end{bmatrix}}_{S}\underbrace{\begin{bmatrix}0.866 & -0.5\\ 0.5 & 0.866\end{bmatrix}}_{V^{\mathrm{T}}} \tag{4}$$

即,

$$\boldsymbol{U}=\begin{bmatrix}0.866 & -0.5\\ 0.5 & 0.866\end{bmatrix},\quad \boldsymbol{S}=\begin{bmatrix}2 & 0\\ 0 & 0.5\end{bmatrix},\quad \boldsymbol{V}=\begin{bmatrix}0.866 & 0.5\\ -0.5 & 0.866\end{bmatrix} \tag{5}$$

⚠ 注意,如果特征值分解和奇异值分解的对象相同,都是可对角矩阵,两个分解得到的结果等价。但是,奇异值分解的强大之处在于,任何实数矩阵都可以奇异值分解。

给定 e_1 和 e_2 两个单位向量：

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{6}$$

e_1 和 e_2 经过 A 转换分别得到：

$$\begin{aligned} Ae_1 &= \begin{bmatrix} 1.625 & 0.6495 \\ 0.6495 & 0.875 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.625 \\ 0.6495 \end{bmatrix} \\ Ae_2 &= \begin{bmatrix} 1.625 & 0.6495 \\ 0.6495 & 0.875 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.6495 \\ 0.875 \end{bmatrix} \end{aligned} \tag{7}$$

图 2 所示为转换前后的结果对比。

⚠ 请大家注意转换前后向量的方向和长度 (模) 的变化。

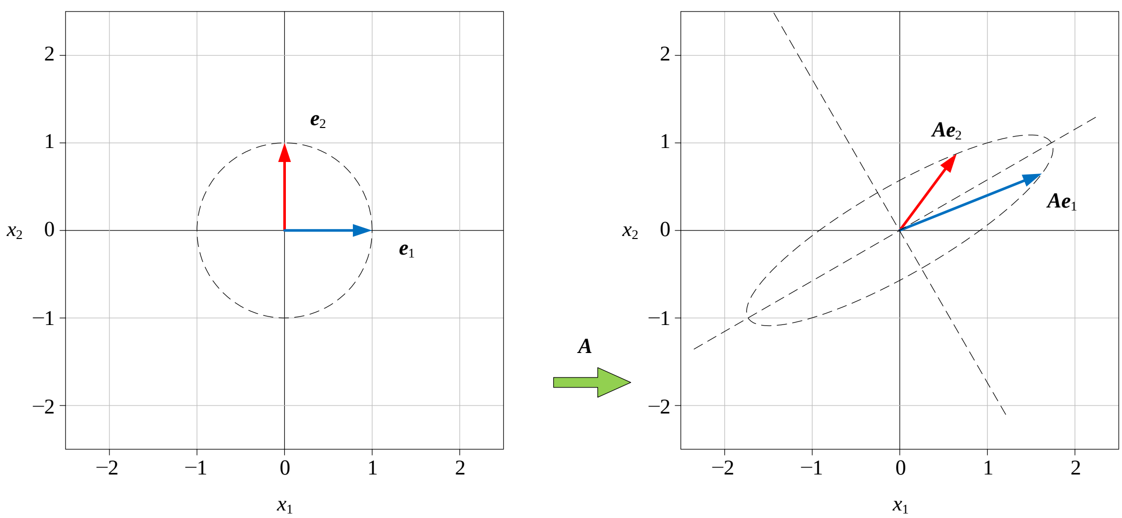


图 2. e_1 和 e_2 经过 A 线性转换

分步几何变换

(7) 等价于“旋转 (V^T) → 缩放 (S) → 旋转 (U)”，具体如图 3 所示。

e_1 和 e_2 两个向量先通过 V^T 进行旋转，得到：

$$\begin{aligned} V^T e_1 &= \begin{bmatrix} 0.866 & 0.5 \\ -0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.866 \\ -0.5 \end{bmatrix} \\ V^T e_2 &= \begin{bmatrix} 0.866 & 0.5 \\ -0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.866 \end{bmatrix} \end{aligned} \tag{8}$$

在 (8) 基础上，再用对角矩阵 S 进行缩放，得到：

$$
\begin{aligned}
\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_1 &= \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}\begin{bmatrix} 0.866 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 1.732 \\ -0.25 \end{bmatrix} \\
\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_2 &= \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}\begin{bmatrix} 0.5 \\ 0.866 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.433 \end{bmatrix}
\end{aligned} \tag{9}
$$

在之前“旋转”和“缩放”两步基础上，最后再利用 $\boldsymbol{U}$ 进行旋转，得到：

$$
\begin{aligned}
\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_1 &= \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}\begin{bmatrix} 1.732 \\ -0.25 \end{bmatrix} = \begin{bmatrix} 1.625 \\ 0.6495 \end{bmatrix} \\
\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_2 &= \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}\begin{bmatrix} 1 \\ 0.433 \end{bmatrix} = \begin{bmatrix} 0.6495 \\ 0.875 \end{bmatrix}
\end{aligned} \tag{10}
$$

