最优化第一次作业

人工智能 222 杨义琦

2024 年 10 月 27 日

注意:

- 作业请用 A4 纸手写或者电脑输入然后打印，本次作业提交时间为 10 月 25 日，请各班学习委员收齐后交给我。
- 该作业通过 latex 生成，如果你需要 latex 源码并在上面直接写作业，请私下联系我。
- 作业内容就是期末考试题库内容，请大家认真独立完成。

1. 计算下面向量或矩阵变量函数的导数 (请写出过程)

   - $f(x)=\frac{1}{2}\|Ax-b\|_2^2$,$A\in\mathbf{R}^{m\times n},x\in\mathbf{R}^n,y\in\mathbf{R}^m$.

     题解：

     首先展开平方范数：

     $$\begin{aligned}f(x)&=\frac{1}{2}\|Ax-b\|_2^2\\&=\frac{1}{2}(Ax-b)^T(Ax-b)\\&=\frac{1}{2}(x^TA^TAx-x^TA^Tb-b^TAx+b^Tb)\end{aligned}$$

     对 x 求导：

     二次型 x^TBx （其中 B 是对称矩阵）的导数是 $2Bx$

     在本题中，A^TA 是一个对称矩阵（因为 $(A^TA)^T=A^TA$）

     所以 x^TA^TAx 是一个二次型

     因此 $\frac{\partial}{\partial x}(x^TA^TAx)=2A^TAx$

     $$\begin{aligned}\frac{\partial f}{\partial x}&=\frac{1}{2}(2A^TAx-A^Tb-A^Tb)\\&=A^TAx-A^Tb\end{aligned}$$

     因此，$\nabla f(x)=A^TAx-A^Tb$

   - $f(X)=a^TXb$, 其中 $X\in\mathbf{R}^{m\times n},a\in\mathbf{R}^m,b\in\mathbf{R}^n$ 为给定的向量。

     题解：

     这是一个标量函数，可以使用矩阵微分的基本性质。

     由矩阵微分的基本结果 $\frac{\partial}{\partial X}Tr(AXB)=A^TB^T$ 可知

对矩阵 X 求导：

$$\frac{\partial f}{\partial X} = ab^T$$

- $f(X) = Tr(X^TAX)$, 其中 $X \in \mathbf{R}^{m\times n}$ 为长方形矩阵,A 为方阵（但不一定对称）

  题解：

  矩阵微分的基本性质：

  (a) $Tr(AB) = Tr(BA)$　迹的循环性质
  (b) $\frac{\partial Tr(AX)}{\partial X} = A^T$　迹对矩阵的导数
  (c) $\frac{\partial Tr(XA)}{\partial X} = A$　迹对矩阵的导数

  考虑函数在方向 H 上的增量：

$$\begin{aligned} f(X+tH) &= Tr((X+tH)^TA(X+tH)) \\ &= Tr((X^T+tH^T)A(X+tH)) \\ &= Tr(X^TAX + tX^TAH + tH^TAX + t^2H^TAH) \end{aligned}$$

  根据导数定义，我们只需要 t 的一次项：

$$\frac{d}{dt}f(X+tH)\bigg|_{t=0} = Tr(X^TAH + H^TAX) = Tr(X^TAH) + Tr(H^TAX)$$

  使用迹的循环性质：

$$\begin{aligned} &= Tr(X^TAH) + Tr(H^TAX) \\ &= Tr(H^TA^TX) + Tr(H^TAX) \quad (\text{迹的性质：}Tr(A) = Tr(A^T)) \\ &= Tr(H^T(AX + A^TX)) \end{aligned}$$

  根据内积的定义 $\langle A, B\rangle = Tr(A^TB)$ 可得：

$$\left\langle \frac{\partial f}{\partial X}, H \right\rangle = Tr(H^T\frac{\partial f}{\partial X})$$

  因此：

$$\frac{\partial f}{\partial X} = \frac{\partial Tr(X^TAX)}{\partial X} = AX + A^TX$$

