2020 级华东师范大学本科生期末考试

《数据科学与工程数学基础》笔试试卷　　2022.9

学院：数据科学与工程学院　　考试形式：闭卷　　所需时间：120 分钟

考生姓名：______　　学号：______　　专业：______　　任课教师：黄定江

题目	一	二	三	四	五	六	七	八	总分
满分	14	12	13	13	13	12	12	11	100
得分									
阅卷人									

题 1　(14') 已知矩阵

$$\boldsymbol{M} = \begin{pmatrix} 1 & 0 \\ -2 & 3 \\ 0 & 1 \end{pmatrix}$$

(1) 计算矩阵 $\boldsymbol{M}$ 的 1 范数；

(2) 计算矩阵 $\boldsymbol{M}$ 的 2 范数；

(3) 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 （1）$\|\boldsymbol{M}\|_1 = \max\{1+\|-2\|, 3+1\} = 4$ **（3 分）**

（2）$\boldsymbol{M}^\top \boldsymbol{M} = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$

$\lambda_{max}(\boldsymbol{M}^\top \boldsymbol{M}) = 14$

$\|\boldsymbol{M}\|_2 = \sqrt{14}$ **（4 分）**

（3）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，$\|\boldsymbol{A}\|_{m_\infty} = 0.$ **（2 分）**

齐次性：

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

（2 分）

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 $\|\boldsymbol{A}+\boldsymbol{B}\|_{m_\infty} \le \|\boldsymbol{A}\|_{m_\infty} + \|\boldsymbol{B}\|_{m_\infty}$ **（3 分）**

题 2 (12’) 已知矩阵

$$\boldsymbol{M} = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 4 & 1 \\ 4 & 2 & 1 \end{pmatrix}$$

(1) 求矩阵 $\boldsymbol{M}$ 的零空间；

(2) 求向量 $\boldsymbol{x} = (1,1,1)^\top$ 在矩阵 $\boldsymbol{M}$ 的列空间上的正交投影；

(3) 利用 Householder 变换，将非零向量 $\boldsymbol{x} = (1,1,1)^\top$ 变为标准向量 $\boldsymbol{e}_1 = (0,1,0)^\top$ 的 $\sqrt{3}$ 倍，请写出这个 Householder 变换矩阵。

解 （1）零空间为 $span\{(1,1,-6)^T\}$ **（3 分）**

（2）列空间为 $span\{(1,2,4)^T, (-1,4,2)^T\}$，若记

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 \\ 2 & 4 \\ 4 & 2 \end{pmatrix}$$

故正交投影为 $\boldsymbol{A}(\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top\boldsymbol{x} = (1/3, 2/3, 4/3)^\top$ **（4 分）**

（3）

$$w = \frac{x - \alpha e_1}{\|x - \alpha e_1\|_2}$$

则，由 $H = \left(I - 2ww^\top\right)$ 可得：

$$\boldsymbol{H} = \begin{pmatrix} \frac{2-\sqrt{3}}{3-\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{3-\sqrt{3}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{3-\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{2-\sqrt{3}}{3-\sqrt{3}} \end{pmatrix}$$

（5 分）

题 3 (13’) 矩阵

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

(1) 计算矩阵 $\boldsymbol{A}$ 的 SVD 分解；

(2) 假设 $\boldsymbol{B}$ 是一个 $n \times d$ 的矩阵，矩阵 $\boldsymbol{M}$ 是 $(n+d) \times (n+d)$ 定义为

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{B}^\top \\ \boldsymbol{B} & \boldsymbol{O} \end{pmatrix}$$

显然 $\boldsymbol{M}$ 是对称矩阵。请证明矩阵 $\boldsymbol{M}$ 的对角化会产生 $\boldsymbol{B}$ 的奇异值分解所需要的所有信息。

