数据科学与工程数学基础

作业提交规范及第 1 次作业

教师：黄定江
助教：陈诺、刘文辉

2022 年 10 月 13 日

作业提交规范

1. 作业提交形式：**使用 Word 或 LaTeX 编写所得到的电子文档**。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为：“**学号 _ 姓名**”。命名示例：52200000000_ 刘某某。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 1 次作业提交传送门**，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 1 次作业

提交截至时间：**2022/09/28 下周四 12:00（中午）**

理论部分 (范数与二次型)

习题 1. 请证明：对任意 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，由

$$\|\boldsymbol{A}\|_{m_\infty} := \max_{1\le i\le m, 1\le j\le n} |a_{ij}|$$

定义的 $\|\cdot\|_{m_\infty}$ 是 $\mathbb{R}^{m\times n}$ 上的（广义）矩阵范数。

解. 非负性：显然

$$\|\boldsymbol{A}\|_{m_\infty} := \max_{1\le i\le m, 1\le j\le n} |a_{ij}| \ge 0,$$

而且仅当 $\boldsymbol{A}=0$ 时，$\|\boldsymbol{A}\|_{m_\infty}=0$. 齐次性：

$$\|c\cdot\boldsymbol{A}\|_{m_\infty} := \max_{1\le i\le m, 1\le j\le n} |c\cdot a_{ij}| = c \max_{1\le i\le m, 1\le j\le n} |a_{ij}| = c\,\|\boldsymbol{A}\|_{m_\infty}$$

三角不等式：考虑 $\|\boldsymbol{A}+\boldsymbol{B}\|_{m_\infty}$，因为

$$|a_{ij}+b_{ij}| \le |a_{ij}|+|b_{ij}| \le \max_{1\le i\le m, 1\le j\le n} |a_{ij}| + \max_{1\le i\le m, 1\le j\le n} |b_{ij}|$$

所以

$$\max_{1\le i\le m, 1\le j\le n} |a_{ij}+b_{ij}| \le \max_{1\le i\le m, 1\le j\le n} |a_{ij}| + \max_{1\le i\le m, 1\le j\le n} |b_{ij}|$$

即 $\|\boldsymbol{A}+\boldsymbol{B}\|_{m_\infty} \le \|\boldsymbol{A}\|_{m_\infty} + \|\boldsymbol{B}\|_{m_\infty}$

习题 2. 求下面矩阵的 *1*-范数、*2*-范数和无穷范数：

$$A_1 = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} -1 & 0 \\ 1 & 2 \end{pmatrix}.$$

解. $A_1^T A_1 = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$

$\lambda_{max}(A_1^T A_1) = 3+\sqrt{5}$

$\|A_1\|_1 = 2, \|A_1\|_2 = \sqrt{3+\sqrt{5}}, \|A_1\|_\infty = 3$

$A_2^T A_2 = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$

$\|A_2\|_1 = 2, \|A_2\|_2 = \sqrt{3+\sqrt{5}}, \|A_2\|_\infty = 3$

习题 3. 阅读完以下补充材料即可：

将一个带有交叉项的二次型转变成没有交叉项的二次型等同于将对称矩阵合同变换为一个对角矩阵。它的几何意义，可以看作是将二次方程所对应的几何图形标准化的过程。例如：在平面上，即椭圆、抛物线和双曲线等二次曲线的标准化过程；在空间上，即二次曲面的标准化过程。下面通过主轴定理及其几何意义窥见一斑。

- **主轴定理**

定理 0.0.1. 令 A 是任意一个 $n\times n$ 的对称矩阵，那么存在一个正交变换 $x=Py$ 使得二次型 x^TAx 转变为 y^TDy，其中 $D=P^TAP$ 是对角矩阵。

定理中 P 的列称为二次型 x^TAx 的主轴。向量 y 是向量 x 在这些主轴下的坐标。

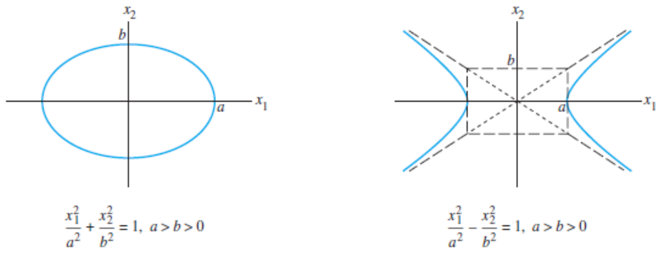
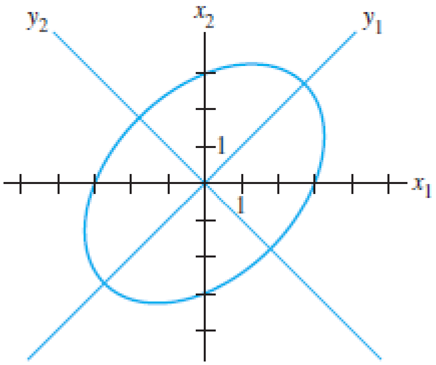


图 1: 对角矩阵时对应标准二次曲线

- **主轴定理的几何意义**

  假设 $Q(x)=x^TAx$，其中 $x\in\mathbb{R}^2$，A 是 2×2 可逆对称矩阵，c 是常数。可以证明 $\mathbb{R}^2$ 中所有满足 $x^TAx=c$ 的 x 的集合可能为椭圆（圆）、双曲线、两相交直线、点或者空集。如果 A 是对角矩阵，那么二次曲线将是标准的二次曲线，如图*1*所示。如果 A 不是对角矩阵，那么二次曲线将是非标准的二次曲线。图*2*是二次方程为 $5x_1^2-4x_1x_2+5x_2^2=48$ 的椭圆



(a) $5x_1^2-4x_1x_2+5x_2^2=48$

图 2: 非对角矩阵对应非标准二次曲线

曲线。显然，它不是一个标准的二次曲线。下面，我们通过寻找该椭圆的主轴将其转化为标准形式，从几何的角度验证主轴定理。

易求得该椭圆方程对应二次型的矩阵表示为

$$A=\begin{pmatrix}5 & -2\\ -2 & 5\end{pmatrix}.$$

可计算出其特征值为 3 和 7，对应的特征向量分别为

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, v_2 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

令 $P = [v_1\ v_2]$，则从图形中可以看出 P 的列向量，正好对应该椭圆主轴所在的方向。现将 $x = Py$ 代入二次型，可得

$$y^T Dy = 3y_1^2 + 7y_2^2.$$

即可得到椭圆的标准方程形式。

习题 4. 矩阵的范数主要包括三种主要类型：诱导范数，元素形式范数和 *Schatten* 范数。诱导范数又称矩阵空间上的算子范数 *(operator norm)*，常用的诱导范数为 *p* 范数, 定义如下

$$\|A\|_p = \sup_{\|x\|_p \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \sup_{\|x\|_p = 1} \|Ax\|_p$$

(1) 设 $A = (a_{ij}) \in C^{m\times n}$, 证明 *1* 范数为列和范数，无穷范数为行和范数

$$\|A\|_\infty = \max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |a_{ij}|, \|A\|_1 = \max_{1\leqslant j\leqslant n} \sum_{i=1}^{m} |a_{ij}|$$

元素形式范数即矩阵按列排成向量，然后采用向量范数的定义得到的矩阵范数，一般称 l_p 范数。

$$l_p : \|A\|_p = \sqrt[P]{\sum_{i,j} |a_{ij}|^P}$$

(2) 试比较 *l1* 范数

$$l_1 : \|A\|_1 = \sum_{i,j} |a_{ij}|^1$$

与诱导范数的关系

解. *(1)*

$$
\begin{aligned}
&A=(a,\ldots,a_n)\\
&\|Ax\|_1=\|\sum_i a_ix_i\|_1\\
&\leqslant\sum_i\|a_ix_i\|_1\\
&=\sum_i\|x_i\|\|a_i\|_1\\
&\leqslant(\max\|a_i\|_1)\left(\sum_i|x_i|\right)\\
&=\max_{1\leq j\leq n}\sum_{i=1}^m|a_{ij}|\,\|x\|_1\\
&\|Ax\|_\infty=\max_i\left|\sum_j a_{ij}x_j\right|\\
&\left|\sum_j a_{ij}x_j\right|\leq\sum_j|a_{ij}x_j|\\
&\leq\sum_j|a_{ij}|\max_j|x_j|\\
&||Ax||_\infty\leqslant\max_{1\leq i\leq m}\sum_{j=1}^n|a_{ij}|\,\|x\|_\infty
\end{aligned}
$$

(2) 以 *l1* 范数与 *1* 范数，无穷范数为例，有

$$\|X\|_1\leqslant\|X\|_1(l1)\leqslant n\|X\|_1$$

$$\|X\|_\infty\leqslant\|X\|_1(l1)\leqslant m\|X\|_\infty$$

习题 5. 阅读完以下补充材料即可：

由 $\|A\|_p=\sup_{\|x\|_p\neq0}\frac{\|Ax\|_p}{\|x\|_p}$,
矩阵 A 的诱导范数可理解为线性变换 Ax 对向量 x 的最大“拉长倍数”
由 $\|A\|_p=\sup_{\|x\|_p=1}\|Ax\|_p$,
矩阵 A 的诱导范数也可理解为 x 在单位范数球上运动时 $\|Ax\|_p$ 的最大值
以下情形可便于理解诱导范数
例: 左图为 $A=\begin{pmatrix}1&1\\1&1\end{pmatrix}$, $p(A)=1,p=2$ 时的情形，在 $x=(\sqrt{2}/2,\sqrt{2}/2)$ 时 $\|Ax\|_2$ 取到最大值 *2*

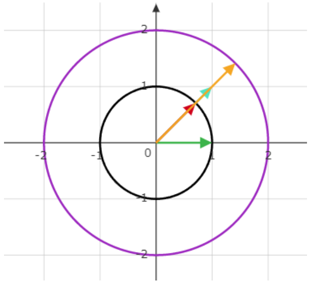


图 3: 诱导范数例 (p=2)

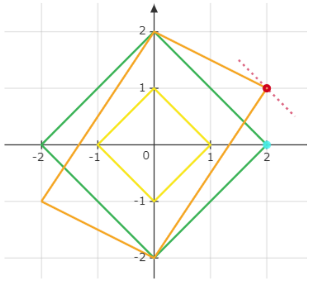


图 4: 诱导范数例 (p=1)

例: 右图绿线为 $A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ *,$p=1$ 时的情形，在 $x=(1,0)$ 时 $\|Ax\|_1$ 取到最大值 2*

例: 右图橙线为 $A = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$ *,$p=1$ 时的情形，在 $x=(1,0)$ 时 $\|Ax\|_1$ 取到最大值 3*

(1) 试证明

$$||A||_2 = \sigma_{max} = \sqrt{\lambda\left(A^\top A\right)}$$

, 其中 σ_{max} 为谱范数，即矩阵 A 的最大奇异值，$\lambda\left(A^\top A\right)$ 表示 $A^\top A$ 的最大特征值。

证: 即证明对于矩阵 $A_{m\times n}$, 对任意向量 x, 在矩阵 A 的变换 (即 Ax) 后, 其长度不大于 $\sigma_{max}\|x\|_2$, 即 $\|Ax\|_2 \le \sigma_{max}\|x\|_2$ 。

$A^\top A$ 是实对称阵，不同特征值对应的特征向量两两正交。不妨令 $B = A^\top A$, 特征向量矩阵为 $\lambda = diag(\lambda_1,...,\lambda_n) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$,$p_1,...,p_n$ 为 B 的一组标准正交特征向量，则 $P = (p_1,...,p_n)$ 为正交矩阵, 故

$$BP = P\lambda \Leftrightarrow B = P\lambda P^{-1} \Leftrightarrow B = P\lambda P^\top$$

