上海师范大学 Shanghai Normal University

《线性代数》

15-奇异值分解 (Singular Value Decomposition)

杨启哲

上海师范大学信机学院计算机系

2024 年 5 月 23 日

主要内容

- SVD 基础
- 用线来拟合数据
- 用 k 维子空间拟合数据
- 再看 $Ax = b$ 的近似解

SVD 基础

对图片的存储

假设一张图片被如下的矩阵表示:

$$A = \begin{bmatrix} a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \end{bmatrix}$$

直接存储需要 $6 \times 6 = 36$ 个数。注意到该矩阵的秩 $\mathrm{rank}(A) = 1$，从而:

$$A = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} a & a & b & b & c & c \end{bmatrix}$$

从而我们只需要存储 12 个数。

实对称矩阵的谱分解 (I)

注意到实对称矩阵可以对角化，即：

定理 1.

令 S 是一个实对称矩阵，$\lambda_1, \cdots, \lambda_n \in \mathbb{R}$ 是其特征值 (可能重复)。从而存在 n 个正交的向量 $v_1, \cdots, v_n \in \mathbb{R}^n$，使得：

$$S = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} v_1^\top \\ \vdots \\ v_n^\top \end{bmatrix} = \lambda_1 v_1 v_1^\top + \cdots + \lambda_n v_n v_n^\top$$

从而：

引理 2.

令 S 是一个 $n \times n$ 的实对称矩阵，S 恰好有 $\mathrm{rank}(S)$ 个非零的特征值 (计算代数重数)。从而 S 可以被 $(n+1)\mathrm{rank}(S)$ 个数字表示。

引理的证明. 注意到 S 是实对称的，从而存在正交矩阵 Q 使得:

$$Q^{\top}SQ = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

因此有:

$$\mathrm{rank}(S) = \mathrm{rank}(\Lambda) = |\{i \in [n] \mid \lambda_i \neq 0\}|$$

令 $s = \mathrm{rank}(S)$，即存在 s 个正交的向量 $v_1, \cdots, v_s \in \mathbb{R}^n$ 使得:

$$S = \lambda_1 v_1 v_1^{\top} + \cdots + \lambda_s v_s v_s^{\top}$$

即存储 S 需要 s 个特征值和 s 个特征向量，共 $(n+1)\mathrm{rank}(S)$ 个数字。 $\square$

实对称矩阵可以完成谱分解，那么对于任意的矩阵 $A \in \mathbb{R}^{m\times n}$，是否存在类似的分解呢?

答案就是奇异值分解 (Singular Value Decomposition, SVD).

定理 3.

令 A 是一个 $m \times n$ 的矩阵，$\mathrm{rank}(A) = r$，定义下列 $m \times n$ 的矩阵 Σ:

$$\Sigma = \begin{bmatrix} D & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix}, \text{其中} D = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

这里, $\sigma_1 \geqslant \cdots \geqslant \sigma_r$ 是 A 的奇异值。则存在 $m \times m$ 的正交矩阵 U 和 $n \times n$ 的正交矩阵 $V =$，使得:

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

特别的，存在 $u_1, \cdots, u_r \in \mathbb{R}^m$ 和 $v_1, \cdots, v_r \in \mathbb{R}^n$ 使得:

$$A = \sigma_1 u_1 v_1^\top + \cdots + \sigma_r u_r v_r^\top$$

$A^\top A$ 和 $AA^\top$

给定一个任意的矩阵 A，我们有：

$$(AA^\top)^\top = AA^\top,\ (A^\top A)^\top = A^\top A$$

即其都是对称矩阵，同时我们有：

引理 4.

$$\mathrm{rank}(AA^\top) = \mathrm{rank}(A^\top A) = \mathrm{rank}(A)$$

证明. 我们只需证明：

$$\mathrm{rank}(A^\top A) = \mathrm{rank}(A)$$

因为：

$$\mathrm{rank}(AA^\top) = \mathrm{rank}((A^\top)^\top A^\top) = \mathrm{rank}(A^\top) = \mathrm{rank}(A) = \mathrm{rank}(A^\top A)$$

def

$\text{rank}(A^\top A) = \text{rank}(A)$ 的证明

由 Rank-Nullity 定理，我们有：

$$\text{rank}(A^\top A) + \dim(N(A^\top A)) = n = \text{rank}(A) + \dim(N(A))$$

从而我们只需证明：

$$\dim(N(A^\top A)) = \dim(N(A))$$

- 任取 $x \in N(A)$，即 $Ax = 0$，从而有：

$$A^\top Ax = A^\top 0 = 0$$

  即$N(A) \subseteq N(A^\top A)$。
- 任取 $x \in N(A^\top A)$，即 $A^\top Ax = 0$，从而有：

$$x^\top A^\top Ax = 0 \implies \|Ax\| = 0 \implies Ax = 0$$

  即$N(A^\top A) \subseteq N(A)$。 $\square$

现在我们来关注一下 $A^\top A$ 和 $AA^\top$ 的特征值的性质。

引理 5.
1. $A^\top A$ 和 $AA^\top$ 的特征值都是非负的。
2. $A^\top A$ 和 $AA^\top$ 是半正定矩阵。

证明. 这里只证明 $A^\top A$ 的特征值是非负的。令 λ 是 $A^\top A$ 的一个特征值，v 是对应的特征向量，即：

$$A^\top Av = \lambda v$$

从而有：

$$\lambda\|v\|^2 = \lambda v^\top v = v^\top(\lambda v) = v^\top A^\top Av = (Av)^\top(Av) = \|Av\|^2 \geqslant 0$$

即：

$$\lambda \geqslant 0$$

□

引理 6.

