

15 奇异值分解

从矩阵的几何作用到低秩近似

1 矩阵的几何作用

从输入方向、伸缩强度和输出方向理解矩阵

■ 输入

- $x \in \mathbb{R}^n$ 描述输入空间中的方向与长度

■ 线性变换

- $A \in \mathbb{R}^{m \times n}$ 对所有输入使用同一线性规则

■ 输出

- $y = Ax \in \mathbb{R}^m$

■ 要解决的问题

- 哪些输入方向最重要，它们被映射到哪里，长度改变多少？

■ 观察对象

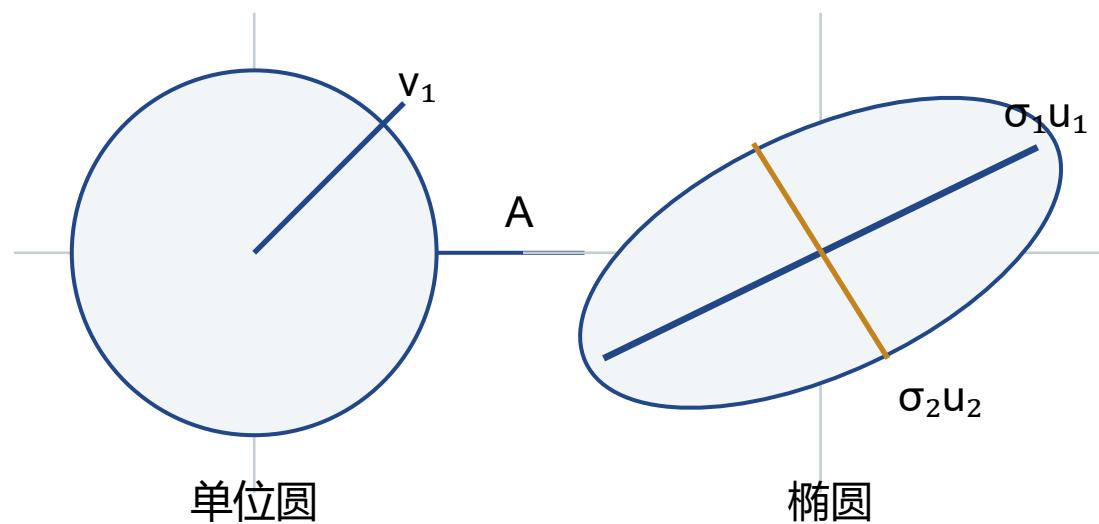
- 单位圆包含所有长度为 1 的二维输入方向

■ 经过矩阵 A

- 一般得到椭圆，也可能退化为线段或一个点

■ 几何读法

- 椭圆半轴方向是左奇异向量
- 半轴长度是奇异值
- 对应半轴的原始输入方向是右奇异向量



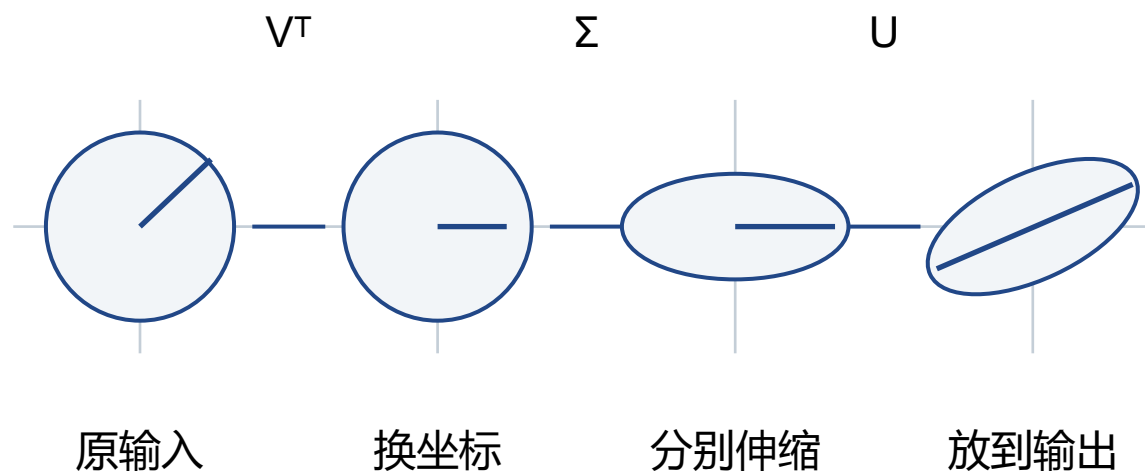
半轴方向: u_i ; 半轴长度: σ_i

■ 实际作用次序从右向左读取

- V^T : 用一组正交基描述输入方向
- Σ : 沿坐标轴分别伸缩, 或把某些方向压到零
- U : 把结果放到输出空间的正交方向

■ 整体关系

- $A = U\Sigma V^T$



观察: V^T 保持圆形; Σ 改变形状; U 改变方向

- 对 $A \in \mathbb{R}^{m \times n}$, 完整 SVD 为
 - $A = U\Sigma V^T$
- **输出正交因子** $U \in \mathbb{R}^{m \times m}$
 - 列向量正交归一, 描述输出方向
- **输入正交因子** $V \in \mathbb{R}^{n \times n}$
 - 列向量正交归一, 描述输入方向
- **伸缩因子** $\Sigma \in \mathbb{R}^{m \times n}$
 - 对角线上为 $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$