2. 求优化目标函数 J 关于变量 F 和 G 的**梯度**和**二阶导数 (海塞矩阵)**

$$J(F,G) = -\sum_{i,j=1}^{n}(s_{ij}\theta_{ij} - \log(1+\exp(\theta_{ij})))$$

其中 $F = [f_1, f_2, \cdots, f_n] \in \mathbf{R}^{c\times n}$, $G = [g_1, g_2, \cdots, g_n] \in \mathbf{R}^{c\times n}$, $\theta_{ij} = \frac{1}{2}f_i^Tg_j$,$s_{ij}$ 为常量

题解：

- 一阶梯度：

对于 F 的第 k 个元素 f_k，令 $p_{ij}=\frac{\exp(\theta_{ij})}{1+\exp(\theta_{ij})}$，一阶偏导数为：

$$\frac{\partial J}{\partial f_k}=\sum_{j=1}^{n}\frac{\partial J}{\partial \theta_{kj}}\frac{\partial \theta_{kj}}{\partial f_k}$$

其中

$$\frac{\partial J}{\partial \theta_{kj}}=-(s_{kj}-\frac{\exp(\theta_{kj})}{1+\exp(\theta_{kj})}) \qquad \frac{\partial \theta_{ij}}{\partial f_k}=\begin{cases}\frac{1}{2}g_j & \text{if } i=k\\ 0 & \text{otherwise}\end{cases}$$

所以我们有

$$\frac{\partial J}{\partial f_k}=-\sum_{j=1}^{n}\left(s_{kj}-\frac{\exp(\theta_{kj})}{1+\exp(\theta_{kj})}\right)\cdot\frac{1}{2}g_j=-\frac{1}{2}\sum_{j=1}^{n}\left(s_{kj}-p_{kj}\right)g_j$$

同理，对于 G 的第 l 个元素 g_l, 一阶偏导数为：

$$\frac{\partial J}{\partial g_l}=-\sum_{i=1}^{n}\left(s_{il}-\frac{\exp(\theta_{il})}{1+\exp(\theta_{il})}\right)\cdot\frac{1}{2}f_i=-\frac{1}{2}\sum_{i=1}^{n}\left(s_{il}-p_{il}\right)f_i$$

- 二阶梯度：

通过求导可以发现：

$$\frac{\partial p_{ij}}{\partial \theta_{ij}}=p_{ij}(1-p_{ij})$$

对于 f_i 的二阶导数：

$$\frac{\partial^2 J}{\partial f_i \partial f_i^T}=-\frac{1}{4}\sum_{j=1}^{n}p_{ij}(1-p_{ij})g_j g_j^T$$

对于 g_j 的二阶导数：

$$\frac{\partial^2 J}{\partial g_j \partial g_j^T}=-\frac{1}{4}\sum_{i=1}^{n}p_{ij}(1-p_{ij})f_i f_i^T$$

对于交叉二阶导数：

$$\frac{\partial^2 J}{\partial f_i \partial g_j}=-\frac{1}{4}p_{ij}(1-p_{ij})g_j$$
$$\frac{\partial^2 J}{\partial g_j \partial f_i}=-\frac{1}{4}p_{ij}(1-p_{ij})f_i$$

完整的海塞矩阵：

$$H=\begin{bmatrix}\frac{\partial^2 j}{\partial f \partial f^t} & \frac{\partial^2 j}{\partial f \partial g^t} & \frac{\partial^2 j}{\partial g \partial f^t} & \frac{\partial^2 j}{\partial g \partial g^t}\end{bmatrix}$$

3. 试证明球及椭球为凸集。（请查看教材 P38 的球和椭球的定义）

定义：

(a) 球：$B(x_c, r) = \{x| \|x - x_c\|_2 \leq r\} = \{x_c + ru| \|u\|_2 \leq 1\}$

(b) 椭球：$\{x|(x - x_c)^T P^{-1}(x - x_c) \leq 1\}$

- 证明球是凸集：

要证明集合 $B(x_c, r)$ 是凸集，需要证明对任意 $x_1, x_2 \in B(x_c, r)$ 和任意 $\theta \in [0, 1]$，都有：$\theta x_1 + (1 - \theta)x_2 \in B(x_c, r)$