(称 λ 合同于 B)

假设对一个向量 x, 在矩阵 A 的变换（即 Ax ）后得到 y, 即满足 $y = Ax$。则

$$\|y\|_2^2 = y^\top y = (Ax)^\top(Ax) = x^\top A^\top Ax = x^\top Bx = x^\top P\lambda P^\top x = (P^\top x)^\top \lambda P^\top x$$

对于二次型，可以看作是一个二次齐次多项式的图形，而正交矩阵 P 的变换，可以保证变换图形的形状和大小不变，仅仅做了位移、旋转或翻转的变换，类似把物体从一个地方移到另一个地方。(可以想象一个三维坐标系，在坐标系上的点构成的图形通过一个非正交的基表示，现坐标系换了一组标准正交基 P，用这组基变换图形不过是移动 *(掰正)* 了图形的位置。记 $z = P^\top x$,

所以 z 不过是一个与 x 一样的（同范数的）向量，只是换了位置，而 $\lambda = diag(\lambda_1, ..., \lambda_n)$ 则进行了掰正位置后的放缩。因而

$$
\begin{aligned}
|y|^2 = z^T \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} z &= \sum_i \lambda_i z_i^2 \\
&= \sum_i \left(\sqrt{\lambda_i} z_i\right)^2 \leq \sum_i \left(\max_j \left(\sqrt{\lambda_j}\right) z_i\right)^2 \\
&= \max \left(\sqrt{\lambda_j}\right)^2 \sum_i z_i^2 = \max\left(\lambda_j\right) |z|^2 == \max\left(\lambda_j\right) |x|^2
\end{aligned}
$$

当且仅当除 $z_{opt\,\max_j(\lambda_j)}$ 以外的其他元素均等于 *0* 时，该不等式的等号成立。

即 $\|Ax\|_2 \leq \sigma_{max}\|x\|_2$ ，证毕

(2) 元素形式下矩阵的 l_2 范数称为 *Frobenius* 范数，即

$$
l_2 : \|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}
$$

试比较 $\|A\|_2$ 与 $\|A\|_F$ 的大小

答:$\|A\|_2 \leq \|A\|_F$

数据科学与工程数学基础

作业提交规范及第 2 次作业

教师：黄定江

助教：陈诺、刘文辉

2022 年 9 月 30 日

作业提交规范

1. 作业提交形式：**使用 Word 或 LaTeX 编写所得到的电子文档**。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为："**学号 _ 姓名**"。命名示例：52200000000_ 刘某某。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 2 次作业提交传送门**，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 2 次作业

提交截至时间：**2022/10/14 周五 12:00（中午）**

理论部分 (范数与二次型)

习题 1. 假设 $M, P \in \mathbb{R}^{n\times n}$ 为对称阵，P 为正交阵，

$$A = \left(\begin{array}{c|c} M & PM \\ \hline MP & PMP \end{array}\right) \in \mathbb{R}^{2n\times 2n}.$$

(i) 证明 $A^{\mathrm{T}} = A$.

(ii) 假设 $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$ 是正交矩阵,$D \in \mathbb{R}^{m\times n}$, 证明 $\|UDV\|_2 = \|D\|_2, \|UDV\|_F = \|D\|_F$.

(iii) 证明 $\|A\|_F = 2\|M\|_F$. $\|A\|_2 \le 2\|M\|_2$. (提示：将 A 分解，并利用 *(ii)* 结论)

(iv) 假设 $n = 4, M = \mathrm{diag}_{4\times 4}(-2, 1, 0, 0), P = (e_4|e_3|e_2|e_1)$. 证明 $\|A\|_p = 2\ \forall p \in [1, \infty)$.

解. *(i)* 解法 1：分解 A

$$A = \begin{pmatrix} I \\ P \end{pmatrix} M \begin{pmatrix} I & P \end{pmatrix} = \begin{pmatrix} I & \\ & P \end{pmatrix} \begin{pmatrix} M & M \\ M & M \end{pmatrix} \begin{pmatrix} I & \\ & P \end{pmatrix} = A^{\mathrm{T}}$$

解法 2：不分解 A

$$A^{\mathrm{T}} = \left(\begin{array}{c|c} M^{\mathrm{T}} & (MP)^{\mathrm{T}} \\ \hline (PM)^{\mathrm{T}} & (PMP)^{\mathrm{T}} \end{array}\right) = \left(\begin{array}{c|c} M^{\mathrm{T}} & P^{\mathrm{T}}M^{\mathrm{T}} \\ \hline M^{\mathrm{T}}P^{\mathrm{T}} & P^{\mathrm{T}}M^{\mathrm{T}}P^{\mathrm{T}} \end{array}\right) = \left(\begin{array}{c|c} M & PM \\ \hline MP & PMP \end{array}\right) = A$$

(ii) 即证明 2 范数与 F 范数满足正交不变性

对 2 范数，即证 $\|UD\|_2 = \|D\|_2, \|DV\|_2 = \|D\|_2$

$$\|UD\|_2 = \sqrt{\lambda_{\max}\left(D^{\mathrm{T}}U^{\mathrm{T}}UD\right)} = \sqrt{\lambda_{\max}\left(D^{\mathrm{T}}D\right)} = \|D\|_2$$

又因为 $\|DV\|_2 = x^{\mathrm{T}}V^{\mathrm{T}}Vx = x^{\mathrm{T}}x = \|x\|_2$

$$\|DV\|_2 = \sup_{\|x\|_2=1} \|DVx\|_2 = \sup_{\|Vx\|_2=1} \|DVx\|_2 = \|D\|_2$$

对 F 范数，

$$\|UDV\|_F = \sqrt{\mathrm{tr}\left(V^{\mathrm{T}}D^{\mathrm{T}}U^{\mathrm{T}}UDV\right)} = \sqrt{\mathrm{tr}\left(VV^{\mathrm{T}}D^{\mathrm{T}}D\right)} = \sqrt{\mathrm{tr}\left(V^{\mathrm{T}}VD^{\mathrm{T}}D\right)} = \sqrt{\mathrm{tr}\left(D^{\mathrm{T}}D\right)} = \|D\|_F$$

(iii) 解法 1：分解 A，由于 $\begin{pmatrix} I & \\ & P \end{pmatrix}$ 为正交阵，由 *(ii)* 得

$$\|A\|_F = \left\|\begin{pmatrix} M & M \\ M & M \end{pmatrix}\right\|_F \qquad \|A\|_2 = \left\|\begin{pmatrix} M & M \\ M & M \end{pmatrix}\right\|_2$$

$$\|A\|_F^2 = \|M\|_F^2 + \|M\|_F^2 + \|M\|_F^2 + \|M\|_F^2 = 4\|M\|_F^2$$

或

$$
\begin{aligned}
\|A\|_F &= \left\|\begin{pmatrix} M & M \\ M & M \end{pmatrix}\right\|_F = \left\|\begin{pmatrix} I \\ I \end{pmatrix} M \begin{pmatrix} I & I \end{pmatrix}\right\|_F \\
&= \sqrt{\operatorname{tr}\left(\left(\begin{pmatrix} I \\ I \end{pmatrix} M \begin{pmatrix} I & I \end{pmatrix}\right)\begin{pmatrix} I \\ I \end{pmatrix} M \begin{pmatrix} I & I \end{pmatrix}\right)} \\
&= \sqrt{2\operatorname{tr}\left(M \begin{pmatrix} I & I \end{pmatrix}\begin{pmatrix} I \\ I \end{pmatrix} M\right)} \\
&= 2\sqrt{\operatorname{tr}\left(M^2\right)} = 2\|M\|_F
\end{aligned}
$$

记$B = \begin{pmatrix} M & M \\ M & M \end{pmatrix}$, $w = \begin{pmatrix} x \\ y \end{pmatrix}$, $\|A\|_2 = \|B\|_2 = \sup_{\|w\|_2 \neq 0} \frac{\|Bw\|_2}{\|w\|_2} = \sup_{\|w\|_2 = 1} \|Bw\|_2$

$$
\begin{aligned}
\left\|B\begin{pmatrix} x \\ y \end{pmatrix}\right\|_2^2 &= \left\|\begin{pmatrix} Mx + My \\ Mx + My \end{pmatrix}\right\|_2^2 \\
&= 2\|Mx + My\|_2^2 \\
&\leqslant 2\left(\|Mx\|_2 + \|My\|_2\right)^2 \\
&\leq 4\|Mx\|_2^2 + 4\|My\|_2^2 \\
&\leq 4\|M\|_2^2\|x\|_2^2 + 4\|M\|_2^2\|y\|_2^2 \text{（相容性）} \\
&= 4\|M\|_2^2\|w\|_2^2 = 4\|M\|_2^2
\end{aligned}
$$

因此 $\|A\|_2 = \sup_{\|w\|_2 = 1} \|Bw\|_2 = 2\|M\|_2$

解法 2：不分解 A

$$
\|A\|_F^2 = \|M\|_F^2 + \|MP\|_F^2 + \|PM\|_F^2 + \|PMP\|_F^2 = 4\|M\|_F^2
$$

$$
\begin{aligned}
\|Aw\|_2^2 &= \left\|\begin{pmatrix} Mx + PMy \\ MPx + PMPy \end{pmatrix}\right\|_2^2 \\
&= \|Mx + PMy\|_2^2 + \|MPx + PMPy\|_2^2 \\
&\leq \left(\|Mx\|_2 + \|PMy\|_2\right)^2 + \left(\|MPx\|_2 + \|PMPy\|_2\right)^2 \\
&\leq 2\|Mx\|_2^2 + 2\|PMy\|_2^2 + 2\|MPx\|_2^2 + 2\|PMPy\|_2^2 \\
&\leq 2\|M\|_2^2\|x\|_2^2 + 2\|PM\|_2^2\|y\|_2^2 + 2\|MP\|_2^2\|x\|_2^2 + 2\|PMP\|_2^2\|y\|_2^2 \\
&= 2\|M\|_2^2\|x\|_2^2 + 2\|M\|_2^2\|y\|_2^2 + 2\|M\|_2^2\|x\|_2^2 + 2\|M\|_2^2\|y\|_2^2 \\
&= 4\|M\|_2^2\left(\|x\|_2^2 + \|y\|_2^2\right) \\
&= 4\|M\|_2^2\|w\|_2^2
\end{aligned}
$$

(iv)

$$M=\begin{pmatrix}-2&&&\\&1&&\\&&0&\\&&&0\end{pmatrix}\quad P=\begin{pmatrix}&&&1\\&&1&\\&1&&\\1&&&\end{pmatrix}\quad A=\left(\begin{array}{cccc|cccc}-2&&&&&&&0\\&1&&&&&0&\\&&0&&&1&&\\&&&0&-2&&&\\\hline&&&-2&0&&&\\&&1&&&0&&\\&0&&&&&1&\\0&&&&&&&-2\end{array}\right)$$

$$\begin{aligned}\|Ax\|_p^p&=\left\|\left(-2x_1,x_2,x_6,-2x_5,-2x_4,x_3,x_7,-2x_8\right)^{\mathrm{T}}\right\|_p^p\\&=|-2x_1|^p+|x_2|^p+|x_6|^p+|-2x_5|^p+|-2x_4|^p+|x_3|^p+|x_7|^p+|-2x_8|^p\\&=2^p\,|x_1|^p+|x_2|^p+|x_3|^p+2^p\,|x_4|^p+2^p\,|x_5|^p+|x_6|^p+|x_7|^p+2^p\,|x_8|^p\\&\le 2^p\left(|x_1|^p+|x_2|^p+|x_3|^p+|x_4|^p+|x_5|^p+|x_6|^p+|x_7|^p+|x_8|^p\right)\\&=2^p\|x\|_p^p.\end{aligned}$$

因此 $\|A\|_p=\sup_{\|x\|_p=1}\|Ax\|_p=2$

(即如果一个变换只将某些维度倍乘并交换顺序，不作维度间相加的操作，那矩阵范数即最大拉伸倍数)

习题 2. 假设 $P\in\mathbb{R}^{n\times n}\backslash\{0\}$ 是一个投影矩阵.