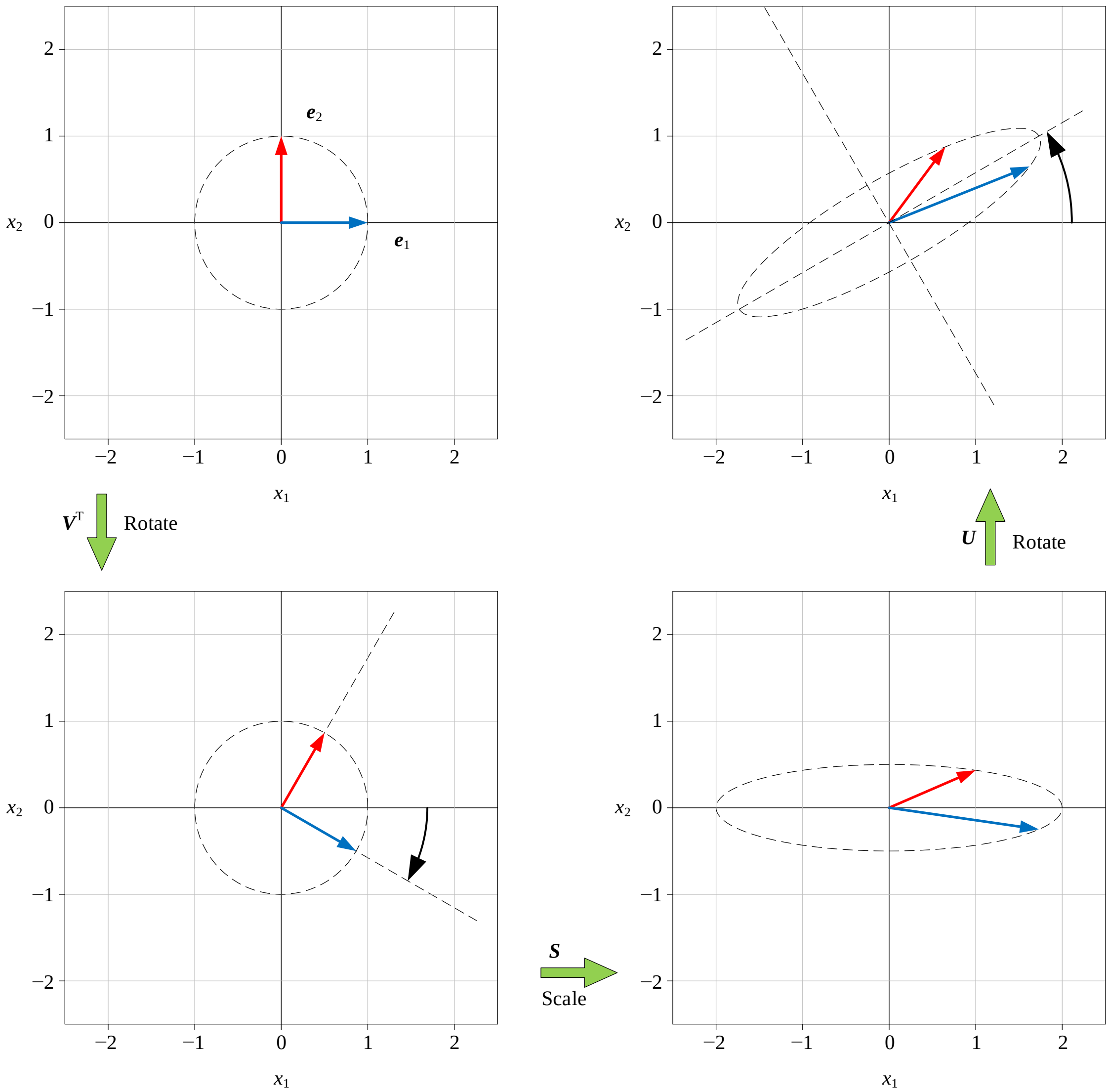


图 3. $\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 分别经过 $\boldsymbol{V}^{\mathrm{T}}$、$\boldsymbol{S}$ 和 $\boldsymbol{U}$ 转换

Bk4_Ch15_01.py 绘制图 3 所有子图。

15.2 不同类型 SVD 分解

SVD 分解分为**完全型** (full)、**经济型** (economy-size, thin)、**紧凑型** (compact) 和**截断型** (truncated)。

本节将简要介绍完全型和经济型两种奇异值分解之间的关系。下一章将深入讲解这四种 SVD 分解。

完全型

图 4 所示为完全型 SVD 分解热图，其中左奇异值矩阵 $\boldsymbol{U}$ 为方阵，形状为 $n \times n$。$\boldsymbol{S}$ 的形状和 $\boldsymbol{X}$ 相同，为 $n \times D$。$\boldsymbol{S}$ 的主对角线元素 s_i 为奇异值，具体形式为：

$$\boldsymbol{S} = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \\ & & & \\ & & & \end{bmatrix} \tag{11}$$

约定俗成，这 D 个奇异值的关系为 $s_1 \geqslant s_2 \geqslant \dots \geqslant s_D$。

如图 4 所示，$\boldsymbol{S}$ 可以分块为上下两个子块——方阵对角矩阵和全 0 矩阵 $\boldsymbol{O}$。右奇异值矩阵 $\boldsymbol{V}$ 形状为 $D \times D$。

