高等代数

一、正交投影(Orthogonal Projection)

1.正交分解定理(Decomposition)

若W是R^n的子空间，那么R^n中每一个向量y都能表示为$\boldsymbol{y} = proj_W\boldsymbol{y} + \boldsymbol{z}$。

其中，$proj_W\boldsymbol{y}$属于W，z属于$W^{\perp}$。

如果$\{u_1, u_2...u_k\}$是W的任意正交基，那么：

$$proj_W\boldsymbol{y} = \frac{y \cdot u_1}{u_1 \cdot u_1}u_1 + ... + \frac{y \cdot u_k}{u_k \cdot u_k}u_k$$

2.正交投影性质

若$y \in W = Span\{u_1, u_2...u_k\}$，那么$proj_W\boldsymbol{y} = y$。由此得到最佳逼近定理：

3.最佳逼近定理(The Best Approximation)

若W是R^n的子空间，那么对于R^n中任一向量y，$proj_W\boldsymbol{y}$是W中最接近y的点。

4.单位正交基情况下的简化

若有单位正交基$\{u_1, u_2...u_k\}$，则：

$$proj_W\boldsymbol{y} = (y \cdot u_1)u_1 + ... + (y \cdot u_k)u_k$$

若记$U = [u_1, u_2...u_k]$，则：

$$proj_W\boldsymbol{y} = UU^Ty$$

二、施密特正交化(Gram-Schmidt Process)

1.正交化方法

对R^n中一个子空间的基$\{x_1, x_2...x_k\}$，按以下序列给出正交基：

$$\begin{cases} v_1 = x_1 \\ v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 \\ v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2 \\ ... \\ v_k = x_k - \frac{x_k \cdot v_1}{v_1 \cdot v_1} v_1 - ... - \frac{x_k \cdot v_{k-1}}{v_{k-1} \cdot v_{k-1}} v_{k-1} \end{cases}$$

上述向量单位化后，即可形成单位正交基/标准正交基(orthonormal basis)。

2.QR分解(QR Factorization)

若矩阵$A_{m\times n}$之列向量线性无关，那么$A = QR$，其中$Q_{m\times n}$的列是$Col\,A$的标准正交基，$R_{n\times n}$是一个上三角可逆的对角线元素为正的矩阵。

显然由于Q是标准正交矩阵，那么$Q^T = Q$,$R = Q^T A$

三、最小二乘问题(Least Squares)

1.定义

若$A_{m\times n}$，$\boldsymbol{b} \in R^m$，那么$A\boldsymbol{x} = \boldsymbol{b}$的最小二乘解是$R^n$中的$\hat{x}$，使得：

$$||b - A\hat{x}|| \le ||b - Ax||$$

总成立。

2.一般最小二乘问题的解

由最佳逼近定理，上述问题的解集与：

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b}$$

的解集一致。

例如：

例 1 求不相容方程 $A\boldsymbol{x}=\boldsymbol{b}$ 的最小二乘解，其中

$$A=\begin{bmatrix}4&0\\0&2\\1&1\end{bmatrix},\qquad \boldsymbol{b}=\begin{bmatrix}2\\0\\11\end{bmatrix}$$

解 利用（3）计算

$$A^{\mathrm{T}}A=\begin{bmatrix}4&0&1\\0&2&1\end{bmatrix}\begin{bmatrix}4&0\\0&2\\1&1\end{bmatrix}=\begin{bmatrix}17&1\\1&5\end{bmatrix}$$

$$A^{\mathrm{T}}\boldsymbol{b}=\begin{bmatrix}4&0&1\\0&2&1\end{bmatrix}\begin{bmatrix}2\\0\\11\end{bmatrix}=\begin{bmatrix}19\\11\end{bmatrix}$$

那么方程 $A^{\mathrm{T}}A\boldsymbol{x}=A^{\mathrm{T}}\boldsymbol{b}$ 变成

$$\begin{bmatrix}17&1\\1&5\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}19\\11\end{bmatrix}$$

行变换可用于解此方程组，但由于 $A^{\mathrm{T}}A$ 是 2×2 可逆矩阵，很快计算得到

$$(A^{\mathrm{T}}A)^{-1}=\frac{1}{84}\begin{bmatrix}5&-1\\-1&17\end{bmatrix}$$

那么可解 $A^{\mathrm{T}}A\boldsymbol{x}=A^{\mathrm{T}}\boldsymbol{b}$ 如下：

$$\begin{aligned}\hat{\boldsymbol{x}}&=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}\\&=\frac{1}{84}\begin{bmatrix}5&-1\\-1&17\end{bmatrix}\begin{bmatrix}19\\11\end{bmatrix}=\frac{1}{84}\begin{bmatrix}84\\168\end{bmatrix}=\begin{bmatrix}1\\2\end{bmatrix}\end{aligned}$$