解（1）易计算

$$AA^\top = \begin{pmatrix} 17 & 8 \\ 8 & 17 \end{pmatrix}$$

求得其特征值为 $25, 9$。对应的特征向量分别为 $(1/\sqrt{2}, 1/\sqrt{2})^\top, (-1/\sqrt{2}, 1/\sqrt{2})^\top$，因此，

$$U = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

（4 分）

由 $A^\top U$ 与 V 的关系可直接计算出

$$V = \begin{pmatrix} 1/\sqrt{2} & -1/3\sqrt{2} & -2/3 \\ 1/\sqrt{2} & 1/3\sqrt{2} & 2/3 \\ 0 & -4/3\sqrt{2} & 1/3 \end{pmatrix}$$

（3 分）

（2）B 的奇异值分解所需要的信息为 B^TB 的特征向量和 BB^T 的特征向量，以及对应的特征值。现假设对应于特征值为 $\lambda^2(\lambda > 0)$ 的 B^TB 的特征向量为 $x_1(\|x_1\|_2 = 1)$,BB^T 的特征向量为 $x_2(\|x_2\|_2 = 1)$. 因此，$B^TBx_1 = \lambda^2 x_1$ 以及 $BB^T x_2 = \lambda^2 x_2$。故

$$(BB^T)Bx_1 = \lambda^2 Bx_1$$

所以存在 k 使得 $Bx_1 = kx_2$。由于 $\|x_1\|_2 = \|x_2\|_2 = 1$，可得 $k = \lambda$，即 $Bx_1 = \lambda x_2$。同理有 $B^T x_2 = \lambda x_1$。（3 分）

下面证明 $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ 是矩阵 M 的特征值为 λ 的特征向量。易计算

$$Mx = \begin{pmatrix} 0 & B^\top \\ B & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} B^T x_2 \\ Bx_1 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

因此，B 的奇异值分解信息包含在 M 的对角化过程中。（3 分）

题 4 (13’) 完成下列函数求导或梯度：

(1) 求激活函数 $\sigma(x)=\frac{1}{1+e^{-x}}$ 的导数；

(2) 求函数 $f(\boldsymbol{X})=Tr(\boldsymbol{X}^{-1})$ 对 $\boldsymbol{X}$ 的梯度，其中 $\boldsymbol{X}\in\mathbb{R}^{m\times n}$。

(3) 若非奇异矩阵 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$，$\boldsymbol{x}\in\mathbb{R}^{n\times 1}$，$\boldsymbol{y}\in\mathbb{R}^{n\times 1}$，求函数 $f(\boldsymbol{A})=\boldsymbol{y}^\top\boldsymbol{A}^{-1}\boldsymbol{x}$ 对 $\boldsymbol{A}$ 的梯度。

解（1）$\sigma'(x)=\frac{e^{-x}}{(1+e^{-x})^2}=\sigma(x)(1-\sigma(x))$ **（4 分）**

（2）因为

$$
\begin{aligned}
0=dI=d(\boldsymbol{X}\boldsymbol{X}^{-1})&=d\boldsymbol{X}\boldsymbol{X}^{-1}+\boldsymbol{X}d\boldsymbol{X}^{-1}\\
\boldsymbol{X}d\boldsymbol{X}^{-1}&=-d\boldsymbol{X}\boldsymbol{X}^{-1}\\
d\boldsymbol{X}^{-1}&=-\boldsymbol{X}^{-1}d\boldsymbol{X}\boldsymbol{X}^{-1}
\end{aligned}
$$

所以

$$
\begin{aligned}
dTr(\boldsymbol{X}^{-1})&=Tr(d\boldsymbol{X}^{-1})\\
&=Tr(-\boldsymbol{X}^{-1}d\boldsymbol{X}\boldsymbol{X}^{-1})\\
&=Tr(-(\boldsymbol{X}^{-1})^2d\boldsymbol{X})
\end{aligned}
$$

即

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

（5 分）

（3）

$$
\begin{aligned}
df(\boldsymbol{A})&=Tr(d\boldsymbol{y}^\top\boldsymbol{A}^{-1}\boldsymbol{x})\\
&=Tr(\boldsymbol{y}^\top d\boldsymbol{A}^{-1}\boldsymbol{x})\\
&=Tr(-\boldsymbol{y}^\top\boldsymbol{A}^{-1}d\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{x})\\
&=Tr(-\boldsymbol{A}^{-1}\boldsymbol{x}\boldsymbol{y}^\top\boldsymbol{A}^{-1}d\boldsymbol{A})
\end{aligned}
$$