令 $\operatorname{rank}(A) = r$，则 $A^\top A$ 和 $AA^\top$ 有相同的 r 个非零特征值 (计算重数)。

证明. 由前面的引理可知:

$$\operatorname{rank}(A^\top A) = \operatorname{rank}(A) = \operatorname{rank}(AA^\top) = r$$

令 $A^\top A$ 有如下 n 个特征值:

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$$

其对应的一组标准正交的特征向量为 $v_1, \cdots, v_n$。从而:

1. $\lambda_1, \cdots, \lambda_r > 0$。
2. $\lambda_{r+1} = \cdots = \lambda_n = 0$。
3. 对任意的 $i \in [n]$ 有 $A^\top A v_i = \lambda_i v_i$ 且 $\|v_i\| = 1$.
4. 对任意的 $1 \leqslant i < j \leqslant n$ 有 $v_i \cdot v_j = v_i^\top v_j = 0$。

$A^\top A$ 和 $AA^\top$ 的特征值 (III)

考察如下的一组向量 $\bar{v}$:

$$\bar{v} := \frac{1}{\sqrt{\lambda_1}}Av_1, \cdots, \frac{1}{\sqrt{\lambda_r}}v_r$$

我们接下来证明:

1. $\bar{v}$ 是标准正交的。
2. $\bar{v}$ 是 $AA^\top$ 的特征值 $\lambda_1, \cdots, \lambda_r$ 的特征向量。
3. $AA^\top$ 和 $A^\top A$ 的特征值的代数重数相同。

从而最终得到:

$A^\top A$ 和 $AA^\top$ 由相同的 r 个非零特征值 (计算重数)。

$\bar{v}$ 是标准正交的

Shanghai Normal University

我们先证明:

$$\frac{1}{\sqrt{\lambda_1}}Av_1, \cdots, \frac{1}{\sqrt{\lambda_r}}v_r$$

是标准正交的。

- 对于 $i \in [r]$. 有:

$$\|\frac{1}{\sqrt{\lambda_i}}Av_i\|^2 = \left(\frac{1}{\sqrt{\lambda_i}}Av_i\right)^\top \frac{1}{\sqrt{\lambda_i}}Av_i = \frac{1}{\lambda_i}v_i^\top A^\top Av_i = \frac{1}{\lambda_i}v_i^\top \lambda_i v_i = \|v_i\|^2 = 1$$

- 对于 $1 \leqslant i < j \leqslant r$. 有:

$$\left(\frac{1}{\sqrt{\lambda_i}}Av_i\right) \cdot \left(\frac{1}{\sqrt{\lambda_j}}Av_j\right) = \left(\frac{1}{\sqrt{\lambda_i}}Av_i\right)^\top \frac{1}{\sqrt{\lambda_j}}Av_j$$

$$= \frac{1}{\sqrt{\lambda_i\lambda_j}}v_i^\top A^\top Av_j = \frac{1}{\sqrt{\lambda_i\lambda_j}}v_i^\top \lambda_j v_j = \sqrt{\frac{\lambda_j}{\lambda_i}}v_i \cdot v_j = 0$$

我们再证明，对任意的 $i \in [r]$ 有：

$$AA^\top\left(\frac{1}{\sqrt{\lambda_i}}Av_i\right) = \lambda_i\left(\frac{1}{\sqrt{\lambda_i}}Av_i\right)$$

事实上，

$$\begin{aligned} AA^\top\left(\frac{1}{\sqrt{\lambda_i}}Av_i\right) &= \frac{1}{\sqrt{\lambda_i}}A(A^\top Av_i) \\ &= \frac{1}{\sqrt{\lambda_i}}A(\lambda_i v_i) \\ &= \lambda_i\left(\frac{1}{\sqrt{\lambda_i}}Av_i\right) \end{aligned}$$

从而对 $A^\top A$ 任意的非零特征值 λ_i，λ_i 也是 $AA^\top$ 的特征值，$\frac{1}{\sqrt{\lambda_i}}Av_i$ 是对应的特征向量。

现在令 $\lambda_{i_1}, \cdots, \lambda_{i_t}$ 是 $A^\top A$ 的两两不同的非零特征值，其中:

$$1 = i_1 < i_2 < \cdots < i_t \leqslant r$$

注意到其也是 $AA^\top$ 的特征值，定义:

1. a_i 是 $A^\top A$ 的特征值 λ_i 的代数重数, a_i' 是 $AA^\top$ 的特征值 λ_i 的代数重数。
2. g_i 是 $A^\top A$ 的特征值 λ_i 的几何重数, g_i' 是 $AA^\top$ 的特征值 λ_i 的几何重数。

并且我们有:

$$\sum_{i\in[t]} a_i = r, \sum_{i\in[t]} a_i' \leqslant r$$

上海师范大学 Shanghai Normal University

注意到 $v_1, \cdots, v_r$ 是一组标准正交的向量，也是 $A^\top A$ 的特征值 $\lambda_1, \cdots, \lambda_r$ 的特征向量。从而我们对任意的 $i \in [t]$ 有:

$$g_i \leqslant g_i'$$

从而由代数重数和几何重数的关系我们有:

$$a_i = g_i \leqslant g_i' \leqslant a_i' \Longrightarrow \sum_{i\in[t]} a_i' \geqslant \sum_{i\in[t]} a_i = r$$

从而 $\sum_{i\in[t]} a_i' = r$，即对所有的 $i \in [t]$:

$$a_i = a_i'$$

从而 $A^\top A$ 和 $AA^\top$ 有相同的 r 个非零特征值 (计算重数)。 $\square$

现在我们来直观的理解一下一般矩阵的分解。令 A 是一个 $m\times n$ 的矩阵，则 $A^{\top}A$ 是一个对称阵，从而存在一个 $n\times n$ 的正交矩阵 Q 使得:

$$A^{\top}A=Q\Lambda Q^{\top}$$

这里 $\Lambda=\begin{bmatrix}D & O\\ O & O\end{bmatrix}$，其中 $D=\mathrm{diag}\{\lambda_1,\cdots,\lambda_r\}$ 是一个对角阵，$r=\mathrm{rank}(A)$。另一方面，在定理3中我们有:

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

从而:

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

我们可以看到:

$$\Lambda=\Sigma^{\top}\Sigma=\begin{bmatrix}D & O_{(m-r)\times r}\\ O_{r\times(n-r)} & O_{(n-r)\times(m-r)}\end{bmatrix}\begin{bmatrix}D & O_{r\times(n-r)}\\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)}\end{bmatrix}=\begin{bmatrix}\sigma_1^2 & & \\ & \ddots & \\ & & \sigma_r^2\end{bmatrix}$$

从而我们有，对任意的 $i \in [r]$ 有:

$$\lambda_i = \sigma_i^2$$

我们称其为奇异值 (Singular Value)。

定义 7.