3.最小二乘误差(error)

$\boldsymbol{b}$到$A\hat{x}$的距离，即$||\boldsymbol{b}-A\hat{x}||$为误差。

4.QR分解

若$A_{m\times n}=QR$具有线性无关的列，那么$A\boldsymbol{x}=\boldsymbol{b}$的唯一最小二乘解是：

$$\hat{\boldsymbol{x}}=R^{-1}Q^T\boldsymbol{b}$$

四、对称矩阵的对角化(Diagonalization of Symmetric Matrices)

对称矩阵是$A^T=A$的方阵。

1.对角化过程

例 2 如果可能，对角化矩阵 $A=\begin{bmatrix}6 & -2 & -1\\ -2 & 6 & -1\\ -1 & -1 & 5\end{bmatrix}$.

解 A 的特征方程是

$$0=-\lambda^3+17\lambda^2-90\lambda+144=-(\lambda-8)(\lambda-6)(\lambda-3)$$

通过标准计算可得到每个特征子空间的一个基：

$$\lambda=8:\boldsymbol{v}_1=\begin{bmatrix}-1\\ 1\\ 0\end{bmatrix};\quad \lambda=6:\boldsymbol{v}_2=\begin{bmatrix}-1\\ -1\\ 2\end{bmatrix};\quad \lambda=3:\boldsymbol{v}_3=\begin{bmatrix}1\\ 1\\ 1\end{bmatrix}$$

这三个向量形成 $\mathbb{R}^3$ 的一个基，事实上，很容易验证 $\{\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3\}$ 是 $\mathbb{R}^3$ 的正交基. 根据第 6 章的经验，正交基对计算可能有用，因此可以单位化 $\boldsymbol{v}_1,\boldsymbol{v}_2$和$\boldsymbol{v}_3$ 得到单位特征向量.

$$\boldsymbol{u}_1=\begin{bmatrix}-1/\sqrt{2}\\ 1/\sqrt{2}\\ 0\end{bmatrix},\quad \boldsymbol{u}_2=\begin{bmatrix}-1/\sqrt{6}\\ -1/\sqrt{6}\\ 2/\sqrt{6}\end{bmatrix},\quad \boldsymbol{u}_3=\begin{bmatrix}1/\sqrt{3}\\ 1/\sqrt{3}\\ 1/\sqrt{3}\end{bmatrix}$$

令

$$P=\begin{bmatrix}-1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3}\\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3}\\ 0 & 2/\sqrt{6} & 1/\sqrt{3}\end{bmatrix},\quad D=\begin{bmatrix}8 & 0 & 0\\ 0 & 6 & 0\\ 0 & 0 & 3\end{bmatrix}$$

那么有 $A=PDP^{-1}$，和平常一样. 由于 P 是方阵且有正交列，所以，P 是一个正交矩阵，而 P^{-1} 就是 P^{T}.（见 6.2 节.） ■

$A=PDP^{-1}=PDP^T$

2.谱分解(Spectral Decompositon)

假设$A=PDP^T$，那么：

$$A = PDP^T = [u_1, ..., u_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix} = [\lambda_1 u_1, ..., \lambda_n u_n] \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix}$$
$$= \lambda_1 u_1 u_1^T + ... + \lambda_n u_n u_n^T$$

五、二次型(Quadratic Form)

1.变量代换

若$\boldsymbol{x}$是R^n中的变量向量，那么$\boldsymbol{x} = P\boldsymbol{y}$，$\boldsymbol{y}$是$R^n$中的新向量。因此，二次型可表示为:

$\boldsymbol{x}^T A\boldsymbol{x} = \boldsymbol{y}^T(P^T AP)\boldsymbol{y}$

$P^T AP$是新的二次型矩阵。

2.二次型的分类

如果一个二次型的所有特征值都是正的，那么它是正定的(positive definite)。

都是负数，它是负定的(negtive definite)。

否则，是不定的(indefinite)。

六、条件优化(Constrained Optimization)

定义如下代数:

$$m = min\{x^T Ax,\ ||x|| = 1\}, \quad M = max\{x^T Ax,\ ||x|| = 1\}$$

1.