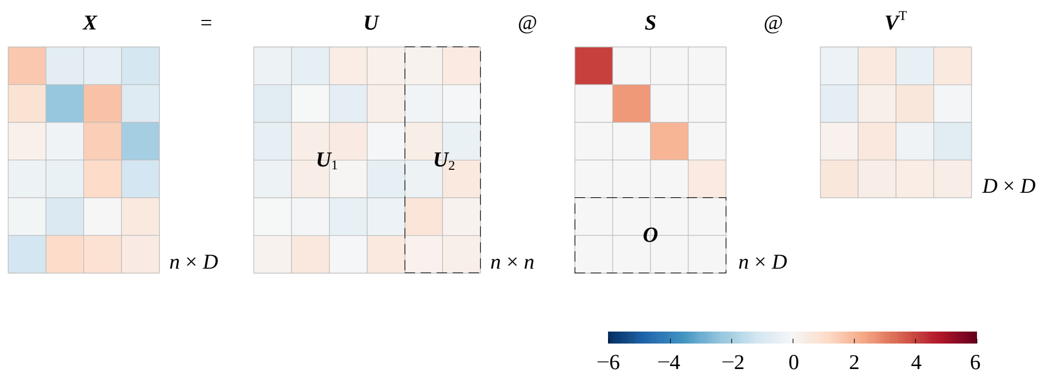


图 4. 矩阵 $\boldsymbol{X}$ 的完全型 SVD 分解

⚠ 注意，一般情况，数据矩阵为“细高”长方形，偶尔大家也会见到“宽矮”长方形的数据矩阵。

(1) 中 $\boldsymbol{X}$ 就是细高长方形，对 $\boldsymbol{X}$ 转置便得到宽矮长方形矩阵 $\boldsymbol{X}^{\mathrm{T}}$。如图 5 所示，相应的，$\boldsymbol{X}^{\mathrm{T}}$ 的 SVD 分解为：

$$\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} \tag{12}$$

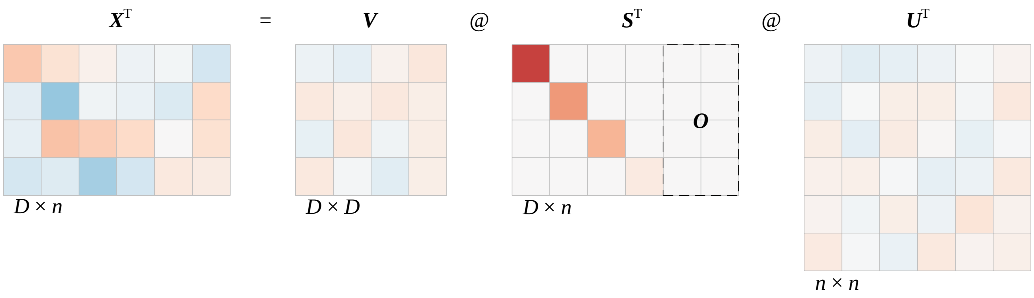


图 5. 矩阵 $\boldsymbol{X}^{\mathrm{T}}$ 的完全型 SVD 分解

经济型

图 6 所示为经济型 SVD 分解结果热图。可以发现，左奇异值矩阵 $\boldsymbol{U}$ 形状和 $\boldsymbol{X}$ 相同，均为为 $n \times D$。而 $\boldsymbol{S}$ 为方阵，形状为 $D \times D$。

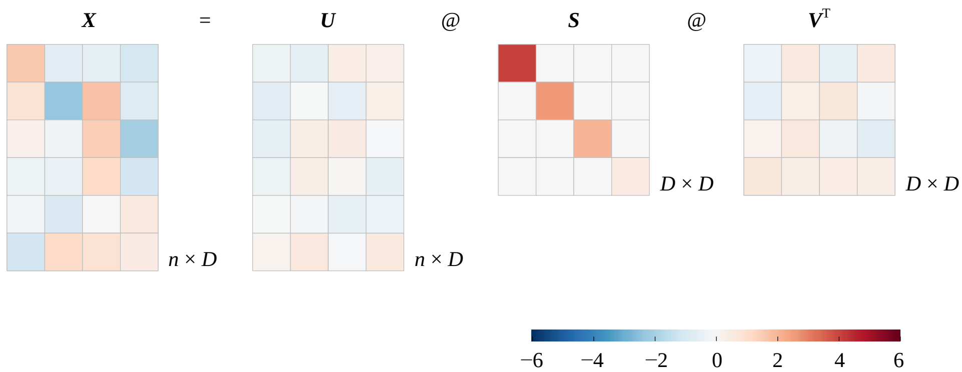


图 6. 经济型 SVD 分解

在经济型 SVD 分解中，$\boldsymbol{S}$ 为对角方阵：

$$\boldsymbol{S} = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} \tag{13}$$

