最优化 2 第一次作业

大数据 001

郅啸淇

学号：2184114639

2022 年 9 月 21 日

目录

1 HW1

由题中条件有

$E(\mathbf{r}) = (\mu)$

$E(R) = E(\mathbf{r}^T\mathbf{x})$

$= \mu^T\mathbf{x}$

对于另一项

$Var(R) = Var(\mathbf{r}^T\mathbf{x})$

$= x^2Var(r) = x^2\sum$

$= \mathbf{x}^T\sum\mathbf{x}$

故原题中两问题等价

2 HW2

从泊松回归推出题中问题

$\bullet b_i \sim P(b_i|\lambda_i)$

$P(b_i|\lambda_i) = \frac{e^{-\lambda_i}\lambda_i^{b_i}}{b_i!}$

$= \frac{e^{-\lambda_i+b_i ln\lambda_i}}{b_i!}$

取分子做优化问题

$\underset{x}{max}\prod_{i=1}^{m}P(b_i|\lambda_i)$

$\Rightarrow \underset{x}{max}\prod_{i=1}^{m}e^{-\lambda_i+b_i ln\lambda_i}$

$\Rightarrow \underset{x}{max}\sum_{i=1}^{m}-\lambda_i+b_i ln\lambda_i$

令 $ln\lambda_i = \langle\alpha_i, x\rangle$

即 $\Rightarrow \underset{x}{max}\sum_{i=1}^{m}-e^{\langle\alpha_i,x\rangle}+b_i\langle\alpha_i, x\rangle$

$\Rightarrow \underset{x}{min}\sum_{i=1}^{m}e^{\langle\alpha_i,x\rangle}-\mathbf{b}^T\mathbf{Ax}$

故推导完成

3 HW3

3.1 Gradient Descent Method

$$\nabla f(x) = (Ax-b)\mathbf{A}^T$$
$$= \mathbf{A}^T(Ax-b)$$
$$x^{t+1} = x^t - St\mathbf{A}^T(Ax-b)$$

因为 $\beta - smooth$ 故 $St = \frac{1}{\beta}$

3.2 Newton-Raphson

$$x^{t+1} = x^t - St(\nabla^2 f(x^t))^{-1}\nabla f(x^t)$$
$$\nabla^2 f(x^t) = A^T A$$

3.3 Sub-Gradient

$$x^{t+1} = x^t - Stg^t$$

其中 $g^t \in \partial f(x^t)$

$$x^{t+1} = x^t - St(A^T(Ax-b)+\lambda)$$

3.4 Proximal Gradient Descent

令 $g(x) = \frac{1}{2}\|Ax-b\|^2, h(x) = \lambda\|x_1\|$

$$x^{k+1} = prox_{th(x)}(x^k - t\nabla(\tfrac{1}{2}\|Ax^k-b\|^2))$$
$$= S_{\lambda t}(x^k - tA^T(Ax-b))$$

其中 $S_{\lambda t} = \begin{cases} x-\lambda t & if x > \lambda t \\ 0 & if|x| \leq \lambda t \\ x+\lambda t & if x < -\lambda t \end{cases}$

4 HW4

分别取 General Optimality Condition 中 y 为 $2x^*, 0$，得到

$\langle \nabla f(x^*), 0 - x^* \rangle \geq 0$

$\langle \nabla f(x^*), 2x^* - x^* \rangle \geq 0$

得到 $-\nabla f(x^*)x^* \geq 0$ 并且 $\nabla f(x^*)x^* \geq 0$

故 $\nabla f(x^*)x^* = 0$

另由 General Optimality Condition 有

$\nabla f(x^*)(y - x^*) \geq 0$

$\Rightarrow \nabla f(x^*)y \geq 0$

因为 $y \in domf$，即 $y \leq 0$

故 $\nabla f(x^*) \leq 0$

得证。

5 HW5

构造 Lagrange 函数

$L(\omega, b, \xi, \alpha, \mu) = \frac{\omega^T\omega}{2} + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i y_i(\omega^T x_i + b) - 1 + \xi_i - \sum_{i=1}^{n}\mu_i\xi_i$

$g(\alpha, \mu) = \underset{\omega,b,\xi}{max}\, L(\omega, b, \xi, \alpha, \mu)$

$s.t. \alpha \geq 0, \mu \geq 0$

对各变量求梯度并令其为 0

$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^{n}\alpha_i y_i x_i = 0$

$\frac{\partial L}{\partial \xi} = C - \alpha - \mu = 0$

$\frac{\partial L}{\partial b} = -\sum_{i=1}^{n}\alpha_i y_i = 0$

将上三式代入对偶函数

$g(\alpha, \mu) = \frac{1}{2}\omega^T\omega - \sum_{i=1}^{n}\alpha_i y_i \omega^T x_i + \alpha$

因为 $\omega = \sum_{i=1} n\alpha_i y_i x_i$

$g(\alpha, \mu) = -\frac{1}{2}\omega^T\omega + \alpha$

$= -\frac{1}{2}\sum_{i=1} n \sum_{j=1} n\alpha_i\alpha_j y_i y_j x_i^T x_j + \alpha$

$= -\frac{1}{2}\alpha^T \mathbf{Q}\alpha + \alpha$

其中 $\mathbf{Q}_{ij} = y_i y_j x_i^T xj$

故 SVM 对偶问题为

$\underset{\alpha}{min}\, \frac{1}{2}\alpha^T \mathbf{Q}\alpha - \alpha$

$s.t. \alpha \geq 0, \mu \geq 0, C - \alpha - \mu = 0, \alpha^T y = x$