数据科学的数学基础

(第1版)

黄定江　编著

华东师范大学
上海

目　录

第2章　习题

习题 2.1　假设向量$\boldsymbol{\beta}$可以经向量组$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$线性表出，证明：表示法是唯一的充分必要条件是$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$线性无关。

习题 2.2　证明$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_r$(其中$\boldsymbol{\alpha}_1 \neq \mathbf{0}$)线性相关的充分必要条件是至少有一$\boldsymbol{\alpha}_i(1 < i \leq s)$可被$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_{i-1}$线性表出。

习题 2.3　把向量$\boldsymbol{\beta}$表成向量$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4$的线性组合：

(1)　$\boldsymbol{\beta}=(1,2,1,1)$,　$\boldsymbol{\alpha}_1=(1,1,1,1)$,　$\boldsymbol{\alpha}_2=(1,1,-1,-1)$,　$\boldsymbol{\alpha}_3=(1,-1,1,-1)$,　$\boldsymbol{\alpha}_4=(1,-1,-1,1)$;

(1)　$\boldsymbol{\beta}=(0,0,0,1)$,　$\boldsymbol{\alpha}_1=(1,1,0,1)$,　$\boldsymbol{\alpha}_2=(2,1,3,1)$,　$\boldsymbol{\alpha}_3=(1,1,0,1)$,　$\boldsymbol{\alpha}_4=(0,1,-1,-1)$;

习题 2.4　设$\boldsymbol{\alpha}_1=(1,-1,2,4),\boldsymbol{\alpha}_2=(0,3,1,2),\boldsymbol{\alpha}_3=(3,0,7,14),\boldsymbol{\alpha}_4=(1,-1,2,0),\boldsymbol{\alpha}_5=(2,1,5,6)$.

(1) 证明：$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$线性无关；

(2) 把$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$扩充成一极大线性无关组。

习题 2.5　设$a_1,a_2,\cdots,a_n$是n个正数，证明：由$\Omega(\boldsymbol{x})=\left(\sum_{i=1}^n a_i x_i^2\right)^{\frac{1}{2}}$定义的函数$\Omega\mathbb{R}^n\to\mathbb{R}$是一个范数。

习题 2.6　证明：当且仅当$\boldsymbol{x}$和$\boldsymbol{y}$线性相关且$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\geq 0$时，才有

$$\|\boldsymbol{x}+\boldsymbol{y}\|_2=\|\boldsymbol{x}\|_2+\|\boldsymbol{y}\|_2$$

习题 2.7　设$\|\cdot\|$是$\mathbb{R}^m$上的一个向量范数，并且设$\boldsymbol{A}\in\mathbb{R}^{m\times n}$。证明：若$\mathrm{rank}(A)=n$，则$\|\boldsymbol{x}\|_{\boldsymbol{A}}:=\|\boldsymbol{A}\boldsymbol{x}\|$是$\mathbb{R}^n$上的一个向量范数。

习题 2.8　计算下列矩阵的秩：

$$(1)\begin{pmatrix}0&1&1&-1&2\\0&2&-2&-2&0\\0&-1&-1&1&1\\1&1&0&1&-1\end{pmatrix},\quad(2)\begin{pmatrix}1&-1&2&1&0\\2&-2&4&-2&0\\3&0&6&-1&1\\0&3&0&0&1\end{pmatrix},\quad(3)\begin{pmatrix}14&12&6&8&2\\6&104&21&9&17\\7&6&3&4&1\\35&30&15&20&5\end{pmatrix}$$

习题 2.9　设$\boldsymbol{A}=\begin{pmatrix}3&1&1\\2&1&2\\1&2&3\end{pmatrix}$，$\boldsymbol{B}=\begin{pmatrix}1&1&-1\\2&-1&0\\1&0&1\end{pmatrix}$

计算$\boldsymbol{AB},\boldsymbol{AB}-\boldsymbol{BA}$

习题 2.10　计算：

$$(1)\quad\begin{pmatrix}1&1\\0&1\end{pmatrix}^n\quad(2)\quad\begin{pmatrix}\cos\phi&-\sin\phi\\\sin\phi&\cos\phi\end{pmatrix}^n\quad(3)\quad\begin{pmatrix}\lambda&1&0\\0&\lambda&1\\0&0&\lambda\end{pmatrix}^n$$

习题 2.11　求$\boldsymbol{A}^{-1}$,设：

$$(1)\quad\boldsymbol{A}=\begin{pmatrix}a&b\\c&d\end{pmatrix}\quad(2)\quad\boldsymbol{A}=\begin{pmatrix}2&2&3\\1&-1&0\\-1&2&1\end{pmatrix}$$

习题 2.12　设

$$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{O}&\boldsymbol{A}\\\boldsymbol{C}\,\boldsymbol{O}&\end{pmatrix}$$

已知$\boldsymbol{A}^{-1},\boldsymbol{C}^{-1}$存在，求$\boldsymbol{X}^{-1}$

习题 2.13　证明：如果$\boldsymbol{A}=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)$是按列分块的，那么

$$\|\boldsymbol{A}\|_F^2=\|Va_1\|_2^2+\|Va_2\|_2^2+\cdots+\|\boldsymbol{a}_n\|_2^2$$

习题 2.14　证明：$\|\boldsymbol{AB}\|_F\leq\|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_F$和$\|\boldsymbol{AB}\|_F\leq\|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_2$

习题 2.15　设$\|\cdot\|$是由向量范数$\|\cdot\|$诱导出的矩阵范数。证明：若$\boldsymbol{A}\in\mathbb{R}^{n\times n}$非奇异，则

$$\|\boldsymbol{A}^{-1}\|^{-1}=\min_{\|\boldsymbol{x}\|=1}\|\boldsymbol{A}\boldsymbol{x}\|$$

习题 2.16 对于如下的两个矩阵：

$$A_1=\begin{pmatrix}9&6\\6&5\end{pmatrix},\qquad A_2=\begin{pmatrix}9&6\\6&3\end{pmatrix}$$

矩阵$\boldsymbol{A}_1$是正定的，因为其对称并且

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}_1\boldsymbol{x}=\begin{pmatrix}x_1x_2\end{pmatrix}\begin{pmatrix}9&6\\6&5\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=9x_1^2+12x_1x_2+5x_2^2=(3x_1+2x_2)^2+x_x^2>0$$

对所有非零向量$\boldsymbol{x}$成立。$\boldsymbol{A}_2$虽然是对称的，但它并不是正定矩阵，因为$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}_2\boldsymbol{x}=9x_1^2+12x_1x_2+3x_2^2=(3x_1+2x_2)^2-x_2^2$可以小于零。比如取$\boldsymbol{x}=(2,-3)^{\mathrm{T}}$

习题 2.17 证明$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\mathrm{Tr}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})$和$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\mathrm{Tr}(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})$

习题 2.18 证明：在$\mathbb{R}^n$上，当且仅当$\boldsymbol{A}$是正定矩阵时，函数$f(\boldsymbol{x})=(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})^{\frac{1}{2}}$是一个向量范数。

