Homework 2

Lecturer: Xiangyu Chang *Scribe:* 赵敬业

Edited by: Junbo Hao

1 HW 1

1.1 问题重述

在项目组合管理示例中，我们将操作问题表示为：

$$\max_{\mathbf{x}} \mathbb{E}(R) - \lambda \operatorname{Var}(R) \tag{1}$$

$$\text{s.t. } x_i \geq 0, i = 1 \ldots, n \tag{2}$$

$$\sum_{i=1}^{n} x_i = 1 \tag{3}$$

其中 $\mathbb{E}(R)$ 是 R 的预期收益率；$\operatorname{Var}(R)$ 是 R 的方差；$\lambda > 0$ 被称为风险厌恶参数，用于平衡投资风险和预期收益。我们进一步假设 $\mathbf{r} = (r_1, \ldots, r_n)^\top$ 是一个期望为 $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ 的随机向量，其协方差为 Σ.

请证明原优化问题等价于：

$$\max_{\mathbf{x}} \boldsymbol{\mu}^\top \mathbf{x} - \lambda \mathbf{x}^\top \Sigma \mathbf{x}, \tag{4}$$

$$\text{s.t. } x_i \geq 0, i = 1 \ldots, n, \tag{5}$$

$$\sum_{i=1}^{n} x_i = 1 \tag{6}$$

1.2 证明

$$\mathbb{E}(r) = \mu \tag{7}$$

$$R = \sum_{i=1}^{n} (r_{it} * x_i) \tag{8}$$

$$\mathbb{E}(R) = \mathbb{E}(\sum_{i=1}^{n} (r_{it} * x_i)) = \sum_{i=1}^{n} (\mathbb{E}(r_{it} * x_i)) = \sum_{i=1}^{n} (x_i * \mathbb{E}(r_{it})) = \sum_{i=1}^{n} (x_i * \mu_i) = \boldsymbol{\mu}^\top \boldsymbol{x} \tag{9}$$

$$\operatorname{Var}(R) = \operatorname{Var}(\sum_{i=1}^{n} (r_{it} * x_i)) = \sum_{i=1}^{n} (\operatorname{Var}(r_{it} * x_i)) = \sum_{i=1}^{n} x_i^2 (\operatorname{Var}(r_{it})) = \sum_{i=1}^{n} x_i^2 \sigma_i = \sum_{i=1}^{n} x_i \sigma_i x_i = \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}. \tag{10}$$

QED

2 HW 2

2.1 问题重述

泊松回归：假设现在 b_i 是计数值，$\mathbf{a}_i$ 是 $\mathbb{R}^p$ 的一些预测值。设 λ_i 为某一时间段内的预期事件数，即得：

$$P\left(b_i \mid \lambda_i\right)=\frac{e^{-\lambda_i} \lambda_i^{b_i}}{b_i!} \tag{11}$$

以广义线性模型为例，推导出下面的最优化形式来表示简单的泊松回归：

$$\min_{\mathbf{x}} \sum_{i=1}^m \exp \left(\left\langle\mathbf{a}_i, \mathbf{x}\right\rangle\right)-\langle\mathbf{b}, A \mathbf{x}\rangle . \tag{12}$$

2.2 证明

假设 $\ln \lambda_i=\left\langle\mathbf{a}_i, \mathbf{x}\right\rangle$ 该模型使用最大似然估计，估计公式如下：

$$\max_\lambda \prod_{i=1}^m \frac{e^{-\lambda_i} \lambda_i^b{}_i}{b_i!} \tag{13}$$

$$\Leftrightarrow \max_\lambda \ln \prod_{i=1}^m \frac{e^{-\lambda_i} \lambda_i^b{}_i}{b_i!} \tag{14}$$

$$=\max_\lambda \sum_{i=1}^m \ln \frac{e^{-\lambda_i} \lambda_i^b{}_i}{b_i!} \tag{15}$$

$$\Leftrightarrow \max_\lambda \sum_{i=1}^m\left(-\lambda_i\right)+b_i \ln \left(\lambda_i\right) \tag{16}$$

$$\Leftrightarrow \min_\lambda \sum_{i=1}^m\left(\lambda_i\right)-b_i \ln \left(\lambda_i\right) . \tag{17}$$

