

作者	生姜 DrGinger
脚本	生姜 DrGinger
视频	崔崔 CuiCui
开源学习资源	https://github.com/Visualize-ML
平台	https://www.youtube.com/@DrGinger_Jiang https://space.bilibili.com/3546865719052873 https://space.bilibili.com/513194466

11.4 格拉姆矩阵



本节你将掌握的核心技能：

- ▶ 格拉姆矩阵为对称矩阵，半正定、正定。
- ▶ 格拉姆矩阵和原矩阵列向量内积关系。
- ▶ 主对角线元素为列向量 L2 范数的平方。
- ▶ 格拉姆矩阵的谱分解。
- ▶ 通过正交矩阵，完成原矩阵的正交化。
- ▶ 根据特征值大小判断特征向量的“解释力”，理解主次之分。
- ▶ 格拉姆矩阵的谱分解和奇异值分解的关系。

在整个线性代数和数据科学的世界里，有一个看似简单却无比核心的角色，那就是格拉姆矩阵 (Gram matrix)。本书前文提过格拉姆矩阵是一类特殊的对称矩阵。

别看格拉姆矩阵的定义朴素，它的影响却无处不在。从内积、向量范数、谱分解到奇异值分解，从正定性到 Cholesky 分解、LDL 分解，再到数据分析中常见的数据矩阵、协方差矩阵、线性相关性系数矩阵，甚至机器学习算法中的核方法 (比如高斯核支持向量机、核主成分分析)——都在某种形式上与格拉姆矩阵息息相关。

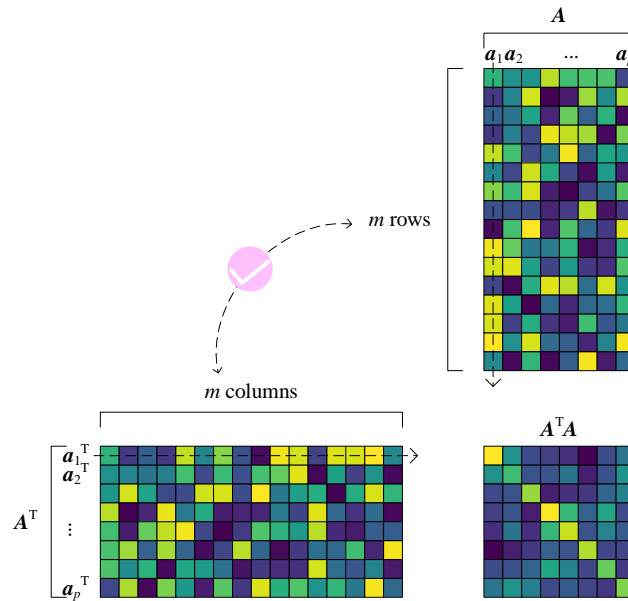
鉴于格拉姆矩阵的重要性，我们专门拿出一节聊聊格拉姆矩阵。

定义

给定矩阵 A 的形状为 $m \times p$ ，即 A 有 m 行、 p 列， A 的格拉姆矩阵为

$$G = A^T A \quad (1)$$

如图 1 所示， A 的格拉姆矩阵形状为 $p \times p$ 。

图 1. 矩阵 A 的格拉姆矩阵 $A^T A$

显然，格拉姆矩阵 G 为对称矩阵

$$G^T = (A^T A)^T = A^T A = G \quad (2)$$

此外，格拉姆矩阵为半正定矩阵 (positive semidefinite)，即对于任意 p 维非零列向量 x ，都有

$$x^T G x = x^T A^T A x = (Ax)^T (Ax) = \|Ax\|^2 \geq 0 \quad (3)$$



本书第 13 章、第 1 节会专门讲解正定性。

格拉姆矩阵的行列式 $A^T A$ 为零时，当且仅当矩阵 A 的列向量线性相关。

$A^T A$ 的秩等于 A 的列秩。可以这样理解， $A^T A$ 只是对原始列向量之间的内积关系做了“压缩整理”，但不会增加或减少信息维度。

用矩阵乘法第一视角展开 $A^T A$ 大家就会看到内积。

矩阵乘法第一视角

用矩阵乘法第一视角展开 $A^T A$

$$A^T A = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_p^T \end{bmatrix} @ \begin{bmatrix} a_1 & a_2 & \cdots & a_p \end{bmatrix} = \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_p \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_p \\ \vdots & \vdots & \ddots & \vdots \\ a_p^T a_1 & a_p^T a_2 & \cdots & a_p^T a_p \end{bmatrix} \quad (4)$$