习题 2.19 t取什么值时，下列二次型是正定的：

(1) $x_1^2+x_2^2+5x_3^2+2tx_1x_2-2x_1x_3+4x_2x_3$

(2) $x_1^2+4x_2^2+x_3^2+2tx_1x_2+10x_1x_3+6x_2x_3$

习题 2.20 设$\boldsymbol{a}_1=\begin{pmatrix}1\\2\\-1\end{pmatrix},\boldsymbol{a}_2=\begin{pmatrix}-1\\3\\1\end{pmatrix},\boldsymbol{a}_3=\begin{pmatrix}4\\-1\\0\end{pmatrix}$，试用正交化过程把这组向量规范正交化。

习题 2.21 在$\mathbb{P}^4$中，求向量η在基$\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4$下的坐标，设

(1) $\varepsilon_1=(1,1,1,1),\varepsilon_2=(1,1,-1,-1),\varepsilon_3=(1,-1,1,-1),\varepsilon_4=(1,-1,-1,1),\eta=(1,2,1,1)$

(2) $\varepsilon_1=(1,1,0,1),\varepsilon_2=(2,1,3,1),\varepsilon_3=(1,1,0,0),\varepsilon_4=(0,1,-1,-1),\eta=(0,0,0,1)$

习题 2.22 在$\mathbb{P}^4$中，求由向量$\boldsymbol{\alpha}_i$生成的子空间与由向量$\boldsymbol{\beta}$生成的子空间的交的基和维数。设

(1) $\begin{cases}\boldsymbol{\alpha}_1=(1,2,1,0)\\\boldsymbol{\alpha}_2=(-1,1,1,1)\end{cases}\quad\begin{cases}\boldsymbol{\beta}_1=(2,-1,0,1)\\\boldsymbol{\beta}_2=(1,-1,3,7)\end{cases}$

(2) $\begin{cases}\boldsymbol{\alpha}_1=(1,1,0,0)\\\boldsymbol{\alpha}_2=(1,0,1,1)\end{cases}\quad\begin{cases}\boldsymbol{\beta}_1=(0,0,1,1)\\\boldsymbol{\beta}_2=(0,1,1,0)\end{cases}$

习题 2.23 设$\boldsymbol{A}=\begin{pmatrix}1 & 4 & 2\\ 0 & -3 & 4\\ 0 & 4 & 3\end{pmatrix}$求$\boldsymbol{A}^k$

习题 2.24 证明：如果$\boldsymbol{A}$可逆，证明：$\boldsymbol{AB}$与$\boldsymbol{BA}$相似

习题 2.25 分别应用幂法于矩阵

$$\boldsymbol{A}=\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix}\text{和}\boldsymbol{B}=\begin{pmatrix}\lambda & 1\\ 0 & -\lambda\end{pmatrix}(\lambda\neq 0)$$

并考察所得序列的特性。

第3章　习题

习题 3.1　构建模型使得预测值与真实值的误差最小常用向量2-范数度量，求解模型过程中需要计算梯度，求梯度：

- $f(\boldsymbol{A})=\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}-\boldsymbol{y}\|_2^2$，求$\frac{\partial f}{\partial \boldsymbol{A}}$
- $f(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}-\boldsymbol{y}\|_2^2$，求$\frac{\partial f}{\partial \boldsymbol{x}}$

习题 3.2　求$\frac{\partial \mathrm{Tr}(\boldsymbol{W}^{-1})}{\partial \boldsymbol{W}}$，利用迹微分法求解

习题 3.3　二次型是数据分析中常用函数，求$\frac{\partial \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}$，$\frac{\partial \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{A}}$

习题 3.4　$f(\boldsymbol{z})=\frac{\exp(\boldsymbol{z})}{\boldsymbol{1}^T\exp(\boldsymbol{z})}$称为softmax函数，$(\exp(\boldsymbol{z}))_i=\exp(\boldsymbol{z}_i)$，如果$\boldsymbol{q}=\frac{\exp(\boldsymbol{z})}{\boldsymbol{1}^T\exp(\boldsymbol{z})}, J=-\boldsymbol{p}^T log(\boldsymbol{q})$，其中$\boldsymbol{p},\boldsymbol{q},\boldsymbol{z}\in\mathbb{R}^n$，并且$\boldsymbol{1}^{\mathrm{T}}\boldsymbol{p}=1$，

- 证：$\frac{\partial J}{\partial \boldsymbol{z}}=\boldsymbol{q}-\boldsymbol{p}$
- 若$\boldsymbol{z}=\boldsymbol{W}\boldsymbol{x}$，其中$\boldsymbol{W}\in\mathbb{R}^{n\times m},\boldsymbol{x}\in\mathbb{R}^m$，$\frac{\partial J}{\partial \boldsymbol{W}}=(\boldsymbol{q}-\boldsymbol{p})\boldsymbol{x}^T$是否成立。

习题 3.5　以下内容是求解正态分布模型的关键步骤：$L=-\frac{Nd}{2}ln(2\pi)-\frac{N}{2}ln|\boldsymbol{\Sigma}|-\frac{1}{2}\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t-\boldsymbol{\mu})$

1) 求$\frac{\partial L}{\partial \boldsymbol{\mu}}$

2) 当$\mu=\frac{1}{N}\sum_t \boldsymbol{x}_t$时求$\frac{\partial L}{\partial \boldsymbol{\Sigma}}$，求$\boldsymbol{\Sigma}$使$\frac{\partial L}{\partial \boldsymbol{\Sigma}}=0,\boldsymbol{x}\in\mathbb{R}^N,\boldsymbol{\Sigma}\in\mathbb{R}^{N\times N}$

习题 3.6　求$\frac{\partial |\boldsymbol{X}^k|}{\partial \boldsymbol{X}}$

习题 3.7　求$\frac{\partial \mathrm{Tr}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^T\boldsymbol{C})}{\partial \boldsymbol{X}}$

第4章　习题

习题 4.1　判定矩阵$\boldsymbol{C}=\begin{bmatrix}3&2&-1\\-1&0&0\\-1&3&0\end{bmatrix}$和$\boldsymbol{B}=\begin{bmatrix}0&2&-1\\-1&4&-1\\1&3&-5\end{bmatrix}$能否进行$\boldsymbol{LU}$分解，为什么？如果能分解，试分解之。

解　对$\boldsymbol{B}$不能进行$\boldsymbol{LU}$分解，因为$\boldsymbol{B}$的一阶顺序主子式为0，面对于$\boldsymbol{C}$可以进行$\boldsymbol{LU}$分解：

$$\boldsymbol{C}=\begin{bmatrix}1&0&0\\-\frac{1}{3}&1&0\\-\frac{1}{3}&\frac{11}{2}&1\end{bmatrix}\begin{bmatrix}\frac{2}{3}&2&-1\\0&\frac{2}{3}&-\frac{1}{3}\\0&0&\frac{3}{2}\end{bmatrix}$$

习题 4.2　对下列矩阵进行杜立特分解：

（1）$\boldsymbol{A}=\begin{bmatrix}2&1&1\\1&3&2\\1&2&2\end{bmatrix}$; (2)$\boldsymbol{B}=\begin{bmatrix}12&-3&3\\-18&3&-1\\1&1&1\end{bmatrix}$