当 $\boldsymbol{S}$ 为对角方阵时，(12) 可以写成：

$$\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^{\mathrm{T}} \tag{14}$$

Bk4_Ch15_02.py 中 Bk4_Ch15_02_A 部分绘制图 4 和图 6。

15.3 左奇异向量矩阵 $\boldsymbol{U}$

$\boldsymbol{U}$ 的列向量称作**左奇异向量** (left singular vector)，$\boldsymbol{U}$ 和自己转置 $\boldsymbol{U}^{\mathrm{T}}$ 的乘积为单位矩阵：

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{I} \tag{15}$$

图 7 所示为运算热图。

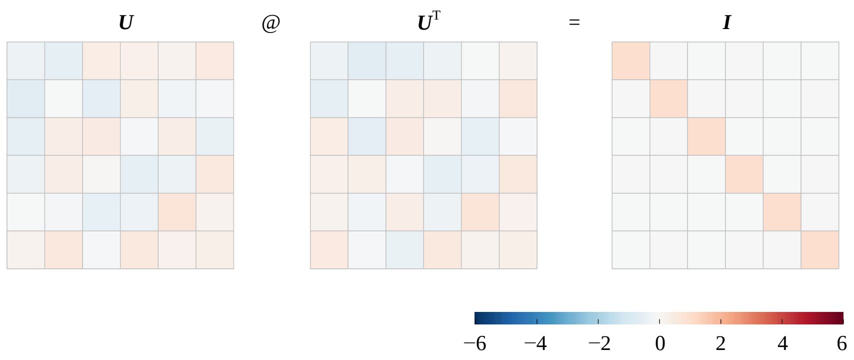


图 7. $\boldsymbol{U}$ 和自己转置 $\boldsymbol{U}^{\mathrm{T}}$ 的乘积为单位矩阵

特征值分解

图 8 所示为 $\boldsymbol{X}$ 和自己转置 $\boldsymbol{X}^{\mathrm{T}}$ 相乘得到 Gram 矩阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 的热图，$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 为 $n \times n$ 方阵。

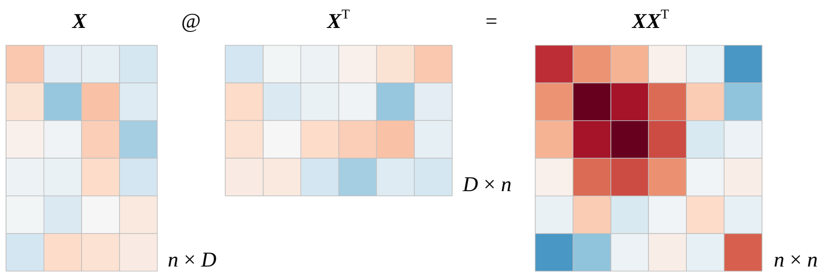


图 8. $\boldsymbol{X}$ 和自己转置 $\boldsymbol{X}^{\mathrm{T}}$ 的乘积热图

对方阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 进行特征值分解，可以发现 $\boldsymbol{U}$ 是特征向量，而 $\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}$ 是特征值矩阵：

$$\begin{aligned}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} &= \left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\right)\left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\right)^{\mathrm{T}} \\ &= \boldsymbol{U}\boldsymbol{S}\left(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\right)\boldsymbol{S}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} \\ &= \boldsymbol{U}\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\end{aligned} \tag{16}$$

图 9 所示为 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 特征值分解热图。

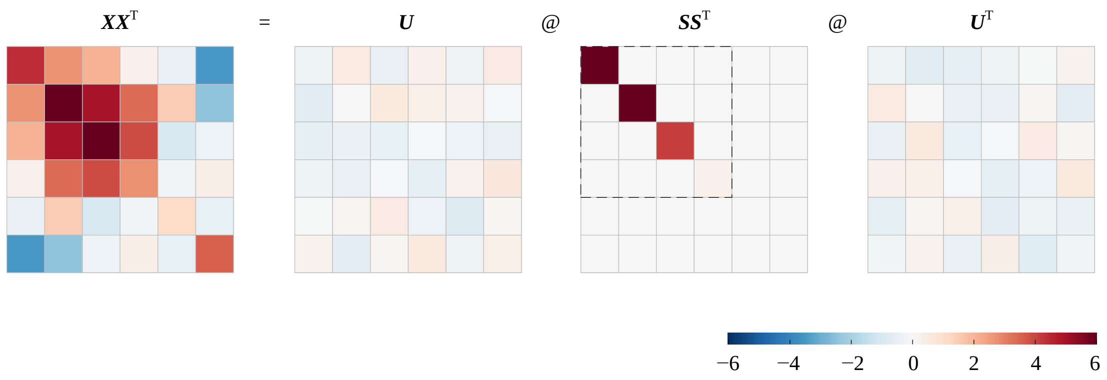


图 9. 对 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 特征值分解

$\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}$ 主对角线为特征值，对 $\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}$ 展开得到：

$$\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \\ & & & \end{bmatrix}\begin{bmatrix} s_1 & & & & \\ & s_2 & & & \\ & & \ddots & & \\ & & & s_D & \end{bmatrix} = \begin{bmatrix} s_1^2 & & & & \\ & s_2^2 & & & \\ & & \ddots & & \\ & & & s_D^2 & \\ & & & & 0 \\ & & & & & 0\end{bmatrix} = \begin{bmatrix} \lambda_1 & & & & & \\ & \lambda_2 & & & & \\ & & \ddots & & & \\ & & & \lambda_D & & \\ & & & & \ddots & \\ & & & & & \lambda_n\end{bmatrix} \tag{17}$$

