第 1 章 第一次作业

本次作业包含三部分内容, 即基础题, 思考题和编程题. 作业提交形式为线上提交, 应使用 LaTeX 独立完成作业电子版.

Deadline: 2022 年 3 月 22 日 23:59.

1.1 基础题

说明：该部分的所有题目均为必做，所有题目及相关变种均可能出现在试卷上。

1.1.1 Sherman-Morrison 公式

已知 A 是一个 $n\times n$ 的可逆矩阵，并且 x 和 y 是两个 $n\times 1$ 向量，使得 $A+xy^{\mathrm{T}}$ 可逆，则有

$$(A+xy^{\mathrm{T}})^{-1}=A^{-1}-\frac{A^{-1}xy^{\mathrm{T}}A^{-1}}{1+y^{\mathrm{T}}A^{-1}x}. \tag{1.1}$$

试验证该公式的正确性.

1.1.2 矩阵的子空间

1. 阅读 [1]，了解矩阵的四个基本子空间：列空间（Column Space）、行空间（Row Space）、零空间（Null Space）、左零空间（Left Null Space），写出对应的定义.
2. 阅读 Wikipedia 或是 Brilliant，基于 Gauss-Jordan 消元法证明秩-零化度定理 (Rank-Nullity Theorem).

1.1.3 矩阵求导

完成一下矩阵求导公式的推导：

1. $\frac{\partial}{\partial X}\mathrm{tr}(X^{\mathrm{T}}AX)=(A+A^{\mathrm{T}})X$.
2. $\frac{\partial}{\partial X}\mathrm{tr}(X^{\mathrm{T}}AY)=AY$.
3. $\frac{\partial}{\partial Y}\mathrm{tr}(X^{\mathrm{T}}AY)=A^{\mathrm{T}}X$.

1.1.4 多元正态分布的等高面

1. 设 $X\sim\mathcal{N}(\mu,\Sigma)=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$, 其中 $\Sigma\succ 0$，X 的密度函数记为 $f(x;\mu,\Sigma)$. 任给 $a>0$，试证明概率密度等高面

$$f(x;\mu,\Sigma)=a \tag{1.2}$$

是一个椭球面.

2. 特别地, 令 $p=2$ 且 $\Sigma=\sigma^2\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}$ $(\rho>0)$，概率密度等高面就是平面上的一个椭圆. 试求该椭圆的方程、长轴和短轴.

1.1.5 多元正态分布的简单性质

设 x 为 p 维随机向量 $x\sim\mathcal{N}(\mu,\Sigma)$, 正态分布的概率密度函数为

$$p(x|\mu,\Sigma)=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

完成以下问题:

1. 证明:
$$\mathbb{E}[x]=\int_{\mathbb{R}^n}xp(x|\mu,\Sigma)dx,$$
2. 计算: $\mathbb{E}[xx^{\mathrm{T}}]=\Sigma+\mu\mu^{\mathrm{T}}$ (提示, 考察与协方差矩阵的关系)
3. 设 A 为对称阵，计算二次型在正态分布下的期望:
$$\mathbb{E}[x^{\mathrm{T}}Ax]=\mathrm{tr}(\Sigma A)+\mu^{\mathrm{T}}A\mu$$
(提示 1. 多变量积分变量代换, 并使用对称矩阵的特征分解和正交矩阵的行列式为 ±1. 提示 2. 另一种方法, 将二次型 $x^{\mathrm{T}}Ax$ 的期望写成矩阵迹的形式, 并利用矩阵的迹的性质进行变形计算.)
4. 当 $\mu=a\mathbf{1}_p$, $A=I_p-\frac{1}{p}\mathbf{1}_{p\times p}$, $\Sigma=\sigma^2I_p$ 时，试利用（1）和（2）的结果证明 $\mathbb{E}(X^{\mathrm{T}}AX)=\sigma^2(p-1)$. 这里 $\mathbf{1}_p$ 表示所有分量均为 1 的列向量, $\mathbf{1}_{p\times p}$ 表示所有元素均为 1 的矩阵.

