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：刘文辉

2022 年 10 月 24 日

作业提交规范

1. 作业提交形式：**pdf 格式文档，可以使用 Word、LaTeX 或手写所得到的电子文档**。建议博士生均使用 LaTeX 编写。
2. 作业命名规范：提交的电子文档必须命名为：“**hw1_ 学号 _ 姓名**”。其中，hw1 表示第 1 次作业。命名示例：hw1_52200000000_ 刘某某。
3. 作业提交途径：点击打开网址：**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，我将批改最新时间提交的作业。

第 1 次作业

提交截至时间：**2021/10/31 12:00（中午）**

理论部分

习题 1. 矩阵 $A^2 = A, B^2 = B$, 并且 B 的列是 A 的列的线性组合。证明 $AB = B$。

解. 由题意知，B 的列是 A 的列的线性组合，因此 $\exists C$, 使得 $B = AC$。又因为 $A^2 = A$，所以

$$AB = AAC = A^2C = AC = B$$

得证。

习题 2. 设 $\boldsymbol{A}, \boldsymbol{B}$ 为任意两个 n 阶方阵，证明：$\boldsymbol{AB}$ 和 $\boldsymbol{BA}$ 具有相同的特征多项式，即 $|\lambda\boldsymbol{E}-\boldsymbol{AB}| = |\lambda\boldsymbol{E}-\boldsymbol{BA}|$。

解. 构造矩阵

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{E} & \boldsymbol{O} \\ -\boldsymbol{A} & \boldsymbol{E} \end{pmatrix}$$

以及

$$\boldsymbol{N} = \begin{pmatrix} \lambda\boldsymbol{E} & \boldsymbol{B} \\ \lambda\boldsymbol{A} & \lambda\boldsymbol{E} \end{pmatrix}$$

计算它们之间的乘积可得

$$\boldsymbol{MN} = \begin{pmatrix} \lambda\boldsymbol{E} & \boldsymbol{B} \\ \boldsymbol{O} & \lambda\boldsymbol{E}-\boldsymbol{AB} \end{pmatrix}$$

和

$$\boldsymbol{NM} = \begin{pmatrix} \lambda\boldsymbol{E}-\boldsymbol{BA} & \boldsymbol{B} \\ \boldsymbol{O} & \lambda\boldsymbol{E} \end{pmatrix}$$

对第一个等式两边求行列式可得：

$$|\boldsymbol{MN}| = |\lambda\boldsymbol{E}| \cdot |\lambda\boldsymbol{E}-\boldsymbol{AB}|$$

同理对第二个等式有：

$$|\boldsymbol{NM}| = |\lambda\boldsymbol{E}-\boldsymbol{BA}| \cdot |\lambda\boldsymbol{E}|$$

由此

$$|\lambda\boldsymbol{E}-\boldsymbol{AB}| = |\lambda\boldsymbol{E}-\boldsymbol{BA}|$$

习题 3. 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 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. 有些平时称之为“距离”的函数其实并不是数学意义上的距离，请判断以下两种所谓的“距离”是否是数学意义上的距离并说明理由。

(1) 假设向量 $a, b \in \mathbb{R}^n$，定义余弦距离为 $d(a,b) = 1 - \cos\langle a,b\rangle$，其中 $\langle a,b\rangle$ 为向量 a,b 间的夹角。

(2) 假设 S_1, S_2 分别表示两个字符串，定义 S_1, S_2 的编辑距离 $d(S_1, S_2)$ 为由 S_1 转成 S_2 所需的最少编辑操作次数。其中一次编辑操作可以是：将 S_1 中的一个字符替换成另一个字符；在 S_1 中插入一个字符；在 S_1 中删除一个字符。例如：*kitten* 和 *sitting* 的编辑距离是 *3*。将 *kitten* 变为 *sitting* 的最小处理方式如下：

kitten → *sitten* （将 *k* 替换为 *s*）

sitten → *sittin* （将 *e* 替换为 *i*）

sittin → *sitting*（尾部插入 *g*）.