观察上式，发现当 $j = 1 \sim D$ 时，特征值 λ_j 和奇异值 s_j 存在如下关系：

$$\lambda_j = s_j^2 \tag{18}$$

特别地，在经济型 SVD 分解中，$\boldsymbol{S}$ 为对角方阵；因此 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 特征值分解可以写作：

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{U}\boldsymbol{S}^2\boldsymbol{U}^{\mathrm{T}} \tag{19}$$

向量空间

本书前文提到过两次，细高的长方形矩阵 $\boldsymbol{X}$ 不能进行特征值分解。但是，它的格拉姆矩阵 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 是对称矩阵，形状为 $n \times n$，可以进行特征值分解。

如图 10 所示，$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 进行特征值分解得到正交矩阵 $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$。前文提到，$\boldsymbol{U}$ 是个规范正交基，张起的空间为 $\mathbb{R}^n$。

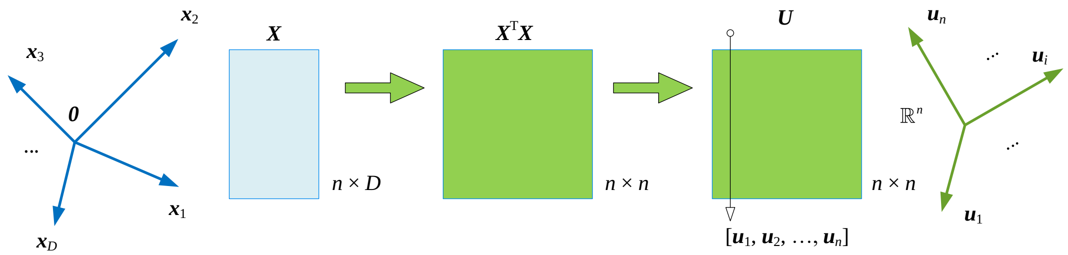


图 10. 对 Gram 矩阵 XX^T 特征值分解得到规范正交基 U

类比 QR 分解

对于数据矩阵 X 进行 QR 分解得到：

$$X = QR \tag{20}$$

对于完全型 QR 分解，Q 为正交矩阵，即一个规范正交基 $[q_1, q_2, \ldots, q_D]$。

对 X 进行完全型 SVD 分解，把结果写成：

$$X = U\left(SV^T\right) \tag{21}$$

对比 (20) 和 (21)，Q 和 U 都是正交矩阵，形状相同；但是，两者显然不同。对于 QR 分解，x_1 和 q_1 平行。打个比方，x_1 像是一个锚，确定了 $[q_1, q_2, \ldots, q_D]$ 的空间位置。而 SVD 分解则引入了一个优化视角——逐个最大化奇异值。对比 (20) 和 (21)，R 则对应 SV^T。特别地，SV^T 结果正交，即 $SV^T(SV^T)^T = SV^TVS = S^2$。

本书第 18 章将深入介绍这个优化视角。

Bk4_Ch15_02.py 中 Bk4_Ch15_02_B 部分绘制图 7。请读者自行编写代码绘制图 8 和图 9。

15.4 右奇异向量矩阵 V

V 的列向量称作**右奇异向量** (right singular vector)，V 和其转置 V^T 的乘积也是单位矩阵：

$$V^TV = I \tag{22}$$

图 11 所示为上式运算对应热图。

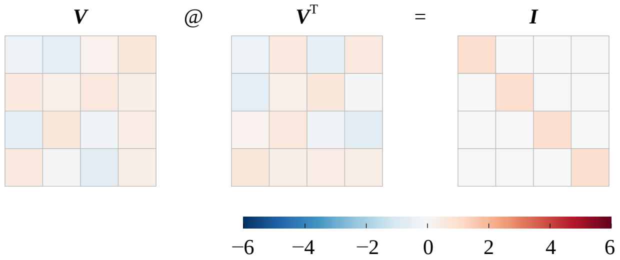


图 11. V 和其转置 V^{T} 的乘积也是单位矩阵

特征值分解

图 12 所示为转置 X^{T} 和 X 相乘得到 $X^{\mathrm{T}}X$ 的热图，$X^{\mathrm{T}}X$ 为 $D \times D$ 方阵。

