西南財經大學
SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS
1925 - 2021
经世济民 孜孜以求

机器学习基础

奇异值分解

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

特征值分解

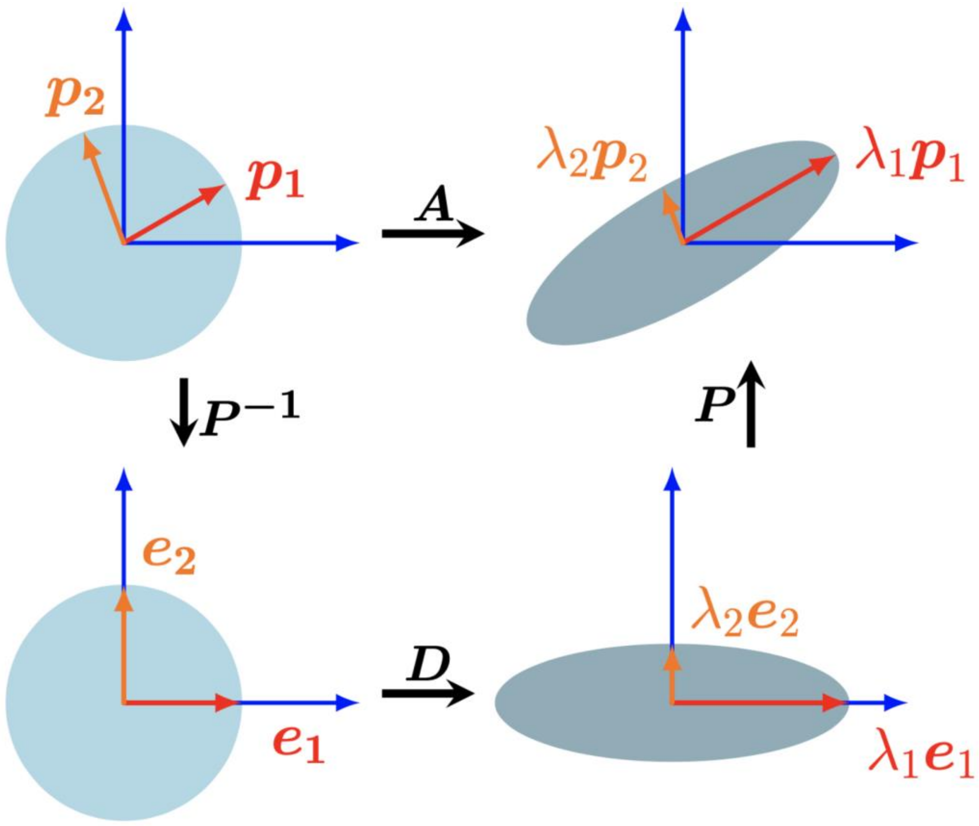
- 假设 n 阶方阵 A 有 n 个不同的特征值 λ_i 和对应的特征向量 P_i
- 令 $P=[P_1,P_2,\cdots,P_n]$ 有，

$$AP=[AP_1,AP_2,\cdots,AP_n]=[\lambda_1 P_1,\lambda_2 P_2,\cdots,\lambda_n P_n]=PD$$

- 其中，$D=\mathrm{diag}\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$， 则 $A=PDP^{-1}$ 是该矩阵的特征值分解
- $V_i=\mathrm{span}\{P_i\}$ 是 A 的不变子空间， 且 $\mathbb{R}^n=V_1\oplus V_2\oplus\cdots\oplus V_n$
- $\forall\eta\in\mathbb{R}^n,\ \eta=\sum_{i=1}^{n}a_i P_i$， 有 $A\eta=\sum_{i=1}^{n}a_i AP_i=\sum_{i=1}^{n}a_i\lambda_i P_i$