设A是对称矩阵，那么m, M分别是最小和最大的特征值，并且在分别对应的特征向量上二次型$x^T Ax$取得最小最大值m, M。

2.

若在$x^Tx=1,x^Tu_1=0$的限制下，x^TAx的最大值是第二大的特征值，并在相应的特征向量下取得此值。这个结论可递推。

七、奇异值分解(Singular Value Decomposition)

1.奇异值

A的奇异值是A^TA的特征值的平方根。一般记为$\sigma_1,...\sigma_n$并且按递减顺序排列。

如果$\{\lambda_1,...,\lambda_n\}$是递减的特征值，
$\{\boldsymbol{v}_1,...,\boldsymbol{v}_n\}$是相应的标准单位正交基。若$A$有$r$个非零奇异值，那么$\{A\boldsymbol{v}_1,...,A\boldsymbol{v}_r\}$是$ColA$的一个正交基，并且$rank(A)=r$。

2.奇异值分解

设有$rank(A_{m\times n})=r$，那么可将其分解为$A=U\Sigma V^T$。

其中，$\Sigma_{m\times n}=\begin{bmatrix}D & 0\\ 0 & 0\end{bmatrix}$，$D$的对角线元素是$A$的前$r$个奇异值；$U_{m\times m}$是正交矩阵，$V_{n\times n}$是正交矩阵。

U的列称为A的左奇异向量，V的列称为A的右奇异向量。

$A_{m\times n}$奇异值的分解分三步：

(1)计算A^TA，将其正交对角化。

(2)计算V,Σ。将特征值降序排列，对应的单位正交特征向量集就是右奇异向量，构成V。之后取前r个非零奇异值够成D，再构成与A同行列的$Sigma$。

(3)计算U。U的前r列是$\{A\boldsymbol{v}_1,...,A\boldsymbol{v}_r\}$构成的单位向量，即$\boldsymbol{u}_k=\dfrac{1}{\sigma_k}A\boldsymbol{v}_k$

例：

例 4 求 $A=\begin{bmatrix}1 & -1\\ -2 & 2\\ 2 & -2\end{bmatrix}$的一个奇异值分解.

解 首先，计算 $A^{\mathrm{T}}A=\begin{bmatrix}9 & -9\\ -9 & 9\end{bmatrix}$，$A^{\mathrm{T}}A$ 的特征值是 18 和 0，相应的单位特征向量是

$$\boldsymbol{v}_1=\begin{bmatrix}1/\sqrt{2}\\ -1/\sqrt{2}\end{bmatrix},\ \boldsymbol{v}_2=\begin{bmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{bmatrix}$$

这两个向量构成 V 的列向量：

$$V=[\boldsymbol{v}_1 \quad \boldsymbol{v}_2]=\begin{bmatrix}-1/\sqrt{2} & 1/\sqrt{2}\\ -1/\sqrt{2} & 1/\sqrt{2}\end{bmatrix}$$

矩阵的奇异值为 $\sigma_1=\sqrt{18}=3\sqrt{2},\ \sigma_2=0$．因为只有一个非零的奇异值，“矩阵” D 可写成单个数值，即 $D=3\sqrt{2}$．矩阵Σ与矩阵 A 的行列数相同，以矩阵 D 为其左上角：

$$\Sigma=\begin{bmatrix}D & 0\\ 0 & 0\\ 0 & 0\end{bmatrix}=\begin{bmatrix}3\sqrt{2} & 0\\ 0 & 0\\ 0 & 0\end{bmatrix}$$

为构造 U ，首先计算 $A\boldsymbol{v}_1$和$A\boldsymbol{v}_2$：

$$A\boldsymbol{v}_1=\begin{bmatrix}2/\sqrt{2}\\ -4/\sqrt{2}\\ 4/\sqrt{2}\end{bmatrix},\quad A\boldsymbol{v}_2=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}$$

为检查计算，验证 $\|A\boldsymbol{v}_1\|=\sigma_1=\sqrt{18}=3\sqrt{2}$．当然，$A\boldsymbol{v}_2=0$，因为 $\|A\boldsymbol{v}_2\|=\sigma_2=0$．目前只找到 U 的一列是

$$\boldsymbol{u}_1=\frac{1}{3\sqrt{2}}A\boldsymbol{v}_1=\begin{bmatrix}1/3\\ -2/3\\ 2/3\end{bmatrix}$$