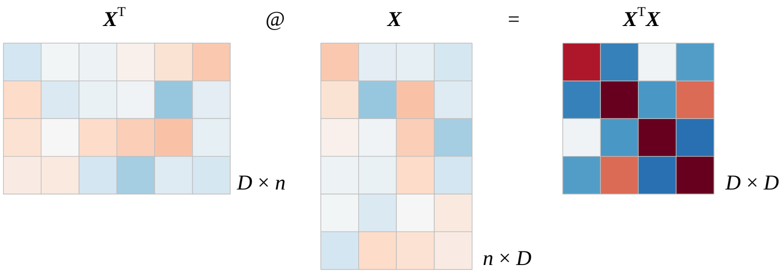


图 12. 转置 X^{T} 和 X 乘积热图

对 $X^{\mathrm{T}}X$ 特征值分解得到：

$$
\begin{aligned}
X^{\mathrm{T}}X &= \left(USV^{\mathrm{T}}\right)^{\mathrm{T}}\left(USV^{\mathrm{T}}\right) \\
&= VS^{\mathrm{T}}\left(U^{\mathrm{T}}U\right)SV^{\mathrm{T}} \\
&= VS^{\mathrm{T}}SV^{\mathrm{T}}
\end{aligned}
\tag{23}
$$

V 是 $X^{\mathrm{T}}X$ 的特征向量，$S^{\mathrm{T}}S$ 为特征值矩阵。图 13 所示为对 $X^{\mathrm{T}}X$ 进行特征值分解热图。

特别地，在经济型 SVD 分解中，S 为对角方阵；因此 $X^{\mathrm{T}}X$ 特征值分解可以写成：

$$
X^{\mathrm{T}}X = VS^2V^{\mathrm{T}} \tag{24}
$$

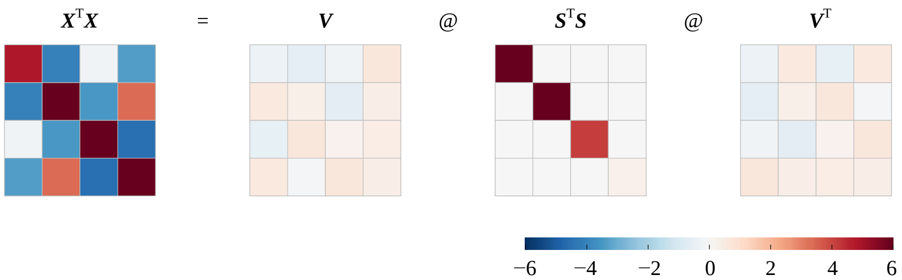


图 13. 对 $X^{\mathrm{T}}X$ 进行特征值分解

前文提过，经济型 SVD 分解中，S 为对角方阵，S 主对角元素 s_i 为奇异值：

$$S=\begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D \end{bmatrix} \tag{25}$$

对 $X^{\mathrm{T}}X$ 进行特征值分解，S^2 为特征值矩阵：

$$S^2=\begin{bmatrix} s_1^2 & & & \\ & s_2^2 & & \\ & & \ddots & \\ & & & s_D^2 \end{bmatrix}=\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix} \tag{26}$$

(25) 和 (26) 奇异值和特征值之间的关系如图 14 所示。

本书第 24 章将总结分解不同对象时，奇异值分解和特征值之间的联系和差异。

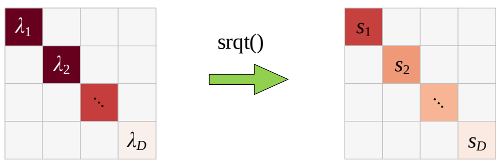


图 14. 奇异值和特征值之间关系

向量空间

如图 10 所示，$X^{\mathrm{T}}X$ 进行特征值分解得到正交矩阵 $V = [v_1, v_2, \dots, v_D]$，它也是个规范正交基，张起的空间为 $\mathbb{R}^D$。

通过本节比较，我们发现奇异值分解强大之处在于，它可以分解一切实数矩阵，而且一次性获得 $U = [u_1, u_2, \dots, u_n]$ 和 $V = [v_1, v_2, \dots, v_D]$ 两个规范正交基。

大家可能好奇，U 和 V 这两个规范正交基有怎么样的联系和应用？这个问题留给本书最后三章回答。

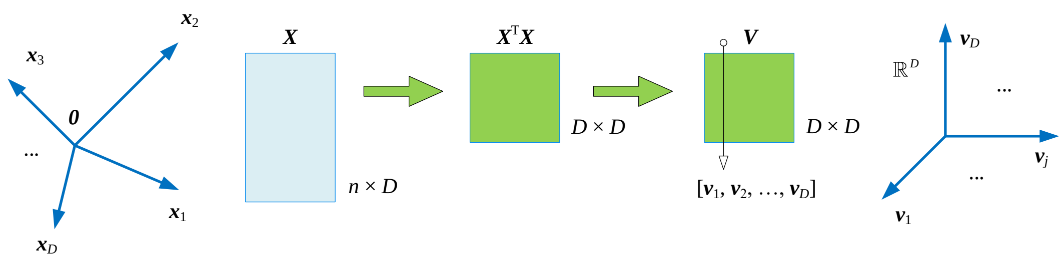


图 15. 对 Gram 矩阵 $X^{\mathrm{T}}X$ 特征值分解得到规范正交基 V

Bk4_Ch15_02.py 中 Bk4_Ch15_02_C 部分绘制图 11。请读者自行编写代码绘制图 12 和图 13。

15.5 两个视角：投影和数据叠加

本节用两个数据视角观察 SVD 分解。

投影

将 (1) 等式左右两侧右乘 $\boldsymbol{V}$，可以得到：

$$\boldsymbol{X}_{n\times D}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S} \tag{27}$$