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

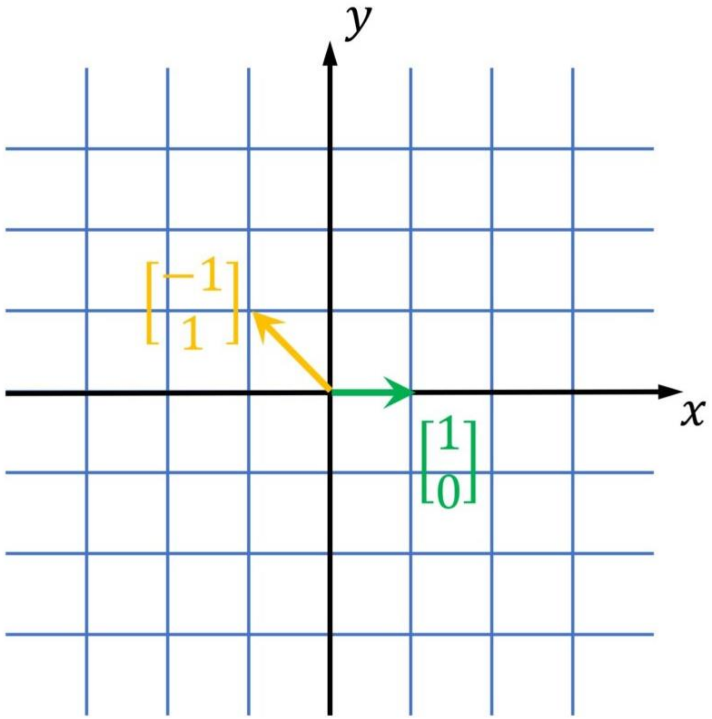
特征值分解的几何意义

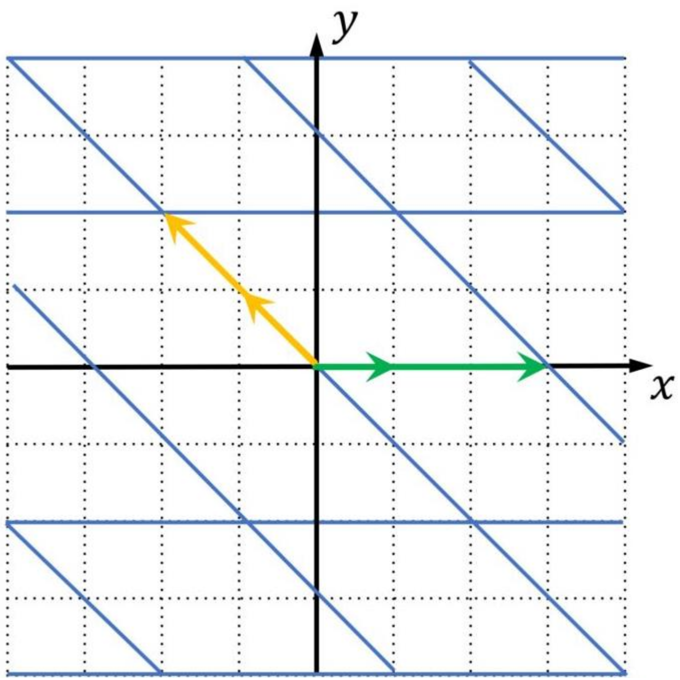


黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

例子

$$\begin{bmatrix}3 & 1\\0 & 2\end{bmatrix}=\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}\begin{bmatrix}3 & 0\\0 & 2\end{bmatrix}\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}^{-1}$$





黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

完全奇异值分解

- 对任意矩阵 $A \in \mathbb{R}^{m\times n}$， 存在 m 阶正交矩阵 U, n 阶正交矩阵 V， 有

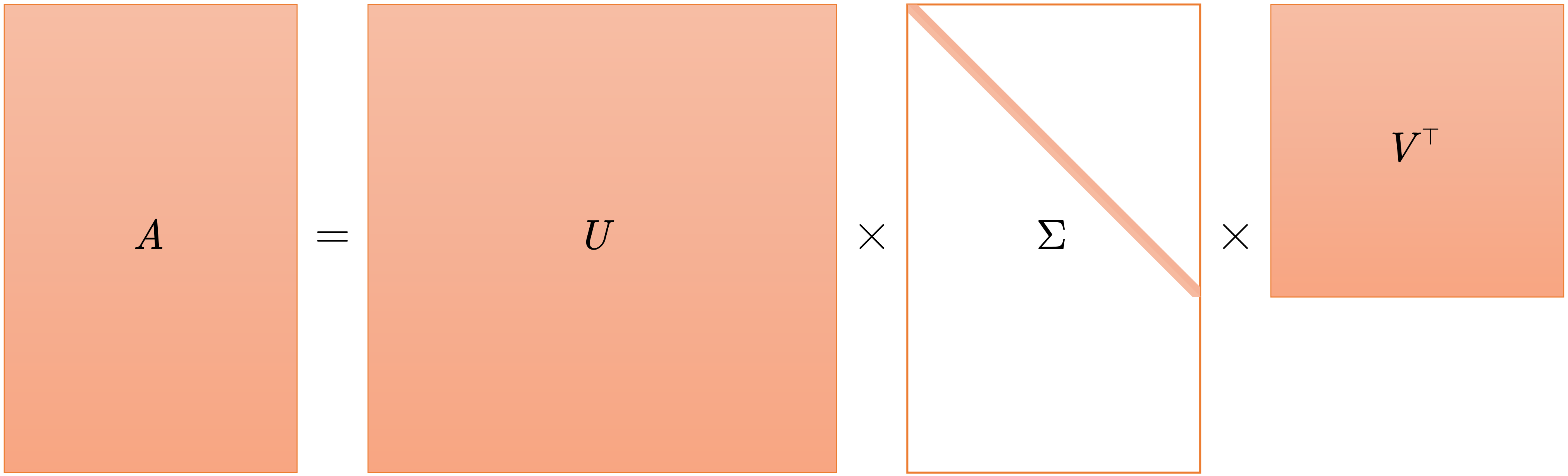
$$A = U\Sigma V^\top$$

其中 Σ 是由降序排列的非负的对角线元素组成的 $m\times n$ 矩阵对角矩阵满足 $\Sigma = \mathrm{diag}\{\sigma_1,\sigma_2,\cdots,\sigma_p\}$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$, $p = \min(m,n)$

