第 29 讲 奇异值分解

Singular value decomposition

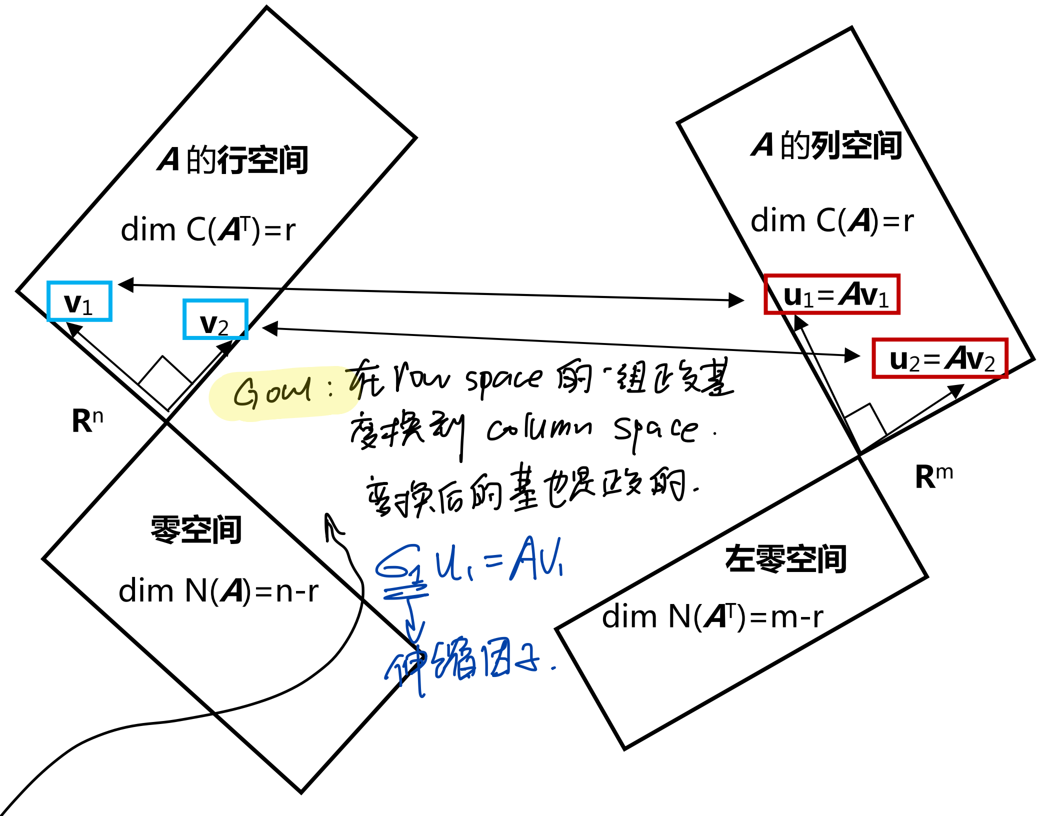
本讲介绍奇异值分解（Singular value decomposition），简称 SVD。这是矩阵最终也是最好的分解，任意矩阵可分解为 $\boldsymbol{A}=\boldsymbol{U\Sigma V}^{\mathrm{T}}$，分解结果为正交矩阵 $\boldsymbol{U}$，对角阵 Σ 和正交矩阵 $\boldsymbol{V}$。

SPD.

如果矩阵 $\boldsymbol{A}$ 是~~正定矩阵~~，它的奇异值分解就是 $\boldsymbol{A}=\boldsymbol{Q\Lambda Q}^{\mathrm{T}}$，一个正交矩阵 $\boldsymbol{Q}$ 就可以满足分解，而不需要两个。而对于可对角化的矩阵有 $\boldsymbol{A}=\boldsymbol{S\Lambda S}^{-1}$，但特征向量矩阵 $\boldsymbol{S}$ 并不是正交矩阵，而 SVD 中的 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 都是正交矩阵。

如何实现

可以将矩阵 $\boldsymbol{A}$ 视为一种线性变换操作，将其行空间中的一个向量 $\mathbf{v}_1$，变为其列空间中的向量 $\mathbf{u}_1=\boldsymbol{A}\mathbf{v}_1$。奇异值分解就是要在行空间中寻找一组正交基，将其通过矩阵 $\boldsymbol{A}$ 线性变换生成列空间中的一组正交基 $\boldsymbol{A}\mathbf{v}_i=\sigma_i\mathbf{u}_i$。



找出矩阵 $\boldsymbol{A}$ 行空间中的正交基很容易，Gram-Schmidt 正交化过程就可以做到，但是随便的一组正交基经过矩阵矩阵 $\boldsymbol{A}$ 变换得到的向量并不一定正交，因此满足此要求的行空间的正交基非常特殊。而矩阵 $\boldsymbol{A}$ 零空间的向量所对应的是矩阵 Σ 对角线上的 0 元素，因此很容易处理。

用矩阵数学语言描述这一过程

→ orthonormal 正交归一.