将 $\boldsymbol{V}$ 和 $\boldsymbol{U}$ 本身分别写成左右排列的列向量：

$$\boldsymbol{X}_{n\times D}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_D\end{bmatrix} = \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_D\end{bmatrix}\begin{bmatrix}s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D\end{bmatrix} \tag{28}$$

(28) 进一步展开得到：

$$\begin{bmatrix}\boldsymbol{X}\boldsymbol{v}_1 & \boldsymbol{X}\boldsymbol{v}_2 & \cdots & \boldsymbol{X}\boldsymbol{v}_D\end{bmatrix} = \begin{bmatrix}s_1\boldsymbol{u}_1 & s_2\boldsymbol{u}_2 & \cdots & s_D\boldsymbol{u}_D\end{bmatrix} \tag{29}$$

因此，

$$\boldsymbol{X}\boldsymbol{v}_j = s_j\boldsymbol{u}_j \tag{30}$$

上式可以理解为 $\boldsymbol{X}$ 向 $\boldsymbol{v}_j$ 投影，结果为 $s_j\boldsymbol{u}_j$。对应运算热图如图 16 所示。

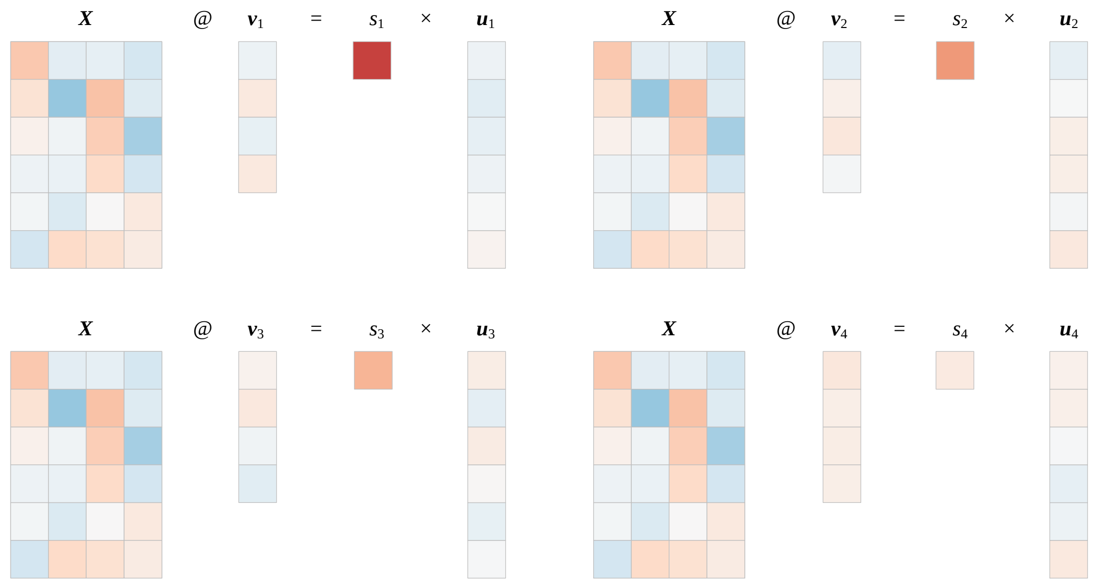


图 16. $\boldsymbol{X}$ 向 $\boldsymbol{v}_j$ 映射结果为 $s_j\boldsymbol{u}_j$

(30) 左右都是向量，等式两侧求模，即 L^2 范数，得到：

$$\left\|\boldsymbol{X}\boldsymbol{v}_j\right\| = \left\|s_j\boldsymbol{u}_j\right\| = s_j \tag{31}$$

也就是说 $\boldsymbol{X}\boldsymbol{v}_j$ 的模为对应奇异值 s_j。由于奇异值 s_1 到 s_4 从大到小排列，也就是说 $\boldsymbol{X}\boldsymbol{v}_1$ 的模最大。这个角度对于理解**主成分分析** (principal component analysis, PCA) 极为重要。

叠加

第二种展开方式如下：

$$\begin{aligned}\boldsymbol{X}_{n\times D} &= \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_D\end{bmatrix}\begin{bmatrix}s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_D\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}}\end{bmatrix} \\ &= \begin{bmatrix}s_1\boldsymbol{u}_1 & s_2\boldsymbol{u}_2 & \cdots & s_D\boldsymbol{u}_D\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_D^{\mathrm{T}}\end{bmatrix} = s_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + s_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots + s_D\boldsymbol{u}_D\boldsymbol{v}_D^{\mathrm{T}}\end{aligned} \tag{32}$$

举个例子，对于 $D = 4$ 时：

$$\boldsymbol{X} = s_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + s_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + s_3\boldsymbol{u}_3\boldsymbol{v}_3^{\mathrm{T}} + s_4\boldsymbol{u}_4\boldsymbol{v}_4^{\mathrm{T}} \tag{33}$$