- σ_i 称为矩阵的奇异值； 奇异值通常递减很快
- 奇异值分解不唯一

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

全奇异值分解



黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

紧奇异值分解

- 令 $rank(A)=r,\ r\leq\min(m,n)$， 则 $A=U_r\Sigma_rV_r^\top$ 为 A 的紧奇异值分解

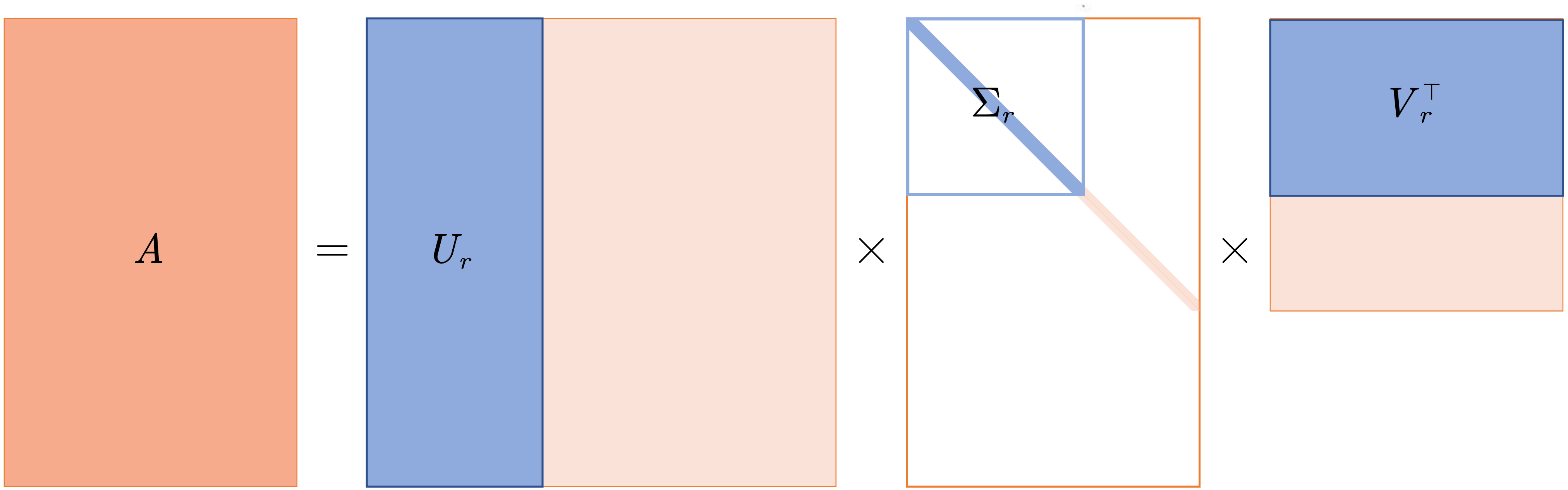
  其中 $U_r\in\mathbb{R}^{m\times r},\ V_r\in\mathbb{R}^{n\times r}$， Σ_r 是 r 阶对角矩阵

