数据科学与工程数学基础

作业提交规范及第 7 次作业

教师：黄定江
助教：陈诺、刘文辉

2022 年 12 月 8 日

作业提交规范

1. 作业提交形式：**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。
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：50000000000_ 刘某某。
3. 作业提交途径：点击打开每次作业的传送门网址：**第 7 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（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。

第 7 次作业

提交截至时间：**2022/11/25 周五 12:00（中午）**

理论部分

习题 1. 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 2. 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 3. 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 4. $(\exp(z))_i=\exp(z_i)$，$(\log(z))_i=\log(z_i)$ $f(z)=\frac{\exp(z)}{\mathbf{1}^T\exp(z)}$ 称为 *softmax* 函数，，如果 $\boldsymbol{q}=f(\boldsymbol{z}), J=-\boldsymbol{p}^Tlog(\boldsymbol{q})$，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。

解. •

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

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

•

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

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

习题 5. 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 $\boldsymbol{\Sigma}$。

解. *1.* $\frac{\partial L}{\partial \boldsymbol{\mu}} = \sum_t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t-\boldsymbol{\mu})$

2.

$$d\boldsymbol{L} = d[\frac{N}{2}ln|\boldsymbol{\Sigma}|] - d[\frac{1}{2}\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t-\boldsymbol{\mu})]$$

第一项为

$$d[\frac{N}{2}ln|\boldsymbol{\Sigma}|] = -\frac{N}{2}d[ln|\boldsymbol{\Sigma}|] = -\frac{N}{2}\,\mathrm{Tr}[\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}]$$

第二项为

$$d[\frac{1}{2}\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t-\boldsymbol{\mu})] = -\frac{1}{2}d\,\mathrm{Tr}[\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t-\boldsymbol{\mu})]$$
$$= -\frac{1}{2}d\,\mathrm{Tr}[\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}]$$
$$= -\frac{1}{2}\,\mathrm{Tr}[\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})(\boldsymbol{x}_t-\boldsymbol{\mu})^T(-\boldsymbol{\Sigma}^{-1}(d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1})]$$
$$= \frac{1}{2}\,\mathrm{Tr}[\boldsymbol{\Sigma}^{-1}\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}]$$

得到

$$\frac{\partial L}{\partial \boldsymbol{\Sigma}} = -\frac{N}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})(\boldsymbol{x}_t-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}$$

令 $\frac{\partial L}{\partial \boldsymbol{\Sigma}} = 0$, 易得 $\boldsymbol{\Sigma} = \frac{1}{N}\sum_t(\boldsymbol{x}_t-\boldsymbol{\mu})(\boldsymbol{x}_t-\boldsymbol{\mu})^T$

习题 6. 求 $\frac{\partial|\boldsymbol{X}^k|}{\partial \boldsymbol{X}}$，其中 $\boldsymbol{X} \in \mathbb{R}^{m\times m}$ 为可逆矩阵。

解.

$$\frac{\partial|\boldsymbol{X}^k|}{\partial \boldsymbol{X}} = \frac{\partial|\boldsymbol{X}^k|}{\partial|\boldsymbol{X}|}\frac{\partial|\boldsymbol{X}|}{\partial \boldsymbol{X}} = k|\boldsymbol{X}|^{k-1}|\boldsymbol{X}|\boldsymbol{X}^{-T} = k|\boldsymbol{X}|^k\boldsymbol{X}^{-T}$$

习题 7. 求 $\frac{\partial\,\mathrm{Tr}(\boldsymbol{AXBX}^T\boldsymbol{C})}{\partial \boldsymbol{X}}$，其中 $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{X} \in \mathbb{R}^{n\times k}, \boldsymbol{B} \in \mathbb{R}^{k\times k}, \boldsymbol{C} \in \mathbb{R}^{n\times m}$

解.

$$\frac{\partial\,\mathrm{Tr}(\boldsymbol{AXBX}^T\boldsymbol{C})}{\partial \boldsymbol{X}} = (\boldsymbol{BX}^T\boldsymbol{CA})^{\mathrm{T}} + \boldsymbol{CAXB}$$