U 的其他列是将集合 $\{\boldsymbol{u}_1\}$ 扩充为 $\mathbb{R}^3$ 的标准正交基而得到的. 此时，我们需要两个单位正交向量且都与 $\boldsymbol{u}_1$ 正交.（见图 7-15.）每个向量必须满足 $\boldsymbol{u}_1^T\boldsymbol{x}=0$，这等价于方程 $x_1-2x_2+2x_3=0$，该方程的解构成的基是

$$\boldsymbol{w}_1=\begin{bmatrix}2\\1\\0\end{bmatrix},\ \boldsymbol{w}_2=\begin{bmatrix}-2\\0\\1\end{bmatrix}$$

（经检验 $\boldsymbol{w}_1$和$\boldsymbol{w}_2$都与$\boldsymbol{u}_1$正交.）应用格拉姆-施密特正交法于 $\{\boldsymbol{w}_1,\boldsymbol{w}_2\}$，可以得到

$$\boldsymbol{u}_2=\begin{bmatrix}2/\sqrt{5}\\1/\sqrt{5}\\0\end{bmatrix},\ \boldsymbol{u}_3=\begin{bmatrix}-2/\sqrt{45}\\4/\sqrt{45}\\5/\sqrt{45}\end{bmatrix}$$

最后，令 $U=[\boldsymbol{u}_1\quad\boldsymbol{u}_2\quad\boldsymbol{u}_3]$,以及由上所得的$V^T$和$\Sigma$，那么

$$A=\begin{bmatrix}1&-1\\-2&2\\2&-2\end{bmatrix}=\begin{bmatrix}1/3&2/\sqrt{5}&-2/\sqrt{45}\\-2/3&1/\sqrt{5}&4/\sqrt{45}\\2/3&0&5/\sqrt{45}\end{bmatrix}\begin{bmatrix}3\sqrt{2}&0\\0&0\\0&0\end{bmatrix}\begin{bmatrix}1/\sqrt{2}&-1/\sqrt{2}\\1/\sqrt{2}&1/\sqrt{2}\end{bmatrix}$$

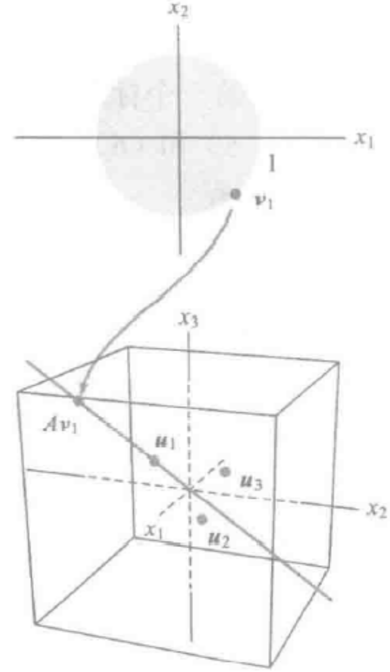

图 7-15 ■

八、习题

(1).7.1.35

35. 若 $\boldsymbol{u}$ 是 $\mathbb{R}^n$ 中的单位向量，取 $B=\boldsymbol{u}\boldsymbol{u}^T$.

a. 对任意属于 $\mathbb{R}^n$ 的 $\boldsymbol{x}$，计算 $B\boldsymbol{x}$ 且证明 $B\boldsymbol{x}$ 是 $\boldsymbol{x}$ 在 $\boldsymbol{u}$ 上的正交投影，正如 6.2 节所描述的情况.

b. 证明 B 是对称矩阵且 $B^2=B$.

c. 证明 $\boldsymbol{u}$ 是 B 的特征向量，并求其对应的特征值.

(a)

即证$B\boldsymbol{x}=\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{x}=\dfrac{\boldsymbol{u}\cdot\boldsymbol{x}}{\boldsymbol{u}\cdot\boldsymbol{u}}\boldsymbol{u}=(\boldsymbol{u}\cdot\boldsymbol{x})\boldsymbol{u}\Leftrightarrow\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{x}=(\boldsymbol{u}\cdot\boldsymbol{x})\boldsymbol{u}$

其中，$\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{x}=\boldsymbol{u}(\boldsymbol{u}^T\boldsymbol{x})=(\boldsymbol{u}^T\boldsymbol{x})\boldsymbol{u}=(\boldsymbol{u}\cdot\boldsymbol{x})\boldsymbol{u}$