- 令 $p = \min(m, n)$, 配对索引为 $i = 1, \dots, p$
- **输入方向:** v_i
- **伸缩强度:** σ_i
- **输出方向:** u_i
- **核心关系**
 - $Av_i = \sigma_i u_i$
 - $A^T u_i = \sigma_i v_i$
- 当 $\sigma_i = 0$ 时, $Av_i = 0$, 因此 $v_i \in \text{null}(A)$
 - 此时不能通过 $u_i = Av_i / \sigma_i$ 构造左奇异向量

完整、薄和截断 SVD 保留的内容不同

■ 完整 SVD

- 保存输入和输出空间的完整正交基

■ 薄 SVD

- 令 $p = \min(m, n)$, 统一写为 $A = U_p \Sigma_p V_p^T$
- $U_p \in \mathbb{R}^{m \times p}, \Sigma_p \in \mathbb{R}^{p \times p}, V_p \in \mathbb{R}^{n \times p}$
- 去掉完整 SVD 中不参与矩阵乘积的额外基向量, 保留 $p = \min(m, n)$ 个可配对方向

■ 截断 SVD

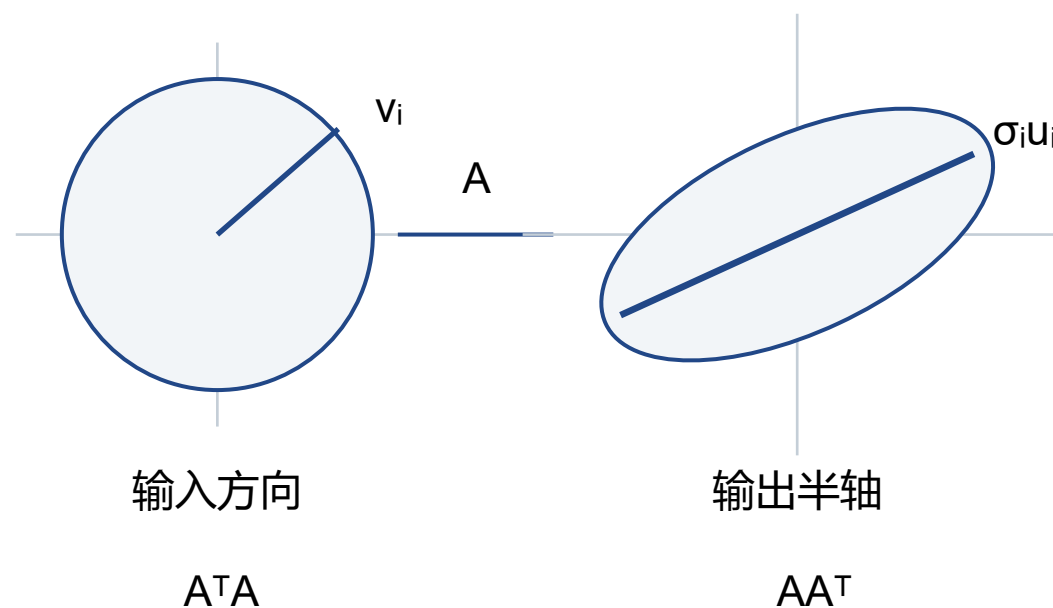
- 只保留最大的 k 个奇异值及对应向量
- 得到秩至多为 k 的近似 A_k

2 奇异方向怎样构造

从输入空间的方向出发，理解伸缩强度与输出方向

两个对称矩阵从两侧寻找同一组伸缩

- **输入侧：** $A^T A v_i = \sigma_i^2 v_i$
 - v_i 是圆上被送往椭圆半轴的原方向
- **输出侧：** $A A^T u_i = \sigma_i^2 u_i$
 - u_i 是椭圆的半轴方向
- **配对：** $A v_i = \sigma_i u_i$
 - 两侧非零特征值相同，都是 σ_i^2