- U_r 是由完全奇异值分解中 U 的前 r 列，V_r 由 V 的前 r 列，Σ_r 由 Σ 的前 r 个对角线元素得到

- 紧奇异值分解不会损失信息，对应无损压缩

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

紧奇异值分解



黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

截断奇异值分解

- 令 $rank(A)=r,\ k<r$， 则 $A \doteq U_k\Sigma_k V_k^\top$ 为 A 的截断奇异值分解

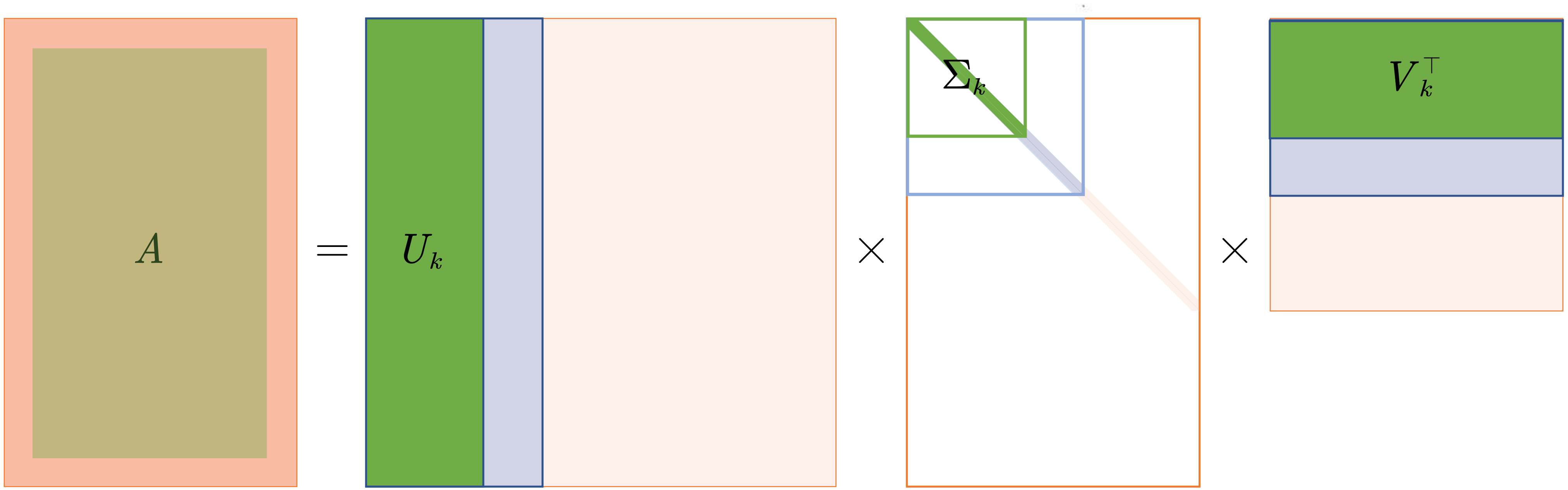
  其中 $U_k \in \mathbb{R}^{m\times k},\ V_k \in \mathbb{R}^{n\times}$， Σ_k 是 k 阶对角矩阵

- U_k 是由完全奇异值分解中 U 的前 k 列，V_k 由 V 的前 k 列，Σ_k 由 Σ 的前 k 个对角线元素得到

- 截断奇异值分解会损失信息，对应有损压缩

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

截断奇异值分解



黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

几何意义

- 矩阵 A 是从空间 $\mathbb{R}^n$ 到空间 $\mathbb{R}^m$ 的线性变换
- 正交矩阵 U, V 对应空间到自身的旋转变换

$$\|U\pi\|^2 = (U\pi)^\top \cdot (U\pi) = \pi^\top U^\top U\pi = \|\pi\|^2$$

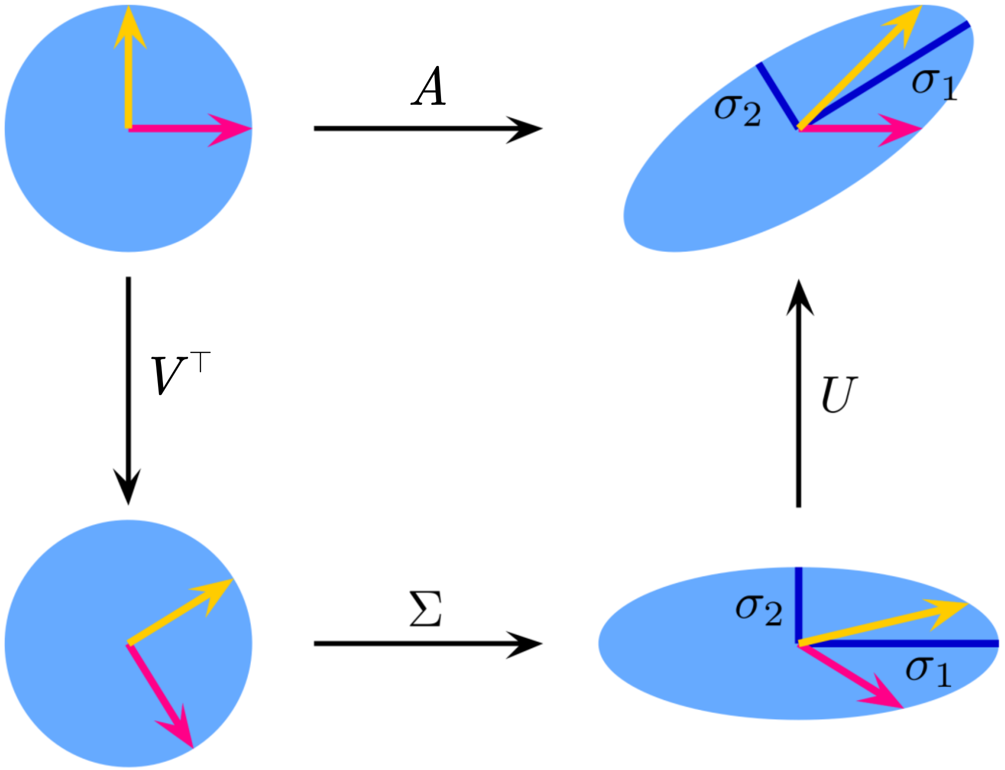
- 对角矩阵 Σ 表示第 i 个坐标放缩 σ_i 倍，是缩放变换

$$\Sigma\pi = (\sigma_1\pi_1, \sigma_2\pi_2, \cdots, \sigma_n\pi_n)$$

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

几何意义

- 对任意向量 $x \in \mathbb{R}^n$，Ax 等价于：
  1. 经过坐标系的旋转变换 $V^\top$
  2. 坐标轴的伸缩变换 Σ
  3. 坐标系的旋转变换 U
- 伸缩变换是唯一的，因为分解中只有一个伸缩变换
- 但旋转可以不唯一，为什么？



黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解的计算

- 求矩阵 $A^\top A$ 的特征值 λ_i $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0)$ 和对应的特征向量
- 将特征向量单位化，并将其构成正交矩阵 $V = [v_1 \ v_2 \ \cdots \ v_n]$
- 构造对角矩阵 $\Sigma = \mathrm{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_n\}$， 其中 $\sigma_i = \sqrt{\lambda_i}$
- 对 A 的前 r 个正特征值，令 $u_j = \dfrac{1}{\sigma_j} v_j,\ j \in \{1, 2, \cdots, r\}$
- 求$A^\top$的零空间的一组标准正交基$\{u_{r+1}, u_{r+2}, \cdots, u_m\}$，令 $U = [u_1 \ u_2 \ \cdots \ u_m]$
- 得到奇异值分解 $A = U \Sigma V^\top$

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

Frobenius范数

- 设矩阵$A=[a_{ij}]_{m\times n}$，则A的Frobenius范数为$\|A\|_F=\left(\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{ij})^2\right)^{\frac{1}{2}}$
- 其等价定义为$\|A\|_F=(\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2)^{\frac{1}{2}}$，其中$\sigma_i$是$A$的奇异值

$$\begin{aligned}\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{ij})^2 &= trace(A^\top A)\\ &= \sum_{i=1}^{n}\lambda_i(A^\top A)\\ &= \sum_{i=1}^{n}\sigma_i^2\end{aligned}$$

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

矩阵的最优近似

- **定理15.3** 设 $A \in \mathbb{R}^{m \times n}$ 且 $rank(A) = r$ ，σ_i，$i \in \{1, 2, \cdots, n\}$ 为其奇异值，并设 $\mathcal{M} = \{M \in \mathbb{R}^{m \times n} : rank(M) = k\}$，则

$$\|A - A'\|_F = \min_{S \in \mathcal{M}} \|A - S\|_F = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_n^2)^{\frac{1}{2}}$$

其中 $rank(A') = k$ 是矩阵 A 的截断奇异值分解

- 奇异值分解是在平方损失（Frobenius范数）意义下对矩阵的最优近似

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

矩阵的外积展开式

- 令 u_i, v_i $(i \in \{1, 2, \cdots, n\})$ 分别是矩阵 U, V 的第 i 列，则 A 的外积展开式为

$$A = U\Sigma V^\top = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top + \cdots + \sigma_n u_n v_n^\top$$

- 令 $A_k = u_k v_k^\top$，有 $A = \sum_{k=1}^{n} \sigma_k A_k$，即将 A 分解为矩阵的有序加权求和
- 并且 A_k 是矩阵 A 在秩为 k 矩阵中在Frobenius范数意义下的 A的最优近似矩阵