(33) 中奇异值 s_1 到 s_4 从大到小排列，即 $s_1 \geqslant s_2 \geqslant s_3 \geqslant s_4$。

⚠ 注意，$s_j\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}$ 的秩为 1。

如图 17 所示，可以发现对应 (33) 等式右侧从左到右的四项相当于逐步还原 $\boldsymbol{X}$。特别地，请大家注意图 17 左侧四副热图由上到下颜色逐渐变浅。下一章会深入介绍通过叠加换换原始数据矩阵。

此外，相信大家看到这里早已发现，本节前两个视角实际上就是矩阵乘法的两个视角。

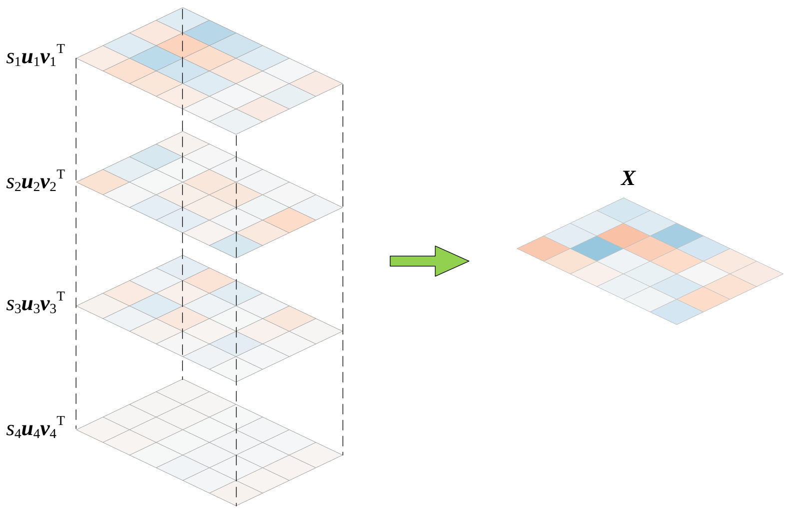


图 17. 四幅热图叠加还原原始图像

张量积

再进一步，利用 (30) 给出的关系，我们将 (33) 写成张量积之和的形式

$$\begin{aligned}\boldsymbol{X} &= s_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + s_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + s_3\boldsymbol{u}_3\boldsymbol{v}_3^{\mathrm{T}} + s_4\boldsymbol{u}_4\boldsymbol{v}_4^{\mathrm{T}} \\ &= \boldsymbol{X}\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{v}_3\boldsymbol{v}_3^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{v}_4\boldsymbol{v}_4^{\mathrm{T}} \\ &= \boldsymbol{X}\left(\boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}} + \boldsymbol{v}_3\boldsymbol{v}_3^{\mathrm{T}} + \boldsymbol{v}_4\boldsymbol{v}_4^{\mathrm{T}}\right) \\ &= \boldsymbol{X}\left(\boldsymbol{v}_1 \otimes \boldsymbol{v}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{v}_2 + \boldsymbol{v}_3 \otimes \boldsymbol{v}_3 + \boldsymbol{v}_4 \otimes \boldsymbol{v}_4\right)\end{aligned} \tag{34}$$

这就是本书之前讲解的思路——数据在不同向量方向上"二次投影"再"层层叠加"。

能完成类似 (34) 投影的规范正交基有无数组，为什么 $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_D]$ 脱颖而出？$\boldsymbol{V}$ 的特殊性体现在哪？回答这个问题需要优化方面的知识，这是本书第 18 章要回答的问题。

Bk4_Ch15_02.py 中 Bk4_Ch15_02_D 部分绘制本节图像。

下图四副子图总结本章主要内容。请大家特别注意，奇异值分解对应“旋转 → 缩放 → 旋转”，不同于特征值分解的“旋转 → 缩放 → 旋转”。

任何实数矩阵都可以进行奇异值分解，但是只有可对角矩阵才能进行特征值分解。此外，奇异值分解得到的两个正交矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 一般形状不同。

Gram 矩阵是奇异值分解和特征值分解的桥梁，这一点本书后面还要提到。

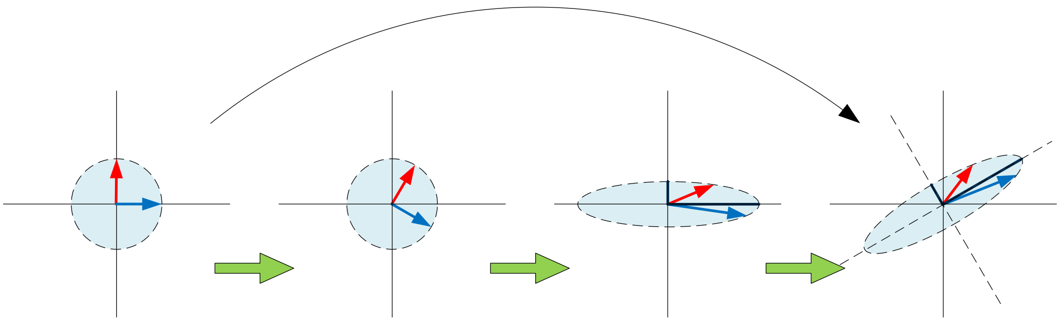


图 18. 总结本章重要内容的四副图