同理，求得其梯度。 **（4 分）**

题 5 (13’) 计算和证明以下问题：

(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 7

试计算事件 A, B, C 发生后所提供的信息量;

(2) 证明联合熵和条件熵有如下关系:

$$H(X,Y) = H(X) + H(Y \mid X)$$

解 （1）$H_A = -\log\frac{5}{18}, H_B = -\log\frac{1}{6}, H_C = -\log\frac{1}{2}$ （6 分） （2）

$$\begin{aligned} H(X,Y) =& \mathbb{E}(\log\frac{1}{p(xy)}) \\ =& \mathbb{E}(\log\frac{1}{p(x)p(y|x)}) \\ =& \mathbb{E}(\log\frac{1}{p(x)} + \log\frac{1}{p(y|x)}) \\ =& \mathbb{E}(\log\frac{1}{p(x)}) + \mathbb{E}(\log\frac{1}{p(y|x)}) \\ =& H(X) + H(Y|X) \end{aligned}$$

（7 分）

题 6 (12’) 求解以下问题:

(1) 假设总体 $X \sim N(\mu, \sigma_0^2)$ (σ_0^2 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 $\hat{\mu}$ 附近, 离 $\hat{\mu}$ 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

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

求 μ 的后验概率分布。

(2) 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（泊松分布的分布律: $P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$，泊松分布的期望为 λ；伽马分布的概率密度函数：$f(x;\alpha,\beta)=\frac{x^{\alpha-1}e^{-\beta x}\beta^\alpha}{\Gamma(\alpha)}$，伽马分布的期望为 α/β）。

解（1）样本分布密度为

$$q(\mathbf{x}\mid\mu)=\frac{1}{\sigma_0^n(2\pi)^{n/2}}\exp\left[-\frac{1}{2\sigma_0^2}\sum_{i=1}^n(x_i-\mu)^2\right]$$

于是后验密度函数为

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

$$\propto\exp\left[-\frac{1}{2\sigma_0^2}\sum_{i=1}^n(x_i-\mu)^2\right]\cdot\exp\left[-\frac{1}{2\hat{\sigma}^2}(\mu-\hat{\mu})^2\right]$$

化简得

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

其中 $t=\frac{\frac{n}{\sigma_0^2}\bar{x}+\frac{1}{\sigma_\mu^2}\hat{\mu}}{\frac{n}{\sigma_0^2}+\frac{1}{\hat{\sigma}^2}}$， $\eta^2=\frac{1}{\frac{n}{\sigma_0^2}+\frac{1}{\sigma_\mu^2}}$, 于是

$$\mu\mid\mathbf{x}\sim N\left(\frac{\frac{n}{\sigma_0^2}\bar{x}+\frac{1}{\sigma_\mu^2}\hat{\mu}}{\frac{n}{\sigma_0^2}+\frac{1}{\hat{\sigma}^2}},\frac{1}{\frac{n}{\sigma_0^2}+\frac{1}{\sigma_\mu^2}}\right).$$

（6 分）

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

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

样木分布密度函数为:

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

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。 **（6 分）**

题 7 (12’) 求证下列与凸集或凸函数相关的问题：

(1) 证明：设 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是仿射变换, 即 $f(\boldsymbol{x})=\boldsymbol{A} \boldsymbol{x}+\boldsymbol{b}, \boldsymbol{A} \in \mathbb{R}^{m \times n}, \boldsymbol{b} \in \mathbb{R}^m$, 则凸集在 f 下的像是凸集:

$$S \subseteq \mathbb{R}^n \text{ 为凸集} \Rightarrow f(S) \stackrel{\text{def}}{=} \{f(\boldsymbol{x}) \mid \boldsymbol{x} \in S\} \text{ 为凸集;}$$