解. 只有第二个才是数学意义上的距离。

（*1*）对于余弦距离，不是距离。

显然不满足非负性。假设 $a = (1,0)$，$b = (2,0)$，不难得到：

$$d(a,b) = 0$$

但是 $a \neq b$，所以不成立。

（*2*）对于编辑距离，是距离。

(i) 编辑次数显然非负，且只有两个字符串完全相同时，所需的编辑次数最少为 0，即满足非负性:

$$d(S_1, S_2) \geq 0，且 \ d(S_1, S_2) = 0 \Leftrightarrow S_1 = S_2$$

(ii) 插入和删除字符互为逆操作、将某个字符 *a* 替换为字符 *b* 的逆操作为将字符 *b* 替换为字符 *a*，不难看出每种编辑操作均可逆，即满足对称性：

$$d(S_1, S_2) = d(S_2, S_1)$$

(iii) 将 S_1 编辑为 S_2 的过程拆解成两部分：将 S_1 编辑为 S_3 以及将 S_3 编辑为 S_2，考虑到其中这两部分可能存在冗余操作。所以直接从 S_1 编辑为 S_2 所需的最优编辑次数只会更少，即满足三角不等式：

$$d(S_1, S_2) \leq d(S_1, S_3) + d(S_3, S_2)$$

综上 *(i)*，*(ii)*，*(iii)* 所述，它是数学意义上的距离。

习题 6. 求向量 $(1,1,1)^T$ 投影到一维子空间 $\mathrm{span}\{(1,-1,1)^T\}$ 的正交投影。

解. 令 $w=(1,-1,1)^T$，投影到一维子空间中投影矩阵记为：$\boldsymbol{P}$，

$$\boldsymbol{P}=\frac{ww^T}{w^Tw}=\frac{1}{3}\begin{pmatrix}1\\-1\\1\end{pmatrix}\begin{pmatrix}1&-1&1\end{pmatrix}=\frac{1}{3}\begin{pmatrix}1&-1&1\\-1&1&-1\\1&-1&1\end{pmatrix}$$

对于给定的向量 $x=(1,-1,1)^T$，投影为：

$$\boldsymbol{P}x=\frac{1}{3}\begin{pmatrix}1&-1&1\\-1&1&-1\\1&-1&1\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}=\begin{pmatrix}\frac{1}{3}\\-\frac{1}{3}\\\frac{1}{3}\end{pmatrix}$$

习题 7. 求向量 $(1,1,1)^T$ 投影到仿射空间 $\mathrm{span}\{(1,-1,1)^T,(1,1,0)^T\}+(1,2,1)^T$ 的正交投影。

解. 令子空间

$$\mathbf{V}=\mathrm{span}\begin{pmatrix}1&1\\-1&1\\1&0\end{pmatrix}$$

以及 $x'=x-x_0=(0,-1,0)^T$。现计算它在 $\mathbf{V}$ 上的投影，记为 $\mathcal{P}_V(x')$。将 $\mathbf{V}$ 上的基底记为 $\boldsymbol{B}$。则

$$\boldsymbol{B}^T\boldsymbol{B}=\begin{pmatrix}1&1\\-1&1\\1&0\end{pmatrix}^T\begin{pmatrix}1&1\\-1&1\\1&0\end{pmatrix}=\begin{pmatrix}3&0\\0&2\end{pmatrix}$$

因此，

$$\mathcal{P}_V(x')=\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^Tx'=\begin{pmatrix}-\frac{1}{6}\\-\frac{5}{6}\\\frac{1}{3}\end{pmatrix}$$

综上，

$$\mathcal{P}(x)=x_0+\mathcal{P}_V(x')=\begin{pmatrix}1\\2\\1\end{pmatrix}+\begin{pmatrix}-\frac{1}{6}\\-\frac{5}{6}\\\frac{1}{3}\end{pmatrix}=\begin{pmatrix}\frac{5}{6}\\\frac{7}{6}\\\frac{4}{3}\end{pmatrix}$$

习题 8. 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 $A \in \mathbb{R}^{n\times m}$，$m \leq n$，$\operatorname{rank}(A)=m$. $P=A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}$ 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 $P=P^2$, $\lambda x=Px=P(Px)=P(\lambda x)=\lambda Px=\lambda^2 x$. 因为 $x\neq 0\in\mathbb{R}^n$, 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)*, $\operatorname{rank}(P)=\operatorname{tr}(P)=\operatorname{tr}(A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}})=\operatorname{tr}(\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}A)=\operatorname{tr}(I_m)=m$.