视听信息系统导论 第四次作业

BobAnkh

December 2020

1. 线性对数几率回归中假设 $f(x) = w^T x + b$，设计一种数据处理方案使得参数 b 可以不用出现

可以在向量 x 的末尾增加一个取值全为 1 的维度；或者也可以将所有的数据都减去某一个特定的数据，如 $x_k - x_0$。

2. 用牛顿法求解无约束优化问题，目标函数为

$$f(x) = \log\ (e^x + e^{-x})$$

分别以初值 $x^{(0)}$=1 和 $x^{(0)} = 1.1$，写出前五步迭代过程，令迭代步长 =1.

目标函数为 $f(x) = \log(e^x + e^{-x})$，则可以得到其一阶导和二阶导为:

$$f^{'}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$f^{''}(x) = \frac{4}{(e^x + e^{-x})^2}$$

由此进行迭代:

1. $x^{(0)} = 1$ 时

$x^{(0)}$	1
$x^{(1)}$	-0.8134
$x^{(2)}$	0.4094
$x^{(3)}$	-0.04730
$x^{(4)}$	7.060e-5
$x^{(5)}$	-2.347e-13

2. $x^{(0)} = 1.1$ 时

$x^{(0)}$	1.1
$x^{(1)}$	-1.129
$x^{(2)}$	1.234
$x^{(3)}$	-1.695
$x^{(4)}$	5.715
$x^{(5)}$	-23021

3. 总变分 (total variation) 是视听处理领域一种经典的去除噪声的方法。

假设输入的被噪声污染的信号用列向量表示为 $x \in R^n$，用 x_i 表示 x^ 的第 i 个元素。则去噪之后的信号 $x \in R^n$ 通过下面的无约束优化问题求得，其中 $\epsilon > 0$,$\mu > 0$ 是两个常数。

$$minimize f(x) = \|x - \hat{x}\|_2^2 + \mu \sum_{i=1}^{n-1} \sqrt{\epsilon^2 + (x_{i+1} - x_i)^2} - \epsilon)$$

1) 若采用梯度下降法求解该问题，写出下降方向 Δx 的数学表达式 (明确给出 Δx 每个元素的表达形式)；

2) 若采用牛顿法求解该问题，写出计算下降方向 Δx 的线性方程组的数学表达式 (明确给出系数矩阵每个元素的表达形式)；

3) 分析求解问题 (2) 中的线性方程组的时间复杂度。

(1) 在梯度下降法中，$\Delta x = -\nabla f(x)$，故有：

$$\frac{\partial f}{\partial x_1} = 2(x_1 - \hat{x_1}) + \mu \frac{x_1 - x_2}{\sqrt{\epsilon^2 + (x_2 - x_1)^2}}$$

$$\frac{\partial f}{\partial x_i} = 2(x_i - \hat{x_i}) + \mu \frac{x_i - x_{i+1}}{\sqrt{\epsilon^2 + (x_{i+1} - x_i)^2}} + \mu \frac{x_i - x_{i-1}}{\sqrt{\epsilon^2 + (x_i - x_{i-1})^2}}, \quad for\ 1 < i < n$$

$$\frac{\partial f}{\partial x_n} = 2(x_n - \hat{x_n}) + \mu \frac{x_n - x_{n-1}}{\sqrt{\epsilon^2 + (x_n - x_{n-1})^2}}$$

(2) 在牛顿法中，$\nabla^2 f(x) \Delta x = -\nabla f(x)$，故 Δx 可通过解此方程求得。$\nabla f(x)$ 上一问中已经计算过，本问中计算海森矩阵 $\nabla^2 f(x)$ 即可：

$\nabla^2 f(x)$ 首行为 $\frac{\partial^2 f(x)}{\partial x_1 \partial x_i}$, 故可知该行第一个元素为 $2 + \mu\epsilon^2(\epsilon^2 + (x_2 - x_1)^2)^{-\frac{3}{2}}$, 第二个元素为 $-\mu\epsilon^2(\epsilon^2 + (x_2 - x_1)^2)^{-\frac{3}{2}}$，其余元素为 0