令 A 是一个 $m \times n$ 的矩阵，$\lambda_1, \cdots, \lambda_r$ 是对称矩阵 $A^\top A$ 的非零特征值，从而其都是非负的。对任意的 $i \in [r]$，定义:

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

我们将 $\sigma_1, \cdots, \sigma_r$ 称为矩阵 A 的奇异值 (Singular Value)。

上海师范大学 Shanghai Normal University

我们现在开始证明定理3。再次回顾一下定理的内容:

定理 3.

令 A 是一个 $m \times n$ 的矩阵，rank$(A) = r$，定义下列 $m \times n$ 的矩阵 Σ:

$$\Sigma = \begin{bmatrix} D & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix}, \text{其中} D = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

这里, $\sigma_1, \cdots, \sigma_r$ 是 A 的奇异值。则存在 $m \times m$ 的正交矩阵 U 和 $n \times n$ 的正交矩阵 $V =$, 使得:

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

特别的，存在 $u_1, \cdots, u_r \in \mathbb{R}^m$ 和 $v_1, \cdots, v_r \in \mathbb{R}^n$ 使得:

$$A = \sigma_1 u_1 v_1^{\top} + \cdots + \sigma_r u_r v_r^{\top}$$

定理3的证明 (I)

令

$$\lambda_1, \cdots, \lambda_r, \lambda_{r+1}, \cdots, \lambda_n$$

是 $A^\top A$ 的特征值，并且我们有：

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$$

并且令 $v_1, \cdots, v_n$ 是其相应的标准正交的特征向量，从而矩阵：

$$V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

是一个 $n \times n$ 的正交矩阵。

定理3的证明 (II)

现在对于任意的 $i \in [r]$，定义：

$$u_i = \frac{1}{\sqrt{\lambda_i}} A v_i \in \mathbb{R}^m$$

则由之前的引理：

$$u_1, \cdots, u_r$$

是 $AA^\top$ 的标准正交的特征向量。利用 Gram–Schmidts 正交化将其扩展成一组标准正交基：

$$u_1, \cdots, u_r, u_{r+1}, \cdots, u_m$$

从而矩阵 $U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}$ 是 $m \times m$ 的正交矩阵。

定理3的证明 (III)

现在对于任意的 $i \in [n]$，考虑 Av_i：

- 若 $i \leqslant r$。则由定义：

$$Av_i = \sqrt{\lambda_i} u_i$$

- 若 $i \geqslant r$，注意到 $(A^\top A)v_i = \lambda_i v_i = \mathbf{0}$，从而：

$$0 = v_i^\top (A^\top A) v_i = (Av_i)^\top (Av_i) = \|Av_i\|^2 \implies Av_i = \mathbf{0}$$

从而：

$$AV = \begin{bmatrix} Av_1 & \cdots & Av_n \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1} u_1 & \cdots & \sqrt{\lambda_r} u_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

$$= \begin{bmatrix} u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & & & \\ & \ddots & & & \\ & & \sqrt{\lambda_r} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} = U\Sigma$$

注意到 V 是正交矩阵，从而 $A = U\Sigma V^\top$。 □

让我们再来关注一下奇异值究竟想表示的内容，注意到：

$$A = U\Sigma V^{\top}，即：AV = U\Sigma$$

则对于 $i \in [n]$ 有：

$$Av_i = \sigma_i u_i = \sqrt{\lambda_i} u_i$$

- u_i 是 $\mathbb{R}^m$ 的一组标准正交基。
- v_i 是 $\mathbb{R}^n$ 的一组标准正交基。
- $AV = U\Sigma$ 与再方阵中 $AX = X\Lambda$ 类似，表达的是不改变方向的变换，只是对一般矩阵而言，其对应的是 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的线性变换，而不是 $\mathbb{R}^n$ 到 $\mathbb{R}^n$。

奇异值和矩阵

由定理3我们知道，对于任意的矩阵 A，我们有：

$$
A = \begin{bmatrix} u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \end{bmatrix}
\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}
\begin{bmatrix} v_1^\top \\ \vdots \\ v_r^\top \\ v_{r+1}^\top \\ \vdots \\ v_n^\top \end{bmatrix}
$$

$$
= \sigma_1 u_1 v_1^\top + \cdots + \sigma_r u_r v_r^\top
$$

推论 9.

正交向量组	作为正交基的对应的向量空间
$u_1, \cdots, u_r$	$C(A)$
$u_{r+1}, \cdots, u_m$	$N(A^\top)$
$v_1, \cdots, v_r$	$C(A^\top)$
$v_{r+1}, \cdots, v_n$	$N(A)$

推论61的证明

由对称性我们只需要证明 $C(A)$ 和 $N(A)$ 的性质。

- 注意到 $\dim(C(A)) = \text{rank}(A)$，我们只要证明：

$$C(A) \subseteq \text{span}\{u_1, \cdots, u_r\}$$

而由 $A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$ 可知，A 的每一列都是 $u_1, \cdots, u_r$ 的线性组合。

- 注意到 $\dim(N(A)) = n - \text{rank}(A) = n - r$，我们只要证明对任意的 $r + 1 \leqslant i \leqslant n$ 有：

$$A v_i = \mathbf{0}$$

而此时有：

$$\begin{aligned} A v_i &= (\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T) v_i \\ &= \sigma_1 u_1 v_1^T v_i + \sigma_2 u_2 v_2^T v_i + \cdots + \sigma_r u_r v_r^T v_i \\ &= \sigma_1 u_1 (v_1 \cdot v_i) + \sigma_2 u_2 (v_2 \cdot v_i) + \cdots + \sigma_r u_r (v_r \cdot v_i) \\ &= \mathbf{0} \end{aligned}$$