上式也可以写成向量内积形式

$$A^T A = \begin{bmatrix} a_1 \cdot a_1 & a_1 \cdot a_2 & \cdots & a_1 \cdot a_p \\ a_2 \cdot a_1 & a_2 \cdot a_2 & \cdots & a_2 \cdot a_p \\ \vdots & \vdots & \ddots & \vdots \\ a_p \cdot a_1 & a_p \cdot a_2 & \cdots & a_p \cdot a_p \end{bmatrix} = \begin{bmatrix} \langle a_1, a_1 \rangle & \langle a_1, a_2 \rangle & \cdots & \langle a_1, a_p \rangle \\ \langle a_2, a_1 \rangle & \langle a_2, a_2 \rangle & \cdots & \langle a_2, a_p \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle a_p, a_1 \rangle & \langle a_p, a_2 \rangle & \cdots & \langle a_p, a_p \rangle \end{bmatrix} \quad (5)$$

如图 2 所示，格拉姆矩阵 $A^T A$ 的主对角线元素为 A 列向量和自身内积，也可以写成 L^2 范数平方，即

$$a_j^T @ a_j = a_j \cdot a_j = \|a_j\|^2 \quad (6)$$

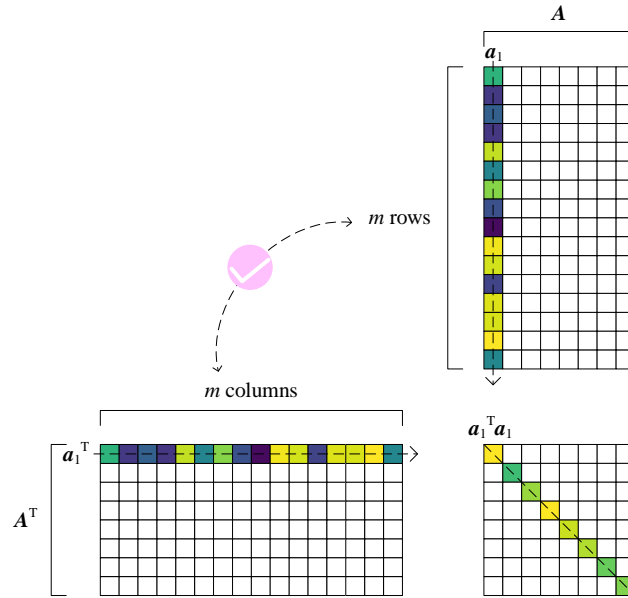


图 2. 矩阵 A 的格拉姆矩阵的对角线元素

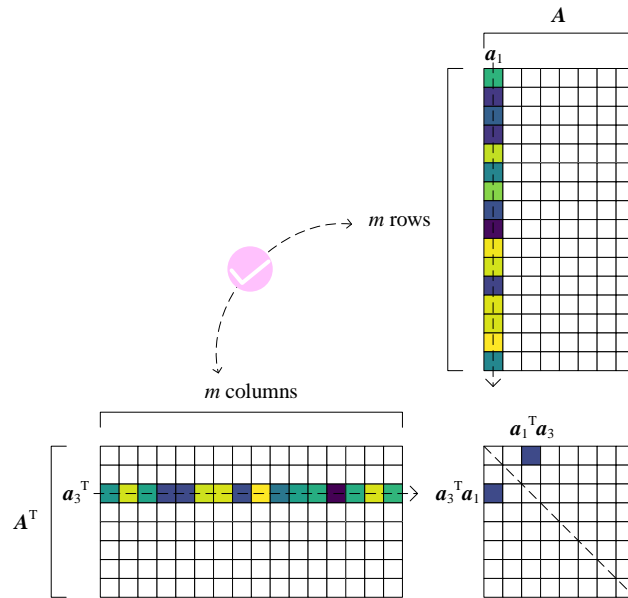
如图 3 所示， $A^T A$ 的非主对角线元素为 A 两两列向量内积。比如，

$$a_3^T a_1 = a_1^T a_3 \quad (7)$$

从内积角度来看上式，

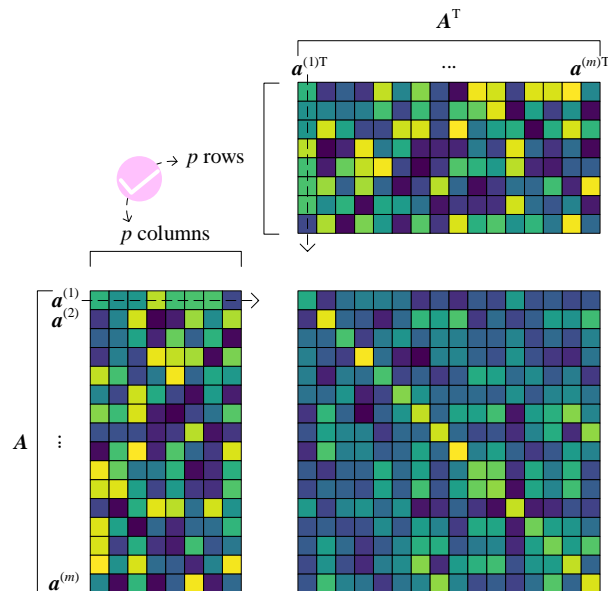
$$a_1 \cdot a_3 = a_3 \cdot a_1 \quad (8)$$

