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：刘文辉、徐艺玮

2023 年 12 月 7 日

作业提交规范

1. 作业提交形式：建议使用 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。若使用手写拍照的方式，请务必规整好各题解答的图片，并整合在一个 PDF 文档中，只发图片格式的作业概不批改！
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地址：**第 5 次作业提交传送门**，无需注册和登录，按要求输入个人学号和姓名，然后上传作业文档即可。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；若发现存在抄袭的作业时，相似的两份作业均会被记为 0 分。

第 5 次作业

提交截至时间：**2023/11/21 下周二 12:00（中午）**

理论部分

习题 1. 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 $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}^T\boldsymbol{b}$

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 $\alpha=-\frac{2\boldsymbol{w}^T\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})}{2\alpha^2\|\boldsymbol{A}\boldsymbol{w}\|_2^2}$ 取得极值点。代入 $\alpha=0$，有

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

又由于 $\boldsymbol{w}$ 的任意性，有

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

习题 2.

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

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

习题 3. 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 4. 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 5. 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 6. $(\exp(\boldsymbol{z}))_i = \exp(\boldsymbol{z}_i)$，$(\log(\boldsymbol{z}))_i = \log(\boldsymbol{z}_i)$ $f(\boldsymbol{z}) = \frac{\exp(\boldsymbol{z})}{\mathbf{1}^T\exp(\boldsymbol{z})}$ 称为 *softmax* 函数，，如果 $\boldsymbol{q} = f(\boldsymbol{z}), J = -\boldsymbol{p}^T log(\boldsymbol{q})$，其中 $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{z} \in \mathbb{R}^n$，并且 $\mathbf{1}^\mathrm{T}\boldsymbol{p} = 1$，

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

解. •

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

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

-

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

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