□

阶段总结

- 矩阵的低秩分解-更少的存储方式。
- 矩阵的奇异值分解 (SVD)、奇异值的概念。
- 奇异值分解的证明。

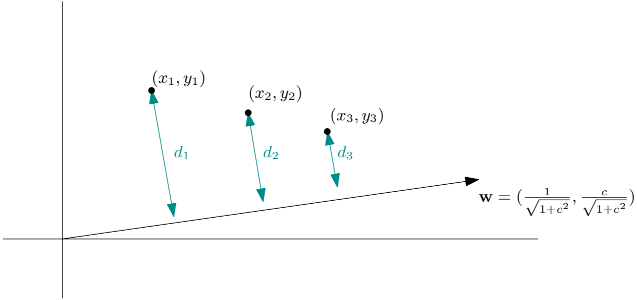
用线来拟合数据

用线来拟合数据 (I)

假设我们现在在 $\mathbb{R}^2$ 中有 m 个数据：

$$(x_1, y_1), \cdots, (x_m, y_m)$$

我们希望找到一条直线：$y = cx (c \in \mathbb{R})$ 使得每个数据点到直线的距稍尽可能的小。

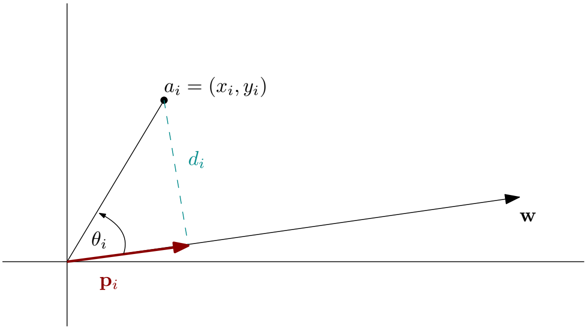


这等价于找到一个单位向量 $w = \begin{bmatrix} \frac{1}{\sqrt{1+c^2}} \\ \frac{c}{\sqrt{1+c^2}} \end{bmatrix}$ 使得：

$$d_1^2 + \cdots d_m^2 \text{ 最小}$$

用线来拟合数据 (II)

对于每个数据点 $a_i(x_i, y_i) \in \mathbb{R}^2$，我们希望寻找到一个单位向量 $w \in \mathbb{R}^2$:



最小化:

$$\sum_{i=1}^{m} d_i^2 = \sum_{i=1}^{m} (\|a_i\|^2 - p_i^2) = \sum_{i=1}^{m} \|a_i\|^2 - \sum_{i=1}^{m} p_i^2$$

即最大化:

$$\sum_{i=1}^{m} p_i^2 = \sum_{i=1}^{m} (\|a_i\| \cos\theta_i)^2 = \sum_{i=1}^{m} \left(\|a_i\| \frac{w \cdot a_i}{\|w\|\|a_i\|} \right)^2 = \sum_{i=1}^{m} (w \cdot a_i)^2$$

n 维的数据

假设现在数据有 n 个特征，即：

$$a_1, \cdots, a_m \in \mathbb{R}^n$$

我们依旧是希望找到一个单位向量 $w \in \mathbb{R}^n$ 最小化：

$$\sum_{i=1}^{m} w \text{ 和 } a_i \text{ 的距离}$$

即最大化：

$$\sum_{i=1}^{m} (w \cdot a_i)^2$$

定义 $m \times n$ 的矩阵:

$$A = \begin{bmatrix} a_1^\top \\ \vdots \\ a_m^\top \end{bmatrix}$$

则我们目标是找到一个单位向量 $w \in \mathbb{R}^n$ 使得:

$$\sum_{i=1}^{m} (w \cdot a_i)^2 = \sum_{i=1}^{m} \left(a_i^\top w\right)^2 = \|Aw\|^2 \text{ 最大}$$

由矩阵的奇异值分解可知，事实上这就是矩阵 A 最大的奇异值的平方。

定理 10.

令 A 是 $m \times n$ 的矩阵，σ_1 是其最大的奇异值，则：

$$\max_{w \in \mathbb{R}^n,\ \|w\|=1} \|Aw\|^2 = \sigma_1^2$$

特比的，该值取到最大的时候恰好是 A 的 SVD 分解中 V 对应 σ_1 的向量，即：

$$A = U\Sigma V^\top = U\Sigma \begin{bmatrix} v_1^\top \\ v_2^\top \\ \vdots \\ v_n^\top \end{bmatrix}$$

在证明这个定理之前，我们先讨论回顾一下基于标准正交基表示的向量的长度。

标准正交基和坐标

考虑 $\mathbb{R}^n$ 上的一组标准正交基 $v_1, \cdots, v_n$。则任何一个 $w \in \mathbb{R}^n$ 都可以被其线性表示:

$$w = c_1 v_1 + \cdots + c_n v_n$$

即 w 关于基 $v_1, \cdots, v_n$ 的坐标为 $(c_1, \cdots, c_n) \in \mathbb{R}^n$。注意到对所有的 $i \in [n]$:

$$v_i \cdot w = v_i \cdot (c_1 v_1 + \cdots + c_n v_n) = c_i (v_i \cdot v_i) = c_i$$

从而我们有:

$$\begin{aligned} w &= (v_1 \cdot w) v_1 + \cdots + (v_n \cdot w) v_n = v_1 v_1^\top w + \cdots + v_n v_n^\top w \\ &= \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} v_1^\top w \\ \vdots \\ v_n^\top w \end{bmatrix} = {\color{red}\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}} \begin{bmatrix} v_1^\top \\ \vdots \\ v_n^\top \end{bmatrix} w = {\color{red}Q} Q^\top w \end{aligned}$$

这其实就是:

$$Qx = b \text{ 的最小二乘解就是 } \hat{x} = Q^\top b$$

回顾：使用正交矩阵投影

令 $q_1, \ldots, q_n \in \mathbb{R}^n$ 是标准正交的，并且：

$$Q = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}$$

显然 Q 是一个正交矩阵。从而 $Qx = b$ 的最小二乘解为：$\hat{x} = Q^\top b$，对应的投影矩阵为 $QQ^\top = I$，从而：

$$b = QQ^\top b = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix} \begin{bmatrix} q_1^\top b \\ \vdots \\ q_n^\top b \end{bmatrix} = q_1 q_1^\top b + \cdots + q_n q_n^\top b$$

也就是：

p 是 b 分别到每条线 $\mathrm{span}(\{q_i\})$ 上的投影的和。

考虑向量空间 $\mathbb{V} = \mathbb{R}^n$ 和其两组基:

$$\bar{e} = e_1, \cdots, e_n \text{ 和 } \bar{v} = v_1, \cdots, v_n$$

从而 $\bar{e}$ 到 $\bar{v}$ 的基变换矩阵 M 满足:

$$\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} e_1 & \cdots & e_n \end{bmatrix} M$$

并且可以得出:

$$M = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

即：对于任意的 $w = (w_1, \cdots, w_n) \in \mathbb{V}$，其在基 $\bar{e}$ 下的坐标就是其自己 w。

回顾坐标变换公式:

定理 11.

