最优化第二次作业

大数据 001

郅啸淇

学号：2184114639

2022 年 6 月 11 日

目录

1 强凸函数相互等价性

1.1 Monotonicity

因为 $\nabla g(x)$ 为单调函数，故有 $(\nabla g(y) - \nabla g(x))^T(y-x) \geq 0$

代入 $g(x) = f(x) - \frac{m}{2}x^Tx$，有 $(\nabla f(y) - my - \nabla f(x) + mx)^T(y-x) \geq 0$

即 $((\nabla f(y) - \nabla f(x)) - m(y-x))^T(y-x) \geq 0$

即 $(\nabla f(y) - \nabla f(x))^T \geq m\|y-x\|_2^2$

Monotonicity 得证

1.2 Jensen's Inequality

由 $g(x)$ 为凸函数，有 $g(\theta x + (1-\theta)y) \leq \theta g(x) + (1-\theta)g(y)$，代入 $g(x) = f(x) - \frac{m}{2}x^Tx$ 有

$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) - \frac{m}{2}(\theta x^Tx + (1-\theta)y^Ty) + \frac{m}{2}(\theta x + (1-\theta)y)^T(\theta x + (1-\theta)y)$

即为 $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) - \frac{m}{2}(1-\theta)\theta(x^Tx + y^Ty - (xy^T + x^Ty))$

即为 $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) - \frac{m}{2}(1-\theta)\theta\|x-y\|_2^2$

Jensen's Inequality 得证

1.3 二阶强凸条件

因为 $g(x)$ 为凸函数，故有 $\nabla^2 g(x) \geq 0$

代入 $g(x) = f(x) - \frac{m}{2}x^Tx$，有 $\nabla^2(f(x) - \frac{m}{2}x^Tx) = \nabla^2 f(x) - mI \geq 0$

即 $\nabla^2 f(x) \geq mI$，得证

2 无穷范数的对偶范数是一范数

对偶范数定义：$\|u\|_* = \sup\limits_{v\neq 0}\frac{u^Tv}{\|v\|} = \sup\limits_{\|v\|=1} u^Tv$

再结合霍尔德不等式，若有 $\frac{1}{p}+\frac{1}{q}=1$

则 $z^Tx \leq \|x\|_p\|z\|_q$

$\|z\|_* = \sup\limits_{x\neq 0}\frac{z^Tx}{\|x\|}_p = \|z\|_q$

当 $p=1$ 时，根据霍尔德不等式，$q=\infty$

得证

3 临近算子

3.1 二次函数

$u = prox^{th}(x)$ 最优性条件为

$x-u \in t\partial h(x)$

即 $x-u \in t(Au+b)$

即 $x-u = tAu+tb$

得到 $u=(x-tb)(I+tA)^{-1}$

3.2 负自然对数的和

有 $h(x) = -\sum_{n=1}^{n} lnx_i$，$u_i = prox_{th}(x_i)$ 最优性条件为

$x-u \in t\partial h(u_i)$

即 $x-u_i+\frac{t}{u_i}=0$，解得 $u_i = \frac{x_i+\sqrt{x_i^2+4t}}{2}$

4 临近算子基本运算规则

4.1 Linear Combination

$prox_f(x) = \arg\min_z\{\frac{1}{2}\|z-x\|_2^2 + f(z)\}$

代入 $f(x) = ag(x) + b$，有 $prox_{f(x)} = \arg\min_z\{\frac{1}{2}\|z-x\|_2^2 + ag(z) + b\}$

即 $f(x) = ag(x)+b$, $prox_{f(x)} = \arg\min_z\{\frac{1}{2}\|z-x\|_2^2+ag(z)\} = prox_{ag}(x)$

得证

4.2 Affine Addition

$prox_f(x) = \arg\min_z\{\frac{1}{2}\|z-x\|_2^2 + a^T z + g(z)\}$

$= \arg\min_z\{\frac{1}{2}\|z\|_2^2 - < z, x > + a^T z + g(z)\}$

$= \arg\min_z\{\frac{1}{2}\|z\|_2^2 - < z, x-a > + g(z)\}$

$= \arg\min_z\{\frac{1}{2}\|z-(x-a)\|_2^2 + g(z)\}$

$= prox_g(x-a)$

得证

4.3 Quadratic Addition

$prox_(f)(x) = \arg\min_z\{\frac{1}{2}\|z-x\|_2^2 + g(z) + \frac{\rho}{2}\|z-a\|_2^2\}$

$= \arg\min_z\{\frac{1+\rho}{2}\|z-x\|_2^2 - < z, x+\rho a > + g(z)\}$

$= \arg\min_z\{\frac{1+\rho}{2}\|z-x\|_2^2 - \frac{1}{1+\rho} < z, x+\rho a > + \frac{1}{1+\rho}g(z)\}$