同一非零谱： σ_i^2 ；方向由 A 配对

- **为什么转向 $A^T A$**

- 要比较单位输入方向的像长度; $\|Av\|^2 = v^T A^T A v$ 把这个问题留在输入空间

- **第一步：求输入方向**

- $A^T A v_i = \lambda_i v_i$

- **第二步：求伸缩强度**

- $A^T A$ 半正定, 所以 $\lambda_i \geq 0$

- $\sigma_i = \sqrt{\lambda_i}$

- **这一步得到**

- 右奇异方向 v_i 和相应伸缩强度 σ_i
- 因为 $v_i^T A^T A v_i = \|A v_i\|^2$, 所以 σ_i 正是该输入方向的像长度

- 对 $\sigma_i > 0$, 把 Av_i 归一化得到
 - $u_i = \frac{Av_i}{\sigma_i}, Av_i = \sigma_i u_i$
- 对 $\sigma_i = 0$, $Av_i = 0$
 - 这个输入方向在零空间中, 不能用除以 σ_i 的公式构造 u_i
- 所有非零方向的作用相加, 就能重构原矩阵
 - $A = \sum_{i=1}^r \sigma_i u_i v_i^T$
 - $r = \text{rank}(A)$, 即非零奇异值的数量

二维算例：从 $A^T A$ 得到奇异三元组

■ 输入与输入侧谱

- $A = \begin{bmatrix} 0 & 3 \\ 2 & 0 \end{bmatrix}, A^T A = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$

■ 按奇异值降序配对

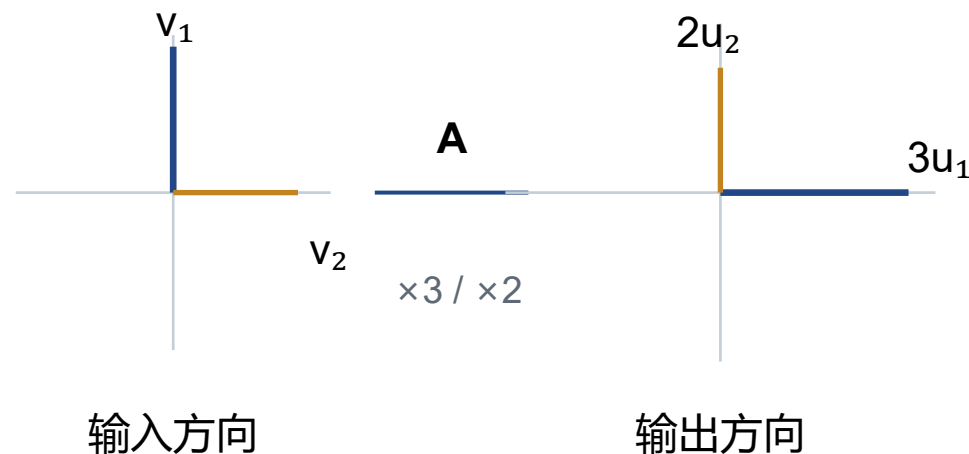
- $\sigma_1 = 3, v_1 = (0,1)^T; \sigma_2 = 2, v_2 = (1,0)^T$