$[u_1, u_2 \dots u_r]$

Goal:

问题的核心就是找到行空间中一组特殊的正交基 s.t.:

用矩阵的形式表示出来.

$$A\underbrace{[\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_r]}_{\text{basis vectors in the row space.}} = [\sigma_1\mathbf{u}_1 \quad \sigma_2\mathbf{u}_2 \quad \cdots \quad \sigma_r\mathbf{u}_r]$$

$$= \underbrace{[\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_r]}_{\text{basis vectors in the column space}} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

multiplying factors.

$AV=U\Sigma$
(前边像矩阵的形式化).

如果加入零空间的部分，等式就变为 $\boldsymbol{AV}=\boldsymbol{U\Sigma}$。其中零空间对应的正交基 $\mathbf{v}_{r+1}\dots\dots\mathbf{v}_n$，经过线性变换得到 $\boldsymbol{A}\mathbf{v}=0$，对应 Σ 矩阵中对角线最后的特征值 $\sigma_{r+1}=\sigma_{r+2}=\dots=0$。在等式 $\boldsymbol{AV}=\boldsymbol{U\Sigma}$ 两侧右乘 $\boldsymbol{V}^{-1}$ 得到 $\boldsymbol{A}=\boldsymbol{U\Sigma V}^{-1}=\boldsymbol{U\Sigma V}^{\mathrm{T}}$。

现在的问题就是怎么找到符合要求的向量 $\mathbf{v}_i$ 和 $\mathbf{u}_i$。→ Goal: 做些处理将U消去. (A^TA)

为了得到这两个正交矩阵，考虑首先解决其中的一个，在等式 $\boldsymbol{A}=\boldsymbol{U\Sigma V}^{\mathrm{T}}$ 两侧分别乘上等式 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{V\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}$ 两侧的项：

A^TA 至少是半正定矩阵.

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{V\Sigma}^{\mathrm{T}}\underbrace{\boxed{\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}}}_{I\ (\text{unitary})}\boldsymbol{\Sigma V}^{\mathrm{T}}$$
$$= \boldsymbol{V\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma V}^{\mathrm{T}}$$
$$= \boldsymbol{V}\begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_r^2 \end{bmatrix}\boldsymbol{V}^{\mathrm{T}}$$

$AV=U\Sigma$
$AVV^{-1}=U\Sigma V^{-1}$
$A=U\Sigma V^{-1}$
$=U\Sigma V^T$ 因为unitary.

同理：消除V就用AA^T.

上式其实是正定矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的正交分解，$\mathbf{v}_i$ 就是矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征向量，σ_i^2 就是矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征值，奇异值 σ_i 要取正平方根。用同样的办法也可以求得 $\boldsymbol{U}$，它的列向量就是矩阵 $\boldsymbol{AA}^{\mathrm{T}}$ 的特征向量。

例 1：矩阵 $\boldsymbol{A}=\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$，求其 SVD 分解。

矩阵为可逆矩阵，秩为 2，则需要在行空间中求得 $\mathbf{v}_1$，$\mathbf{v}_2$，列空间中求得 $\mathbf{u}_1$，$\mathbf{u}_2$，以及伸缩因子 σ_1，σ_2。

计算得 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix}\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}=\begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$，它的特征向量为 $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ 和 $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$。标准化得到 $\mathbf{v}_1=\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$，$\mathbf{v}_2=\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$，求得 $\sigma_1^2=32$，$\sigma_2^2=18$。

V_1, V_2 是经过标准化的特征向量.

求 $\boldsymbol{U}$ 的过程可以利用矩阵 $\boldsymbol{AA}^{\mathrm{T}}$。 $\because AA^T=U\Sigma V^T\cdot(V\Sigma^TU^T)=U\Sigma\Sigma^TU^T$

也至少是半正定矩阵.

$\boldsymbol{AA}^{\mathrm{T}}=\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}\begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix}=\begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$，求得 $\sigma_1^2=32$，$\sigma_2^2=18$，它的特征向量为 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$。→ 已经归一化了.

AA^T的特征值与A^TA的特征值完全相同.

$$\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{32} & 0 \\ 0 & \sqrt{18} \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

希望对角线上的元素都为正的.