解　（1）$\boldsymbol{A}=\begin{bmatrix}1&&\\\frac{1}{2}&1&\\\frac{1}{2}&\frac{3}{5}&1\end{bmatrix}\begin{bmatrix}2&1&1\\&\frac{5}{2}&\frac{3}{2}\\&&\frac{3}{5}\end{bmatrix}$　（2）$\boldsymbol{B}=\begin{bmatrix}1&&\\-\frac{3}{2}&1&\\\frac{1}{12}&-\frac{5}{6}&1\end{bmatrix}\begin{bmatrix}12&-3&3\\&-\frac{3}{2}&\frac{7}{2}\\&&\frac{11}{3}\end{bmatrix}$

习题 4.3　求矩阵$\boldsymbol{A}=\begin{bmatrix}2&1&1\\1&2&1\\1&1&0\end{bmatrix}$的$\boldsymbol{LU}$分解。

解 对$\left[\begin{array}{c:c} A & I \end{array}\right] = \left[\begin{array}{c:c} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 00 & 1 & \end{bmatrix} \end{array}\right] \to \left[\begin{array}{c:c} \begin{bmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & -\frac{2}{3} \end{bmatrix} & \begin{bmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & -\frac{2}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & 1 \end{bmatrix} \end{array}\right]$

可得$P = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & 1 \end{bmatrix}$, $L = P^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix}$, $U = \begin{bmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & -\frac{2}{3} \end{bmatrix}$即得分解式。

习题 4.4 求对称正定矩阵

$$A = \begin{bmatrix} 5 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 5 \end{bmatrix}$$

的不带平方根的乔累斯基分解。

解

$$A = \begin{bmatrix} \sqrt{5} & & \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \\ \frac{-4}{\sqrt{5}} & \frac{-2}{\sqrt{5}} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & \frac{2}{\sqrt{5}} & \frac{-4}{\sqrt{5}} \\ & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ & & 1 \end{bmatrix}$$

习题 4.5 对A进行LU分解

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

解 首先将矩阵第一列对角线上元素A_{11}下面的元素通过矩阵初等行变换变为0。$\begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & -2 \\ -3 & 1 & 1 \end{bmatrix}$ →第2行减去第1行的2倍→ $\begin{bmatrix} 1 & 2 & -1 \\ 0 & -3 & 0 \\ -3 & 1 & 1 \end{bmatrix}$ →第3 行减去第1行的-3倍→ $\begin{bmatrix} 1 & 2 & -1 \\ 0 & -3 & 0 \\ 0 & 7 & -2 \end{bmatrix}$

然后再将矩阵第二列对角线上元素A_{22}下面的元素通过矩阵初等行变换变为0。→第3行减去第2行的$-\frac{7}{3}$倍→ $\begin{bmatrix} 1 & 2 & -1 \\ 0 & -3 & 0 \\ 0 & 0 & -2 \end{bmatrix} = U$

则得到的上三角矩阵就是$\boldsymbol{U}$。这个时候，$\boldsymbol{L}$也已经求出来了。通过将下三角主对角线上的元素都置为1，乘数因子放在下三角相应的位置（放在消元时将元素变为0的那个元素的位置），就可以得到下三角矩阵$\boldsymbol{L}$。如下：

$$\boldsymbol{L}=\begin{bmatrix}1&0&0\\2&1&0\\-3&\frac{7}{3}&1\end{bmatrix}$$

对于$\boldsymbol{L}$的构造，举个例子，如将第一列的元素2变为0时，第二行减去第一行乘以2，于是$\boldsymbol{A}_{21}$就变成了0。这个乘数因子将元素$\boldsymbol{A}_{21}$变成了0，对应的，下三角矩阵$\boldsymbol{L}$中对应位置的元素$\boldsymbol{L}_{21}$就为乘数因子2。其他的情况类似。

习题 4.6　做Cholesky分解

$$\begin{bmatrix}25&15&-5\\15&18&0\\-5&0&11\end{bmatrix}$$

$$\begin{bmatrix}25&15&-5\\15&18&0\\-5&0&11\end{bmatrix}=\begin{bmatrix}R_{11}&0&0\\R_{12}&R_{22}&0\\R_{13}&R_{23}&R_{33}\end{bmatrix}\begin{bmatrix}R_{11}&R_{12}&R_{13}\\0&R_{22}&R_{23}\\0&0&R_{33}\end{bmatrix}=\begin{bmatrix}5&0&0\\3&3&0\\-1&1&3\end{bmatrix}\begin{bmatrix}5&3&-1\\0&3&1\\0&0&3\end{bmatrix}$$

习题 4.7　求下列矩阵的正交三角分解（UR）表达式：

$$\boldsymbol{A}=\begin{bmatrix}0&1&1\\1&1&0\\1&0&1\end{bmatrix}$$

解　记$\boldsymbol{\alpha}_1=(0,1,1)^T,\boldsymbol{\alpha}_2=(1,1,0)^T,\boldsymbol{\alpha}_3=(1,0,1)^T$,由史密特正交化方法可得

$$\boldsymbol{\beta}_1=\boldsymbol{\alpha}_1=(0,1,1)^T,|\boldsymbol{\beta}_1|=\sqrt{2},\boldsymbol{\eta}_1=\frac{1}{\sqrt{2}}(0,1,1)^T;$$

$$(\boldsymbol{\eta}_1,\boldsymbol{\alpha}_2)=\frac{1}{\sqrt{2}},\boldsymbol{\beta}_2=\boldsymbol{\alpha}_2-(\boldsymbol{\alpha}_2,\boldsymbol{\eta}_1)\boldsymbol{\eta}_1=(1,\frac{1}{2},-\frac{1}{2})^T,|\boldsymbol{\beta}_2|=\frac{\sqrt{6}}{2},\boldsymbol{\eta}_2=\frac{1}{\sqrt{6}}(2,1,-1)^T;$$

$$(\boldsymbol{\eta}_1,\boldsymbol{\alpha}_3)=\frac{1}{\sqrt{2}},(\boldsymbol{\eta}_2,\boldsymbol{\alpha}_3)=\frac{1}{6},\boldsymbol{\beta}_3=\boldsymbol{\alpha}_2-(\boldsymbol{\alpha}_3,\boldsymbol{\eta}_1)\boldsymbol{\eta}_1-(\boldsymbol{\alpha}_3,\boldsymbol{\eta}_2)\boldsymbol{\eta}_2=(\frac{2}{3},-\frac{2}{3},\frac{2}{3})^T,$$

$$|\boldsymbol{\beta}_3|=\frac{2}{\sqrt{3}},\boldsymbol{\eta}_3=\frac{1}{\sqrt{3}}(1,-1,1)^T$$

. 于是取

$$U=(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\eta}_3)=\begin{bmatrix}0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}}\end{bmatrix}$$

$$\boldsymbol{R}=\begin{bmatrix}|\boldsymbol{\beta}_1| & (\boldsymbol{\alpha}_2,\boldsymbol{\eta}_1) & (\boldsymbol{\alpha}_3,\boldsymbol{\eta}_1)\\ 0 & |\boldsymbol{\beta}_2| & (\boldsymbol{\alpha}_3,\boldsymbol{\eta}_2)\\ 0 & 0 & |\boldsymbol{\beta}_3|\end{bmatrix}=\begin{bmatrix}\sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ 0 & \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}}\\ 0 & 0 & \frac{2}{\sqrt{3}}\end{bmatrix}$$