(i) 证明 $Py=y\ \forall y\in\mathcal{R}(P)$. $Px-x\in\mathcal{N}(P)\ \forall x\in\mathbb{R}^n$.

(即证明投影 P 沿着零空间 $\mathcal{N}(P)$ 投影到列空间 $\mathcal{R}(P)$)

(ii) 证明 P 的特征值 $\lambda\in\Lambda(P)\subseteq\{0,1\}$. 假设 $\mathcal{R}(P)=\operatorname{span}(u_1,\ldots,u_r)$, $\mathcal{N}(P)=\operatorname{span}(v_{r+1},\ldots,v_n)$，试找到 P 的特征分解 $P=XDX^{-1}$ 并证明 $\operatorname{tr}(P)=\operatorname{rank}(P)$. *(提示：利用 (i) 结论.)*

(iii) 证明当 $P\neq I_n$，$\det(P)=0$.

(iv) 证明当 P 是正交投影矩阵（$P^2=P=P^{\mathrm{T}}$）时，I_n-2P 是正交矩阵.

(v) 假设 $A\in\mathbb{R}^{n\times m}$，$m\le n$，$\operatorname{rank}(A)=m$. $P=A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}$ 证明 P 是正交投影矩阵，$\operatorname{rank}(P)=m$. *(提示：利用 (ii) 结论.)*

解. *(i)* $\forall\, y\in\mathcal{R}(P)$ 即对 $x\in\mathbb{R}^n$，$y=Px$，$Py=P^2x=Px=y$.

$\forall x\in\mathbb{R}^n$, $P(Px-x)=P^2x-Px=Px-Px=0$.

(ii) 对 $\lambda\in\Lambda(P)$,$x\in\mathbb{R}^n\backslash\{0\}$, 有 $Px=\lambda x$. 由于 $P=P^2$, $\lambda x=Px=P(Px)=P(\lambda x)=\lambda Px=\lambda^2x$. 因为 $x\neq0\in\mathbb{R}^n$, 故 $\lambda=\lambda^2$, $\lambda\in\{0,1\}$. 因此 $\Lambda(P)\subseteq\{0,1\}$.

由 *(i)* 可知，$\forall i=1,...,r, u_i\in\mathcal{R}(P), Pu_i=u_i.\forall j=r+1,...,n, v_j\in\mathcal{N}(P), Pv_j=0$.

故令 $X:=(u_1|\cdots|u_r|v_{r+1}|\cdots|\,v_n)\in\mathbb{R}^{n\times n}$,$D:=\operatorname{diag}_{n\times n}(\underbrace{1,\ldots,1}_{r\ times},0\ldots,0)\in\mathbb{R}^{n\times n}$，此时

$P = XDX^{-1}$

(注：也可理解为 *SVD(*后续课程会讲*)*，即 $P = UDV^{\mathrm{T}}, U \in \mathcal{R}(P), V \in \mathcal{N}(P)$)

$\operatorname{tr}(P) = \operatorname{tr}\left(XDX^{-1}\right) = \operatorname{tr}(D) = r.$

(iii) 反证 $\det(P) \neq 0 \Longrightarrow P = I_n$. 由于 $\det(P) \neq 0$, P 可逆. 故由 $P^2 = P$，得 $P^{-1}P^2 = P^{-1}P$, $P = I_n$.

(iv) 由于 P 是正交投影矩阵, $P^2 = P = P^{\mathrm{T}}$. 令 $Q := I_n - 2P$,$Q^{\mathrm{T}} = I_n - 2P^{\mathrm{T}} = Q$, $Q^2 = I_n - 4P + 4P^2 = I_n$. 因此, $Q^{\mathrm{T}}Q = QQ^{\mathrm{T}} = I_n$.

(v)$P^2 = A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}} = A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}} = P$

$P^{\mathrm{T}} = A\left(\left(A^{\mathrm{T}}A\right)^{-1}\right)^{\mathrm{T}}A^{\mathrm{T}} = A\left(\left(A^{\mathrm{T}}A\right)^{\mathrm{T}}\right)^{-1}A^{\mathrm{T}} = A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}} = P.$

由 *(ii)*, $\operatorname{rank}(P) = \operatorname{tr}(P) = \operatorname{tr}(A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}) = \operatorname{tr}(\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}A) = \operatorname{tr}(I_m) = m.$

数据科学与工程数学基础

作业提交规范及第 3 次作业

教师：黄定江

助教：陈诺、刘文辉

2022 年 11 月 2 日

作业提交规范

1. 作业提交形式：**使用 Word 或 LaTeX 编写所得到的电子文档**。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为：“**学号 _ 姓名**”。命名示例：52200000000_ 刘某某。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 3 次作业提交传送门**，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 3 次作业

提交截至时间：**2022/10/21 周五 12:00（中午）**

理论部分 (范数与二次型)

习题 1. 证明：（**关联矩阵的四个子空间性质**）设图 $G=<\mathbb{V},\mathbb{E}>$ 是一个具有 m 个顶点和 n 条边的连通图，其对应的关联矩阵为 $\boldsymbol{B}$，则关于 $\boldsymbol{B}$ 的四个基本子空间具有以下性质：

- $\boldsymbol{B}$ 的左零空间维数为 1，即 $Null\left(\boldsymbol{B}^{\mathrm{T}}\right)=1$，且 $Null\left(\boldsymbol{B}^{\mathrm{T}}\right)=\mathbf{span}\{\mathbf{1}\}$。
- $\boldsymbol{B}$ 的行空间维数为 $m-1$，即 $Col\left(\boldsymbol{B}^{\mathrm{T}}\right)=m-1$，且 $Col\left(\boldsymbol{B}^{\mathrm{T}}\right)$ 可由 $\boldsymbol{B}^{\mathrm{T}}$ 任意 $m-1$ 个列向量生成。
- $\boldsymbol{B}$ 的列空间维数为 $m-1$，即 $Col\left(\boldsymbol{B}\right)=m-1$，且若 T 是图 G 的一棵生成树，那么 $Col\left(\boldsymbol{B}\right)$ 可由 T 的关联矩阵的 $m-1$ 个列向量生成。
- $\boldsymbol{B}$ 的零空间维数为 $n-m+1$，即 $Null\left(\boldsymbol{B}\right)=n-m+1$，这个数等于图 G 中小圈的个数。

解. 因为 $\boldsymbol{B}^{\mathrm{T}}$ 的行恰好对应有向图的边，因此，这些行线性相关等价于这些边是否存在环。又因为一个顶点为 m 的连通图，不形成环当且仅当其边数为 $m-1$。因此，$\boldsymbol{B}^{\mathrm{T}}$ 的秩为 $m-1$。

- 易知 $Null\left(\boldsymbol{B}^{\mathrm{T}}\right)=m-(m-1)=1$。易知，每一行仅有一个元素恰为 *1*，一个元素恰为 *-1*. 因此，$\boldsymbol{B}^{\mathrm{T}}\mathbf{1}=\mathbf{0}$。所以 $Null\left(\boldsymbol{B}^{\mathrm{T}}\right)=\mathbf{span}\{\mathbf{1}\}$。
- 易知 $Col\left(\boldsymbol{B}^{\mathrm{T}}\right)=m-1$ 成立。下面利用反证法说明 $\boldsymbol{B}^{\mathrm{T}}$ 任意 $m-1$ 个列向量线性无关。假设 $\boldsymbol{B}^{\mathrm{T}}$ 的 m 个列向量分别为 $b_1,\cdots,b_m$。不妨设前 $m-1$ 个列向量线性相关，且 $b_1=k_2b_2+\cdots+k_{m-1}b_{m-1}$。从而，$(-1,k_2,\cdots,k_{m-1},0)^{\top}\in Null\left(\boldsymbol{B}^{\mathrm{T}}\right)$。这与第一问的结论矛盾，故得证。（或者不利用第一问的结论去证：又因为根据上述左零空间的性质知 $b_m=-b_1-\cdots-b_{m-1}$。所以 $b_m=-(k_2+1)b_2-\cdots-(k_{m-1}+1)b_{m-1}$。因此，秩小于等于 $m-2$，矛盾。故得证。）
- 易知 $Col\left(\boldsymbol{B}\right)=m-1$ 成立。若 T 是图 G 的一棵生成树，则加入 G 中的其他的任一条边，必然会形成一个环。因此，$\boldsymbol{B}$ 中除 T 的关联矩阵的 $m-1$ 列向量以外的其他列向量，可由这 $m-1$ 个列向量生成。
- 易知 $Null\left(\boldsymbol{B}\right)=n-(m-1)=n-m+1$。小圈的数目恰好是图变成一棵生成树所需删除边的数目，即 $n-(m-1)$。因此，相等。

习题 2. 求向量 $(1,1,1)^T$ 投影到一维子空间 $span\{(1,-1,1)^T\}$ 的正交投影。

解. 首先求得投影矩阵

$$P_\pi=\frac{1}{3}\begin{pmatrix}1&-1&1\\-1&1&-1\\1&-1&1\end{pmatrix}$$

向量 $(1,1,1)^T$ 投影到一维子空间 $span\{(1,-1,1)^T\}$ 的正交投影为

$$P_\pi\begin{pmatrix}1\\1\\1\end{pmatrix}=\frac{1}{3}\begin{pmatrix}1&-1&1\\-1&1&-1\\1&-1&1\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}=\frac{1}{3}\begin{pmatrix}1\\-1\\1\end{pmatrix}$$

习题 3. 完成以下阅读材料即可。

- **为什么需要 *LU* 分解？**

  为什么不直接求解方程组 $Ax=b$？如果了解 LU 分解所需要的计算量与高斯消去法解方程组所需的计算量，便可知计算复杂度均为 $O(n^3)$。似乎 LU 分解并没有任何优势。然而，在实际的工程问题中，我们经常会遇到的是这样的问题，需要求解这样一系列的方程组（例如时序的）：

  $$Ax=b_1,\ Ax=b_2,\ \cdots,\ Ax=b_r.$$

  其中的系数矩阵是不变的，只是右边的常数项在改变。这时，如果直接使用高斯消去法求解各个方程组的话，则对应的计算复杂度为 $O(rn^3)$。若使用 LU 分解，则只需要的运算量级为 $O(n^3+rn^2)$，这是因为 LU 分解在解方程组时只需要 $O(n^2)$。

  或许，你会提出直接计算出 A^{-1}，然后再利用它去求解各个方程组。这时所需要的计算量的效果似乎与 LU 分解是相似的。事实上，在实际计算中往往不会直接计算矩阵的逆。这里给出其中一个原因：通常实际工程中矩阵 A 的维度很大，但有一个优点是它是稀疏的，例如一个"带状的"矩阵。使用 LU 分解，可以保持矩阵的稀疏性（因此，还可以提高运算速度）。然而，直接求解矩阵的逆，则会丢失矩阵的稀疏性。因此，在数据的存储上直接求逆存在明显的劣势。