带入 $\ln \lambda_i=\left\langle\mathbf{a}_i, \mathbf{x}\right\rangle$，得

$$\min_{\mathbf{x}} \sum_{i=1}^m\left(\exp \left\langle\mathbf{a}_i, \mathbf{x}\right\rangle\right)-b_i\left\langle\mathbf{a}_i, \mathbf{x}\right\rangle . \tag{18}$$

注意到：

$$\sum_{i=1}^m b_i\left\langle\mathbf{a}_i, \mathbf{x}\right\rangle=\langle\mathbf{b}, \mathbf{A x}\rangle . \tag{19}$$

原式 $=$

$$\min_{\mathbf{x}} \sum_{i=1}^m\left(\exp \left\langle\mathbf{a}_i, \mathbf{x}\right\rangle\right)-\langle\mathbf{b}, \mathbf{A x}\rangle . \tag{20}$$

QED

3 HW 3

3.1 问题重述

设

$$f(\mathbf{x})=\sum_{i=1}^{m}\log\left(1+\exp\left(\langle\mathbf{a}_i,\mathbf{x}\rangle\right)\right)-\langle\mathbf{b},A\mathbf{x}\rangle \tag{21}$$

为逻辑回归的目标函数，请计算梯度和 $f(\mathrm{x})$ 的 Hessian 矩阵。

3.2 证明

$$f(\mathbf{x})=\sum_{i=1}^{m}\log\left(1+\exp\left(\langle\mathbf{a}_i,\mathbf{x}\rangle\right)\right)-\langle\mathbf{b},A\mathbf{x}\rangle. \tag{22}$$

假设：

$$z_i=\langle\mathbf{a}_i,\mathbf{x}\rangle\, g(z_i)=\sum_{i=1}^{m}\log\left(1+\exp\left(z_i\right)\right)-b_i z_i \tag{23}$$

$$G(z)=\sum_{i=1}^{m}g(z_i) \tag{24}$$

$$\nabla G(z)_i=\left(\frac{\exp\left(z_i\right)}{1+\exp\left(z_i\right)}-b_i\right). \tag{25}$$

由通用公式:

$$\nabla f(x)=A^{\top}\nabla G(z)=(a_1,a_2,\ldots,a_m)\begin{pmatrix}\frac{\exp(\langle\mathbf{a}_1,\mathbf{x}\rangle)}{1+\exp(\langle\mathbf{a}_1,\mathbf{x}\rangle)}-b_1\\ \frac{\exp(\langle\mathbf{a}_2,\mathbf{x}\rangle)}{1+\exp(\langle\mathbf{a}_2,\mathbf{x}\rangle)}-b_2\\ \vdots\\ \frac{\exp(\langle\mathbf{a}_m,\mathbf{x}\rangle)}{1+\exp(\langle\mathbf{a}_m,\mathbf{x}\rangle)}-b_m\end{pmatrix}. \tag{26}$$

其中第 i 个

$$\frac{\partial f(x)}{\partial x}=\sum_{j=1}^{m}a_{ij}\left(\frac{\exp\left(\langle\mathbf{a}_j,\mathbf{x}\rangle\right)}{1+\exp\left(\langle\mathbf{a}_j,\mathbf{x}\rangle\right)}-b_j\right)=\sum_{j=1}^{m}a_{ij}\frac{\exp\left(\langle\mathbf{a}_j,\mathbf{x}\rangle\right)}{-}(1+\exp\left(\langle\mathbf{a}_j,\mathbf{x}\rangle\right)-\sum_{j=1}^{m}a_{ij}b_j. \tag{27}$$