那么$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{R}$即为所求表达式。

习题 4.8　求矩阵

$$\boldsymbol{A}=\begin{bmatrix}1 & \frac{1}{2} & 5\\ 1 & -\frac{1}{2} & 2\\ -1 & \frac{1}{2} & -2\\ 1 & -\frac{3}{2} & 0\end{bmatrix}$$

的QR分解。

解　令$\boldsymbol{\alpha}_1=(1,1,-1,1)^T,\boldsymbol{\alpha}_2=(\frac{1}{2},-\frac{1}{2},\frac{1}{2},-\frac{3}{2})^T,\boldsymbol{\alpha}_3=(5,2,-2,0)^T$,由史密特正交化方法得

$$\boldsymbol{\beta}_1=\boldsymbol{\alpha}_1=(1,1,-1,1)^T,\boldsymbol{\eta}_1=\frac{1}{2}(1,1,-1,1)^T$$

$$\boldsymbol{\beta}_2=\boldsymbol{\alpha}_2-(\boldsymbol{\eta}_1,\boldsymbol{\alpha}_2)\boldsymbol{\eta}_1=(1,0,0,-1)^T,\boldsymbol{\eta}_2=\frac{1}{\sqrt{2}}(1,0,0,-1)^T$$

$$\boldsymbol{\beta}_3=\boldsymbol{\alpha}_3-(\boldsymbol{\eta}_1,\boldsymbol{\alpha}_3)\boldsymbol{\eta}_1-(\boldsymbol{\eta}_2,\boldsymbol{\alpha}_3)\boldsymbol{\eta}_2=\frac{1}{4}(1,-1,1,1)^T,\boldsymbol{\eta}_3=\frac{1}{2}(1,-1,1,1)^T$$

从而有

$$\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{R}=\begin{bmatrix}\frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2}\\ \frac{1}{2} & 0 & -\frac{1}{2}\\ -\frac{1}{2} & 0 & \frac{1}{2}\\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}2 & -1 & \frac{9}{2}\\ 0 & \sqrt{2} & \frac{5}{\sqrt{2}}\\ 0 & 0 & \frac{1}{2}\end{bmatrix}$$

习题 4.9　求矩阵$\boldsymbol{A}=\begin{bmatrix}1 & 0 & 0 & -1\\ 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix}$的奇异值分解。

解　由$\boldsymbol{A}^H\boldsymbol{A}=\begin{bmatrix}1&0&0&-1\\0&1&0&0\\0&1&0&0\\-1&1&0&2\end{bmatrix}$,特征值为$\lambda_1=1,\lambda_2=3,\lambda_3=\lambda_4=0$,特征向量依次为

$$\boldsymbol{\alpha}_1=\begin{bmatrix}1\\1\\0\\0\end{bmatrix},\boldsymbol{\alpha}_2=\begin{bmatrix}-1\\1\\0\\2\end{bmatrix},\boldsymbol{\alpha}_3=\begin{bmatrix}0\\0\\1\\0\end{bmatrix},\boldsymbol{\alpha}_4=\begin{bmatrix}1\\-1\\0\\1\end{bmatrix}$$

它们已两两正交，于是可得

$$\mathrm{rank}(\boldsymbol{A})=2,\sum=\begin{bmatrix}1&0\\0&\sqrt{3}\end{bmatrix},\boldsymbol{V}=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{6}}&0&\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{6}}&0&\frac{-1}{\sqrt{3}}\\0&0&1&0\\0&\frac{2}{\sqrt{6}}&0&\frac{1}{\sqrt{3}}\end{bmatrix}\tag{4.1}$$

此时

$$\boldsymbol{V}_1=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{6}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{6}}\\0&0\\0&\frac{2}{\sqrt{6}}\end{bmatrix},\boldsymbol{U}_1=\boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1}=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\0&0\\0&0\end{bmatrix}\tag{4.2}$$

取

$$\boldsymbol{U}_2=\begin{bmatrix}0\\0\\0\end{bmatrix},\boldsymbol{U}_1=[\boldsymbol{U}_1\quad\boldsymbol{U}_2]=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}&0\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}&0\\0&0&1\end{bmatrix}\tag{4.3}$$

则$\boldsymbol{A}$的奇异值分解

$$\boldsymbol{A}=\boldsymbol{U}\begin{bmatrix}1&0&0&0\\0&\sqrt{3}&0&0\\0&0&0&0\end{bmatrix}\boldsymbol{V}^H\tag{4.4}$$

习题 4.10　求$\boldsymbol{A}=\begin{bmatrix}1&0\\0&1\\1&1\end{bmatrix}$的奇异值分解.

解: $\boldsymbol{A}^H\boldsymbol{A}=\begin{bmatrix}2&1\\1&2\end{bmatrix}$,得$\lambda_1=3,\lambda_2=1$,相应的特征向量为$(1,1)^T,(-1,1)^T$.于是有

$$\operatorname{rank}(\boldsymbol{A})=2,\boldsymbol{\Sigma}=\begin{bmatrix}\sqrt{3}&0\\0&1\end{bmatrix},\boldsymbol{V}=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{bmatrix}\tag{4.5}$$

取

$$\boldsymbol{V}_1=\boldsymbol{V},\boldsymbol{U}_1=\boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1}=\begin{bmatrix}\frac{1}{\sqrt{6}}&\frac{-1}{\sqrt{2}}\\\frac{1}{\sqrt{6}}&\frac{1}{\sqrt{2}}\\\frac{2}{\sqrt{6}}&0\end{bmatrix},\boldsymbol{U}_2=\begin{bmatrix}\frac{-1}{\sqrt{3}}\\\frac{-1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{bmatrix}\tag{4.6}$$

构造正交阵$\boldsymbol{U}=[\boldsymbol{U}_1,\boldsymbol{U}_2]=\begin{bmatrix}\frac{1}{\sqrt{6}}&\frac{-1}{\sqrt{2}}&\frac{-1}{\sqrt{3}}\\\frac{1}{\sqrt{6}}&\frac{1}{\sqrt{2}}&\frac{-1}{\sqrt{3}}\\\frac{2}{\sqrt{6}}&0&\frac{1}{\sqrt{3}}\end{bmatrix}$,则$\boldsymbol{A}$的奇异值为

$$\boldsymbol{A}=\boldsymbol{U}\begin{bmatrix}\sqrt{3}&0\\0&1\\0&0\end{bmatrix}\boldsymbol{V}^H\tag{4.7}$$

习题 4.11　设$\boldsymbol{A}=\begin{bmatrix}2&0&1\\1&2&0\end{bmatrix}$,求$\boldsymbol{A}$的奇异值分解.

解: 第一步计算$\boldsymbol{A}^T\boldsymbol{A}$的奇异值,注意到矩阵

$$\boldsymbol{A}^T\boldsymbol{A}=\begin{bmatrix}5&2&2\\2&4&0\\2&0&1\end{bmatrix}\tag{4.8}$$

由于$\left|\lambda\boldsymbol{I}-\boldsymbol{A}^T\boldsymbol{A}\right|=\lambda(\lambda-3)(\lambda-7)$,故$\boldsymbol{A}^T\boldsymbol{A}$的特征值为7,3,0,所以$\boldsymbol{A}$的奇异值为$\sigma_1=\sqrt{7},\sigma_2=\sqrt{3}$

第二步　求$\boldsymbol{A}^T\boldsymbol{A}$的对应于三个特征值的3个标准正交特征向量.