- ***Gauss* 消去与 *LU* 分解**

  还记得初中学过的求解二元一次方程组的消元法吗？这就是 LU 分解与 *Cholesky* 分解的全部。假设我们有如下二元一次方程组

  $$\begin{cases} 2x_1+4x_2=5 \\ 4x_1+5x_2=3 \end{cases}$$

  现在我们考虑方程组的求解。在中学时，我们会通过将第一个等式乘以 -2 加到第二个等式，则得到

  $$\begin{cases} 2x_1+4x_2=5 \\ 0x_1-3x_2=-7 \end{cases}$$

  这个行变换对应的矩阵为 $\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$，它的逆矩阵就是 $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$。另外，变换后的系数矩阵为 $\begin{pmatrix} 2 & 4 \\ 0 & -3 \end{pmatrix}$。

  如果我们使用 LU 分解，则可得到

  $$\begin{pmatrix} 2 & 4 \\ 4 & 5 \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 4 \\ 0 & -3 \end{pmatrix}$$

从上可以看出，高斯消去的变换矩阵的逆恰好对应 LU 分解的 L 矩阵，变换后的系数矩阵恰好对应 U 矩阵。

习题 4. 对矩阵 $\begin{pmatrix} 5 & 3 & -1 & 3 \\ 0 & 1 & 1 & -2 \\ -5 & -3 & 4 & -4 \\ 0 & 1 & 1 & 0 \end{pmatrix}$ 进行 LU 分解。

解. （具体计算过程可参考教材或课件中的例题。）

$$\boldsymbol{A} = \begin{pmatrix} 5 & 3 & -1 & 3 \\ 0 & 1 & 1 & -2 \\ -5 & -3 & 4 & -4 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 & 3 & -1 & 3 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 3 & -1 \\ 0 & 0 & 0 & 2 \end{pmatrix} = \boldsymbol{LU}$$

习题 5. 设 A 对称且 $a_{11} \neq 0$, 并假经过一步 *Gauss* 消去之后, A 具有如下形式

$$\begin{bmatrix} a_{11} & a_1^{\mathrm{T}} \\ \mathbf{0} & A_2 \end{bmatrix}$$

证明 A_2 仍是对称阵.

解. 记矩阵 A 和高斯变换矩阵分别为：

$$A = \begin{bmatrix} a_{11} & a_1^{\mathrm{T}} \\ a_1 & A_1 \end{bmatrix}, G = \begin{bmatrix} 1 & \mathbf{0}^{\mathrm{T}} \\ l_1 & I \end{bmatrix}$$

则

$$GA = \begin{bmatrix} a_{11} & a_1^{\mathrm{T}} \\ -a_{11}l_1 + a_1 & -l_1a_1^{\mathrm{T}} + A_1 \end{bmatrix}$$

易知 $-a_{11}l_1 + a_1 = \mathbf{0}$ 以及 $A_2 = -l_1a_1^{\mathrm{T}} + A_1$。由第一个等式可得 $l_1 = 1/a_{11}a_1$，代入第二个等式知

$$A_2 = -\frac{1}{a_{11}}a_1a_1^{\mathrm{T}} + A_1.$$

故可知 A_2 仍是对称矩阵。

习题 6. 证明上三角矩阵与上三角矩阵的乘积仍是上三角矩阵。

解. 假设 A, B 为上三角矩阵，且其乘积为 C。下证 C 为上三角矩阵，只需证 $C_{ij}, (i > j)$ 时为 0。易知，

$$C_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} = \sum_{k=1}^{j} A_{ik}B_{kj} + \sum_{k=j+1}^{n} A_{ik}B_{kj}$$

因为 A, B 为上三角矩阵，所以右边第一项为中 $A_{ik} = 0$，右边第二项中 $B_{kj} = 0$。故得证。

习题 7. 用 *Householder* 方法求矩阵 $A=\begin{bmatrix}1&1\\2&0\\2&1\end{bmatrix}$ 的 QR 分解。

解. 令 $\alpha_1=(1,2,2)^T$，$a_1=\|\alpha_1\|_2=3$，则

$$w_1=\frac{\alpha_1-a_1\boldsymbol{e}_1}{\|\alpha_1-a_1\boldsymbol{e}_1\|_2}=\frac{1}{2\sqrt{3}}(-2,2,2)^T$$

故有

$$H_1=I-2w_1w_1^T=\begin{bmatrix}1/3&2/3&2/3\\2/3&1/3&-2/3\\2/3&-2/3&1/3\end{bmatrix}$$

此时，

$$H_1A=\begin{bmatrix}3&1\\0&0\\0&1\end{bmatrix}$$

再令 $\beta_1=(0,1)^T$，$b_1=\|\beta_1\|_2=1$，则

$$w_2=\frac{\beta_1-b_1\boldsymbol{e}_1}{\|\beta_1-b_1\boldsymbol{e}_1\|_2}=\frac{1}{\sqrt{2}}(-1,1)^T$$

故有

$$\hat{H}_2=I-2w_2w_2^T=\begin{bmatrix}0&1\\1&0\end{bmatrix}$$

记

$$H_2=\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}$$

此时

$$H_2H_1A=\begin{bmatrix}3&1\\0&1\\0&0\end{bmatrix}\triangleq R$$

因此，$A=QR$，其中 $Q=H_1^TH_2^T$。

数据科学与工程数学基础

作业提交规范及第 4 次作业

教师：黄定江
助教：陈诺、刘文辉

2022 年 10 月 22 日

作业提交规范

1. 作业提交形式：**使用 Word 或 LaTeX 编写所得到的电子文档**。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为："**学号 _ 姓名**"。命名示例：10175501112_ 陈诺。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 4 次作业提交传送门**，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 4 次作业

提交截至时间：**2022/10/28 周五 12:00（中午）**

理论部分

习题 1. 定义

$$A=\begin{pmatrix} 2 & -2 & 1 \\ -1 & 1 & -1 \\ -3 & -1 & 1 \end{pmatrix}$$

(i) 给出矩阵 $A^{\mathrm{T}}A$ 的 *Cholesky* 分解 $A^{\mathrm{T}}A = GG^{\mathrm{T}}$

(ii) 试说明 $\|A^{\mathrm{T}}A\|_2 = \|A\|_2^2 = \|G\|_2^2$

解. *(i)* 记

$$M = A^{\mathrm{T}}A = \begin{pmatrix} 14 & -2 & 0 \\ -2 & 6 & -4 \\ 0 & -4 & 3 \end{pmatrix}$$

消除 M 的第一列中的对角线条目

$$L_1M = \begin{pmatrix} \sqrt{14} & -\frac{2}{\sqrt{14}} & 0 \\ 0 & \frac{40}{7} & -4 \\ 0 & -4 & 3 \end{pmatrix} \qquad L_1 = \begin{pmatrix} \frac{1}{\sqrt{14}} & 0 & 0 \\ \frac{1}{7} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

L_1M 右乘 L_1^{T}

$$L_1ML_1^{\mathrm{T}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{40}{7} & -4 \\ 0 & -4 & 3 \end{pmatrix}$$

消除 $L_1ML_1^{\mathrm{T}}$ 的第二列中的对角线条目

$$L_2L_1ML_1^{\mathrm{T}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{2\sqrt{70}}{7} & -\frac{\sqrt{70}}{5} \\ 0 & 0 & \frac{1}{5} \end{pmatrix} \qquad L_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{70}}{20} & 0 \\ 0 & \frac{7}{10} & 1 \end{pmatrix}$$

L_1M 右乘 $L_2L_1ML_1^{\mathrm{T}}L_2^{\mathrm{T}}$

$$L_2L_1ML_1^{\mathrm{T}}L_2^{\mathrm{T}} = \mathrm{diag}_{3\times 3}\left(1, 1, \frac{1}{5}\right).$$

令 $L_3 := \mathrm{diag}_{3\times 3}(1, 1, \sqrt{5})$ 使得 $L_3L_2L_1ML_1^{\mathrm{T}}L_2^{\mathrm{T}}L_3^{\mathrm{T}} = I_3$. 我们有 $M = A^{\mathrm{T}}A = GG^{\mathrm{T}}$ 其中

$$G = L_1^{-1}L_2^{-1}L_3^{-1} = \begin{pmatrix} \sqrt{14} & 0 & 0 \\ -\frac{\sqrt{14}}{7} & \frac{2\sqrt{70}}{7} & 0 \\ 0 & -\frac{\sqrt{70}}{5} & \frac{\sqrt{5}}{5} \end{pmatrix}$$

(ii) 令 $G = U\Sigma V^{\mathrm{T}}$ ，其中 $U, V \in \mathbb{R}^{n\times n}$ 正交，$\Sigma = \mathrm{diag}_{n\times n}(\sigma_1, \ldots, \sigma_n) \in \mathbb{R}^{n\times n}$ 且 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. 故 $A^{\mathrm{T}}A = GG^{\mathrm{T}} = U\Sigma V^{\mathrm{T}}V\Sigma U^{\mathrm{T}} = U\Sigma^2U^{\mathrm{T}}$ ，$A^{\mathrm{T}}A$ 的奇异值为 $\sigma_1^2, \ldots, \sigma_n^2$. 因此 $\|A^{\mathrm{T}}A\|_2 = \sigma_1^2 = \|G\|_2^2$. 同样的，我们可以得出 $\|A^{\mathrm{T}}A\|_2 = \|A\|_2^2$.

习题 2. 对 $k \in \mathbb{N}_0$, 定义

$$A=\begin{pmatrix}-8&5&1\\-4&7&5\\-8&5&1\\-4&7&5\end{pmatrix},\quad \gamma_k=\inf_{\substack{M\in\mathbb{R}^{3\times 4}\\ \mathrm{rk}(M)\le k}}\left\|A^{\mathrm{T}}-M\right\|_2$$

(i) 计算矩阵 A 的 *SVD* 分解 $A=U\Sigma V^{\mathrm{T}}$, 并使 $2U$ 为 *Hadamard* 矩阵

(ii) 使用 *(i)* 中的结论, 求 $\operatorname{rank}(A), \mathcal{R}(A), \mathcal{N}(A), \|A\|_2, \|A\|_F$

(iii) 对每个 $k\in\mathbb{N}_0$, 计算 γ_k 并找出矩阵 $A_k\in\mathbb{R}^{3\times 4}$ 使得 $\operatorname{rank}(A_k)\le k$ 且 $\left\|A^{\mathrm{T}}-A_k\right\|_2=\gamma_k$

解. *(i)*

$$A^{\mathrm{T}}A=4\begin{pmatrix}40&-34&-14\\-34&37&20\\-14&20&13\end{pmatrix}$$

特征多项式 $p_{A^{\mathrm{T}}A}(z)=\det\left(zI_3-A^{\mathrm{T}}A\right)=z(z-36)(z-324)$. $A^{\mathrm{T}}A$ 特征值为

$$\lambda_1=324,\quad \lambda_2=36,\quad \lambda_3=0$$

$\lambda_i, i\in\{1,2,3\}$ 对应 $A^{\mathrm{T}}A$ 的特征空间 $E_{\lambda_i}=\mathcal{N}\left(\lambda_i I_3-A^{\mathrm{T}}A\right)$ 分别为 $E_{\lambda_1}=\operatorname{span}\left((-2,2,1)^{\mathrm{T}}\right), E_{\lambda_2}=\operatorname{span}\left((2,1,2)^{\mathrm{T}}\right), E_{\lambda_3}=\operatorname{span}\left((1,2,-2)^{\mathrm{T}}\right)$. 取归一化特征向量

$$v_1=\frac{1}{3}(-2,2,1)^{\mathrm{T}},\quad v_2=\frac{1}{3}(2,1,2)^{\mathrm{T}},\quad v_3=\frac{1}{3}(1,2,-2)^{\mathrm{T}}$$