$\nabla^2 f(x)$ 中间第 i 行 ($1 < i < n$) 为 $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}(1 \le j \le n)$，故可知对于这第 i 行，第 i-1 个元素为 $-\mu\epsilon^2(\epsilon^2 + (x_i - x_{i-1})^2)^{-\frac{3}{2}}$，第 i 个元素为 $2 + \mu\epsilon^2(\epsilon^2 + (x_{i+1} - x_i)^2)^{-\frac{3}{2}} + \mu\epsilon^2(\epsilon^2 + (x_i - x_{i-1})^2)^{-\frac{3}{2}}$，第 i+1 个元素为 $-\mu\epsilon^2(\epsilon^2 + (x_{i+1} - x_i)^2)^{-\frac{3}{2}}$，其余元素为 0

$\nabla^2 f(x)$ 最后一行 (即第 n 行) 为 $\frac{\partial^2 f(x)}{\partial x_1 \partial x_i}$，故可知该行第 n-1 个元素为 $-\mu\epsilon^2(\epsilon^2 + (x_n - x_{n-1})^2)^{-\frac{3}{2}}$，第 n 个元素为 $2 + \mu\epsilon^2(\epsilon^2 + (x_n - x_{n-1})^2)^{-\frac{3}{2}}$，其余元素为 0

(3) 由 (2) 中分析过程可知，对于本问题中的海森矩阵，除首尾两行各 2 个元素非 0 外，每行只有两个元素非 0，故可直接使用高斯消元法求解，时间复杂度是 O(n) 级别的。

4. x 为二维非负向量, 证明:

双曲集合 $\{x \in R_+^2 | x_1 x_2 \geq 1\}$ **是凸集合**

假设 (x_1, x_2) 和 (y_1, y_2) 都属于该双曲集合，那么对于 $\theta \in [0,1]$，有:

$$z_1 = \theta x_1 + (1-\theta) y_1 \ , \ z_2 = \theta x_2 + (1-\theta) y_2$$

若 $x_1 \geq y_1$ 且 $x_2 \geq y_2$，那么显然有 $z_1 \geq y_1$, $z_2 \geq y_2$，则可得 $z_1 z_2 \geq y_1 y_2 \geq 1$

若 $y_1 \geq x_1$ 且 $y_2 \geq x_2$，那么显然有 $z_1 \geq x_1$, $z_2 \geq x_2$，则可得 $z_1 z_2 \geq x_1 x_2 \geq 1$

而另外的两种情况必然会有 $(y_1 - x_1)(y_2 - x_2) < 0$，则可求得:

$$\begin{aligned}
z_1 z_2 &= (\theta x_1 + (1-\theta) y_1)(\theta x_2 + (1-\theta) y_2) \\
&= \theta^2 x_1 x_2 + (1-\theta)^2 y_1 y_2 + \theta(1-\theta)(y_1 x_2 + x_1 y_2) \\
&= \theta x_1 x_2 + (1-\theta) y_1 y_2 - \theta(1-\theta)(y_1 - x_1)(y_2 - x_2) \\
&\geq 1
\end{aligned}$$

综上，可以证明该双曲集合是凸集合。

5. 已知 R^{m+n} 上的两个凸集合 S_1 和 S_2

证明它们的部分和集合 S 也是凸集合: $S = \{(x, y_1 + y_2) | x \in R^m, y_1, y_2 \in R^n, (x, y_1) \in S_1, (x, y_2) \in S_2\}$

取 $(\bar{x}, \bar{y_1}) \in S_1$, $(\bar{x}, \bar{y_2}) \in S_2$, $(\hat{x}, \hat{y_1}) \in S_1$, $(\hat{x}, \hat{y_2}) \in S_2$，则

$(\bar{x}, \bar{y_1} + \bar{y_2}) \in S$, $(\hat{x}, \hat{y_1} + \hat{y_2}) \in S$

对于 $\theta \in [0,1]$, 有 $\theta(\bar{x}, \bar{y_1} + \bar{y_2}) + (1-\theta)(\hat{x}, \hat{y_1} + \hat{y_2}) = (\theta\bar{x} + (1-\theta)\hat{x} \ , (\theta\bar{y_1} + (1-\theta)\hat{y_1}) + (\theta\bar{y_2} + (1-\theta)\hat{y_2}))$