■ 映到输出方向

- $u_1 = Av_1/\sigma_1 = (1,0)^T; u_2 = Av_2/\sigma_2 = (0,1)^T$

■ 重构闭环

- $A = 3u_1 v_1^T + 2u_2 v_2^T$

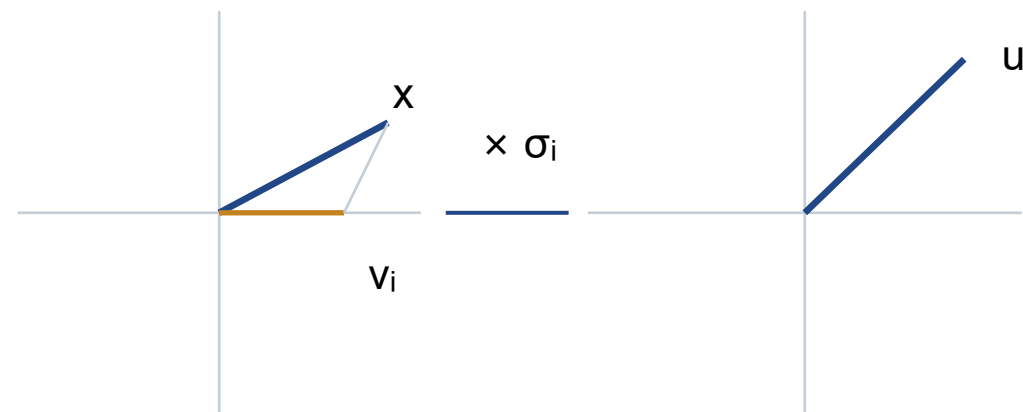


v_1 映到 u_1 ; v_2 映到 u_2

3 秩一分量与低秩近似

矩阵可以看成按奇异值排序的简单作用之和

- $A = \sum_{i=1}^r \sigma_i u_i v_i^T$
- $v_i^T x$ 读取输入在方向 v_i 上的坐标
- σ_i 调整这个坐标的强度
- u_i 把结果放到输出方向
- 每个 $\sigma_i u_i v_i^T$ 的秩至多为 1



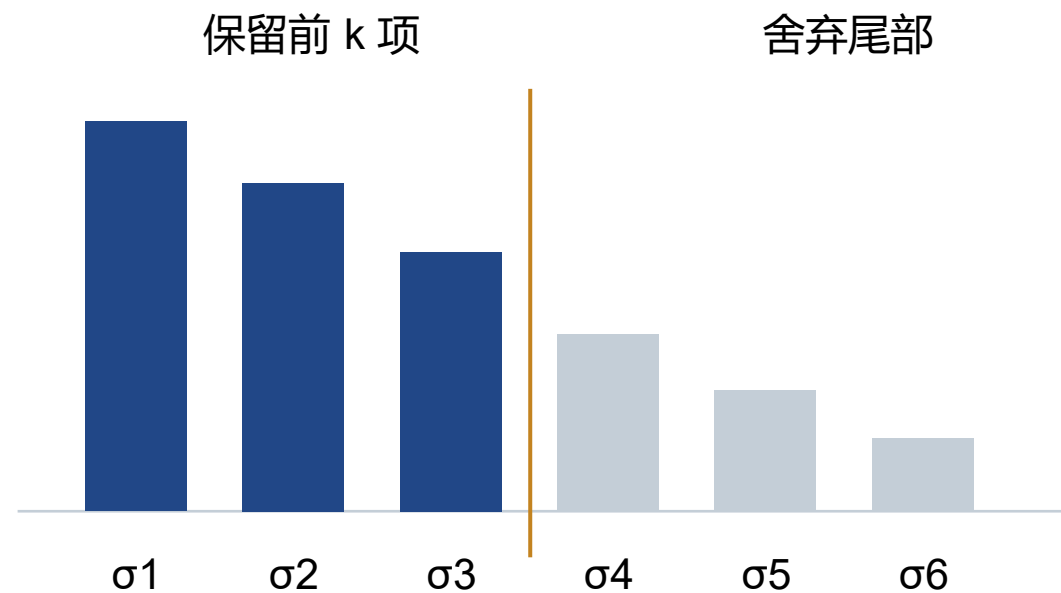
投影: $v_i^T x$

输出的一项

只沿一个输出方向产生贡献

截断 SVD 只保留最强的 k 个方向

- $A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$
- **保留**
 - 最大奇异值对应的输入输出关系
- **舍弃**
 - 其余较弱方向及其细节
- **结果**
 - A_k 的秩至多为 k
 - 存储从约 mn 个数降为约 $k(m + n + 1)$ 个数



示意谱，非具体矩阵数据

- 各秩一项相互正交，舍弃代价由奇异值决定；因此保留最大的 k 项

- **谱范数误差**

- $\min_{rank(B) \leq k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$

- **Frobenius 范数误差**

- $\min_{rank(B) \leq k} \|A - B\|_F^2 = \|A - A_k\|_F^2 = \sum_{i>k} \sigma_i^2$