这也印证了 $A^T A$ 为对称矩阵。

图 3. 矩阵 A 的格拉姆矩阵的非对角线元素

第二个格拉姆矩阵

值得注意的是，如图 4 所示， AA^T 也是一个格拉姆矩阵。 AA^T 的主对角线为 A 行向量和自身的内积 (L^2 范数平方)； AA^T 的非主对角线元素为 A 两两行向量内积。显然， AA^T 也是对称矩阵。

图 4. 矩阵 A^T 的格拉姆矩阵 AA^T

谱分解格拉姆矩阵

如图 5 所示，对 $G = A^T A$ 进行谱分解得到

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。

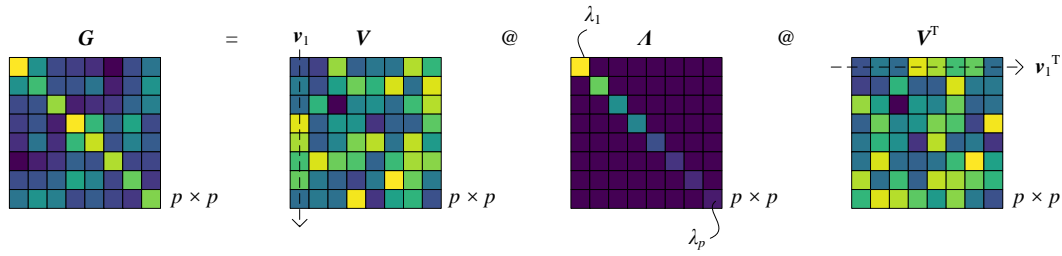
版权归清华大学出版社所有，请勿商用，引用请注明出处。

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

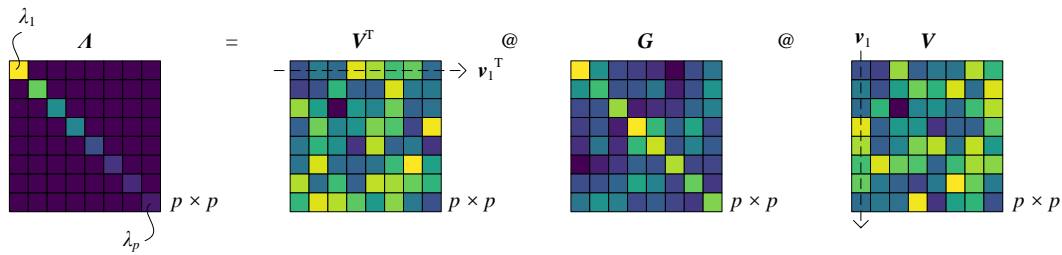
$$G = VAV^T \quad (9)$$

图 5. 对格拉姆矩阵 $A^T A$ 谱分解

对角化

如图 6 所示, G 对角化得到

$$V^T G V = A \quad (10)$$

图 6. 格拉姆矩阵 $A^T A$ 对角化

把 V 写成 $[v_1, v_2, \dots, v_p]$, (10) 展开写成

$$\underbrace{\begin{bmatrix} v_1^T G v_1 & v_1^T G v_2 & \cdots & v_1^T G v_p \\ v_2^T G v_1 & v_2^T G v_2 & \cdots & v_2^T G v_p \\ \vdots & \vdots & \ddots & \vdots \\ v_p^T G v_1 & v_p^T G v_2 & \cdots & v_p^T G v_p \end{bmatrix}}_{V^T G V} = \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix}}_A \quad (11)$$

图 7 所示为计算对角方阵 A 主对角线元素示例。

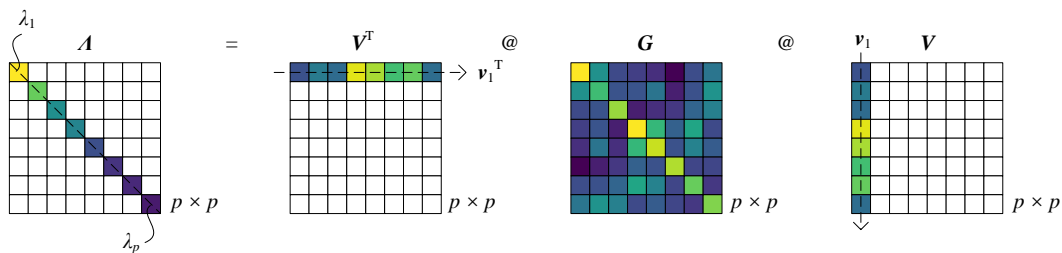
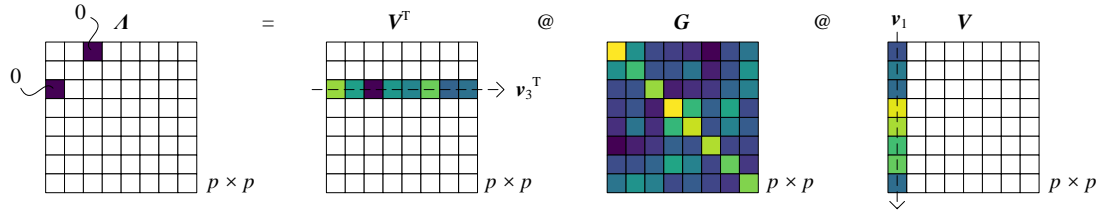
图 7. 对角方阵 A 主对角线元素