对于$\lambda_1=7$,求得$\boldsymbol{\alpha}_1=(3,2,1)^T,\boldsymbol{\varepsilon}_1=\frac{1}{\sqrt{14}}(3,2,1)^T$；

对于$\lambda_2 = 3$,求得$\boldsymbol{\alpha}_2 = (1,-2,1)^T, \boldsymbol{\varepsilon}_1 = \frac{1}{\sqrt{6}}(1,-2,1)^T$；

对于$\lambda_3 = 0$,求得$\boldsymbol{\alpha}_1 = (2,-1,-4)^T, \boldsymbol{\varepsilon}_1 = \frac{1}{\sqrt{21}}(2,-1,-4)^T$；

可得酉矩阵$\boldsymbol{V} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3)$,取$\boldsymbol{V}_1 = (\boldsymbol{\varepsilon}_1, \varepsilon_2)$

第三步　计算

$$\boldsymbol{U}_1 = \boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Lambda}^{-1} = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{3}{14} & \frac{1}{6} \\ \frac{2}{14} & \frac{-2}{6} \\ \frac{1}{14} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} \frac{1}{7} & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \tag{4.9}$$

最后, 计算$\boldsymbol{U} = \boldsymbol{U}_1$, $\boldsymbol{V} = (\varepsilon_1, \varepsilon_2, \varepsilon_3)$,则有

$$A = U \begin{bmatrix} \sqrt{7} & & \\ & \sqrt{3} & \\ & & 0 \end{bmatrix} V^H \tag{4.10}$$

习题 4.12　已知

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \\ 0 & 2 \\ 1 & 0 \end{bmatrix}$$

求$\boldsymbol{A}$的奇异分解表达式。

解：由于$rank\boldsymbol{A} = 2, |\lambda\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A}| = (\lambda-5)(\lambda-2)$，于是$\boldsymbol{A}$的奇异值$\sigma_1 = \sqrt{5}, \sigma_2 = \sqrt{2}$.

对应于λ_1与λ_2的标准正交向量分别为

$$\varepsilon_1 = (0,1)^{\mathrm{T}}, \quad \varepsilon_2 = (1,0)^{\mathrm{T}}$$

记$\boldsymbol{V}_r = (\varepsilon_1, \varepsilon_2), \Delta = \mathrm{diag}(\sigma_1, \sigma_2)$，得

$$\boldsymbol{U}_r = \boldsymbol{A}\boldsymbol{V}\Delta^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \\ 0 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & \frac{-1}{\sqrt{\sqrt{2}}} \\ \frac{2}{\sqrt{5}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

于是

$$A = U_r \Delta V_r^H = \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & \frac{-1}{\sqrt{2}} \\ \frac{2}{\sqrt{5}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^H$$

习题 4.13　已知$\boldsymbol{A} \in \mathcal{C}_r^{m\times n}$（秩为$r > 0$）的奇异值分解表达式为

$$\boldsymbol{A} = \boldsymbol{U} \begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{V}^H$$

试求矩阵$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{A} \end{bmatrix}$的奇异值分解表达式。

解：由$\boldsymbol{A}$的奇异值分解$\boldsymbol{A} = \boldsymbol{U} \begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{V}^H$，其中$\Delta = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$，注意到

$$\boldsymbol{B}\boldsymbol{B}^H = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{A} \end{bmatrix} (\boldsymbol{A}^H . \boldsymbol{A}^H) = \begin{bmatrix} \boldsymbol{A}\boldsymbol{A}^H & \boldsymbol{A}\boldsymbol{A}^H \\ \boldsymbol{A}\boldsymbol{A}^H & \boldsymbol{A}\boldsymbol{A}^H \end{bmatrix}$$

如果将$\boldsymbol{U}$的前r列向量的矩阵记为$\boldsymbol{U}_1$，而$\boldsymbol{U}$的后$n-r$列向量组成的矩阵记为$\boldsymbol{U}_2$，既有

$$\boldsymbol{U} = [\boldsymbol{U}_1, \boldsymbol{U}_2]$$

取

$$\widetilde{\boldsymbol{U}} = \begin{bmatrix} \frac{1}{\sqrt{2}}\boldsymbol{U}_1 & \frac{1}{\sqrt{2}}\boldsymbol{U}_1 & \boldsymbol{U}_1 & 0 \\ \frac{1}{\sqrt{2}}\boldsymbol{U}_1 & -\frac{1}{\sqrt{2}}\boldsymbol{U}_1 & 0 & \boldsymbol{U}_2 \end{bmatrix}, \quad \widetilde{\boldsymbol{V}} = \boldsymbol{V}$$

那么

$$\boldsymbol{B} = \widetilde{\boldsymbol{U}} \begin{bmatrix} \sqrt{2}\Delta & 0 \\ 0 & 0 \end{bmatrix} \widetilde{\boldsymbol{V}}^H$$

即为矩阵$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{A} \end{bmatrix}$的奇异值分解表达式。

第四章内容

习题 4.14 已知矩阵

$$A = \begin{bmatrix} 0 & 2 & 4 \\ \frac{1}{2} & 0 & 2 \\ \frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix}$$

验证$\boldsymbol{A}$是可对角化矩阵，并求$\boldsymbol{A}$的谱分解表达式。

解 考察$\boldsymbol{A}$的特征多项式$|\lambda \boldsymbol{I} - \boldsymbol{A}| = (\lambda+1)^2(\lambda-2)$，所以其特征值为$\lambda_1 = \lambda_2 = -1, \lambda_3 = 2$，它们对应的三个线性无关的特征向量为$\boldsymbol{\alpha}_1 = (-2,1,0)^{\mathrm{T}}, \boldsymbol{\alpha}_2 = (-4,0,1)^{\mathrm{T}}, \alpha_3 = (4,2,1)^{\mathrm{T}}$，于是

$$\boldsymbol{P} = (\alpha_1, \alpha_2, \alpha_3) = \begin{bmatrix} -2 & -4 & 4 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{P}^{-1} = \begin{bmatrix} -\frac{1}{6} & \frac{2}{3} & \frac{-2}{3} \\ -\frac{1}{12} & -\frac{1}{6} & \frac{2}{3} \\ \frac{1}{12} & \frac{1}{6} & \frac{1}{3} \end{bmatrix}$$

而且有$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} -1 & & \\ & -1 & \\ & & 2 \end{bmatrix}$，所以$\boldsymbol{A}$是单纯矩阵（可对角化）。

由于

$$(\boldsymbol{P}^{-1})^{\mathrm{T}} = \begin{bmatrix} -\frac{1}{6} & -\frac{1}{12} & \frac{1}{12} \\ \frac{2}{3} & -\frac{1}{6} & \frac{1}{6} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix} = (\beta_1, \beta_2, \beta_3)$$