但是我们直接将$\begin{bmatrix}1&0\\0&1\end{bmatrix}$作为矩阵 U 代入，会产生错误 $\boldsymbol{U\Sigma V}^{\mathrm{T}}=\begin{bmatrix}4&4\\3&-3\end{bmatrix}$。这是因为确定特征向量的过程中，特征向量反向仍然符合要求，通过现在的方法无法确认向量的符号，但是一旦我们确认 **v** 的方向之后，**u** 的方向也就随之确定，将 **v** 代入 $\boldsymbol{AV}=\boldsymbol{U\Sigma}$ 计算 **u** 可以避免这种问题。**u** 和 **v** 之间的符号联系在进行 $\boldsymbol{AA}^{\mathrm{T}}$ 的计算时被切断了，而用 $\boldsymbol{AV}=\boldsymbol{U\Sigma}$ 可以避免此问题。

例 2：奇异阵 $\boldsymbol{A}=\begin{bmatrix}4&3\\8&6\end{bmatrix}$，求其 SVD 分解。

矩阵的秩为 1，行空间和列空间都是 1 维的。行空间和零空间可以找到一组正交基转换得到列空间和左零空间的一组正交基。

很容易确定 $\mathbf{v}_1=\begin{bmatrix}0.8\\0.6\end{bmatrix}$，$\mathbf{v}_2=\begin{bmatrix}0.6\\-0.8\end{bmatrix}$，$\mathbf{u}_1=\frac{1}{\sqrt{5}}\begin{bmatrix}1\\2\end{bmatrix}$，$\mathbf{u}_2=\frac{1}{\sqrt{5}}\begin{bmatrix}2\\-1\end{bmatrix}$。

$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\begin{bmatrix}4&8\\3&6\end{bmatrix}\begin{bmatrix}4&3\\8&6\end{bmatrix}=\begin{bmatrix}80&60\\60&45\end{bmatrix}$，秩 1 矩阵，很容易求得$\sigma_1^2=125$，$\sigma_2^2=0$。

矩阵的 SVD 分解为：$\boldsymbol{A}=\frac{1}{\sqrt{5}}\begin{bmatrix}1&2\\2&-1\end{bmatrix}\begin{bmatrix}\sqrt{125}&0\\0&0\end{bmatrix}\begin{bmatrix}0.8&0.6\\0.6&-0.8\end{bmatrix}$

做奇异值分解就是在矩阵的四个子空间中寻找到合适的基：

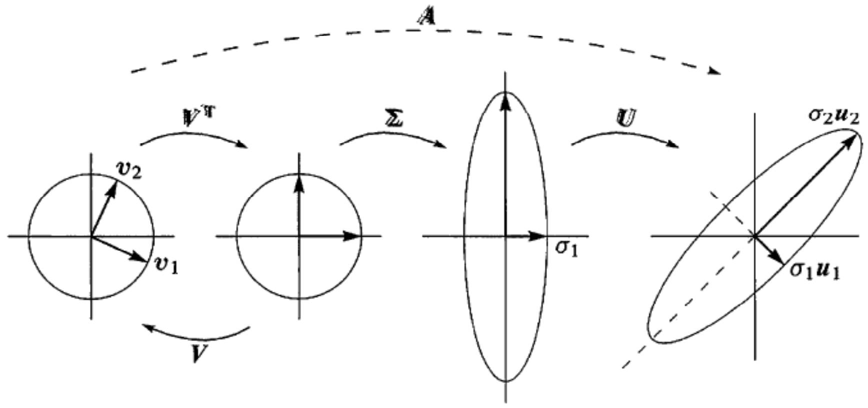
$\mathbf{v}_1$，$\mathbf{v}_2$……$\mathbf{v}_r$ 为行空间的标准正交基。 A 乘以每个 v 对应一个 u 的方向.

$\mathbf{u}_1$，$\mathbf{u}_2$……$\mathbf{u}_r$ 为列空间的标准正交基。

$\mathbf{v}_{r+1}$，$\mathbf{v}_{r+2}$……$\mathbf{v}_n$ 为零空间的标准正交基。

$\mathbf{u}_{r+1}$，$\mathbf{u}_{r+2}$……$\mathbf{u}_m$ 为左零空间的标准正交基。

奇异值分解在最小二乘法问题中有重要应用，因为在实际问题中常碰到矩阵 $\boldsymbol{A}$ 不是列满秩的状态，因此 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 不可逆，无法用之前的方法求最优解。即使是列满秩的情况当矩阵是超大型矩阵时，$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的计算量太大，用奇异值分解的办法会降低计算量。



图为 GS 给出的二阶方阵 SVD 的几何意义。