图 8 所示为计算对角方阵 Λ 非主对角线元素示例。图 8. 对角方阵 Λ 非主对角线元素

把 $G = A^T A$ 带入上式并展开得到

$$\underbrace{\begin{bmatrix} \mathbf{v}_1^T A^T A \mathbf{v}_1 & \mathbf{v}_1^T A^T A \mathbf{v}_2 & \cdots & \mathbf{v}_1^T A^T A \mathbf{v}_p \\ \mathbf{v}_2^T A^T A \mathbf{v}_1 & \mathbf{v}_2^T A^T A \mathbf{v}_2 & \cdots & \mathbf{v}_2^T A^T A \mathbf{v}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_p^T A^T A \mathbf{v}_1 & \mathbf{v}_p^T A^T A \mathbf{v}_2 & \cdots & \mathbf{v}_p^T A^T A \mathbf{v}_p \end{bmatrix}}_{\mathbf{v}^T G \mathbf{v}} = \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix}}_A \quad (12)$$

整理上式得到

$$\underbrace{\begin{bmatrix} (\mathbf{A} \mathbf{v}_1)^T (\mathbf{A} \mathbf{v}_1) & (\mathbf{A} \mathbf{v}_1)^T (\mathbf{A} \mathbf{v}_2) & \cdots & (\mathbf{A} \mathbf{v}_1)^T (\mathbf{A} \mathbf{v}_p) \\ (\mathbf{A} \mathbf{v}_2)^T (\mathbf{A} \mathbf{v}_1) & (\mathbf{A} \mathbf{v}_2)^T (\mathbf{A} \mathbf{v}_2) & \cdots & (\mathbf{A} \mathbf{v}_2)^T (\mathbf{A} \mathbf{v}_p) \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{A} \mathbf{v}_p)^T (\mathbf{A} \mathbf{v}_1) & (\mathbf{A} \mathbf{v}_p)^T (\mathbf{A} \mathbf{v}_2) & \cdots & (\mathbf{A} \mathbf{v}_p)^T (\mathbf{A} \mathbf{v}_p) \end{bmatrix}}_{\mathbf{v}^T G \mathbf{v}} = \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_D \end{bmatrix}}_A \quad (13)$$

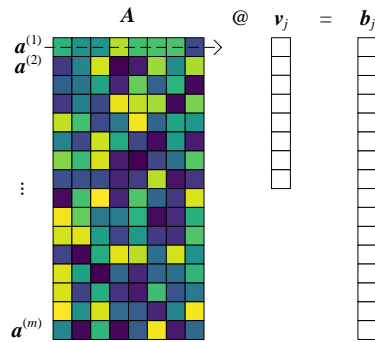
投影

令 $\mathbf{b}_j = \mathbf{A} \mathbf{v}_j$ ($j = 1, \dots, p$), 上式可以写成

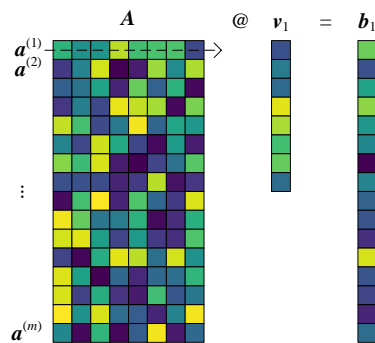
$$\underbrace{\begin{bmatrix} \mathbf{b}_1^T \mathbf{b}_1 & \mathbf{b}_1^T \mathbf{b}_2 & \cdots & \mathbf{b}_1^T \mathbf{b}_p \\ \mathbf{b}_2^T \mathbf{b}_1 & \mathbf{b}_2^T \mathbf{b}_2 & \cdots & \mathbf{b}_2^T \mathbf{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_p^T \mathbf{b}_1 & \mathbf{b}_p^T \mathbf{b}_2 & \cdots & \mathbf{b}_p^T \mathbf{b}_p \end{bmatrix}}_{\mathbf{B}^T \mathbf{B}} = \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix}}_A \quad (14)$$

如图 9 所示, 由于 \mathbf{v}_j 是单位列向量, 矩阵乘积 $\mathbf{A} \mathbf{v}_j$ 相当于矩阵 \mathbf{A} 向 $\text{span}(\mathbf{v}_j)$ 投影结果为 \mathbf{b}_j 。