而已知 S_1 和 S_2 都是凸集,故有 $(\theta\bar{x} + (1-\theta)\hat{x}, \theta\bar{y_1} + (1-\theta)\hat{y_1}) \in S_1$, $(\theta\bar{x} + (1-\theta)\hat{x}, \theta\bar{y_2} + (1-\theta)\hat{y_2}) \in S_2$

所以可以得到 $(\theta\bar{x} + (1-\theta)\hat{x} \ , (\theta\bar{y_1} + (1-\theta)\hat{y_1}) + (\theta\bar{y_2} + (1-\theta)\hat{y_2})) \in S$，由此得证。

6. 连续函数 $f : R^n \to R$

证明 f 为凸函数当且仅当对任意 $x, y \in R^n$**, 下式成立:** $\int_0^1 f(x + \lambda(y - x)) \mathrm{d}\lambda \leq \frac{f(x) + f(y)}{2}$

先证 f 为凸函数时，该式成立:

根据 Jensen's 不等式，对于 $\forall \lambda \in [0,1]$ 有

$$f(\lambda y + (1-\lambda) x) \leq \lambda f(y) + (1-\lambda) f(x)$$

两边在 $[0,1]$ 上积分可得:

$$\int_0^1 f(\lambda y+(1-\lambda)x)d\lambda \le \int_0^1 (\lambda f(y)+(1-\lambda)f(x))d\lambda = \frac{f(x)+f(y)}{2}$$

此即 $\int_0^1 f(x+\lambda(y-x))d\lambda \le \frac{f(x)+f(y)}{2}$

再证 f 不是凸函数时，该式一定不成立:

若 f 不是凸函数，则 Jensen's 不等式不成立，即 $\exists x, y, \lambda_0 \in (0,1)$，使得 $f(\lambda_0 x+(1-\lambda_0)y) > \lambda_0 f(x)+(1-\lambda_0)f(y)$

考虑连续函数 $F(\lambda)=f(\lambda x+(1-\lambda)y)-\lambda f(x)-(1-\lambda)f(y)$, 显然仅当 $\lambda=0$ 或 $\lambda=1$ 时, $F(\lambda)=0$, 对于 λ_0，有 $F(\lambda_0)>0$

若记 α , β 分别为 λ_0 左侧和右侧最近的零点，则在区间 (α,β) 上，有

$$f(\lambda x+(1-\lambda)y) > \lambda f(x)+(1-\lambda)f(y)$$

令 $u=\alpha x+(1-\alpha)y$, $v=\beta x+(1-\beta)y$，则对于 $\lambda\in(0,1)$ 有

$$f(\lambda u+(1-\lambda)v) > \lambda f(u)+(1-\lambda)f(v)$$

两边在 $[0,1]$ 上积分可得:

$$\int_0^1 f(v+\lambda(u-v))d\lambda > \int_0^1 (f(v)+\lambda(f(u)-f(v)))d\lambda = \frac{f(u)+f(v)}{2}$$

这与题设条件相矛盾，由此可以说明在 f 不是凸函数时，该式一定不成立。

综上可知，当且仅当 f 为凸函数是，该式成立。

7. 两个元素取值为正数的向量 x, y 满足:

$x, y\in R^n_{++}$，其相对熵定义如下，其中 x_k, y_k 分别是向量 x,y 的第 k 个元素，log 表示自然对数:

$$\sum_{k=1}^{n} x_k \log(\frac{x_k}{y_k})$$

1) 证明相对熵是变量 x,y 的凸函数;

2) 假设 y 是一个给定的常数向量，定义如下的等式约束优化问题，其中 $A\in R^{m\times n}$, $b\in R^m$:

$$minimize\ \sum_{k=1}^{n} x_k \log(\frac{x_k}{y_k})$$

$$subject\ to\ Ax=b$$

$$\sum_{k=1}^{n} x_k = 1$$

证明该优化问题的对偶问题是如下形式，其中 a_k 是矩阵 A 的第 k 列:

$$minimize\ b^T z + \log \sum_{k=1}^{n} y_k e^{-a_k^T z}$$

1. 先证明相对熵是 x 的凸函数。记相对熵为 $f(x,y) = \sum_{k=1}^{n} x_k \log \frac{x_k}{y_k}$

   令 $g(x) = x \log \frac{x}{y}$，其中 y 此时是常数参量，则有 $g^{''}(x) = \frac{1}{x}$，显然在 $x > 0$ 时，$g(x)$ 是凸函数