令 $V=(v_1\,|v_2|\,v_3)$

$$\sigma_1=\sqrt{\lambda_1}=18,\quad \sigma_2=\sqrt{\lambda_2}=6,\quad \sigma_3=\sqrt{\lambda_3}=0$$

并令 $\Sigma=\operatorname{diag}_{4\times 3}(\sigma_1,\sigma_2,\sigma_3)$. 然后找到正交矩阵 $U=(u_1\,|u_2|\,u_3\mid u_4)\in\mathbb{R}^{4\times 4}$ 使得 $Av_i=\sigma_i u_i$ *for all* $i\in\{1,2,3\}$. 其中

$$u_1=\frac{Av_1}{\sigma_1}=\frac{1}{2}(1,1,1,1)^{\mathrm{T}},\quad u_2=\frac{Av_2}{\sigma_2}=\frac{1}{2}(-1,1,-1,1)^{\mathrm{T}}.$$

计算 $\mathcal{N}\left((u_1\mid u_2)^{\mathrm{T}}\right)=\operatorname{span}\left((1,0,-1,0)^{\mathrm{T}},(0,1,0,-1)^{\mathrm{T}}\right)$ 并得到

$$u_3=\frac{1}{2}(1,1,-1,-1)^{\mathrm{T}}$$

(我们希望 $2U$ 是一个 Hadamard 矩阵, 需要在 $\mathcal{N}\left((u_1\mid u_2)^{\mathrm{T}}\right)$ 找个单位向量，其元素只含 $\pm\frac{1}{2}$.)

最后，计算 $\mathcal{N}\left((u_1\,|u_2|\,u_3)^{\mathrm{T}}\right)=\operatorname{span}\left((-1,1,1,-1)^{\mathrm{T}}\right)$ 并得到

$$u_4=\frac{1}{2}(-1,1,1,-1)^{\mathrm{T}}.$$

(我们希望 $2U$ 是一个 Hadamard 矩阵, 需要在 $\mathcal{N}\left((u_1\,|u_2|\,u_3)^{\mathrm{T}}\right)$ 找个单位向量, 其元素只含 $\pm\frac{1}{2}$.)

因此 A 的 *SVD* 分解为:

$$A=\begin{pmatrix}-8&5&1\\-4&7&5\\-8&5&1\\-4&7&5\end{pmatrix}=\begin{pmatrix}\frac12&-\frac12&\frac12&-\frac12\\\frac12&\frac12&\frac12&\frac12\\\frac12&-\frac12&-\frac12&\frac12\\\frac12&\frac12&-\frac12&-\frac12\end{pmatrix}\begin{pmatrix}18&0&0\\0&6&0\\0&0&0\\0&0&0\end{pmatrix}\begin{pmatrix}-\frac23&\frac23&\frac13\\\frac23&\frac13&\frac23\\\frac13&\frac23&-\frac23\end{pmatrix}^{\mathrm{T}}=U\Sigma V^{\mathrm{T}}.$$

(ii) A 有两个非零奇异值故 $\operatorname{rank}(A)=2$.

$$\mathcal{R}(A)=\operatorname{span}(u_1,\ldots,u_r)=\operatorname{span}(u_1,u_2)=\operatorname{span}\left(\frac{1}{2}(1,1,1,1)^{\mathrm{T}},\frac{1}{2}(-1,1,-1,1)^{\mathrm{T}}\right),$$

$$\mathcal{N}(A)=\operatorname{span}(v_{r+1},\ldots,v_3)=\operatorname{span}(v_3)=\operatorname{span}\left(\frac{1}{3}(1,2,-2)^{\mathrm{T}}\right)$$

$\|A\|_2=\sigma_1=18$, $\|A\|_F=\sqrt{\sum_{i=1}^r\sigma_i^2}=\sqrt{\sigma_1^2+\sigma_2^2}=6\sqrt{10}$.

(iii) 例用 *(i)* 中 A 的 *SVD* 分解，记 $A^{\mathrm{T}}=\tilde{U}\tilde{\Sigma}\tilde{V}^{\mathrm{T}}$ ，其中 $\tilde{U}=V,\tilde{\Sigma}=\Sigma^{\mathrm{T}},\tilde{V}=U$:

$$A^{\mathrm{T}}=\begin{pmatrix}-\frac{2}{3}&\frac{2}{3}&\frac{1}{3}\\\frac{2}{3}&\frac{1}{3}&\frac{2}{3}\\\frac{1}{3}&\frac{2}{3}&-\frac{2}{3}\end{pmatrix}\begin{pmatrix}18&0&0&0\\0&6&0&0\\0&0&0&0\end{pmatrix}\begin{pmatrix}\frac{1}{2}&-\frac{1}{2}&\frac{1}{2}&-\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&-\frac{1}{2}&-\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}&-\frac{1}{2}&-\frac{1}{2}\end{pmatrix}^{\mathrm{T}}=\tilde{U}\tilde{\Sigma}\tilde{V}^{\mathrm{T}}.$$

记 $\tilde{U}=(\tilde{u}_1\,|\tilde{u}_2|\,\tilde{u}_3)$, $\tilde{V}=(\tilde{v}_1\,|\tilde{v}_2|\,\tilde{v}_3\mid\tilde{v}_4)$, $\tilde{\Sigma}=\operatorname{diag}_{3\times4}(\tilde{\sigma}_1,\tilde{\sigma}_2,\tilde{\sigma}_3)$, 其中 $\tilde{u}_1,\tilde{u}_2,\tilde{u}_3\in\mathbb{R}^3$,$\tilde{v}_1,\tilde{v}_2,\tilde{v}_3,\tilde{v}_4\in\mathbb{R}^4$,$\tilde{\sigma}_1\geq\tilde{\sigma}_2\geq\tilde{\sigma}_3\geq0$.

$k=0$: 注意到 $\left\{M\in\mathbb{R}^{3\times4}:\operatorname{rk}(M)\leq0\right\}=\{0_{3\times4}\}$. 记 $A_0:=0_{3\times4}$ 并有 $\gamma_0=\left\|A^{\mathrm{T}}\right\|_2=\tilde{\sigma}_1=18$

$k=1$: 利用 *Eckhart-Young-Mirsky* 定理,

$$A_1=\tilde{\sigma}_1\tilde{u}_1\tilde{v}_1^{\mathrm{T}}=18\begin{pmatrix}-\frac{2}{3}\\\frac{2}{3}\\\frac{1}{3}\end{pmatrix}\begin{pmatrix}\frac{1}{2}&\frac{1}{2}&\frac{1}{2}&\frac{1}{2}\end{pmatrix}=\begin{pmatrix}-6&-6&-6&-6\\6&6&6&6\\3&3&3&3\end{pmatrix}$$

并有 $\gamma_1=\left\|A^{\mathrm{T}}-A_1\right\|_2=\tilde{\sigma}_2=6.$

$k\geq2$: 因为 $\operatorname{rank}\left(A^{\mathrm{T}}\right)=2$, 记 $A_k:=A^{\mathrm{T}}$, 对任意 $k\in\mathbb{N}_{\geq2}$ 都有 $\gamma_k=0.$

习题 3.

(i) 假设 A 可逆，根据 A 的 *SVD* 结果给出 A^{-1} 的 *SVD* 分解 *(*提示:$Av_i=\sigma_iu_i\ \forall i\in\{1,\ldots,n\}$*)*

(ii) 假设 Q 是正交阵. 给出 Q 的 *SVD* 分解及其奇异值

(iii) 假设 $A=QBQ^{\mathrm{T}}$ ，其中 Q 是正交阵, 说明 A 和 B 有相同奇异值

解. *(i)* 记 A 的 *SVD* 分解为

$$A=U\Sigma V^{\mathrm{T}}=(u_1|\cdots|u_n)\left[\operatorname{diag}_{n\times n}(\sigma_1,\ldots,\sigma_n)\right](v_1|\cdots|v_n)^{\mathrm{T}}$$

其中 $U,V\in\mathbb{R}^{n\times n}$ 正交, $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_n\geq0$. 因为 A 可逆，有 $\operatorname{rank}(A)=n$, $\sigma_i>0$, $\forall i\in\{1,\ldots,n\}$. 又由 $A=U\Sigma V^{\mathrm{T}}$ 有 $Av_i=\sigma_iu_i\ \forall i\in\{1,\ldots,n\}$ ，因此 $A^{-1}u_i=A^{-1}\left(\frac{1}{\sigma_i}Av_i\right)=\frac{1}{\sigma_i}v_i$(其中 $\frac{1}{\sigma_n}\geq\cdots\geq\frac{1}{\sigma_2}\geq\frac{1}{\sigma_1}>0$) 故

$$A^{-1}=(v_n|\cdots|v_2\mid v_1)\left[\operatorname{diag}_{n\times n}\left(\frac{1}{\sigma_n},\ldots,\frac{1}{\sigma_2},\frac{1}{\sigma_1}\right)\right](u_n|\cdots|u_2\mid u_1)^{\mathrm{T}}=(VP)\left(P\Sigma^{-1}P\right)(UP)^{\mathrm{T}},$$

记 $P := (e_n|\cdots|e_2 \mid e_1) \in \mathbb{R}^{n\times n}$. *(由于 $PP^{\mathrm{T}} = P^2 = I_n$，$P$ 是正交阵，且 VP, UP 是正交阵.)*

(ii) 由于 $Q \in \mathbb{R}^{n\times n}$ 正交，$Q = QI_nI_n^T$ 即 Q 的 *SVD* 分解. 显然，所有奇异值为 *1*.

(iii) 记 B 的 *SVD* 分解为 $B = U\Sigma V^{\mathrm{T}}$，则 $A = QBQ^{\mathrm{T}} = QU\Sigma V^{\mathrm{T}}Q^{\mathrm{T}} = (QU)\Sigma(QV)^{\mathrm{T}}$. 显然 A 和 B 有相同奇异值.

数据科学与工程数学基础

作业提交规范及第 5 次作业

教师：黄定江
助教：陈诺、刘文辉

2022 年 11 月 20 日

作业提交规范

1. 作业提交形式：**使用 Word 或 LaTeX 编写所得到的电子文档**。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为："**学号 _ 姓名**"。命名示例：52200000000_ 刘某某。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 5 次作业提交传送门**，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 5 次作业

提交截至时间：**2022/11/07 周一 12:00（中午）**

理论部分 (范数与二次型)

习题 1. 前面已经介绍过矩阵的 LU 分解，于是我们可以对线性方程组的系数矩阵线进行 LU 分解，再解方程组。例如下二元一次方程组

$$\begin{cases} 2x_1 + 4x_2 = 5 \\ 4x_1 + 5x_2 = 3 \end{cases},$$

如果我们对 $A = \begin{pmatrix} 2 & 4 \\ 4 & 5 \end{pmatrix}$ 使用 LU 分解，则可得到

$$\begin{pmatrix} 2 & 4 \\ 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 0 & -3 \end{pmatrix}$$

利用该结果表示线性方程组的解。

解.

$$A = LU$$

$$Ax = b$$

$$LUx = b$$

$$x = U^{-1}L^{-1}b = \begin{pmatrix} 2 & 4 \\ 0 & -3 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 5 \\ 3 \end{pmatrix} = \begin{pmatrix} -13/6 \\ 7/3 \end{pmatrix}$$

习题 2. 写出一种 LU 分解不能分解的矩阵 A，并分析该矩阵在线性方程组 $Ax = b$ 中时方程解集不可能出现的情况。对于这种情况，应该做怎样的处理才能使用 LU 分解。

解.