$= \arg\min_z\{\frac{1+\rho}{2}\|z-(\frac{1}{1+\rho}x + \frac{\rho}{1+\rho}a)\|_2^2 + \frac{1}{1+\rho}g(z)\}$

$= prox_{\frac{1}{1+\rho}g}(\frac{1}{1+\rho}x + \frac{\rho}{1+\rho}a)$

得证

4.4 Scaling and Translation

如果 $f(\mathbf{x}) = g(a\mathbf{x}+\mathbf{b})+$ 有 $a \neq 0$ 那么

$prox_f(\mathbf{x}) = arg\min_z \left\{\frac{1}{2}\|\mathbf{z}-\mathbf{x}\|_2^2 + f(\mathbf{z})\right\}$

$= arg\min_z \left\{\frac{1}{2}\|\mathbf{z}-\mathbf{x}\|_2^2 + g(a\mathbf{z}+\mathbf{b})\right\}$

$= arg\min_z \left\{\frac{1}{2}\|\mathbf{z}\|_2^2 - \langle\mathbf{z},\mathbf{x}\rangle + g(a\mathbf{z}+\mathbf{b})\right\}$

若 $\mathbf{t} = a\mathbf{z}+b$ 那么

$\mathbf{t} = arg\min_z \left\{\frac{1}{2}\|\frac{\mathbf{t}-\mathbf{b}}{a}-\mathbf{x}\|_2^2 + g(\mathbf{t})\right\}$

$= arg\min_z \left\{\frac{1}{2a^2}\|\mathbf{t}-(a\mathbf{x}+\mathbf{b})\|_2^2 + g(\mathbf{t})\right\}$

$= arg\min_z \left\{\frac{1}{2}\|\mathbf{t}-(a\mathbf{x}+\mathbf{b})\|_2^2 + a^2 g(\mathbf{t})\right\}$

$= prox_{a^2g}(a\mathbf{x}+\mathbf{b})$

$prox_f(\mathbf{x}) = \mathbf{z} = \frac{\mathbf{t}-\mathbf{b}}{a} = \frac{1}{a}\left[prox_{a^2g}(a\mathbf{x}+\mathbf{b}) - \mathbf{b}\right]$

得证

5 优化方法总结

	梯度法	次梯度法	近似点梯度法
步数	$O\left(\frac{1}{k}\right)$	$O\left(k^{\frac{1}{2}}\right)$	$O\left(\frac{1}{k}\right)$
收敛速率	$O\left(\frac{1}{\varepsilon}\right)$	$O\left(\frac{1}{\varepsilon^2}\right)$	$O\left(\frac{1}{\varepsilon}\right)$
若强凸	$O\left(log\left(\frac{1}{\varepsilon}\right)\right)$	$O\left(\frac{1}{\varepsilon}\right)$	$O\left(log\left(\frac{1}{\varepsilon}\right)\right)$

6 共轭函数总是凸函数

可以看成是一系列关于 y 的凸函数取上确界

等价于要证明：$f^*(\theta t_1 + (1-\theta)t_2) \le \theta f^*(t_1) + (1-\theta) f^*(t_2)$

代入共轭函数的定义，上述不等式等价于:

$\underset{x\in dom(f)}{max}\{(\theta t_1 + (1-\theta)t_2)x - f(x)\} \le \theta \underset{x\in dom(f)}{max}\{t_1 x - f(x)\} + (1-\theta)\underset{x\in dom(f)}{max}\{t_2 x - f(x)\}$

左式等价于：

$\underset{x\in dom(f)}{max}\{\theta(t_1 x - f(x)) + (1-\theta)(t_2 x - f(x))\}$

假设在 x_0 取到最大值，即：$\theta\left(t_1x_0-f\left(x_0\right)\right)+\left(1-\theta\right)\left(t_2x_0-f\left(x_0\right)\right)$

又：

$\theta\left(t_1x_0-f\left(x_0\right)\right)\leq \underset{x\in dom(f)}{max}\left\{\theta\left(t_1x_0-f\left(x_0\right)\right)\right\}$

$\left(1-\theta\right)\left(t_2x_0-f\left(x_0\right)\right)\leq \underset{x\in dom(f)}{max}\left\{\left(1-\theta\right)\left(t_2x_0-f\left(x_0\right)\right)\right\}$

所以原式得证

7 Lagrange 对偶问题总是凸优化问题

原问题：

$\underset{x}{min}f\left(x\right)$

$s.t.\ g_i\left(x\right)\leq 0, i=1,2,\cdots\cdots,m$

$h_j\left(x\right)=0, j=1,2,\cdots\cdots,p$

进行一次拉格朗日乘子法，转换为:

$\underset{x,\lambda,\nu}{min}L\left(x,\lambda,\nu\right)=f\left(x\right)+\sum_{i=1}^{m}\lambda_ig_i\left(x\right)+\sum_{i=1}^{p}\nu_ih_i\left(x\right)$