矩阵 \mathbf{A} 的每个行向量相当于数据点。

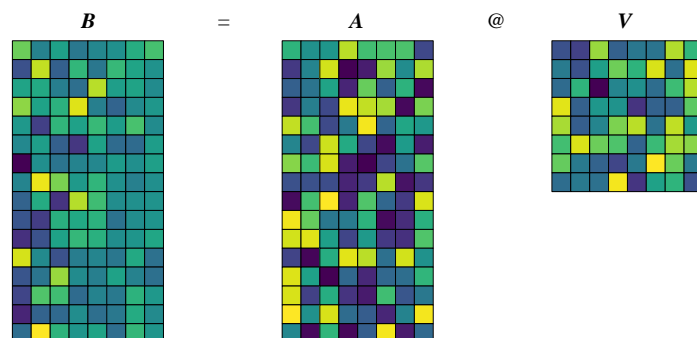
图 9. 矩阵 A 向 $\text{span}(v_j)$ 投影结果为 b_j

比如，图 10 所示为矩阵 A 向 $\text{span}(v_1)$ 投影结果为 b_1 。

图 10. 矩阵 A 向 $\text{span}(v_1)$ 投影结果为 b_1

正交化

图 11 所示为矩阵乘法 $B = AV$ 。几何角度来看，矩阵 A 向 $\text{span}(v_1, v_1, \dots, v_p)$ 投影结果为 B 。

图 11. 矩阵 A 向 $\text{span}(v_1, v_1, \dots, v_p)$ 投影结果为 B

将 $B = AV$ 带入(14)，得到

$$\mathbf{B}^T \mathbf{B} = \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_p \end{bmatrix}}_{\mathbf{A}} \quad (15)$$

如图 12 所示，上式相当于 \mathbf{B} 的格拉姆矩阵。

大家是否发现， \mathbf{B} 的列向量两两正交？！

也就是说， $\mathbf{B} = \mathbf{A}\mathbf{V}$ 相当于完成了矩阵 \mathbf{A} 的正交化。

➡ 本书前文介绍的 QR 分解也是正交化，请大家回顾。

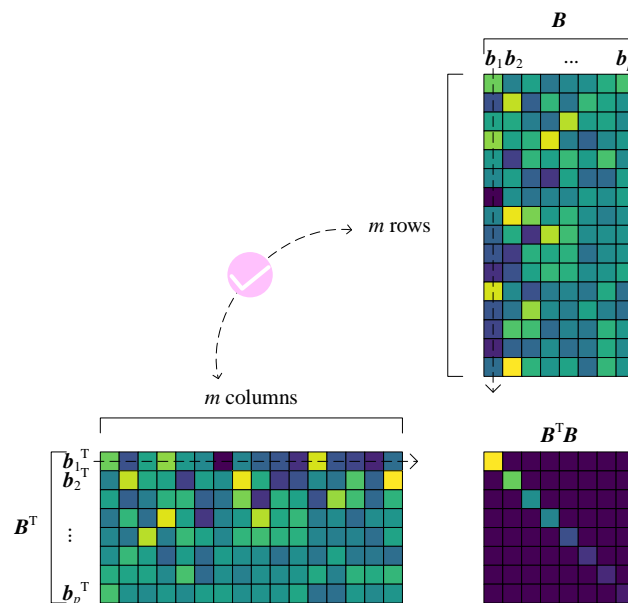
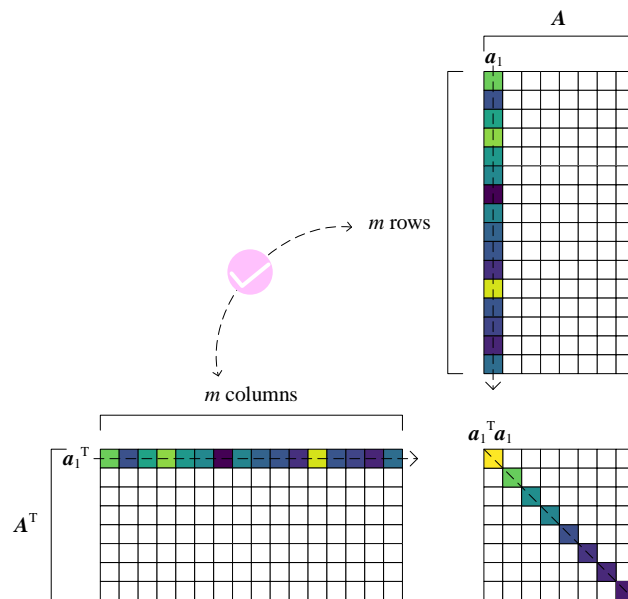