1.1.6 多元正态分布的极大似然估计

设 $X_{(i)}$ $(i=1,2,\cdots,n)$ 为 p 元正态总体 $N_p(\mu,\Sigma)$ 的随机样本，记 $\overline{X}$ 为样本均值，A 为样本离差阵，有

$$\overline{X}=\frac{1}{n}\sum_{t=1}^{n}X_{(t)},\ A=\sum_{t=1}^{n}\left(X_{(t)}-\overline{X}\right)\left(X_{(t)}-\overline{X}\right)^{\mathrm{T}}. \tag{1.3}$$

完成以下问题

1. 推导 μ,Σ 的极大似然估计分别是 $\hat{\mu}=\overline{X},\hat{\Sigma}=\frac{1}{n}A$.
2. 证明 $\overline{X}\sim N\left(\mu,\frac{1}{n}\Sigma\right)$.
3. 对于参数 θ 的一个估计 $\hat{\theta}$，我们称 $\hat{\theta}$ 是一个无偏估计，如果 $\mathbb{E}(\hat{\theta})=\theta$ 对于参数 θ 的任意取值都成立（通俗的说，就是假设真实分布中对应的参数为 θ，我们利用观测到的样本构造一个估计的方法 $\hat{\theta}$，这个样本满足 $\mathbb{E}(\hat{\theta})=\theta$）. 证明 $\overline{X}$ 是 μ 的无偏估计.

1.1.7 范数的严格凸性

求证范数的严格凸性等价于下列条件:

$$\|x+y\| = \|x\| + \|y\| \ (\forall x \neq \theta, y \neq \theta) \implies x = cy \ (c > 0). \tag{1.4}$$

1.1.8 凸函数极小值的性质

设 $\mathscr{X}$ 为线性赋范空间，函数 $\varphi : \mathscr{X} \to \mathbb{R}^1$ 称为凸的，如果不等式

$$\varphi(\lambda x + (1-\lambda)y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y) \ (\forall 0 \leq \lambda \leq 1) \tag{1.5}$$

成立. 求证凸函数的局部极小值必然是全空间的极小值.

1.1.9 投影的最优性条件

给定数据集 $\{\boldsymbol{x}_i\}_{i=1}^N$, $\boldsymbol{x}_i \in \mathbb{R}^n$, 设其均值为 $\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{x}_i$. 设以 $\boldsymbol{w}$ 为方向向量, 过点 $\bar{\boldsymbol{x}}$ 的直线为

$$\boldsymbol{L} = \{\boldsymbol{y} \in \mathbb{R}^n | \boldsymbol{y} = \bar{\boldsymbol{x}} + t\boldsymbol{w},\ t \in \mathbb{R}\}. \tag{1.6}$$

1. 对任何 $\boldsymbol{x} \in \mathbb{R}^n$, 定义 $\tilde{\boldsymbol{x}}_i = \arg\min_{\boldsymbol{y} \in \boldsymbol{L}} \|\boldsymbol{x}_i - \boldsymbol{y}\|$ 为 $\boldsymbol{x}$ 在 $\boldsymbol{L}$ 上的投影. 请计算数据 $\boldsymbol{x}_i$ 在 $\boldsymbol{L}$ 的投影所对应的点 $\tilde{\boldsymbol{x}}$.
2. 对于目标函数
$$\min_{\boldsymbol{w}} J(\boldsymbol{w}) = \frac{1}{N}\sum_{i=1}^N \|\boldsymbol{x}_i - \tilde{\boldsymbol{x}}_i\|^2 \tag{1.7}$$
如果 $\|\boldsymbol{w}\| = 1$, 求最优点所满足的最优性条件. (提示: 参考瑞利商)
3. 设 $t_i^* = \arg\min_t \|\boldsymbol{x}_i - (\bar{\boldsymbol{x}} + t\boldsymbol{w})\|$, 则 t_i^* 依赖于方向 $\boldsymbol{w}$. 对于目标函数
$$\max_{\boldsymbol{w}} K(\boldsymbol{w}) = \frac{1}{N}\sum_{i=1}^N (t_i^*)^2, \tag{1.8}$$
如果 $\|\boldsymbol{w}\| = 1$, 求最优点所满足的最优性条件.

1.1.10 KL 散度的简单介绍

对于空间 $\mathbb{R}^n$ 中的两个概率分布 p, q，我们定义对应的 KL 散度如下：

$$\mathrm{D}_{\mathrm{KL}}(p\|q) = \int_{\mathbb{R}^n} p(x)\left(\log p(x) - \log q(x)\right) dx. \tag{1.9}$$

试完成以下问题

1. 验证 KL 散度是否满足对称性（即 $\forall p, q, \mathrm{D}_{\mathrm{KL}}(p\|q) = \mathrm{D}_{\mathrm{KL}}(q\|p)$）.
2. 利用 $\log x$ 的凸性与 Jensen 不等式验证 KL 散度是否满足非负性（即 $\forall p, q, \mathrm{D}_{\mathrm{KL}}(p\|q) \geq 0$）.