以上就是其梯度，下面计算海森矩阵：
假设：

$$t_i=\langle\mathbf{a}_i,\mathbf{x}\rangle\,\phi(t_i)=\sum_{j=1}^{m}a_{ij}\left(\frac{\exp\left(t_j\right)}{1+\exp\left(t_j\right)}-b_j\right) \tag{28}$$

$$\Phi(t)=\sum_{j=1}^{m}\phi(t_j). \tag{29}$$

由通用公式，得：

$$\nabla^2 f(x)_i = A^\top \nabla\Phi(t) = (a_1, a_2, \dots, a_m)\begin{pmatrix} a_{i1}\frac{\exp\langle \mathbf{a}_1,\mathbf{x}\rangle}{(1+\exp\langle \mathbf{a}_1,\mathbf{x}\rangle)^2} \\ a_{i2}\frac{\exp\langle \mathbf{a}_2,\mathbf{x}\rangle}{(1+\exp\langle \mathbf{a}_2,\mathbf{x}\rangle)^2} \\ \vdots \\ a_{im}\frac{\exp\langle \mathbf{a}_m,\mathbf{x}\rangle}{(1+\exp\langle \mathbf{a}_m,\mathbf{x}\rangle)^2} \end{pmatrix}. \tag{30}$$

特别地：

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \sum_{j=1}^{m} a_{ij}^2 \left(\frac{\exp\left(\langle \mathbf{a}_j, \mathbf{x}\rangle\right)}{\left(1+\exp\left(\langle \mathbf{a}_j, \mathbf{x}\rangle\right)\right)^2}\right). \tag{31}$$

4 HW 4

4.1 问题重述

考虑函数 $f(x) = x_1^2 + x_2^2, x = (x_1, x_2) \in \mathbb{R}^2$, 以及迭代点列 $x^k = (1+\frac{1}{2^k})(\cos k, \sin k)^{\mathrm{T}}, k = 1, 2, \cdots,$ 请说明：

(a) $\left\{f\left(x^{k+1}\right)\right\}$ 是否收敛? 若收敛，给出 Q-收敛速度;

(b) $\left\{x^{k+1}\right\}$ 是否收敛? 若收敛，给出 Q-收敛速度.

4.2 解

(a) 收敛

$$\lim_{k\to\infty} \frac{\left\|f(x^{k+1}) - f(x^*)\right\|}{\left\|f(x^k) - f(x^*)\right\|} \tag{32}$$

$$= \frac{\left\|\left(\left(1+\frac{1}{2^{k+1}}\right)\cos(k+1)\right)^2 + \left(\left(1+\frac{1}{2^{k+1}}\right)\sin(k+1)\right)^2 - 1\right\|}{\left\|\left(1+\frac{1}{2^k}\right)\cos^2 k + \left(1+\frac{1}{2^{k+1}}\right)\sin^2 k - 1\right\|} \tag{33}$$

$$= \frac{\left(1+\frac{1}{2^{k+1}}\right)^2 - 1}{\left(1+\frac{1}{2^k}\right)^2 - 1} = \frac{2+\frac{1}{2^{k+1}}}{2+\frac{1}{2^k}}\frac{1}{2} = \frac{1}{2}. \tag{34}$$

收敛速度 $\frac{1}{2}$
(b) 不收敛

$$\lim_{k\to\infty} x^k = (\cos k, \sin k)^\top \tag{35}$$

不收敛

5 HW 5

5.1 问题重述

请通过 Jacobi 和 Gauss-Seidel 迭代算法求解以下方程, 并讨论这两个方程的收敛条件：

$$A\mathrm{x} = \mathrm{b} \tag{36}$$

(a) $A = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{pmatrix}$ and $\mathbf{b} = (2, 12, 2)^\top$.

(b) $A = \begin{pmatrix} 8 & -3 & 2 \\ 4 & 11 & -1 \\ 2 & 1 & 4 \end{pmatrix}$ and $\mathbf{b} = (20, 33, 12)^\top$.

5.2 证明

使用 python 代码
为主文档简洁起见，请移步附录观看源代码及其求解结果。出现了如下结果：
当使用 jacobi 迭代时，第一问不收敛，第二问收敛, 当使用 gauss 迭代时，均收敛至 x^*
经过计算，发现雅各比迭代在第一问中的谱半径为 1.5>1，不能保证它收敛，具体还需进一步讨论；第二问谱半径 0.3592<1，收敛；
而高斯迭代第一问谱半径 0.25, 第二问谱半径 0.13 均小于 1，故收敛。

参考文献