数据科学与工程理论基础

总复习题（数学）

教师：黄定江
助教：刘文辉

2023 年 1 月 19 日

习题 1. 求下面矩阵的 *1*-范数、*2*-范数和无穷范数：

$$A_1 = \begin{pmatrix} 2 & 4 \\ 1 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} -2 & 0 \\ 3 & 6 \end{pmatrix}.$$

习题 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

习题 3. 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。

习题 4. 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.

习题 5. 证明上三角矩阵与上三角矩阵的乘积仍是上三角矩阵。

习题 6. 用 *Householder* 方法求矩阵 $A = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 2 & 1 \end{bmatrix}$ 的 *QR* 分解。

习题 7. 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 $\|A^{\mathrm{T}}A\|_2 = \|A\|_2^2 = \|G\|_2^2$

习题 8. 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）:

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

习题 9. 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 $A \in R^{m\times n}$，$x \in R^n$，$b, y \in R^m$

习题 10. $(\exp(z))_i = \exp(z_i)$，$(\log(z))_i = \log(z_i)$ $f(z) = \frac{\exp(z)}{\mathbf{1}^T \exp(z)}$ 称为 *softmax* 函数，，如果 $\boldsymbol{q} = f(\boldsymbol{z}), J = -\boldsymbol{p}^T log(\boldsymbol{q})$，其中 $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{z} \in \mathbb{R}^n$，并且 $\mathbf{1}^{\mathrm{T}}\boldsymbol{p} = 1$，

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

习题 12. 求 $\frac{\partial |\boldsymbol{X}^k|}{\partial \boldsymbol{X}}$，其中 $\boldsymbol{X} \in \mathbb{R}^{m\times m}$ 为可逆矩阵。

习题 13. 求 $\frac{\partial\, \mathrm{Tr}(\boldsymbol{AXBX}^T\boldsymbol{C})}{\partial \boldsymbol{X}}$，其中 $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{X} \in \mathbb{R}^{n\times k}, \boldsymbol{B} \in \mathbb{R}^{k\times k}, \boldsymbol{C} \in \mathbb{R}^{n\times m}$

习题 14. 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。

习题 15. *(*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 $I(X_1; X_2, \ldots, X_n)$

习题 16. 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。

习题 17. 写出下述非线性规划的 *KKT* 条件并求解

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

习题 18. 考虑等式约束的最小二乘问题

$$\begin{aligned}
& \textit{minimize} && \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2\\
& \textit{suject to} && \mathbf{G}\mathbf{x}=\mathbf{h}
\end{aligned}$$

其中 $\mathbf{A}\in\mathbb{R}^{m\times n}, \operatorname{rank}(\mathbf{A})=n, \mathbf{G}\in\mathbb{R}^{p\times n}, \operatorname{rank}(\mathbf{G})=p$. 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.

习题 19. 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。

习题 20. 用 *Lagrange* 乘子法求欠定方程 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的最小二范数解，其中 $\boldsymbol{A}\in\mathbb{R}^{m\times n}, m\leq n, \operatorname{rank}(\boldsymbol{A})=m$

习题 21. 使用梯度下降法和固定步长 $\lambda=0.01$ 计算 $\min f(x)=(x_1-1)^2+16(x_2-2)^2$ ，初始点 $x^{(0)}=(2,3)^T$，迭代两步后终止.

习题 22. 考虑问题

$$\min f(x)=3x_1^2+3x_2^2-x_1^2x_2.$$

从初始点 $x^{(0)}=(1.5,1.5)^{\mathrm{T}}$ 出发, 用 *Newton* 方法求迭代两步后该问题的解（可编写程序辅助计算）.