令

$$\boldsymbol{G}_1 = \alpha_1\beta_1^{\mathrm{T}} + \alpha_2\beta_2^{\mathrm{T}} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & -\frac{4}{3} \\ -\frac{1}{6} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{1}{12} & -\frac{1}{6} & \frac{2}{3} \end{bmatrix}$$

$$\boldsymbol{G}_2 = \alpha_3\beta_3^{\mathrm{T}} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{4}{3} \\ \frac{1}{6} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{12} & \frac{1}{6} & \frac{1}{3} \end{bmatrix}$$

故$\boldsymbol{A} = -\boldsymbol{G} + 2\boldsymbol{G}_2$为$\boldsymbol{A}$的谱分解表达式。

第5章　习题

习题 5.1　设$\boldsymbol{A}=\begin{bmatrix}1&2\\3&4\\5&6\end{bmatrix}$，$\boldsymbol{b}=\begin{bmatrix}1\\1\\1\end{bmatrix}$用正则化方法求对应的LS问题的解。

习题 5.2　设$\boldsymbol{A}=\begin{bmatrix}1&3&1&1\\2&0&0&0\\1&0&0&0\end{bmatrix}$，$\boldsymbol{b}=\begin{bmatrix}1\\1\\1\end{bmatrix}$求对应的LS问题的全部解。

习题 5.3　设$\boldsymbol{A}\in\mathbb{R}^{m\times n}$且存在$\boldsymbol{X}\in\mathbb{R}^{n\times m}$使得对每一个$\boldsymbol{b}\in\mathbb{R}^m,\boldsymbol{x}=\boldsymbol{X}\boldsymbol{b}$均极小化$\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2$. 证明$\boldsymbol{A}\boldsymbol{X}\boldsymbol{A}=\boldsymbol{A}$和$(\boldsymbol{A}\boldsymbol{X})^T=\boldsymbol{A}\boldsymbol{X}$.

习题 5.4　利用等式

$$\|\boldsymbol{A}(\boldsymbol{x}+\alpha\boldsymbol{w})-\boldsymbol{b}\|_2^2=\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2+2\alpha\boldsymbol{w}^T\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})+\alpha^2\|\boldsymbol{A}\boldsymbol{w}\|_2^2$$

证明：如果$\boldsymbol{x}\in\boldsymbol{X}_{LS}$,那么$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}^T\boldsymbol{b}$

习题 5.5　给定点集$\boldsymbol{p}_1,\cdots,\boldsymbol{p}_m\in\mathbb{R}^n$构成的$m\times n$矩阵$\boldsymbol{P}=[\boldsymbol{p}_1,\cdots,\boldsymbol{p}_m]$.考虑问题

$$\min_{\boldsymbol{X}} F(\boldsymbol{X})=\sum_{i=1}^{m}\|\boldsymbol{x}_i-\boldsymbol{p}_i\|_2^2+\frac{\lambda}{2}\sum_{1\le i,j\le m}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_2^2$$

其中$\lambda\ge 0$为参数，变量是一个$m\times n$矩阵$\boldsymbol{X}=[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m]$，其中$\boldsymbol{x}_i\in\mathbb{R}^n$是$\boldsymbol{X}$的第$i$列，$i=1\cdots,m$.上述问题尝试聚类点集$\boldsymbol{p}_i$，第一项鼓励聚类中心$\boldsymbol{x}_i$靠近对应的点$\boldsymbol{p}_i$，第二项鼓励$\boldsymbol{x}_i$们之间彼此靠近，当$\lambda$增大的时候，对应更高的组群影响。

1.请说明这个问题属于最小二乘类问题。不需要明确阐述这个问题的形式。

2.证明$\frac{1}{2}\sum_{1\le i,j\le m}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_2^2=trac\boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^T$,其中$\boldsymbol{H}=m\boldsymbol{I}_m-\boldsymbol{1}\boldsymbol{1}^T$是一个$m\times m$矩阵，$\boldsymbol{I}_m$是$m\times m$单位矩阵，$\boldsymbol{1}$是$\mathbb{R}^n$中的单位向量。

3.证明$\boldsymbol{H}$是半正定的。

4.证明函数F在矩阵$\boldsymbol{X}$处的梯度是一个$n \times m$矩阵，为：

$$\nabla F(\boldsymbol{X}) = 2(\boldsymbol{X} - \boldsymbol{P} + \lambda \boldsymbol{X} \boldsymbol{H})$$

提示：对于第二项，找到函数的一阶展式，$\triangle \to trace((\boldsymbol{X} + \triangle)\boldsymbol{H}(\boldsymbol{X} + \triangle)^T)$,其中$\triangle \mathbb{R}^{n,m}$。

5.依据最小二乘问题的最优条件为目标函数的梯度为零。证明最优点集的形式为:

$$\boldsymbol{x}_i = \frac{1}{m\lambda + 1}\boldsymbol{p}_i + \frac{m\lambda}{m\lambda + 1}\hat{\boldsymbol{p}}, i = 1, \cdots, m,$$

其中$\hat{\boldsymbol{p}} = (1/m)(\boldsymbol{p}_1 + \cdots + \boldsymbol{p}_m)$是给定点集的中心。

6.阐述你的结果，你认为这是聚类点集的一个好的模型么？

习题 5.6　判断$[1, 3, 4]$的转置是否在$\boldsymbol{A}$的零空间中

$$\boldsymbol{A} = \begin{bmatrix} 3 & 5 & -3 \\ 6 & -2 & 0 \\ -8 & 4 & 1 \end{bmatrix}$$

习题 5.7　求矩阵

$$\begin{bmatrix} 5 & 21 & 19 \\ 13 & 23 & 2 \\ 8 & 14 & 1 \end{bmatrix}$$

的行空间和列空间

习题 5.8　简答：阐述非负矩阵分解和主成分分析的相同点和不同点

第6章　习题

习题 6.1　已知三阶张量$\mathbf{X} \in R^{3\times4\times2}$的正面切片矩阵为

$$\begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix} \qquad \begin{bmatrix} 15 & 18 & 21 & 24 \\ 16 & 19 & 22 & 25 \\ 17 & 20 & 23 & 26 \end{bmatrix}$$

令$U = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$，求$\mathbf{Y} = \mathbf{X} \times_1 \boldsymbol{U}$的正面切片矩阵$\boldsymbol{Y}_1$ 和$\boldsymbol{Y}_2$。

习题 6.2　证明：$x^{I\times JK}_{i,(k-1)J+j} = x_{ijk} \Leftrightarrow \boldsymbol{X}^{(I\times JK)} = [\boldsymbol{X}_{::1}, \cdots, \boldsymbol{X}_{::K}] = AG^{(P\times QR)}(C\otimes B)^T$

习题 6.3　令

$$X_{i::} = b_1 c_1^T a_{i1} + \cdots + b_R c_R^T a_{iR}$$

证明水平展开的CP分解

$$\boldsymbol{X}^{(J\times KI)} = [\boldsymbol{X}_{1::}, \cdots, \boldsymbol{X}_{I::}] = \boldsymbol{B}(\boldsymbol{A}\odot \boldsymbol{C})^{\mathrm{T}}$$

和垂直展开的CP分解

$$\boldsymbol{X}^{(IJ\times K)} = \begin{bmatrix} \boldsymbol{X}_{1::} \\ \vdots \\ \boldsymbol{X}_{I::} \end{bmatrix} = (\boldsymbol{A}\odot \boldsymbol{B})\boldsymbol{C}^T$$

第7章　习题

习题 7.1　证明$x^* = (1, 1/2, -1)$是如下优化问题的最优解

$$\begin{array}{ll} \texttt{minimize} & (1/2)x^T P x + q^T x + r \\ \texttt{subject to} & -1 \leq x_i \leq 1, \quad i = 1,2,3 \end{array}$$

其中

$$P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, \quad q = \begin{bmatrix} -22.0 \\ -14.5 \\ 13.0 \end{bmatrix}, \quad r = 1$$

证明　最优点x^*处的梯度:

$$\nabla f_0(x^*) = (-1, 0, 2)$$

对任意y满足$-1 \leq y_i \leq 1$

$$\nabla f_0(x^*)^T (y - x) = -1(y_1 - 1) + 0(y_2 - 1/2) + 2(y_3 + 1) \geq 0$$

因此x^*满足最优性条件, 是目标函数的最优点.