1.2 思考题

说明：该部分的所有题目均可以不做，给出精彩的解答可以弥补在基础题区域的失分，特别精彩的解答可能可以折算为平时成绩的额外加分（待定）。部分题目及相关变种均可能出现在试卷上。

1.2.1 Sherman-Morrison 公式（接1.1.1）

试推导 Sherman-Morrison 公式.

1.2.2 矩阵的子空间（接1.1.2）

1. 阅读 Wikipedia，了解正交补（Orthogonal Complement），并说明矩阵的四个基本子空间直接的正交关系.
2. 基于正交补关系与秩-零化度定理证明矩阵的行秩等于列秩.
3. 阅读 Wikipedia，了解正交基（Orthonormal Basis），并写出 Bessel 不等式与 Parseval 等式）.
4. 阅读 [2] 中关于等周问题相关章节，并与通过变分法给出的证明进行比较.

1.2.3 矩阵求导（接1.1.3）

完成以下矩阵求导公式的推导：

1. $$\frac{\partial \det(X)}{\partial X} = \det(X) X^{-\mathrm{T}}.$$
2. $$\frac{\partial Y^{-1}}{\partial x} = -Y^{-1}\frac{\partial Y}{\partial x}Y^{-1}.$$

1.2.4 多元正态分布的极大似然估计（接1.1.6）

1. 说明 A 可以写成 $\sum_{t=1}^{n-1} Z_t Z_t^{\mathrm{T}}$ 的形式，其中 $Z_i \overset{\text{i.i.d.}}{\sim} N_p(0,\Sigma)$ $(i=1,2,\cdots,n-1)$.
2. 说明 Σ 的极大似然估计 $\hat{\Sigma} = \frac{1}{n}A$ 不是无偏估计.
3. 验证样本协方差阵 $S = \frac{1}{n-1}A$ 是 Σ 的无偏估计.

1.3 编程题

说明：该部分注明为“思考题”的题目均可以不做，给出精彩的解答可以弥补在基础题区域的失分，特别精彩的解答可能可以折算为平时成绩的额外加分（待定）。

非“思考题”部分必做，主要考察思考过程，不会因为给出算法的效率差等扣分（不排除后期用 OJ 收代码，可能部分性能极差的代码会扣分）。

除特殊要求的题目外，所有题目的编程语言不限；除特殊申明，不限库函数的使用；担心提交代码无法编译的，可以备注开发环境。

代码可能会抽样查重，发现代码抄袭现象时（尤其是逐字符相同、最后修改日期不正常等），一切处理手段解释保留。

1.3.1 Gauss-Jordan 消元法

试用 C++ 实现，Gauss-Jordan 消元法，要求可以 AC洛谷 P3389，允许参考题解，但请自己写一遍（除源码外，请额外提交带帐号名的 AC 截图）.

1.3.2 凸包的简单应用

文件 ‘2021JSCPC 热身赛题面.pdf’ 中有四道题，针对其中的 C 题，考虑基于凸包的算法（允许参考题解）。

1. 请用 algorithm2e 宏包给出对应的伪代码.
2. 给出 C++ 代码实现（请设置编译选项为 ‘-std=c++14 -O2’）.

1.3.3 BiCGSTAB（思考题）

参考稳定双共轭梯度法，利用 algorithm2e 宏包给出对应的伪代码. 感兴趣的同学可以利用 Matlab 或 Python 提供的函数进行测试，与其它算法的性能进行比较.

1.3.4 个人绩点分析（思考题）

1. 将教务处上的课程绩点，复制并保存为 csv 文件，自设一些可能的变量（如喜爱的程度、是否有预备基础等），并存入 MySQL 中.
2. 利用 Python 中 mysql.connector 等方法连接数据库，并保存为 DataFrame 文件（主要是练习数据库基础，也可直接导入 csv）.
3. 利用 statsmodels 提供的 ols 方法，提交 OLS Regression Results，并解释 R-squared, F-statistic, p-values 等变量的统计学含义（本小问仅作帮助理解假设检验用）.

Bibliography

[1] Gilbert Strang. The Four Fundamental Subspaces: 4 Lines. Website. https://web.mit.edu/18.06/www/Essays/newpaper_ver3.pdf (cit. on p. 1).

[2] Elias M Stein and Rami Shakarchi. Fourier analysis: an introduction. Vol. 1. Princeton University Press, 2011 (cit. on p. 4).