即得证。

(b)

因$B^T=(uu^T)^T=(u^T)^T(u)^T=uu^T=B$

故B为对称矩阵。

因$B^2=(uu^T)(uu^T)=u(u^Tu)u^T=uu^T=B$

故$B^2=B$

(c)

即证存在λ使$B\boldsymbol{u}=\lambda\boldsymbol{u}$成立，即$\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{u}=\lambda\boldsymbol{u}$

而$\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{u}=\boldsymbol{u}(\boldsymbol{u}^T\boldsymbol{u})=\boldsymbol{u}$

从而特征值$\lambda=1$

(2).7.4.12

求奇异值分解：$\begin{bmatrix}1 & 1\\0 & 1\\-1 & 1\end{bmatrix}$

解：

$$A^TA=\begin{bmatrix}2 & 0\\0 & 3\end{bmatrix}$$

显然特征值为$\lambda_1=3,\lambda_1=2$，分别对应特征向量：

$$\boldsymbol{v}_1=[0,1]^T,\boldsymbol{v}_2=[1,0]^T$$

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

$$\Sigma=\begin{bmatrix}\sqrt{3} & 0\\0 & \sqrt{2}\\0 & 0\end{bmatrix}$$

$$Av_1=\begin{bmatrix}1\\1\\1\end{bmatrix},Av_2=\begin{bmatrix}1\\0\\-1\end{bmatrix}$$

故$A=\begin{bmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\\\frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}}\\\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\end{bmatrix}\begin{bmatrix}\sqrt{3} & 0\\0 & \sqrt{2}\\0 & 0\end{bmatrix}\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$

(3).7.4.19

19. 证明：V 的列是 $A^{T}A$ 的特征向量，U 的列是 AA^{T} 的特征向量，而 Σ 对角线上的元素是 A 的奇异值.（提示：利用奇异值分解计算 $A^{T}A$ 和 AA^{T}.）

因$A = U\Sigma V^T$

从而$A^TA = (U\Sigma V^T)^T U\Sigma V^T = (V\Sigma^T U^T)U\Sigma V^T$

其中，U, V都是正交矩阵，从而：

$A^TA = V\Sigma^T(U^TU)\Sigma V^T = V(\Sigma^T\Sigma)V^T$

说明V将A^TA对角化且其为单位特征向量组。同时：

$$\Sigma^T\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} D^T & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} DD^T & 0 \\ 0 & 0 \end{bmatrix}$$

这样，D对角线元素之平方就是A^TA的特征值，也即D对角线元素是A的奇异值。

同理，$AA^T = U\Sigma V^T(U\Sigma V^T)^T = U\Sigma V^T V\Sigma^T U^T = U(\Sigma\Sigma^T)U^T$

说明U是AA^T的单位特征向量组。

(4).7.4.23

23. 如果 $U = [\boldsymbol{u}_1 \quad \cdots \quad \boldsymbol{u}_m]$和$V = [\boldsymbol{v}_1 \quad \cdots \quad \boldsymbol{v}_n]$，证明：

$$A = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{T} + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{T} + \cdots + \sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^{T}$$

$A = U\Sigma V^T$

而$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sigma_r \end{bmatrix}$

$$U\Sigma = [u_1, u_2, ..., u_m] \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1[1] & \sigma_2 u_2[1] & \cdots & \sigma_r u_r[1] \\ \sigma_1 u_1[2] & \sigma_2 u_2[2] & & \vdots \\ \vdots & & \ddots & \\ \sigma_1 u_1[?] & \cdots & & \sigma_r u_r[?] \end{bmatrix}$$

$= [\sigma_1 u_1, \sigma_2 u_2, ..., \sigma_r u_r, 0]$

故$A = U\Sigma V^T = [\sigma_1 u_1, \sigma_2 u_2, ..., \sigma_r u_r, 0][v_1, v_2, ..., v_n]^T = \sigma_1 u_1 v_1 + \sigma_2 u_2 v_2 + ... + \sigma_r u_r v_r$

九、前置

1.矩阵的秩

(1)$0 \le rank(A_{m\times n}) \le min\{m, n\}$

(2)$max\{R(A), R(B)\} \le R(A, B) \le R(A) + R(B)$

(3)$R(A + B) \le R(A) + R(B)$

(4)$R(AB) \le min\{R(A), R(B)\}$

(5)若$A_{m\times n}B_{n\times l} = O$，则$R(A) + R(B) \le n$