习题 7.2　考虑极小化二次函数

$$f_0(\boldsymbol{x}) = (1/2)\boldsymbol{x}^{\mathrm{T}} \boldsymbol{P} \boldsymbol{x} + \boldsymbol{q}^{\mathrm{T}} \boldsymbol{x} + r$$

其中，$\boldsymbol{P} \in \mathbb{S}_+^n$($n$阶半正定矩阵)。给出$\boldsymbol{x}$为$f_0$最小解的重要条件，并说明$\boldsymbol{x}$何时无解，有唯一解，有多个解。

习题 7.3　计算$f(x)$的共轭函数，以及共轭函数的定义域。

- $f(x) = -\log x$

- $f(x) = e^x$

习题 7.4 证明: Gauss概率密度函数的累积分布函数

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

是对数-凹函数.

证明: 由题意得,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

$$\Phi^{'}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\Phi^{''}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}(-x)$$

$$(\Phi^{'}(x))^2 = \frac{1}{2\pi} e^{-x^2}$$

$$\Phi(x) \log \Phi^{''}(x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{-u^2/2} du \cdot e^{-x^2/2}(-x)$$

当$x \geq 0$时, $(\Phi^{'}(x))^2 \geq 0 \geq \Phi(x)\Phi^{''}(x)$.
当$x < 0$时, 由于$\frac{u^2}{2}$是凸函数, 则

$$\frac{u^2}{2} \geq \frac{x^2}{2} + (u-x)x \geq xu - \frac{x^2}{2}$$

所以,

$$\int_{-\infty}^{x} e^{-u^2/2} du \leq \int_{-\infty}^{x} e^{\frac{x^2}{2} - xu} du$$

$$= e^{\frac{x^2}{2}} \cdot \frac{e^{-xu}}{-x} \Big|_{u=-\infty}^{x}$$

$$= e^{\frac{x^2}{2}} \cdot \frac{e^{-x^2}}{-x}$$

因此$\Phi(x)\Phi^{''}(x) \leq \frac{1}{2\pi} e^{-x^2} = (\Phi^{'}(x))^2$, $\Phi(x)$是对数凹函数.

习题 7.5 考虑优化问题

$$\begin{array}{ll} \texttt{minimize} & f_0(x_1, x_2) \\ \texttt{subject to} & 2x_1 + x_2 \geq 1 \end{array}$$

$$x_1 + 3x_2 \geq 1$$

$$x_1 \geq 0, x_2 \geq 0$$

给出以下函数最优集和最优值

$$(a) \quad f_0(x_1, x_2) = x_1 + x_2$$

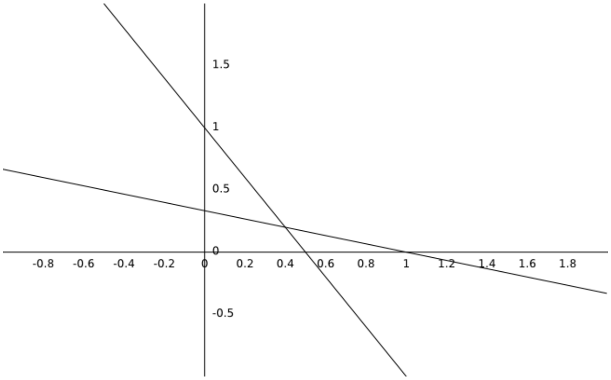
$$(b) \quad f_0(x_1, x_2) = -x_1 - x_2$$

$$(c) \quad f_0(x_1, x_2) = x_1$$

$$(d) \quad f_0(x_1, x_2) = \max\{x_1, x_2\}$$

$$(e) \quad f_0(x_1, x_2) = x_1^2 + 9x_2^2$$

解：作出约束函数图像得，



(a)

$$\begin{cases} 2x_1 + x_2 = 1 \\ x_2 + 3x_2 = 1 \end{cases}$$

$$x_1^* = \frac{2}{5}, x_2^* = \frac{1}{5}, f_0^* = \frac{3}{5}.$$

(b)无最优值.

(c)最优集为$\{(x_1, x_2) | x_1 = 0, x_2 \geq 1\}$.

(d)当$x_1 \geq x_2$时, 在可行域上, $\min\{x_1\} = \frac{1}{3}, x_1 = \frac{1}{3}, x_2 = \frac{1}{3}$;

当$x_2 \geq x_1$时, 在可行域上, $\min\{x_2\} = \frac{1}{3}, x_1 = \frac{1}{3}, x_2 = \frac{1}{3}$.

因此$x_1^* = \frac{1}{3}, x_2^* = \frac{1}{3}, f_0^* = \frac{1}{3}$.

(e)梯度为$\nabla f_0 = (2x_1, 18x_2)^T$.

若梯度方向正交于$2x_1+x_2=1$, 有$\frac{2x_1}{2}=\frac{18x_2}{1}$, 得$x_1=\frac{18}{37}, x_2=\frac{251}{37}$, 但不在可行域中.
若梯度方向正交于$x_1+3x_2=1$, 有$\frac{2x_1}{1}=\frac{18x_2}{3}$, 得$x_1=\frac{1}{2}, x_2=\frac{1}{6}$, 在可行域中.
因此最优点$x_1=\frac{1}{2}, x_2=\frac{1}{6}$, 最优值为$\frac{1}{2}$.

习题 7.6　写出下述非线性规划的KKT条件并求解

$$
\begin{aligned}
(1)\quad \texttt{maximize}\quad & f(x)=(x-3)^2\\
\texttt{suject to}\quad & 1\le x\le 5\\
(2)\quad \texttt{minimize}\quad & f(x)=(x-3)^2\\
\texttt{suject to}\quad & 1\le x\le 5
\end{aligned}
$$

(1)解: 原问题等价于

$$
\begin{cases}
\texttt{minimize}\quad -f(x)=(x-3)^2\\
g_1(x)=-x+1\le 5\\
g_2(x)=x-5\le 0
\end{cases}
$$

求目标函数和约束函数的梯度得,

$$\nabla_x f(x)=-2(x-3), \nabla_x g_1(x)=-1, \nabla_x g_2(x)=1$$

将约束引入广义Lagrange乘子v_1, v_2, 在KKT条件上有

$$
\begin{cases}
-2(x^*-3)-v_1^*+v_2^*=0\\
v_1^*(-x^*+1)=0\\
v_2^*(x^*-5)=0\\
v_1^*\ge 0, v_2^*\ge 0
\end{cases}
$$

若$v_1^*\ne 0, v_2^*\ne 0$, 无解.
若$v_1^*=0, v_2^*\ne 0$, 得$x^*=5, v_2^*=4, f(x^*)=-4$.
若$v_1^*\ne 0, v_2^*=0$, 得$x^*=1, v_1^*=4, f(x^*)=-4$.
若$v_1^*=0, v_2^*=0$, 得$x^*=3, f(x^*)=0$.
因此最优点$x^*=1$或$x^*=5$, $\texttt{maximize} f(x)=4$.