令 $\overline{\mathrm{v}}\mathrm{n}$ 和 $\overline{\mathrm{v}'}$ 是 $\mathbb{V}$ 的两组基，则对于任意的 $\mathrm{u} \in \mathbb{V}$，我们有:

$$T_{\overline{v}}(u) = MT_{\overline{v'}}(u)$$

从而 w 在基 $\overline{\mathrm{v}}$ 下的坐标为:

$$M^{-1}w$$

如果基 $\overline{\mathrm{v}}$ 是标准正交的，即:

$$M = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \text{是正交矩阵}$$

则我们有:

$$M^{-1}w = M^{\top}w, \quad \text{即:}\ w = MM^{\top}w = v_1v_1^{\top}w + \cdots + v_nv_n^{\top}w$$

关于 w 的长度

注意到:

定理 12.
令 Q 是 $n \times n$ 的正交矩阵，并且 $x, y \in \mathbb{R}^n$，则:

1. $\|Qx\| = \|x\|$，特别的 $\|Q^\top x\| = \|x\|$。
2. $Qx \cdot Qy = x \cdot y$，从而 $Q^\top x \cdot Q^\top y = x \cdot y$。

从而考虑任意的 $w \in \mathbb{R}^n$，我们有:

1. 其在标准正交基 $v_1, \cdots, v_n$ 下的坐标为 $Q^\top w$，这里 $Q = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$，即:

$$(v_1^\top w, \cdots, v_n^\top w)$$

2. 对应坐标的长度和 w 的长度是相同的，即:

$$\|w\| = \|Q^\top w\| = \sqrt{(v_1^\top w)^2 + \cdots + (v_n^\top w)^2}$$

现在我们来证明定理10。回顾 A 的 SVD 分解，令 $A^\top A$ 的正特征值为 $\lambda_1 > \cdots > \lambda_r$。对任意的 $i \in [r]$，我们定义 $\sigma_i = \sqrt{\lambda_i}$。则 $\sigma_1, \cdots, \sigma_r$ 是 A 的奇异值。并且存在标准正交基 $u_1, \cdots, u_m \in \mathbb{R}^m$ 和 $v_1, \cdots, v_n \in \mathbb{R}^n$:

$$A = \begin{bmatrix} u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} v_1^\top \\ \vdots \\ v_r^\top \\ v_{r+1}^\top \\ \vdots \\ v_n^\top \end{bmatrix}$$

$$= \sigma_1 u_1 v_1^\top + \cdots + \sigma_r u_r v_r^\top$$

考虑 $\mathbb{R}^n$ 的一个单位向量 w，存在 $c_1, \cdots, c_n$ 使得 $w = c_1 v_1 + \cdots + c_n v_n$，并且由前面的讨论：

$$c_1^2 + \cdots + c_n^2 = \|w\|^2 = 1$$

从而：

$$\begin{aligned} Aw &= (\sigma_1 u_1 v_1^\top + \cdots + \sigma_r u_r v_r^\top)(c_1 v_1 + \cdots + c_n v_n) \\ &= \sum_{i\in[r], j\in[n]} \sigma_i c_j u_i v_j^\top v_j = \sum_{i\in[r]} \sigma_i c_i u_i \end{aligned}$$

从而 Aw 在标准正交基 $u_1, \cdots, u_r, u_{r+1}, \cdots, u_m$ 下的坐标为 $(\sigma_1 c_1, \cdots, \sigma_r c_r, 0, \cdots, 0)$，因此我们有：

$$\|Aw\| = \sigma_1^2 c_1^2 + \cdots + \sigma_r^2 c_r^2 \leqslant \sigma_1^2 \sum_{i\in[r]} c_i^2 \leqslant \sigma_1^2 \sum_{i\in[n]} c_i^2 = \sigma_1^2$$

当 $w = v_1$ 时取到上述不等式等号。 $\square$

用 k 维子空间拟合数据

固定一个 $k \geqslant 1$。

$$a_1, \cdots, a_m \in \mathbb{R}^n$$

我们希望找到一个 k 维的子空间 $\mathbb{W} \subseteq \mathbb{R}^n$ 最小化:

$$\sum_{i \in [m]} (a_i \text{ 到 } \mathbb{W} \text{ 的距离})^2$$

我们可以选取 $\mathbb{W}$ 中的一组标准正交基 $w_1, \cdots, w_k \in \mathbb{W}$，从而 a_i 到 $\mathbb{W}$ 的距离是:

a_i 和 a_i 到 $\mathbb{W}$ 的投影 p_i 的距离。

回顾：投影的做法

1. 我们的目标是计算 b 到下列空间：

$$V = \text{span}(\{a_1, \ldots, a_n\})$$

的投影 p，其中 $a_1, \ldots, a_n$ 是线性无关的，$p \in V$.