- **“最佳” 的准确含义**

- 固定近似秩并使用指定矩阵范数衡量误差，不保证下游预测效果最佳

- 保留的 Frobenius 能量比例：当 $\sum_i \sigma_i^2 > 0$ 时

- $$\rho_k = \frac{\sum_{i \leq k} \sigma_i^2}{\sum_i \sigma_i^2}$$

- 阈值选取规则

- $k = \min\{j: \rho_j \geq \eta\}$
- 给定阈值 η , 取累计能量首次达到阈值的最小 k
- 能量比例只反映矩阵近似误差, 最终 k 仍应结合验证集性能或任务约束选择

4 SVD 可以怎样用

保留、舍弃或反演奇异方向产生不同用途

- 灰度图像是像素矩阵 $A \in \mathbb{R}^{m \times n}$, 截断 SVD 用 $k(m + n + 1)$ 个参数表示 A_k
- 图下注明的能量保留率与相对 Frobenius 误差均由当前图像实际计算
- 奇异值衰减越快, 较小的 k 越可能保留主要结构



原图

160×240
38,400 像素



k=5

参数 2,005
能量 91.9%
相对误差 28.4%



k=30

参数 12,030
能量 97.5%
相对误差 15.7%

同一张图、同一次实际 SVD 计算

■ 建模假设

- 有用结构集中在少数强方向，部分扰动分散在弱方向

■ 操作

- 对观测矩阵做截断 SVD，舍弃选定的弱方向

■ 输出

- 得到较低秩、结构更集中的矩阵

■ 必要边界

- 小奇异值不必然是噪声
- 信号也可能位于弱方向，阈值需要结合任务验证

- 约定 $X \in \mathbb{R}^{N \times d}$ ：行是样本，列是特征
- v_i ：特征空间中的方向
- $\sigma_i u_i = X v_i$ ：所有样本在该方向上的得分
- $\sigma_i u_i v_i^T$ ：一个“样本得分 $\times$ 特征模式”的秩一分量

- 约定 $X \in \mathbb{R}^{N \times d}$: 样本为行, 特征为列, 且 $N > 1$
- **中心化**
 - $X_c = X - \mathbf{1}\mu^T$
- **分解**
 - $X_c = U\Sigma V^T$
- **降维坐标**
 - $Z = X_c V_k = U_k \Sigma_k$
- 特征协方差为 $X_c^T X_c / (N - 1)$, 第 i 个主方向的样本方差为
 - $Var(Z_i) = \frac{\sigma_i^2}{N-1}$
- PCA 的方差最大化、协方差解释和选维方法留到下一单元

- $A^+ = V\Sigma^+U^T$
- U^T 把目标分解到输出奇异方向
- Σ^+ 对每个非零奇异值取倒数
- V 把结果放回输入空间
- 当 $Ax = b$ 没有精确解时, $x = A^+b$ 给出最小二乘解中的最小范数解

小奇异值的倒数会放大相应方向上的扰动

- 若 b 在方向 u_i 上有分量 $u_i^T b$
- 伪逆在对应输入方向产生
 - $\frac{u_i^T b}{\sigma_i} v_i$
- 当 σ_i 很小时
 - b 的微小变化可能造成解的明显变化
- 令 $p = \min(m, n)$ 。当 A 满秩时
 - $\kappa_2(A) = \frac{\sigma_1}{\sigma_p}$
- 当 $\text{rank}(A) < p$ 时，普通线性求解的稳定性分析通常把条件数视为无穷大
- 若只考察可恢复的非零奇异方向， σ_1/σ_r 描述这些方向之间的伸缩差异
 - 它不表示原映射一定可逆；矩形矩阵通常不存在双侧逆

■ 压缩

- 保留前 k 项，以较少参数近似原矩阵

■ 降噪

- 在结构假设支持时舍弃候选弱扰动方向

■ PCA

- 投影到已去均值数据的主要右奇异方向

■ 最小二乘

- 对非零奇异值取倒数，反演可恢复方向

- 正交因子可能包含反射，“旋转”只是一种简化直觉
- 奇异值唯一，奇异向量的符号不唯一；重复奇异值的子空间基也不唯一
- 截断 SVD 的最优性针对固定秩和指定矩阵范数
- 小奇异值表示弱伸缩方向，不自动表示噪声或无用信息

■ 分解

- V^T 读取输入方向, Σ 给出伸缩强度, U 指定输出方向

■ 构造

- $A^T A$ 的特征向量给出右奇异方向, 特征值平方根给出奇异值

■ 表示

- 矩阵是按奇异值排序的秩一分量之和

■ 使用

- 截断支持低秩近似, 投影支持 PCA, 取倒数支持伪逆