   对于 $\forall \theta \in [0,1]$ 有

$$\begin{aligned}
\theta f(x_1, y) + (1-\theta) f(x_2, y) &= \theta \sum_{k=1}^{n} x_{1k} \log \frac{x_{1k}}{y_k} + (1-\theta) \sum_{k=1}^{n} x_{2k} \log \frac{x_{2k}}{y_k} \\
&= \sum_{k=1}^{n} (\theta x_{1k} \log \frac{x_{1k}}{y_k} + (1-\theta) x_{2k} \log \frac{x_{2k}}{y_k}) \\
&\geq \sum_{k=1}^{n} [\theta x_{1k} + (1-\theta) x_{2k}] \log \frac{[\theta x_{1k} + (1-\theta) x_{2k}]}{y_k} \\
&= f(\theta x_1 + (1-\theta) x_2, y)
\end{aligned}$$

   由此可说明 $f(x,y)$ 是 x 的凸函数。

   接下来证明相对熵是 y 的凸函数。此时可以将相对熵写作如下形式:

$$f(x,y) = \sum_{k=1}^{n} x_k \log \frac{x_k}{y_k} = \sum_{k=1}^{n} x_k \log x_k - \sum_{k=1}^{n} x_k \log y_k$$

   令 $h(y) = -x \log y$，其中 x 此时是常数参量，则有 $h^{''}(y) = \frac{1}{y^2}$，显然在 $y > 0$ 时，$h(x)$ 是凸函数

   对于 $\forall \theta \in [0,1]$ 有

$$\begin{aligned}
\theta f(x, y_1) + (1-\theta) f(x, y_2) &= \sum_{k=1}^{n} (x_k \log x_k - \theta x_k \log y_{1k} - (1-\theta) x_k \log y_{2k}) \\
&\geq \sum_{k=1}^{n} (x_k \log x_k - x_k \log(\theta y_{1k} + (1-\theta) y_{2k})) \\
&= f(x, \theta y_1 + (1-\theta) y_2)
\end{aligned}$$

   由此可说明 $f(x,y)$ 是 y 的凸函数。

2. 对于等式约束优化问题，有 $L(x,\lambda,\mu) = \sum_{k=1}^{n} x_k \log \frac{x_k}{y_k} + \lambda^T (Ax - b) + \mu(\sum_{k=1}^{n} x_k - 1)$

   令 $\frac{\partial L}{\partial x_k} = 0$, $\frac{\partial L}{\partial \lambda} = 0$, $\frac{\partial L}{\partial \mu} = 0$，则有

$$\log \frac{x_k}{y_k} = -1 - \mu - \lambda^T a_k \tag{1}$$

$$Ax - b = 0 \tag{2}$$

$$\sum_{k=1}^{n} x_k - 1 = 0 \tag{3}$$

将 (1)(2)(3) 三式代入 $L(x,\lambda,\mu)$ 可得原问题的对偶形式：

$$\min_{x}\sum_{k=1}^{n}x_k\log\frac{x_k}{y_k}\iff\max_{\lambda,\mu}g(\lambda,\mu)$$

其中，$g(\lambda,\mu)$ 满足：

$$\begin{aligned}g(\lambda,\mu)=L(x,\lambda,\mu)&=\sum_{k=1}^{n}x_k(-1-\mu-\lambda^Ta_k)\\&=-\sum_{k=1}^{n}x_ka_K^T\lambda-(1+\mu)\\&=-(AX)^T\lambda-(1+\mu)\end{aligned}$$

又由 (1) 知：$x_k=y_ke^{-1-\mu-\lambda^Ta_k}$，由此可得 $\mu+1=\log\sum_{k=1}^{n}y_ke^{-a_K^T\lambda}$

故 $g(\lambda,\mu)=-b^T\lambda-\log\sum_{k=1}^{n}y_ke^{-a_K^T\lambda}$

由此可得，原问题等价于 $\min_\lambda(-b^T\lambda-\log\sum_{k=1}^{n}y_ke^{-a_K^T\lambda})$，此即 $\min_\lambda(b^T\lambda+\log\sum_{k=1}^{n}y_ke^{-a_K^T\lambda})$，证毕。