2. 我们令 $p \in V$ 是满足其误差 $e = b - p$ 与 V 垂直的向量。我们证明了，对于任意的 $v \in V$:

$$\|b - v\| = \min_{u \in V} \|b - u\| \quad \Longleftrightarrow \quad v = p$$

3. 我们得到了相应的投影矩阵 $P = A(A^\top A)^{-1}A^\top$，即：

$$p = A\hat{x} = A(A^\top A)^{-1}A^\top b$$

并且我们证明了当 $\text{rank}(A) = n$ 时 $(A^\top A)^{-1}$ 是存在的，这也说明了 p 的唯一性。

我们利用如下的标准正交的向量来表示 $\mathbb{W}$:

$$w_1, \cdots, w_k \in \mathbb{R}^n$$

特别的，令

$$Q = \begin{bmatrix} w_1 & \cdots & w_k \end{bmatrix}$$

则:

$$Q^\top Q = \begin{bmatrix} w_1^\top \\ \vdots \\ w_k^\top \end{bmatrix} \begin{bmatrix} w_1 & \cdots & w_k \end{bmatrix} = \begin{bmatrix} w_1^\top w_1 & \cdots & w_1^\top w_k \\ \vdots & \ddots & \vdots \\ w_k^\top w_1 & \cdots & w_k^\top w_k \end{bmatrix} = I_{k\times k}$$

$$QQ^\top = \begin{bmatrix} w_1 & \cdots & w_k \end{bmatrix} \begin{bmatrix} w_1^\top \\ \vdots \\ w_k^\top \end{bmatrix} = w_1 w_1^\top + \cdots + w_k w_k^\top$$

则 a_i 到 $\mathbb{W}(=\text{span}(\{w_1,\cdots,w_k\}))=C(Q)$ 的投影为：

$$
\begin{aligned}
p_i &= Q(Q^\top Q)Q^\top a_i \\
&= QQ^\top a_i \\
&= w_1 w_1^\top a_i + \cdots + w_k w_k^\top a_i \\
&= (w_1 \cdot a_i)w_1 + \cdots + (w_k \cdot a_i)w_k \\
&= \sum_{j\in[k]} (w_j \cdot a_i)w_j
\end{aligned}
$$

即：

a_i 到 $\mathbb{W}$ 的投影 $= a_i$ 到直线 $\text{span}(\{w_1\}),\cdots,\text{span}(\{w_k\})$ 的投影之和

最小化距离之和

另一方面，注意到 a_i 到 W 的距离可以表示为：

$$d_i^2 = \|a_i - p_i\|^2 = \|a_i\|^2 - \|p_i\|^2 = \|a_i\|^2 - \sum_{j\in[k]} (w_j \cdot a_i) w_j$$

从而最小化 $d_1^2 + \cdots + d_m^2$ 等价于最大化：

$$\sum_{i\in[m]} \|p_i\|^2 = \sum_{i\in[m]} \left\| \sum_{j\in[k]} (w_j \cdot a_i) w_j \right\|^2 = \sum_{i\in[m]} \sum_{i\in[k]} (w_j \cdot a_i)^2$$

注意到之前用线去拟合的情况就是该例中 $k = 1,\ w_1 = w$ 的特殊情况。

我们依旧用一个矩阵来表示这 m 个数据:

$$A = \begin{bmatrix} a_1^\top \\ \vdots \\ a_m^\top \end{bmatrix} \in \mathbb{R}^{m\times n}$$

从而我们的目标是寻找到一组标准正交的向量 $w_1, \cdots, w_k \in \mathbb{R}^n$ 最大化:

$$\sum_{i\in[m]} \sum_{j\in[k]} (w_j \cdot a_i)^2 = \sum_{j\in[k]} \sum_{i\in[k]} (w_j \cdot a_i)^2 = \sum_{j\in[k]} \|Aw_j\|^2$$

利用 SVD 最大化

注意到:

$$A = \begin{bmatrix} u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} v_1^\top \\ \vdots \\ v_r^\top \\ v_{r+1}^\top \\ \vdots \\ v_n^\top \end{bmatrix}$$

$$= \sigma_1 u_1 v_1^\top + \cdots + \sigma_r u_r v_r^\top$$

定理 13.

假设 $k \leqslant r$，则对任意的标准正交的 $w_1, \cdots, w_k \in \mathbb{R}^n$ 有:

$$\sum_{j\in[k]} \|Aw_j\|^2 \leqslant \sum_{j\in[k]} \sigma_j^2$$

当 $w_1 = v_1, \cdots, w_k = v_k$ 时等号成立。

定理13的证明思路

由定理10，我们知道存在 $w_1 \in \mathbb{R}^n$，使得对于任意的 $w \in \mathbb{R}^n$：

$$\|Aw\|^2 \leqslant \|Aw_1\|^2 = \sigma_1^2$$

从而 w_2 的选取应当满足 $w_2 \perp w_1$，之后的情况可以类似归纳。为此我们先证明两个引理。

引理 14.

令 $k \in [r]$，并且 $w \in \mathbb{R}^n$ 是一个单位向量，并且满足对任意的 $i \in [k-1]$ 有 $w \perp v_i$，则我们有：

$$\|Aw\|^2 \leqslant \|Av_k\|^2 = \sigma_k^2$$

引理 15.

令 $2 \leqslant k \leqslant n$，$\mathbb{W} \subseteq \mathbb{R}^n$ 是一个 k 维的子空间，则存在一个单位向量 $w \in \mathbb{R}^n$，使得：对任意的 $i \in [k-1]$ 有 $w \perp v_i$

引理 14.

令 $k \in [r]$，并且 $w \in \mathbb{R}^n$ 是一个单位向量，并且满足对任意的 $i \in [k-1]$ 有 $w \perp v_i$，则我们有：

$$\|Aw\|^2 \leqslant \|Av_k\|^2 = \sigma_k^2$$

证明. 令 $w = c_1 v_1 + \cdots + c_n v_n$，则由题目假设我们有：

1. $c_1 = c_2 = \cdots = c_{k-1} = 0$
2. $c_k^2 + \cdots + c_n^2 = 1$

从而：

$$Aw = (\sigma_1 u_1 v_1^\top + \cdots + \sigma_r u_r v_r^\top)(c_1 v_1 + \cdots + c_n v_n) = \sigma_k c_k u_k + \cdots + \sigma_r c_r u_r$$

因此：

$$\|Aw\|^2 = \|(0, \cdots, \sigma_k c_k, \cdots, \sigma_r c_r, 0, \cdots, 0)\| = \sum_{k \leqslant j \leqslant r} \sigma_j^2 c_j^2 \leqslant \sigma_k^2 \sum_{k \leqslant j \leqslant r} c_j^2 \leqslant \sigma_k^2 \sum_{j \in [n]} c_j^2 = \sigma_k^2$$

$\square$

引理 16.