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

原图：1080 * 608

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

K = 1

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

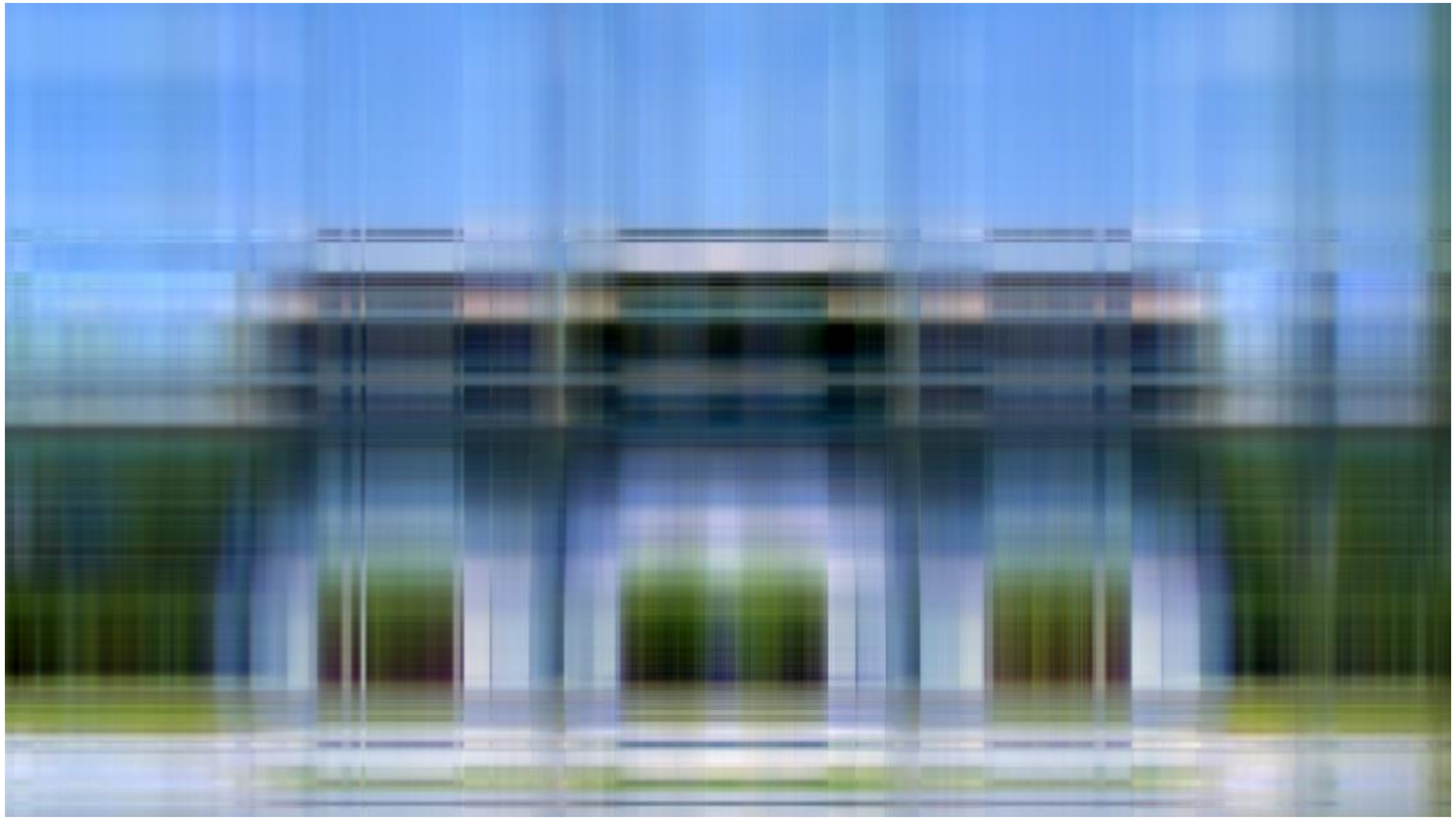
奇异值分解应用：图像压缩

K = 2

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

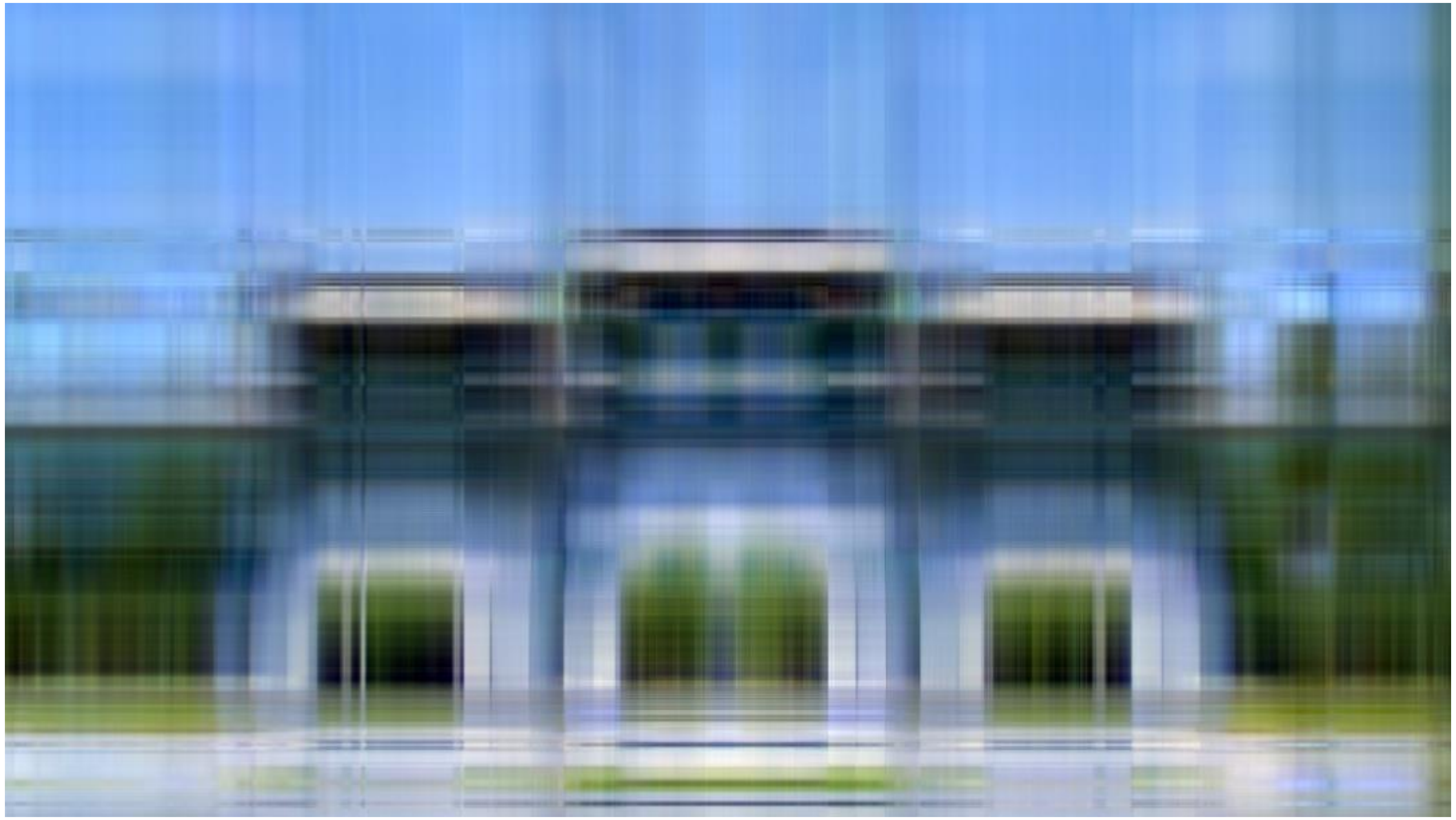
K = 3



黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

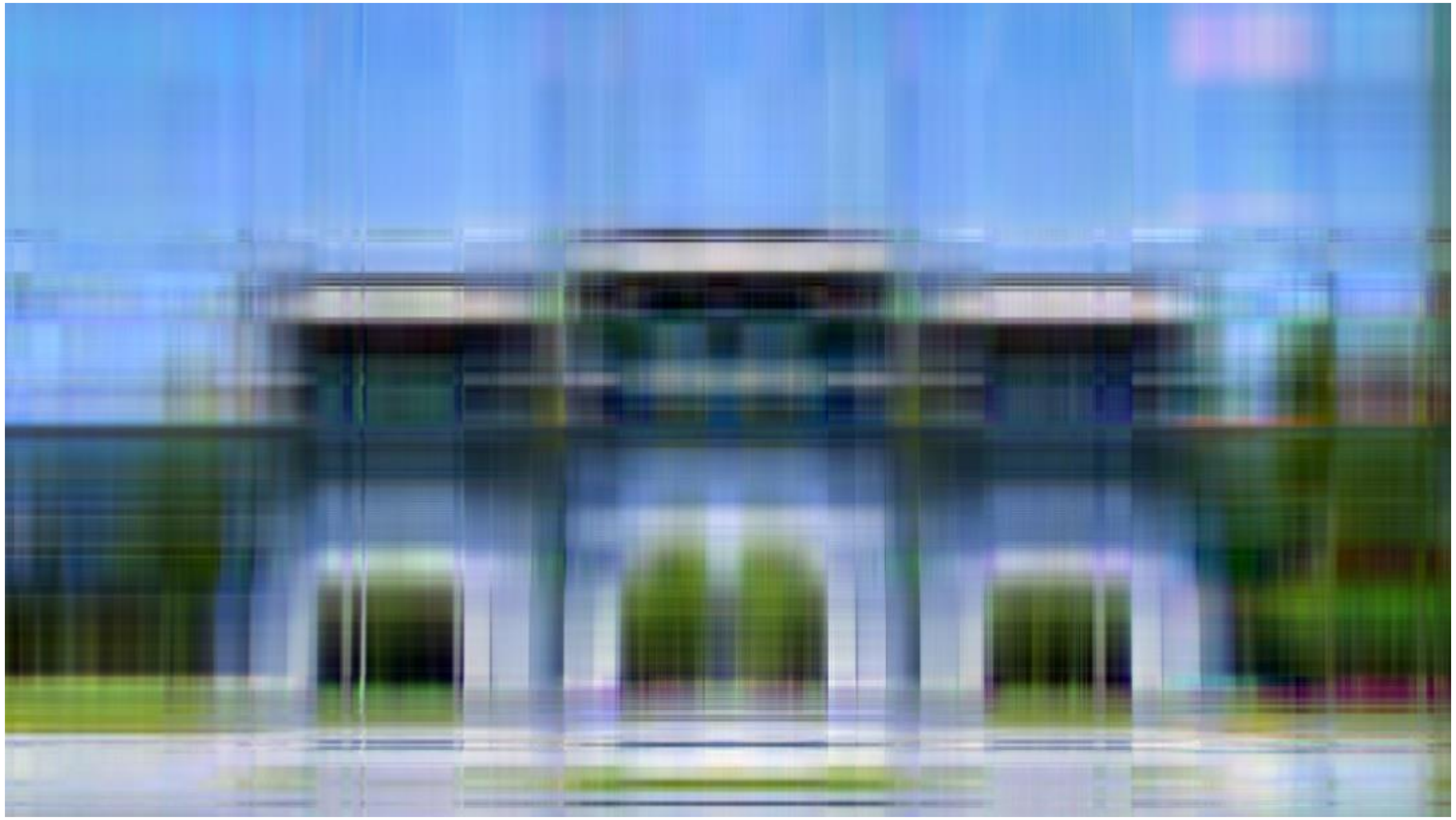
K = 4



黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

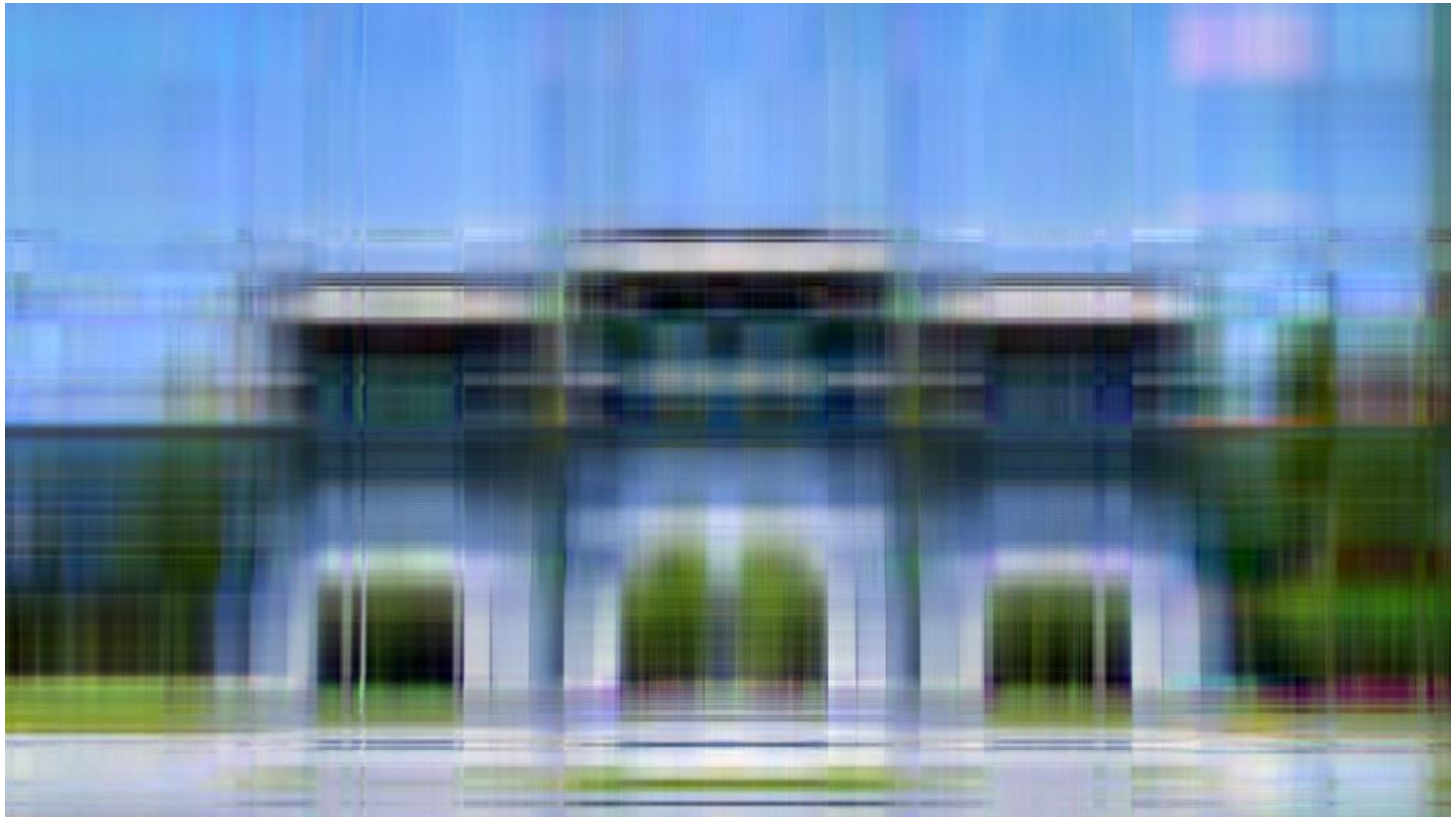
K = 5



黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

K = 5



黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

K = 6

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

K = 10

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

K = 15

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

K = 30

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

K = 50

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

K = 100

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

K = 300

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

K = 608

黄迟 计算机与人工智能学院
huangchi@swufe.edu.cn

奇异值分解应用：图像压缩

- 原图需要存储数据

$$1080 \times 608 \times 3 = 1.96992 \times 10^6$$

- 采用截断奇异值分解，取 $K = 100$， 需要存储数据

$$(1080 \times 100 + 100 + 608 \times 100) \times 3 = 506700$$

- 取 $K = 300$， 需要存储数据

$$(1080 \times 300 + 300 + 608 \times 300) \times 3 = 1.5201 \times 10^6$$