假设 $x_1, x_2 \in B(x_c, r)$，即:

$$\|x_1 - x_c\|_2 \leq r \quad \text{和} \quad \|x_2 - x_c\|_2 \leq r.$$

我们需要证明对于任意 $\lambda \in [0, 1]$，点 $x_\lambda = \lambda x_1 + (1 - \lambda)x_2$ 也在球内，即 $x_\lambda \in B(x_c, r)$，满足:

$$\|x_\lambda - x_c\|_2 \leq r.$$

首先，我们展开 $x_\lambda - x_c$:

$$x_\lambda - x_c = \lambda x_1 + (1 - \lambda)x_2 - x_c = \lambda(x_1 - x_c) + (1 - \lambda)(x_2 - x_c).$$

然后使用三角不等式:

$$\|x_\lambda - x_c\|_2 = \|\lambda(x_1 - x_c) + (1 - \lambda)(x_2 - x_c)\|_2 \leq \lambda\|x_1 - x_c\|_2 + (1 - \lambda)\|x_2 - x_c\|_2.$$

由于 $\|x_1 - x_c\|_2 \leq r$ 和 $\|x_2 - x_c\|_2 \leq r$，因此:

$$\|x_\lambda - x_c\|_2 \leq \lambda r + (1 - \lambda)r = r.$$

因此，$\|x_\lambda - x_c\|_2 \leq r$，说明 $x_\lambda \in B(x_c, r)$，即球是凸集，证毕。

- 证明椭球是凸集：

假设 $x_1, x_2 \in E$，即:

$$(x_1 - x_c)^T P^{-1}(x_1 - x_c) \leq 1 \quad \text{和} \quad (x_2 - x_c)^T P^{-1}(x_2 - x_c) \leq 1.$$

我们需要证明对于任意 $\lambda \in [0, 1]$，点 $x_\lambda = \lambda x_1 + (1 - \lambda)x_2$ 也在椭球内，即:

$$(x_\lambda - x_c)^T P^{-1}(x_\lambda - x_c) \leq 1$$

首先，$x_\lambda - x_c$ 可以展开为：

$$x_\lambda - x_c = \lambda x_1 + (1 - \lambda)x_2 - x_c = \lambda(x_1 - x_c) + (1 - \lambda)(x_2 - x_c)$$

所以，我们可以得到：

$$[\lambda(x_1 - x_c) + (1 - \lambda)(x_2 - x_c)]^T P^{-1}[\lambda(x_1 - x_c) + (1 - \lambda)(x_2 - x_c)]$$

继续展开得到：

$$\lambda^2(x_1-x_c)^TP^{-1}(x_1-x_c)+2\lambda(1-\lambda)(x_1-x_c)^TP^{-1}(x_2-x_c)+(1-\lambda)^2(x_2-x_c)^TP^{-1}(x_2-x_c)$$

由柯西 - 施瓦茨不等式和椭球的定义可知：

$$(x_1-x_c)^TP^{-1}(x_2-x_c)\le\sqrt{(x_1-x_c)^TP^{-1}(x_1-x_c)}\cdot\sqrt{(x_2-x_c)^TP^{-1}(x_2-x_c)}\le\sqrt{1}\cdot\sqrt{1}\le 1$$

所以，对之前的展开式进行放缩：

$$\text{before}\le\lambda^2(x_1-x_c)^TP^{-1}(x_1-x_c)+2\lambda(1-\lambda)\cdot 1+(1-\lambda)^2(x_2-x_c)^TP^{-1}(x_2-x_c)$$

根据椭球的定义放缩可得：

$$\text{before}\le\lambda^2\cdot 1+2\lambda(1-\lambda)\cdot 1+(1-\lambda)^2\cdot 1=(\lambda+(1-\lambda))^2=1$$

因此，$(x_\lambda-x_c)TP^{-1}(x_\lambda-x_c)\le 1$，说明 x_λ 也属于椭球 E，即椭球是一个凸集，证毕。