令 $2 \leqslant k \leqslant n$，$\mathbb{W} \subseteq \mathbb{R}^n$ 是一个 k 维的子空间，则存在一个单位向量 $w \in \mathbb{R}^n$，使得：对任意的 $i \in [k-1]$ 有 $w \perp v_i$

证明. 对任意的 $i \in [k-1]$，定义 v_i' 是 v_i 到 $\mathbb{W}$ 的投影，显然我们有 $(v_i - v_i') \perp \mathbb{W}$。注意到：

$$\mathbb{W}_0 = \mathrm{span}(\{v_1', \cdots, v_{k-1}'\}) \subseteq \mathbb{W}(\text{由于 } \dim(\mathbb{W}_0) \leqslant k-1)$$

令 $\dim(\mathbb{W}_0) = l$，则我们可以构造出一组 $\mathbb{W}$ 的标准正交基：

$$w_1, \cdots, w_l, \cdots, w_{k-1}, w_k$$

使得 $w_1, \cdots, w_l$ 是 $\mathbb{W}$ 中的一组标准正交基，定义 $w = w_k$，则：

1. $\|w\| = \|w_k\| = 1$。
2. $w \perp \mathbb{W}_0$，从而对任意的 $i \in [k-1]$ 有：

$$w \cdot v = w^\top v_i = w^\top v_i' + w^\top (v_i - v_i') = 0 + 0 = 0$$

□

定理 13.

假设 $k \leqslant r$，则对任意的标准正交的 $w_1, \cdots, w_k \in \mathbb{R}^n$ 有：

$$\sum_{j \in [k]} \|Aw_j\|^2 \leqslant \sum_{j \in [k]} \sigma_j^2$$

当 $w_1 = v_1, \cdots, w_k = v_k$ 时等号成立。

证明. 定义：

$$\mathbb{V}_k = \mathrm{span}(\{v_1, \cdots, v_k\})$$

从而：

$$\begin{aligned}
\sum_{i \in [m]} (a_i \text{ 到 } \mathbb{V}_k \text{ 的投影})^2 &= \sum_{j \in [k]} \|Av_j\|^2 \\
&= \sum_{j \in [k]} \left\| (\sigma_1 u_1 v_1^\top + \cdots + \sigma_r u_r v_r^\top) v_j \right\|^2 \\
&= \sum_{j \in [k]} \|\sigma_j u_j\|^2 = \sum_{j \in [k]} \sigma_j^2
\end{aligned}$$

上海师范大学 Shanghai Normal University

下面我们对 k 作归纳，证明对任意的 k 维子空间 $\mathbb{W}$, 我们有:

$$\sum_{i\in[m]}(a_i \text{ 到 } \mathbb{W} \text{ 的投影})^2 \leqslant \sum_{i\in[m]}(a_i \text{ 到 } \mathbb{V}_k \text{ 的投影})^2 = \sum_{j\in[k]}\sigma_j^2$$

即 $\mathbb{V}_k$ 是最佳的 k 维子空间。

$k=1$ 的时候就是定理 10，即对任意的 $w\in\mathbb{R}^n$ 且 $\|w\|=1$，我们有:

$$\|Aw^2\|^2 \leqslant \|Av_1\|^2 = \sigma_1^2$$

当 $k \geqslant 2$ 的时候，令 $\mathbb{W} \subseteq \mathbb{R}^n$，且 $\dim(\mathbb{W}) = k$。由引理15存在一个单位向量 $w \in \mathbb{W}$，使得对任意的 $i \in [k-1]$ 有 $w \perp v_i$。利用 Gram–Schmidts 正交化，我们可以获得 $\mathbb{W}$ 的一组标准正交基：

$$w_1, \cdots, w_{k-1}, w_k = w$$

从而由归纳假设我们有：

$$\sum_{j \in [k-1]} \|Aw_j\|^2 \leqslant \sum_{j \in [k-1]} \sigma_j^2$$

利用引理14我们有：$\|Aw\|^2 \leqslant \sigma_k^2$，从而：

$$\sum_{j \in [k]} \|Aw_j\|^2 = \|Aw\|^2 + \sum_{j \in [k-1]} \|Aw_j\|^2 \leqslant \sigma_k^2 + \sum_{j \in [k-1]} \sigma_j^2 = \sum_{j \in [k]} \sigma_j^2$$

$\square$

再看 $A\mathrm{x} = \mathrm{b}$ 的近似解

$Ax = b$ 的近似解

令 A 是 $m \times n$ 的矩阵，$b \in \mathbb{R}^n$，考虑方程组：

$$Ax = b$$

可能有无数解，也可能没有任何解。

- $Ax = b$ 当且仅当：

$$\mathrm{rank}(A) = \mathrm{rank}(\begin{bmatrix} A & b \end{bmatrix})$$

- 其最小二乘解 $\hat{x}$ 满足：

$$\|b - A\hat{x}\| = \min_{x \in \mathbb{R}^n} \|b - Ax\|$$

最小二乘解-rank$(A) = n$

1. 我们知道如果 $\hat{x} \in \mathbb{R}$ 满足:

$$A^T(A\hat{x} - b) = \mathbf{0} \quad \text{即} \quad A^TA\hat{x} = A^Tb$$

   则 $A\hat{x} - b$ 与 $C(A)$ 正交。

2. 我们可以得到 $\hat{x}$ 的表达式:

$$\hat{x} = (A^TA)^{-1}A^Tb$$

   并且我们知道 $\hat{x}$ 是唯一的。

3. 我们称 $\hat{x}$ 就是最小二乘解（least square solution），因为其误差的长度 $\|e\|$

$$e = b - A\hat{x}$$

   是所有 $b - Ax$ 中最小的。

1. 选择 $C(A)$ 的一组基 $\{a_{i_1}, \cdots, a_{i_r}\}$，其中 $r = \mathrm{rank}(A)$。
2. 定义 $m \times r$ 的矩阵：
$$A' = \begin{bmatrix} a_{i_1} & \cdots & a_{i_r} \end{bmatrix}$$
显然我们有：$C(A) = C(A')$
3. $\mathrm{rank}(A')$ 是列满秩的，所以我们可以利用前面的方法来找到 $A'x' = b$ 的最优近似解，即：
$$\hat{x}' = (A'^T A')^{-1} A'^T b \in \mathbb{R}^n$$
显然其误差 $e = b - A'\hat{x}'$ 是所有 $b - Ax$ 中长度最小的，即：
$$\|e\| = \min_{x \in \mathbb{R}^n} \|b - Ax\| = \min\{\|b - v\| \mid v \in C(A')\} == \min\{\|b - v\| \mid v \in C(A)\}$$
4. 我们需要的 $\hat{x} \in \mathbb{R}^n$ 只要满足：
$$A\hat{x} = A'\hat{x}'$$

注意此时 $\hat{x}$ 并不是唯一的。

现在我们利用 SVD 来获取最小二乘解。令 $m \times n$ 的矩阵 A 的 SVD 分解为:

$$A = U\Sigma V^{\top} = U \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} V^{\top}$$

定义 19 **[A 的广义逆 (Pseudoinverse)].**

定义 A 的广义逆 A^+ 为:

$$A^+ = V\Sigma^+ U^{\top} = V \begin{bmatrix} \sigma_1^{-1} & & & & & \\ & \ddots & & & & \\ & & \sigma_r^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} U^{\top}$$

这里 Σ^+ 是一个 $n \times m$ 的矩阵。

定义:

$$x^+ = A^+ b$$

我们说明 x^+ 就是 $Ax = b$ 的最小二乘解。

定理 20.

x^+ 是 $Ax = b$ 的最小二乘解，即:

$$\|b - Ax^+\| = \min_{x \in \mathbb{R}^n} \|b - Ax\|$$

特别的，对于该方程的每个最小二乘解 $x \neq x^+$，即:

$$\|b - Ax^+\| = \|b - Ax\|$$

我们有:

$$\|x^+\| < \|x\|$$

定理20的证明 (I)

注意到:

$$A^+ = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \sigma_1^{-1} & & & & & \\ & \ddots & & & & \\ & & \sigma_r^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} u_1^\top \\ \vdots \\ u_m^\top \end{bmatrix} = \sigma_1^{-1} v_1 u_1^\top + \cdots + \sigma_r^{-1} v_r u_r^\top$$

从而:

$$x^+ = A^+ b = \sigma_1^{-1} v_1 u_1^\top b + \cdots + \sigma_r^{-1} v_r u_r^\top b$$

同时注意到:

$$A = \sigma u_1 v_1^\top + \cdots + \sigma_r u_r v_r^\top$$

因此:

$$\begin{aligned} Ax^+ &= (\sigma u_1 v_1^\top + \cdots + \sigma_r u_r v_r^\top)(\sigma_1^{-1} v_1 u_1^\top b + \cdots + \sigma_r^{-1} v_r u_r^\top b) \\ &= u_1 u_1^\top b + \cdots + u_r u_r^\top b \\ &= (u_1 \cdot b) u_1 + \cdots + (u_r \cdot b) u_r \end{aligned}$$

注意到 $u_1, \cdots, u_r$ 是标准正交的，从而：

$$Ax^+ = b \text{ 到 } \mathrm{span}(\{u_1, \cdots, u_r\}) \text{ 的投影}$$

由推论61：

正交向量组	作为正交基的对应的向量空间
$u_1, \cdots, u_r$	$C(A)$
$u_{r+1}, \cdots, u_m$	$N(A^\top)$
$v_1, \cdots, v_r$	$C(A^\top)$
$v_{r+1}, \cdots, v_n$	$N(A)$

Ax^+ 是 b 到 $C(A)$ 上的投影，即 x^+ 是 $Ax = b$ 的最小二乘解。

定理20的证明 (III)

现在考虑 $Ax = b$ 的另一个最小二乘解 $x(\neq x^+)$，即：

$$\|b - Ax\| = \|b - Ax^+\|$$

我们有：

$$Ax = Ax^+ \Longrightarrow A(x - x^+) = \mathbf{0},\ \text{即}\ x - x^+ \in N(A)$$

从而

$$x^+ = \sigma_1^{-1} v_1 u_1^T b + \cdots + \sigma_r^{-1} v_r u_r^T b \in \mathrm{span}(\{v_1, \cdots, v_r\}) = C(A^T)$$

$$x - x^+ \in N(A) = \mathrm{span}(\{v_{r+1}, \cdots, v_n\})$$

从而 $x^+ \perp (x - x^+)$，并且：

$$\|x\|^2 = \|x^+\|^2 + \|x - x^+\|^2 > \|x^+\|^2$$

最后一个严格大于号是由于 $x - x^+ \neq \mathbf{0}$. $\square$

矩阵 A 的奇异值分解并不是唯一的，即：

$$A = U_1\Sigma_1V_1^T, \qquad A = U_2\Sigma_2V_2^T$$

则从定义上来讲，其可能存在两个广义逆：

$$A_1^+ = V_1\Sigma_1^+U_1^+, \qquad A_2^+ = V_2\Sigma_2^+U_2^+$$

但我们可以证明，这两个是相等的，即 A 的广义逆是唯一的。

定理 21.

$A_1^+ = A_2^+$

由定理20，对于该方程的每个最小二乘解 $x \neq x^+$，即：

$$\|b - Ax^+\| = \|b - Ax\|$$

我们有：

$$\|x^+\| < \|x\|$$

从而对任意的 $b \in \mathbb{R}^m$，我们有：

$$A_1^+ b = A_2^+ b$$

考察所有的 $b = e_i$，则有：

$$A_1^+ = A_2^+$$

$\square$

阶段总结

我们介绍了 SVD 的一些应用，以及其几何意义。

- 利用 k 维子空间拟合数据-SVD 的最大 k 个奇异值是最好的结果。
- 矩阵的广义逆，最小的二乘解。