图 12. 矩阵 \mathbf{B} 的格拉姆矩阵 $\mathbf{B}^T \mathbf{B}$

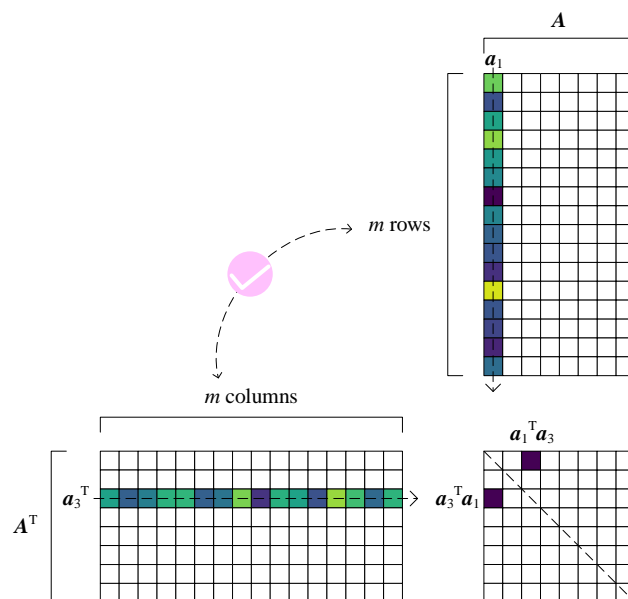
如图 13 所示，(14) 对角线元素显然可以写成 L^2 范数的平方：

$$\|\mathbf{b}_j\|_2^2 = \|\mathbf{A}\mathbf{v}_j\|_2^2 = \lambda_j \quad (16)$$

⚠ 注意，看到矩阵乘积结果为标量时，一定要想一想矩阵乘积能否写成 L^2 范数。

图 13. 矩阵 B 的格拉姆矩阵的对角线元素

如图 14 所示，矩阵 B 的格拉姆矩阵的非对角线元素为 0 。

图 14. 矩阵 B 的格拉姆矩阵的非对角线元素

几何视角

该怎么理解 (16)?

我们还是要拿出看家本领——几何视角。

如图 15 所示， A 的每个行向量为一个散点 \bullet 。这些散点 \bullet 向 $\text{span}(\mathbf{v}_j)$ 投影结果为 \mathbf{b}_j ，即图中直线上的 \bullet 。 \mathbf{b}_j 中的每个值就直线上的 \bullet 到原点的距离。

图 15 中黄色 \bullet 背景点代表矩阵 A 的第 i 行行向量为 $\mathbf{a}^{(i)}$ 。 $\mathbf{a}^{(i)}$ 向 \mathbf{v}_j 投影结果 $b_j^{(i)}$ 就是 $\mathbf{a}^{(i)}$ 在 $\text{span}(\mathbf{v}_j)$ 的坐标：

$$b_j^{(i)} = \mathbf{a}^{(i)} \mathbf{v}_j \quad (17)$$

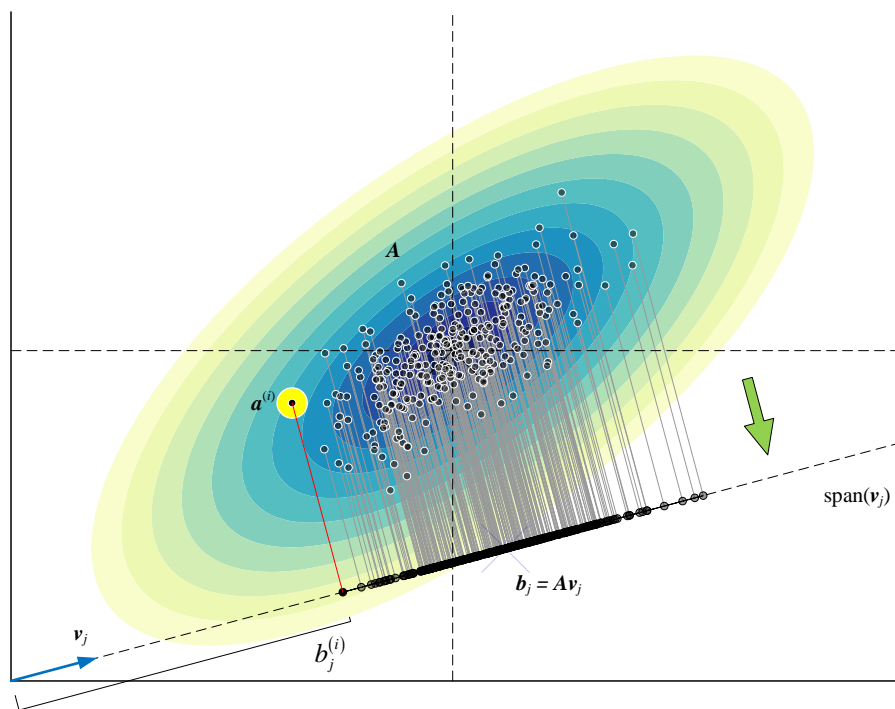


图 15. A 向 $\text{span}(\mathbf{v}_j)$ 正交投影结果为 \mathbf{b}_j ，几何视角



图 15 中椭圆实际上是另外一种距离度量——马氏距离。这是第 13 章、第 4 节要讲解的话题。

有了这个视角，我们知道 (16) 中 $\|\mathbf{b}_j\|_2^2$ 代表 $b_j^{(i)}$ 到原点距离 (有正负) 的平方和，即：

$$\|\mathbf{b}_j\|_2^2 = (b_j^{(1)})^2 + (b_j^{(2)})^2 + \dots + (b_j^{(n)})^2 = \lambda_j \quad (18)$$