定理 0.0.1.

矩阵$A \in \mathbb{R}^{n\times n}$ 能够进行LU 分解的充分必要条件是 A 的前 n 阶顺序主子式不为 0。

A 能进行 LU 分解需满足 A 的前 n 阶顺序主子式不为 0，即 A 可逆。

一个不能进行 LU 分解的矩阵如

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

如果我们对其使用 LU 分解，能得到

$$\begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \ell_{21} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ \ell_{21}u_{11} & \ell_{21}u_{12} + u_{22} \end{bmatrix}$$

求解该方程我们需要同时满足

$$u_{11} = 0$$

$$\ell_{21}u_{11} = 2$$

该方程无解，所以 A 无法进行 LU 分解。实际上，A 的一阶顺序主子式为 0。（当然，对于存在顺序主子式为 0 的矩阵 A，也可以进行 LUP 分解）。

习题 3. 利用 QR 分解求解下述线性方程组的解（最终结果可只需写出具体矩阵与向量的乘积形式即可）：

$$\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

解.

$$A = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{6} & \sqrt{6} & \frac{7}{\sqrt{6}} \\ 0 & \sqrt{3} & \frac{1}{\sqrt{3}} \\ 0 & 0 & \sqrt{2} \end{bmatrix} = QR$$

因此，

$$\boldsymbol{x} = \begin{bmatrix} \sqrt{6} & \sqrt{6} & \frac{7}{\sqrt{6}} \\ 0 & \sqrt{3} & \frac{1}{\sqrt{3}} \\ 0 & 0 & \sqrt{2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^{\top} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

数据科学与工程数学基础

作业提交规范及第 6 次作业

教师：黄定江
助教：陈诺、刘文辉

2022 年 11 月 4 日

作业提交规范

1. 作业提交形式：**使用 Word 或 LaTeX 编写所得到的电子文档**。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为："**学号 _ 姓名**"。命名示例：10175501112_ 陈诺。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 6 次作业提交传送门**，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 6 次作业

提交截至时间：**2022/11/14 周一 12:00（中午）**

理论部分

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

解. 该 LS 问题的解就是下列正规化方程组的解：

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

即

$$\begin{bmatrix}35&44\\44&56\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1\\\boldsymbol{x}_2\end{bmatrix}=\begin{bmatrix}9\\12\end{bmatrix}$$

解得：$\boldsymbol{x}=(-1,1)^T$ 对于非满秩的 A，也可以先行变换后消去多余行再对 LS 问题求解。

习题 2. 设 $\boldsymbol{A}=\begin{bmatrix}1&3&1&1\\2&0&0&0\\1&0&0&0\end{bmatrix}$, $\boldsymbol{b}=\begin{bmatrix}1\\1\\1\end{bmatrix}$ 用任意一种方法求对应的 LS 问题的全部解。

解. 该 LS 问题的解就是下列正规化方程组的解:

$$\begin{bmatrix}6&3&1&1\\3&9&3&3\\1&3&1&1\\1&3&1&1\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1\\\boldsymbol{x}_2\\\boldsymbol{x}_3\\\boldsymbol{x}_4\end{bmatrix}=\begin{bmatrix}4\\3\\1\\1\end{bmatrix}$$

初等行变换得到同解方程组

$$\begin{bmatrix}1&3&1&1\\0&15&5&5\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1\\\boldsymbol{x}_2\\\boldsymbol{x}_3\\\boldsymbol{x}_4\end{bmatrix}=\begin{bmatrix}1\\2\end{bmatrix}$$

从而

$$\begin{bmatrix}1&3\\0&15\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1\\\boldsymbol{x}_2\end{bmatrix}=\begin{bmatrix}1-\boldsymbol{x}_3-\boldsymbol{x}_4\\2-5\boldsymbol{x}_3-5\boldsymbol{x}_5\end{bmatrix}$$

即

$$\begin{bmatrix}\boldsymbol{x}_1\\\boldsymbol{x}_2\end{bmatrix}=\begin{bmatrix}\frac{3}{5}\\\frac{2}{15}-\frac{1}{3}\boldsymbol{x}_3-\frac{1}{3}\boldsymbol{x}_4\end{bmatrix}$$

其中 $\boldsymbol{x}_3,\boldsymbol{x}_4\in R$

习题 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 $(\boldsymbol{A}\boldsymbol{X})^T=\boldsymbol{A}\boldsymbol{X}$.

证明. 由 $\boldsymbol{b}$ 的任意性, 取 $\boldsymbol{b}$ 分别为 $\boldsymbol{A}$ 的每一列 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$, 则显然, 若 $\boldsymbol{x}$ 极小化 $\|\boldsymbol{Ax}-\boldsymbol{a}_i\|_2$, $\boldsymbol{x}$ 可以取第 i 个元素为 1, 其余元素为 0 的向量, 因此 $\boldsymbol{X}$ 使得 $\boldsymbol{x}=\boldsymbol{Xa}_i$ 最小化的 $\|\boldsymbol{AXa}_i-\boldsymbol{a}_i\|_2=0$, 这样 $\boldsymbol{AXa}_i=\boldsymbol{a}_i$, 即

$$\boldsymbol{AXA}=\boldsymbol{A}$$

因为对每一个 $\boldsymbol{b}\in\mathbb{R}^m, \boldsymbol{x}=\boldsymbol{Xb}$ 均极小化 $\|\boldsymbol{Ax}-\boldsymbol{b}\|_2$。有 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{AXb}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$。

由于 $\boldsymbol{b}$ 的任意性，有 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{AX}=\boldsymbol{A}^{\mathrm{T}}$，等式两边同时乘以 $\boldsymbol{X}^{\mathrm{T}}$，有

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{AX}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$$

即

$$(\boldsymbol{AX})^{\mathrm{T}}(\boldsymbol{AX})=(\boldsymbol{AX})^{\mathrm{T}}$$

所以

$$\boldsymbol{AX}=(\boldsymbol{AX})^{\mathrm{T}}(\boldsymbol{AX})=(\boldsymbol{AX})^{\mathrm{T}}$$

证毕。 □

习题 4. 利用等式

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

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

解. 设 $f(\alpha)=\|\boldsymbol{A}(\boldsymbol{x}+\alpha\boldsymbol{w})-\boldsymbol{b}\|_2^2$，由于 $\boldsymbol{x}\in\boldsymbol{X}_{LS}$，说明当 $\alpha=0$ 时，函数取极小点。由于 f 是关于 α 的二次函数，故在 $\alpha=-\frac{2\boldsymbol{w}^T\boldsymbol{A}^T(\boldsymbol{Ax}-\boldsymbol{b})}{2\alpha^2\|\boldsymbol{Aw}\|_2^2}$ 取得极值点。代入 $\alpha=0$，有

$$\boldsymbol{w}^T\boldsymbol{A}^T(\boldsymbol{Ax}-\boldsymbol{b})=0$$

又由于 $\boldsymbol{w}$ 的任意性，有

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

习题 5.

$$A:=\begin{pmatrix}5&-1&1\\-1&2&0\\1&0&3\end{pmatrix}$$

记 $\Lambda(A)=\{\lambda_1,\lambda_2,\lambda_3\}\subseteq\mathbb{C}$ *with* $|\lambda_1|\geq|\lambda_2|\geq|\lambda_3|$.

(i) 使用 *Gerschgorin* 圆盘定理, 证明 $\frac{|\lambda_1|}{|\lambda_3|}\leq 7$.（注：由于 A 为对称矩阵，$\frac{|\lambda_1|}{|\lambda_3|}$ 为 A 的条件数）

(ii)（编程题，提交代码）使用幂法与反幂法计算 $\frac{|\lambda_1|}{|\lambda_3|}$

解. *(i)* 令 $a\in\mathbb{C}$, $r\in\mathbb{R}$, $r>0$, 记 $D(a,r):=\{z\in\mathbb{C}:|z-a|\leq r\}\subseteq\mathbb{C}$. 对 A 和 A^{T} 使用 *Gerschgorin* 圆盘定理，有 $\Lambda(A)=\Lambda\left(A^{\mathrm{T}}\right)\subseteq\tilde{G}_1\cup\tilde{G}_2\cup\tilde{G}_3$，其中

$$\tilde{G}_1:=D(5,2),\tilde{G}_2:=D(2,1),\tilde{G}_3:=D(3,1)$$

可得 $|\lambda_1|\leq 7$, $|\lambda_3|\geq 1$. 故 $\frac{|\lambda_1|}{|\lambda_3|}\leq 7$.

(ii) $\frac{|\lambda_1|}{|\lambda_3|}\approx 3.4823$

数据科学与工程数学基础

作业提交规范及第 7 次作业

教师：黄定江
助教：陈诺、刘文辉

2022 年 12 月 8 日

作业提交规范

1. 作业提交形式：**使用 Word 或 LaTeX 编写所得到的电子文档**。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为："**学号 _ 姓名**"。命名示例：50000000000_ 刘某某。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 7 次作业提交传送门**，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 7 次作业

提交截至时间：**2022/11/25 周五 12:00（中午）**

理论部分

习题 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：

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

其中 $A\in R^{m\times n}$，$x\in R^n$，$b,y\in R^m$

解.

$$\begin{aligned}\frac{\partial}{\partial A}f &= \frac{\partial}{\partial A}\frac{1}{2}(x^TA^TAx+2(b-y)^TAx+(b-y)^T(b-y))\\ &= \frac{\partial}{\partial A}\frac{1}{2}(x^TA^TAx+2(b-y)^TAx)\\ &= Axx^T+(b-y)x^T\end{aligned}$$

$$\frac{\partial}{\partial x}f=A^TAx+A^T(b-y)$$

习题 2. 二次型是数据分析中常用函数，求 $\frac{\partial x^TAx}{\partial x}$，$\frac{\partial x^TAx}{\partial A}$，其中 $A\in R^{m\times m}$，$x\in R^m$

解. $\frac{\partial x^TAx}{\partial x}=(A+A^T)x$
$\frac{\partial x^TAx}{\partial A}_{ij}=x_ix_j,\frac{\partial x^TAx}{\partial A}=xx^T$

习题 3. 利用迹微分法求解 $\frac{\partial Tr(W^{-1})}{\partial W}$，其中 $W\in R^{m\times m}$

解. 因为

$$\begin{aligned}0=dI=d(WW^{-1})&=dWW^{-1}+WdW^{-1}\\ WdW^{-1}&=-dWW^{-1}\\ dW^{-1}&=-W^{-1}dWW^{-1}\end{aligned}$$

所以

$$\begin{aligned}dTr(W^{-1})&=Tr(dW^{-1})\\ &=Tr(-W^{-1}dWW^{-1})\\ &=Tr(-(W^{-1})^2dW)\end{aligned}$$

即

$$\frac{\partial Tr(W^{-1})}{\partial W}=-(W^{-T})^2$$

习题 4. $(\exp(z))_i=\exp(z_i)$，$(\log(z))_i=\log(z_i)$ $f(z)=\frac{\exp(z)}{\mathbf{1}^T\exp(z)}$ 称为 *softmax* 函数，，如果 $\boldsymbol{q}=f(\boldsymbol{z}),J=-\boldsymbol{p}^Tlog(\boldsymbol{q})$，其中 $\boldsymbol{p},\boldsymbol{q},\boldsymbol{z}\in\mathbb{R}^n$，并且 $\mathbf{1}^{\mathrm{T}}\boldsymbol{p}=1$，