(2)解: 原问题等价于

$$\begin{cases} \texttt{minimize} \quad f(x) = (x-3)^2 \\ g_1(x) = -x+1 \leq 5 \\ g_2(x) = x-5 \leq 0 \end{cases}$$

求目标函数和约束函数的梯度得,

$$\nabla_x f(x) = -2(x-3), \nabla_x g_1(x) = -1, \nabla_x g_2(x) = 1$$

将约束引入广义Lagrange乘子v_1, v_2, 在KKT条件上有

$$\begin{cases} 2(x^*-3) - v_1^* + v_2^* = 0 \\ v_1^*(-x^*+1) = 0 \\ v_2^*(x^*-5) = 0 \\ v_1^* \geq 0, v_2^* \geq 0 \end{cases}$$

若$v_1^* \neq 0, v_2^* \neq 0$, 无解.
若$v_1^* = 0, v_2^* \neq 0$, 得$x^* = 5, v_2^* = -4 < 0$, 不是KKT点.
若$v_1^* \neq 0, v_2^* = 0$, 得$x^* = 1, v_1^* = -4 < 0$, 不是KKT点.
若$v_1^* = 0, v_2^* = 0$, 得$x^* = 3, f(x^*) = 0$.
因此最优点$x^* = 3$, $\texttt{minimize} f(x) = 0$.

习题 7.7 考虑等式约束的最小二乘问题

$$\texttt{minimize} \quad \|\mathbf{Ax} - \mathbf{b}\|_2^2$$

$$\texttt{suject to} \quad \mathbf{Gx} = \mathbf{h}$$

其中$\mathbf{A} \in \mathbb{R}^{m\times n}, \mathrm{rank}(\mathbf{A}) = n, \mathbf{G} \in \mathbb{R}^{p\times n}, \mathrm{rank}(\mathbf{G}) = p$. 给出KKT 条件, 推导原问题最优解$x^*$以及对偶问题最优解$v^*$的表达式.

解 求得Lagrangian函数为

$$\begin{aligned} L(x,v) =& \|Ax-b\|_2^2 + v^T(Gx-h) \\ =& x^TA^TAx + (G^Tv - 2A^Tb)^Tx - v^Th \end{aligned}$$

可通过如下最优性条件得到函数最小值. 令梯度为0得,

$$\nabla_x L(x,v) = 2A^T Ax + G^T v - 2A^T b = 0$$

因此当$x = \frac{1}{2}(A^T A)^{-1}(G^T v - 2A^T b)$时, Lagrangian函数取得最小值.
对偶函数为$g(x) = -\frac{1}{4}(G^T v - 2A^T b)^T (A^T A)^{-1}(G^T v - 2A^T b) - v^T h$.
最优性条件为

$$\begin{cases} 2A^T(Ax^* - b) + G^T v^* = 0 \\ Gx^* = h \end{cases}$$

解方程得,

$$\begin{cases} v^* = 2(G(A^T A)^{-1} G^T)^{-1}(G(A^T A)^{-1} A^T b - h) \\ x^* = (A^T A)^{-1}(A^T b - G^T (G(A^T A)^{-1} G^T)^{-1}(G(A^T A)^{-1} A^T b - h)) \end{cases}$$

习题 7.8 用Lagrange乘子法证明：矩阵$\boldsymbol{A} \in \mathbb{R}^{m\times n}$的2范数

$$\|\boldsymbol{A}\|_2 = \max_{\|\boldsymbol{x}\|_2=1, \boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x}\|_2$$

的平方是$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$的最大特征值。

证明 优化问题为

$$\texttt{maximize} \quad f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$$

$$\texttt{suject to} \quad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 1$$

Lagrange函数为:

$$L(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \lambda(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - 1)$$

$$\frac{\partial L}{\partial \boldsymbol{x}} = 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - 2\lambda\boldsymbol{x}$$

令$\frac{\partial L}{\partial \boldsymbol{x}} = 0$，有：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \lambda$$

这表示在$f(\boldsymbol{x})$的极大值点，$\boldsymbol{x}$是$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$的特征向量，$\lambda$是对应的特征值。此时，

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\lambda\boldsymbol{x} = \lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \lambda$$

因此说明，为使$f(\boldsymbol{x})$最大，$f(\boldsymbol{x}) = \lambda_{max}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$，其中$\lambda_{max}$表示最大特征值。即

$$\|\boldsymbol{A}\|_2^2 = \lambda_{max}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$$

习题 7.9 求欠定方程$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$的最小二范数解，其中$\boldsymbol{A} \in \mathbb{R}^{m\times n}, m \le n, \mathrm{rank}(\boldsymbol{A}) = m$

习题 7.10 用最速下降法和精确线搜索计算$\min f(x) = x_1^2 + x_2^2 + x_3^2$ ，初始点$x^{(0)} = (2,2,1)^T$. 当$(f(x^{(n+1)}) - f(x^{(n)})) < 0.001$ 时迭代终止.

解: 由题意得, $f(x) = x^T x, \nabla_x f(x) = 2x$, 设最速下降法的步长为$\lambda$, 那么

$$\begin{aligned} f(x - \lambda\nabla_x f(x)) =&(x - \lambda\nabla_x f(x))^T(x - \lambda\nabla_x f(x)) \\ =&x^T x - 2\lambda\nabla_x f(x)^T x + \lambda^2\nabla_x f(x)^T\nabla_x f(x) \end{aligned}$$

在$x - \lambda\nabla_x f(x)$方向上, 使$f(x)$最小的λ满足

$$\frac{\partial f(x - \lambda\nabla_x f(x))}{\partial\lambda} = -2\nabla_x f(x)^T x + 2\lambda\nabla_x f(x)^T\nabla_x f(x)$$

得

$$\lambda = \frac{\nabla_x f(x)^T x}{\nabla_x f(x)^T\nabla_x f(x)} = \frac{1}{2}$$

所以,

$$x^{(1)} = x^{(0)} - \frac{1}{2}\nabla_x f(x^{(0)}) = (0,0,0)^T$$

$$f(x^{(1)}) = 0$$

$$x^{(2)} = x^{(1)} - \frac{1}{2}\nabla_x f(x^{(1)}) = (0,0,0)^T$$

$$f(x^{(2)}) = 0$$

同理可得, $f(x^{(n)}) = 0 (n > 0)$, 因此当$|(f(x^{(n+1)}) - f(x^{(n)})| = 0 < 0.001$时, 迭代终止.