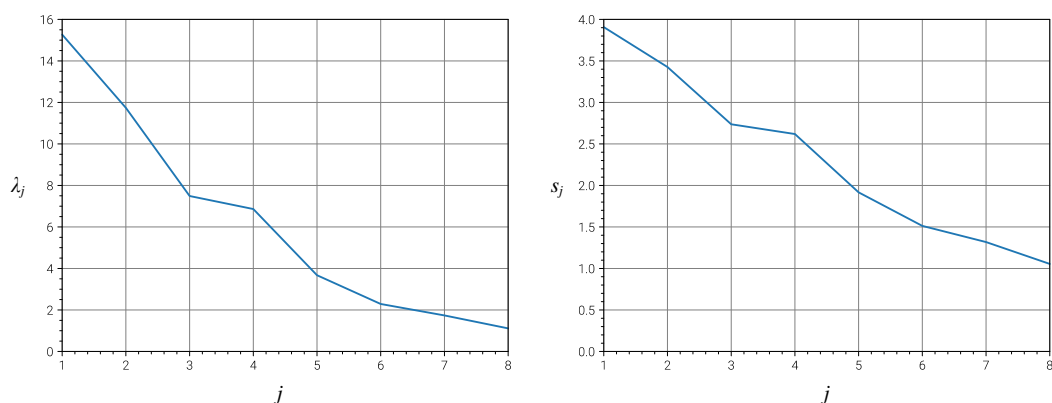
这些距离的平方和恰好等于特征值 λ_j 。

若 (14) 中特征值 λ_j 按大小排列，即 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_D$ 。这说明特征向量 \mathbf{v}_j 也有主次之分。

这图 16 所示，矩阵 A 朝不同特征向量 \mathbf{v} 投影，得到的 $\|\mathbf{b}\|_2^2 = \|A\mathbf{v}\|_2^2$ 有大有小，大小就对应特征值。这也意味着特征向量也有主次重要性之分。在 \mathbb{R}^D 有无数个 \mathbf{v} 中， A 朝第一特征向量 \mathbf{v}_1 投影对应的 $\|\mathbf{b}_1\|_2^2 = \|A\mathbf{v}_1\|_2^2$ 最大，最大值为 λ_1 。



这便是本书第 12 章要介绍的主成分分析的核心思路。

图 16. 特征值、 s_j (特征值平方根) 决定了特征向量的“解释力”大小

单位化

(18) 开方得到

$$\|\mathbf{b}_j\| = \sqrt{(b_j^{(1)})^2 + (b_j^{(2)})^2 + \dots + (b_j^{(n)})^2} = \sqrt{\lambda_j} \quad (19)$$

令 $\|\mathbf{b}_j\| = \sqrt{\lambda_j} = s_j$, 上式可以写成

$$\|\mathbf{b}_j\| = s_j \quad (20)$$

如果特征值大于 0, 我们可以对 \mathbf{b}_j 单位化

$$\mathbf{u}_j = \frac{\mathbf{b}_j}{s_j} \quad (21)$$

也就是说

$$\mathbf{b}_j = s_j \mathbf{u}_j \quad (22)$$

把 (22) 带入 $\mathbf{b}_j = \mathbf{A}\mathbf{v}_j$ 得到

$$\mathbf{A}\mathbf{v}_j = s_j \mathbf{u}_j \quad (23)$$

层层叠加

由于 \mathbf{V} 为正交矩阵, 矩阵 \mathbf{A} 可以写成

$$\begin{aligned}
 \underset{I}{A} &= \underset{I}{A} V V^T = A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_p^T \end{bmatrix} \\
 &= A \underbrace{(\mathbf{v}_1 \mathbf{v}_1^T + \mathbf{v}_2 \mathbf{v}_2^T + \cdots + \mathbf{v}_p \mathbf{v}_p^T)}_I \\
 &= A \mathbf{v}_1 \mathbf{v}_1^T + A \mathbf{v}_2 \mathbf{v}_2^T + \cdots + A \mathbf{v}_p \mathbf{v}_p^T
 \end{aligned} \tag{24}$$

上式中的 $\mathbf{v}_j \mathbf{v}_j^T$ 都是投影矩阵！

将 (23) 带入 (24)，整理

$$\begin{aligned}
 \mathbf{A} &= s_1 \mathbf{u}_1 \mathbf{v}_1^T + s_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + s_p \mathbf{u}_p \mathbf{v}_p^T \\
 &= \sqrt{\lambda_1} \mathbf{u}_1 \mathbf{v}_1^T + \sqrt{\lambda_2} \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sqrt{\lambda_p} \mathbf{u}_p \mathbf{v}_p^T
 \end{aligned} \tag{25}$$