- 证：$\frac{\partial J}{\partial z}=\boldsymbol{q}-\boldsymbol{p}$
- 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。

解. •

$$J = -\boldsymbol{p}^T \log(\frac{\exp(\boldsymbol{z})}{\mathbf{1}^T \exp(\boldsymbol{z})})$$
$$= -\boldsymbol{p}^T \boldsymbol{z} + \boldsymbol{p}^T \log\left(\mathbf{1}^T \exp(\boldsymbol{z})\right) \mathbf{1}$$
$$= -\boldsymbol{p}^T \boldsymbol{z} + \boldsymbol{p}^T \mathbf{1} \log\left(\mathbf{1}^T \exp(\boldsymbol{z})\right)$$
$$= -\boldsymbol{p}^T \boldsymbol{z} + \log\left(\mathbf{1}^T \exp(\boldsymbol{z})\right)$$

$$\frac{\partial J}{\partial \boldsymbol{z}} = -\boldsymbol{p} + \frac{\partial \log(\mathbf{1}^T \exp(\boldsymbol{z}))}{\partial \boldsymbol{z}}$$
$$= -\boldsymbol{p} + \frac{\partial \mathbf{1}^T \exp(\boldsymbol{z})}{\partial \boldsymbol{z}} \frac{1}{\mathbf{1}^T \exp(\boldsymbol{z})}$$
$$= -\boldsymbol{p} + \frac{\exp(\boldsymbol{z})}{\mathbf{1}^T \exp(\boldsymbol{z})}$$
$$= -\boldsymbol{p} + \boldsymbol{q}$$

•

$$dJ = d\,\mathrm{Tr}(J) = \mathrm{Tr}(dJ) = \mathrm{Tr}[(-\boldsymbol{p}+\boldsymbol{q})^T d\boldsymbol{W}\boldsymbol{x}] = \mathrm{Tr}[\boldsymbol{x}(-\boldsymbol{p}+\boldsymbol{q})^T d\boldsymbol{W}]$$

$$\frac{\partial J}{\partial \boldsymbol{W}} = (-\boldsymbol{p}+\boldsymbol{q})\boldsymbol{x}^T$$

习题 5. 以下内容是利用极大似然估计求解多元正态分布模型的关键步骤：$L = -\frac{Nd}{2}ln(2\pi) - \frac{N}{2}ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_t(\boldsymbol{x}_t - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu})$，$L$ 是对数似然，N 为样本数，d 为样本维数，$\Sigma \in \mathbb{R}^{d\times d}$ 为协方差矩阵，$\mu \in \mathbb{R}^d$ 为期望向量。

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

2) 当 $\mu = \frac{1}{N}\sum_t \boldsymbol{x}_t$ 时，求 $\frac{\partial L}{\partial \boldsymbol{\Sigma}}$，并求使 $\frac{\partial L}{\partial \boldsymbol{\Sigma}} = 0$ 成立的 $\boldsymbol{\Sigma}$。

解. *1.* $\frac{\partial L}{\partial \boldsymbol{\mu}} = \sum_t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu})$

2.

$$d\boldsymbol{L} = d[\frac{N}{2}ln|\boldsymbol{\Sigma}|] - d[\frac{1}{2}\sum_t(\boldsymbol{x}_t - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu})]$$

第一项为

$$d[\frac{N}{2}ln|\boldsymbol{\Sigma}|] = -\frac{N}{2}d[ln|\boldsymbol{\Sigma}|] = -\frac{N}{2}\,\mathrm{Tr}[\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}]$$

第二项为

$$
\begin{aligned}
d[\frac{1}{2}\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t-\boldsymbol{\mu})] &= -\frac{1}{2}d\operatorname{Tr}[\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t-\boldsymbol{\mu})] \\
&= -\frac{1}{2}d\operatorname{Tr}[\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}] \\
&= -\frac{1}{2}\operatorname{Tr}[\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})(\boldsymbol{x}_t-\boldsymbol{\mu})^T(-\boldsymbol{\Sigma}^{-1}(d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1})] \\
&= \frac{1}{2}\operatorname{Tr}[\boldsymbol{\Sigma}^{-1}\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}]
\end{aligned}
$$

得到

$$
\frac{\partial L}{\partial \boldsymbol{\Sigma}} = -\frac{N}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}
$$

令 $\frac{\partial L}{\partial \boldsymbol{\Sigma}}=0$, 易得 $\boldsymbol{\Sigma}=\frac{1}{N}\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})(\boldsymbol{x}_t-\boldsymbol{\mu})^T$

习题 6. 求 $\frac{\partial|\boldsymbol{X}^k|}{\partial \boldsymbol{X}}$，其中 $\boldsymbol{X}\in\mathbb{R}^{m\times m}$ 为可逆矩阵。

解.

$$
\frac{\partial|\boldsymbol{X}^k|}{\partial \boldsymbol{X}} = \frac{\partial|\boldsymbol{X}^k|}{\partial|\boldsymbol{X}|}\frac{\partial|\boldsymbol{X}|}{\partial \boldsymbol{X}} = k|\boldsymbol{X}|^{k-1}|\boldsymbol{X}|\boldsymbol{X}^{-T} = k|\boldsymbol{X}|^k\boldsymbol{X}^{-T}
$$

习题 7. 求 $\frac{\partial\operatorname{Tr}(\boldsymbol{AXBX}^T\boldsymbol{C})}{\partial \boldsymbol{X}}$，其中 $\boldsymbol{A}\in\mathbb{R}^{m\times n},\boldsymbol{X}\in\mathbb{R}^{n\times k},\boldsymbol{B}\in\mathbb{R}^{k\times k},\boldsymbol{C}\in\mathbb{R}^{n\times m}$

解.

$$
\frac{\partial\operatorname{Tr}(\boldsymbol{AXBX}^T\boldsymbol{C})}{\partial \boldsymbol{X}} = (\boldsymbol{BX}^T\boldsymbol{CA})^{\mathrm{T}} + \boldsymbol{CAXB}
$$

数据科学与工程数学基础

作业提交规范及第 8 次作业

教师：黄定江

助教：陈诺、刘文辉

2022 年 11 月 25 日

作业提交规范

1. 作业提交形式：**使用 Word 或 LaTeX 编写所得到的电子文档**。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为："**学号_姓名**"。命名示例：10175501112_陈诺。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 8 次作业提交传送门**，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 8 次作业

提交截至时间：**2022/12/5 周一 12:00（中午）**

理论部分

习题 1. 同时抛 2 颗骰子，事件 A, B, C 分别表示为

(A) 仅有一个骰子是 3

(B) 至少一个骰子是 4

(C) 骰子上点数总和为偶数。

试计算事件 A, B, C 发生后所提供的信息量

解. $H_A = -\log\frac{5}{18}, H_B = -\log\frac{11}{36}, H_C = -\log\frac{1}{2}$

习题 2. 一个容器里面装有 a 个红球和 a 个白球，若从容器中取出 $k, (k \geq 2)$ 个球。对于有放回和无放回两种情况，哪种情况的熵更大？请回答并给予说明。

解. 考虑集合 $\{(x_1, x_2, ..., x_n)|x_i = 0 \ or \ 1\}$ 如果 $x_i = 0$ 则代表第 i 次取出红球否则取出白球。在有放回的情况下，取得该集合里面任意元素的概率都是相同的，且概率和为 1. 而在无放回的情况下，则取得不同元素的概率是有可能不同的。且概率和也为 1. 根据熵的极致性得，有放回的情况下熵更大。

习题 3. 证明：在多分类问题中，利用交叉熵函数作为损失函数和用 KL 散度作为损失函数是等价的。

解. 真实分布：设第 i 个样本 x_i 属于 y_i 类，真实标签分布为 p_i，p_i 是第 y_i 个分量为 1 的 *one-hot* 向量。预测分布：对于第 i 个样本 x_i，预测标签分布是 $q_i = f(x_i; \theta)$，θ 是要学习的参数。

KL 散度 $=(p_i^T \log p_i - p_i^T \log q_i)$

交叉熵 $=(-p_i^T \log q_i)$

由于真实标签是真实存在的，不变的。所以 $\arg\min\limits_{\theta}$ KL 散度 $=\arg\min\limits_{\theta}$ 交叉熵。

习题 4. *(互信息)* 假设 $X_1 \to X_2 \to X_3 \to \cdots \to X_n$ 是一个马尔科夫链，即

$$p(x_1, x_2, \ldots, x_n) = p(x_1) p(x_2 \mid x_1) \cdots p(x_n \mid x_{n-1})$$

试化简 $I(X_1; X_2, \ldots, X_n)$

解.

$$
\begin{aligned}
I\left(X_1; X_2, \ldots, X_n\right) &= H\left(X_1\right) - H\left(X_1 \mid X_2, \ldots, X_n\right) \\
&= H\left(X_1\right) - \left[H\left(X_1, X_2, \ldots, X_n\right) - H\left(X_2, \ldots, X_n\right)\right] \\
&= H\left(X_1\right) - \left[\sum_{i=1}^{n} H\left(X_i \mid X_{i-1}, \ldots, X_1\right) - \sum_{i=2}^{n} H\left(X_i \mid X_{i-1}, \ldots, X_2\right)\right] \\
&= H\left(X_1\right) - \left[\left(H\left(X_1\right) + \sum_{i=2}^{n} H\left(X_i \mid X_{i-1}\right)\right) - \left(H\left(X_2\right) + \sum_{i=3}^{n} H\left(X_i \mid X_{i-1}\right)\right)\right] \\
&= H\left(X_2\right) - H\left(X_2 \mid X_1\right) \\
&= I\left(X_2; X_1\right) \\
&= I\left(X_1; X_2\right)
\end{aligned}
$$

数据科学与工程数学基础

作业提交规范及第 9 次作业

教师：黄定江
助教：陈诺、刘文辉

2023 年 1 月 30 日

作业提交规范

1. 作业提交形式：**使用 Word 或 LaTeX 编写所得到的电子文档**。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为："**学号 _ 姓名**"。命名示例：50000000000_ 刘某某。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 9 次作业提交传送门**，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 9 次作业

提交截至时间：**2022/12/12 周一 12:00（中午）**

理论部分

习题 1. 设某种电子器件的寿命（以 h 计）T 服从双参数的指数分布，其概率密度为

$$f(t)=\begin{cases}\frac{1}{\theta}e^{-(t-c)/\theta} & t\geq c\\ 0 & \text{其他}\end{cases}$$

其中 $c,\theta(c,\theta>0)$ 为未知参数．自一批这种器件中随机地取 n 件进行寿命试验．设它们的失效时间依次为 $x_1\leqslant x_2\leqslant\cdots\leqslant x_n$。

(1) 求 θ 与 c 的最大似然估计值．

(2) 求 θ 与 c 的矩估计量

解. *(1)* 易知似然函数为

$$L(\theta,c)=\frac{1}{\theta^n}\exp\left\{\frac{nc-\sum_{i=1}^n x_i}{\theta}\right\}.$$

所以

$$\ln L(\theta,c)=-n\ln\theta+\frac{nc-\sum_{i=1}^n x_i}{\theta}.$$

对 θ 求偏导，并令导数为 0，可得 $\frac{n}{\theta}+\frac{nc-\sum_{i=1}^n x_i}{\theta^2}=0$. 可得 $\theta=\frac{\sum_{i=1}^n x_i}{n}-c$. 对该函数求二阶导并将 θ 代入, 可得

$$\frac{n}{\theta^2}+\frac{2nc-2\sum_{i=1}^n x_i}{\theta^3}=-\frac{n}{\theta^2}<0,$$

这说明求得的 θ 确实是极大值点。因此, θ 的最大似然估计值为 $\theta=\frac{\sum_{i=1}^n x_i}{n}-c$.