(2) 判定函数 $f(x)=\max (\|\boldsymbol{A x}+\boldsymbol{b}\|_2,\|\boldsymbol{x}^{\top} \boldsymbol{x}\|_1), \boldsymbol{A} \in \mathbb{R}^{m \times n}, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{b} \in \mathbb{R}^m$ 是否为凸函数，并说明理由。

解 （1）令 $\boldsymbol{y}_1, \boldsymbol{y}_2$ 表示再像集合中的任意两个点，它们的原象分别为 $\boldsymbol{x}_1, \boldsymbol{x}_2$。因此，$\boldsymbol{y}_1=\boldsymbol{A} \boldsymbol{x}_1+\boldsymbol{b}\, \boldsymbol{y}_2=\boldsymbol{A} \boldsymbol{x}_2+\boldsymbol{b}$ 因此，有

$$(1-\lambda) \boldsymbol{y}_1+\lambda \boldsymbol{y}_2=\boldsymbol{A}((1-\lambda) \boldsymbol{x}_1+\lambda \boldsymbol{x}_2)+\boldsymbol{b}$$

故得证。 **（6 分）**

（2）由仿射函数的任意范数为凸函数可知 $\|\boldsymbol{A x}+\boldsymbol{b}\|_2$ 为凸函数，另外 $\|\boldsymbol{x}^{\top} \boldsymbol{x}\|_1$ 为欧几里得范数的平方，显然为凸函数。根据逐点取最大值具有保凸性，可知 $f(x)$ 是凸函数。 **（6 分）**

题 8 (11') 考虑优化问题

$$\min_{\boldsymbol{x}} \quad \boldsymbol{x}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{x}$$

其中 $\boldsymbol{W} \in S_{+}^{n}$ 是对称半正定矩阵。

(1) 求出在任意点 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\top}$ 处沿负梯度方向迭代的最佳步长 α。

(2) 若对上述优化问题增加约束条件

$$x_i^2 = 1, \quad i = 1, \cdots, n$$

请写出此时约束优化问题的拉格朗日函数 $L(\boldsymbol{x}, \boldsymbol{\lambda})$，并计算出约束优化问题的拉格朗日对偶函数。

解 （1）可求得 $f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{x}$ 的梯度为 $2\boldsymbol{W}\boldsymbol{x}$。因此，为寻求得最佳步长，需求解

$$\min_{\boldsymbol{x}} \quad (\boldsymbol{x} - 2\alpha \boldsymbol{W}\boldsymbol{x})^{\mathrm{T}} \boldsymbol{W} (\boldsymbol{x} - 2\alpha \boldsymbol{W}\boldsymbol{x})$$

（2 分）

求极值可得：

$$\alpha = \frac{\boldsymbol{x}^{\top} \boldsymbol{W}^2 \boldsymbol{x}}{2\boldsymbol{x}^{\top} \boldsymbol{W}^3 \boldsymbol{x}}.$$

（3 分）

（2）若增加约束条件，可得 Lagrange 函数为

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\nu}) &= \boldsymbol{x}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{x} + \sum_{i=1}^{n} \nu_i \left(x_i^2 - 1\right) \\ &= \boldsymbol{x}^{\mathrm{T}} (\boldsymbol{W} + \mathrm{diag}(\boldsymbol{\nu})) \boldsymbol{x} - \mathbf{1}^{\mathrm{T}} \boldsymbol{\nu} \end{aligned}$$

（3 分）

对 $\boldsymbol{x}$ 求极小得到 Lagrange 对偶函数

$$\begin{aligned} g(\boldsymbol{\nu}) &= \inf_{\boldsymbol{x}} \boldsymbol{x}^{\mathrm{T}} (\boldsymbol{W} + \mathrm{diag}(\boldsymbol{\nu})) \boldsymbol{x} - \mathbf{1}^{\mathrm{T}} \boldsymbol{\nu} \\ &= \begin{cases} -\mathbf{1}^{\mathrm{T}} \boldsymbol{\nu} & \boldsymbol{W} + \mathrm{diag}(\boldsymbol{\nu}) \succeq \mathbf{0} \\ -\infty & \text{其他情况} \end{cases} \end{aligned}$$

（3 分）