也就是说我们把矩阵 \mathbf{A} 写成了若干形状相同矩阵的层层叠加，而 s_j (特征值平方根) 决定了特征向量的“解释力”大小，具体如图 16 所示。

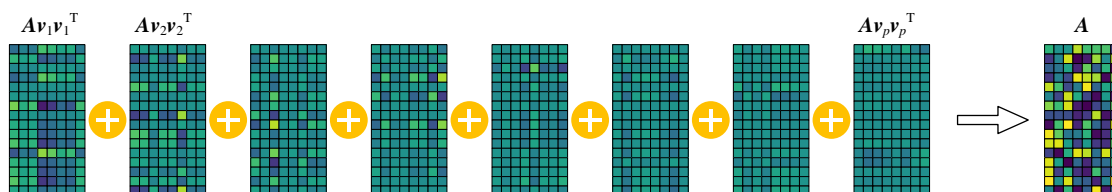


图 17. 层层叠加

一个全新的矩阵分解

当 j 取 $1 \sim p$ 整数值，(23) 对应一系列矩阵乘法

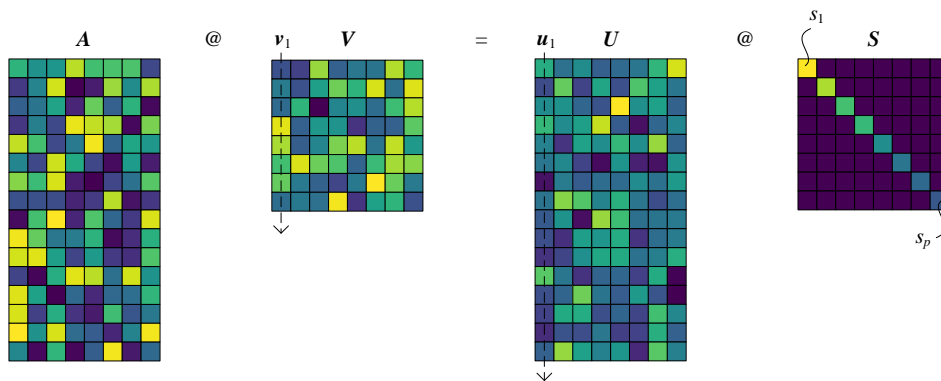
$$\begin{aligned}
 A \mathbf{v}_1 &= s_1 \mathbf{u}_1 \\
 A \mathbf{v}_2 &= s_2 \mathbf{u}_2 \\
 &\vdots \\
 A \mathbf{v}_p &= s_p \mathbf{u}_p
 \end{aligned} \tag{26}$$

把这一系列乘法写成

$$A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix} \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_p \end{bmatrix} \tag{27}$$

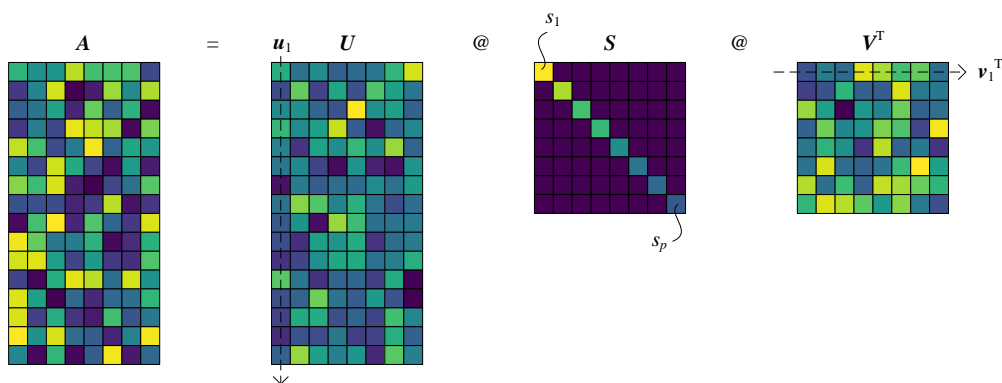
即

$$\mathbf{A} \mathbf{V} = \mathbf{U} \mathbf{S} \tag{28}$$

图 18. $AV = US$

由于 V 为正交矩阵， V 的逆等于其转置，(28) 可以写成

$$A = USV^T \quad (29)$$

图 19. 矩阵 A 的 SVD 分解

这实际上引出了本书第 14 章要讲解的奇异值分解，如图 19 所示。 S 的主对角线元素就是奇异值。大家很快就会发现，矩阵 U 的列向量也均为单位向量，且两两正交！

实践应用时还有一种更为常见的格拉姆矩阵——协方差矩阵——这是我们下一章要介绍的内容。



请大家用 DeepSeek/ChatGPT 等工具完成本节如下习题。

Q1. 请计算如下矩阵的格拉姆矩阵

► $\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$

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

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Q2. 对 **Q1** 中格拉姆矩阵谱分解。