另外，在上式中, 通过简单地观测可以发现 $\ln L(\theta,c)$ 的数值会随着 c 的增加而增加, 故为使得上式最大, 应当使 c 最大. 故 c 的最大似然估计值为 $c=x_{(1)}$.

(2) 该分布的期望和二阶矩分别为

$$E(X)=\int_c^{+\infty}\frac{x}{\theta}\exp\left\{-\frac{x-c}{\theta}\right\}\mathrm{d}x=\theta+c.$$

$$E\left(X^2\right)=\int_c^{+\infty}\frac{x^2}{\theta}\exp\left\{-\frac{x-c}{\theta}\right\}\mathrm{d}x=c^2+2c\theta+2\theta^2.$$

该分布的方差 $\operatorname{Var}(X)=E\left(X^2\right)-E(X)^2=\theta^2$. 通过联立方程组，可求得 θ 的矩估计为 $\theta=\sqrt{\frac{\sum_{i=1}^n(x_i-\bar{x})}{n}}$，$c$ 的矩估计为 $c=\bar{x}-\sqrt{\frac{\sum_{i=1}^n(x_i-\bar{x})}{n}}$.

习题 2. 设总体 X 的概率密度为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}x^{(1-\theta)/\theta} & 0<x<1\\ 0 & \text{其他}\end{cases}\quad 0<\theta<+\infty$$

$X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本。

(1) 验证 θ 的最大似然估计量是 $\hat{\theta}=\frac{-1}{n}\sum_{i=1}^n\ln X_i$

(2) 证明 $\hat{\theta}$ 是 θ 的无偏估计量。

解. (1) 易知似然函数为

$$L(\theta)=\frac{1}{\theta^n}\left(\prod_{i=1}^n x_i\right)^{\frac{1-\theta}{\theta}}.$$

所以 $\ln L(\theta)=-n\ln\theta+\frac{1-\theta}{\theta}\ln\left(\prod_{i=1}^n x_i\right)$. 对 θ 求导，并令导数为 0，可得

$$-\frac{n}{\theta}-\frac{1}{\theta^2}\ln\left(\prod_{i=1}^n x_i\right)=0.$$

可得 $\hat{\theta}=-\frac{1}{n}\Sigma_{i=1}^n\ln X_i$. 对该函数求二阶导并将 $\hat{\theta}$ 代入, 可得 $-\frac{n^3}{\left(\sum_{i=1}^n\ln X_i\right)^2}<0$, 这说明求得的 θ 确实是极大值点。故原命题得证。

(2) 首先求得

$$E(\ln x)=\int_0^1\ln x\frac{1}{\theta}x^{\frac{1-\theta}{\theta}}\mathrm{d}x=-\theta.$$

所以 $E(\hat{\theta})=-\frac{-n\theta}{n}=\theta$. 故原命题得证。

习题 3. 假设总体 $X\sim N\left(\mu,\sigma^2\right)$ (σ^2 已知), $X_1,X_2,\ldots,X_n$ 为来自总体 X 的样本, 由过去的经验和知识, 我们可以确定 μ 的取值比较集中在 μ_0 附近, 离 μ_0 越远, μ 取值的可能性越小, 于是我们假定 μ 的先验分布为正态分布

$$\pi(\mu)=\frac{1}{\sqrt{2\pi\sigma_\mu^2}}\exp\left[-\frac{1}{2\sigma_\mu^2}\left(\mu-\mu_0\right)^2\right]\quad(\mu_0,\sigma_\mu\text{已知})$$

求 μ 的后验概率分布。

解. 样本分布密度为

$$q(\mathbf{x}\mid\mu)=\frac{1}{\sigma^n(2\pi)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n\left(x_i-\mu\right)^2\right]$$

于是后验密度函数为

$$h(\mu\mid\mathbf{x})=\frac{q(\mathbf{x}\mid\mu)\cdot\pi(\mu)}{f_{\mathbf{x}}(\mathbf{x})}=\frac{q(\mathbf{x}\mid\mu)\cdot\pi(\mu)}{\int_{-\infty}^{+\infty}q(\mathbf{x}\mid\mu)\cdot\pi(\mu)d\mu}$$

$$\propto\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n\left(x_i-\mu\right)^2\right]\cdot\exp\left[-\frac{1}{2\sigma_\mu^2}\left(\mu-\mu_0\right)^2\right]$$

化简得

$$h(\mu\mid\mathbf{x})\propto\exp\left[-\frac{(\mu-t)^2}{2\eta^2}\right]$$

其中 $t=\frac{\frac{n}{\sigma^2}\bar{x}+\frac{1}{\sigma_\mu^2}\mu_0}{\frac{n}{\sigma^2}+\frac{1}{\sigma_\mu^2}}$， $\eta^2=\frac{1}{\frac{n}{\sigma^2}+\frac{1}{\sigma_\mu^2}}$，于是

$$\mu\mid\mathbf{x}\sim N\left(\frac{\frac{n}{\sigma^2}\bar{x}+\frac{1}{\sigma_\mu^2}\mu_0}{\frac{n}{\sigma^2}+\frac{1}{\sigma_\mu^2}},\frac{1}{\frac{n}{\sigma^2}+\frac{1}{\sigma_\mu^2}}\right).$$

习题 4. 假设总体 $X \sim P(\lambda), X_1, X_2, \ldots, X_n$ 为来自总体 X 的样本, 假定 λ 的先验分布为伽玛分布 $\Gamma(\alpha, \beta)$, 求 λ 的后验期望估计（平方损失下的贝叶斯估计）。

解. 因为 λ 的先验密度函数 $\pi(\lambda)$ 为伽玛分布 $\Gamma(\alpha, \beta)$, 即

$$\pi(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda}$$

样本分布密度函数为:

$$q(\mathbf{x} \mid \lambda) = \frac{\lambda^{\sum_{i=1}^n x_i}}{x_1! x_2! \ldots, x_n!} e^{-n\lambda} \propto \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}$$

所以

$$h(\lambda \mid \mathbf{x}) \propto \lambda^{\alpha + \sum_{i=1}^n x_i - 1} e^{-(\beta+n)\lambda}$$

即

$$\lambda \mid \mathbf{x} \sim \Gamma\left(\alpha + \sum_{i=1}^n x_i, \beta + n\right)$$

故 λ 的后验期望估计为

$$\hat{\lambda} = \frac{\alpha + \sum_{i=1}^n x_i}{\beta + n} = \frac{n}{\beta + n} \bar{x} + \frac{\beta}{\beta + n} \frac{\alpha}{\beta}$$

它是样本均值 $\bar{x}$ 和先验分布 $\Gamma(\alpha, \beta)$ 的均值 $\frac{\alpha}{\beta}$ 的加权平均。

数据科学与工程数学基础

作业提交规范及第 10 次作业

教师：黄定江

助教：陈诺、刘文辉

2022 年 12 月 10 日

作业提交规范

1. 作业提交形式：**使用 Word 或 LaTeX 编写所得到的电子文档**。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为："**学号 _ 姓名**"。命名示例：10175501112_ 陈诺。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 10 次作业提交传送门**，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 10 次作业

提交截至时间：**2022/12/19 周一 12:00（中午）**

理论部分

习题 1. 考虑以下概率图模型

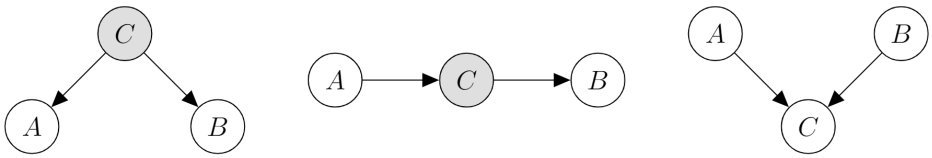


图 1: 概率图模型

i) 对左图, 证明 $A \perp\!\!\!\perp B \mid C$; (即 A 和 B 在 C 的条件下独立)
ii) 对中图, 证明 $A \perp\!\!\!\perp B \mid C$;
iii) 对右图, 证明 $A \perp\!\!\!\perp B \mid \emptyset$.

解. *i)*

$$p(A, B \mid C) = \frac{p(A, B, C)}{p(C)} = \frac{p(A \mid C)p(B \mid C)p(C)}{p(C)} = p(A \mid C)p(B \mid C)$$

ii)

$$\begin{aligned} p(A, B \mid C) &= \frac{p(A, B, C)}{p(C)} = \frac{1}{p(C)} p(B \mid C)p(C \mid A)p(A) \\ &= p(B \mid C)\frac{p(C \mid A)p(A)}{p(C)} = p(B \mid C)p(A \mid C), \end{aligned}$$

iii)

$$p(A, B) = \sum_C p(A, B, C) = \sum_C p(C \mid A, B)p(A)p(B) = p(A)p(B),$$

习题 2. 下面的函数哪些是凸函数? 请说明理由。

1. $f(x) = e^x + 1$, $x \in \mathbb{R}$
2. $f(x) = \max(\|Ax + b\|_2, \|x^T x\|_1), A \in \mathbb{R}^{m\times n}, x \in \mathbb{R}^n, b \in \mathbb{R}^m$
3. $f(x) = -\cos x, x \in [-\pi/2, \pi/2]$

解. 1. $f''(x) = e^x > 0$ 所以是凸函数
2. 因为 $\|\|_2, \|\|_1$ 是凸函数, 所以 $\|Ax + b\|_2, \|x^T x\|_1$ 是凸函数, max 是保凸运算, 所以 $f(x)$ 是凸函数。
3. $f''(x) = \cos x > 0 x \in [-\pi/2, \pi/2]$ 所以是凸函数

习题 3. 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. 即 $\log(\Phi(x))$ 是凹函数。

解. 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,

$$\begin{aligned}\int_{-\infty}^{x}e^{-u^2/2}du&\leq\int_{-\infty}^{x}e^{\frac{x^2}{2}-xu}du\\&=e^{\frac{x^2}{2}}\cdot\frac{e^{-xu}}{-x}\Big|_{u=-\infty}^{x}\\&=e^{\frac{x^2}{2}}\cdot\frac{e^{-x^2}}{-x}\end{aligned}$$

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 4. 计算函数 $f(x)$ 的共轭函数，以及共轭函数的定义域。
(1) $f(x)=-\log x$
(2) $f(x)=e^x$

解. *(1)*$f(x)=-\log x$，定义域为 $domf=x|x>0$。当 $y<0$ 时，函数 $xy+\log x$ 无上界，当 $y\leq 0$ 时，在 $x=-1/y$ 处函数达到最大值。因此，定义域为 $domf^*=\{y|y<0\}$，共轭函数为 $f^*(y)=-\log(-y)-1(y<0)$

(2)$f(x)=e^x$。当 $y<0$ 时，函数 $xy-e^x$ 无界。当 $y>0$ 时，函数 $xy-e^x$ 在 $x=\log y$ 处达到最大值。因此，$f^*(y)=y\log y-y$。当 $y=0$ 时，$f^*(y)=\sup_x -e^x=0$，综上，$domf^*=\{y|y\geq 0\}$，$f^*(y)=y\log y-y$。*(*规定 $0\log 0=0$*)* 。