再求一次对偶问题：

$\underset{\lambda,\nu}{max}\underset{x}{min}L\left(x,\lambda,\nu\right)$

$s.t.\ \lambda\geq 0$

首先约束条件一定是凸函数，因为它是线性的，也就是说其实我们现在只要证明:$\underset{x}{min}L\left(x,\lambda,\nu\right)$ 是凸函数即可

$-\underset{x}{min}L\left(x,\lambda,\nu\right)=-k_1\lambda-k_2\nu-b$ if $k_1=g\left(x\right),k_2=h\left(x\right),b=f\left(x\right)$

上面跟 x xx 有关的全部是一个定值。这就是一个线性变化，或者说仿射变换。然后就用到了我们前面用到的那个重要定理：仿射集一定是凸集。所以说这是一个凸函数。

8 等式约束范数极小化

$f_0^*\left(y\right)=\underset{x}{sup}\left(y^Tx-\left\|x\right\|\right)$

根据：$\|v\|_* = sup_{\|u\|\leq 1} u^T visdualnormof \|\cdot\|$

又：$\|x\| = \underset{x}{sup}\left(y^T x\right), \|y\|_* \leq 1$

如果 $\|y\|_* \leq 1, y^T x \leq \|x\|$, 显然 x=0 时取等号

又：$\|y\|_* = \underset{x}{sup}\left(x^T y\right) > 1, \|x\| \leq 1$

如果 $\|y\|_* > 1, \exists x, \|x\| \leq 1, then x^T y > 1$

$f_0^*\left(y\right) \geq y^T\left(tx\right) - \|tx\| = t\left(y^T x - \|x\|\right) > 0$

$t \to \infty, f_0^*\left(y\right) \to \infty$

所以:$f_0^*\left(y\right) = \begin{cases} 0 & \|y\|_* \leq 1 \\ \infty & otherwise \end{cases}$ 等式约束范数极小化得证

9 SVM 的 KKT 条件

原问题

$$\begin{aligned} &\min_{\beta,\beta_0,\xi} \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n}\xi_i \\ &s.t. \xi_i \geq 0 i = 1, 2, 3...n \\ &y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i \end{aligned} \tag{1}$$

稳定性条件:$0 \in \partial_{\beta,\xi}(\frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n}\xi_i + u_i(1 - \xi_i - y_i(x_i^T\beta + \beta_0) + \xi_i))$

互补松弛条件:$u_i(1 - \xi_i - y_i(x_i^T\beta + \beta_0)) = 0$

原可行性:$1 - \xi_i - y_i(x_i^T\beta + \beta_0) \leq 0$

对偶可行性:$u_i \geq 0$

10 Rigde Regression 的对偶问题

原岭回归问题，引入约束条件

$$\min_x \frac{1}{2}\|Ax-b\|_2^2+\|x\|_2^2$$
$$s.t. x=z \tag{2}$$

有 $g(u)=\min_x \frac{1}{2}\|Ax-b\|_2^2+\|z\|_2^2+u^T(z-x)$

$=\min_x \frac{1}{2}\|Ax-b\|_2^2-u^Tx+\|z\|_2^2+u^Tz$

$=\min_x \frac{1}{2}\|Ax-b\|_2^2-((A^{-1})^Tu)^T(Ax-b)+\|z\|_2^2+u^Tz$

$=-\|(A^{-1})^Tu\|_{2*}^2-\|-u\|_{2*}^2-u^TA^{-1}b$

若有 $\|(A^{-1})^Tu\|_*\le 1, \|-u\|_*\le 1$

则 $g(u)=-u^TA^{-1}b$

对偶问题则为 $\max_u -u^TA^{-1}b$

11 逻辑回归 loss 为凸函数

逻辑岭回归

$\min_{\beta_0,\beta}\sum_{i=1}^n[-y_i(\beta_0+\beta^Tx_i)+log(1+e^{\beta_0+\beta^Tx_i})]+\lambda\|\beta\|_2^2$

其中 $y_i\in\{0,1\}, x_i\in R^p$

记 $R=-y_i(\beta_0+\beta^Tx_i)+log(1+e^{\beta_0+\beta^Tx_i})$

$\frac{\partial R}{\partial \beta_0}=-\sum_{i=1}^n y_i+\sum_{i=1}^n\frac{e^{\beta_0+\beta^Tx_i}}{1+e^{\beta_0+\beta^Tx_i}}$

$\frac{\partial^2 R}{\partial \beta_0^2}=\sum_{i=1}^n\frac{e^{\beta_0+\beta^Tx_i}}{(1+e^{\beta_0+\beta^Tx_i})^2}$

$\frac{\partial R}{\partial \beta}=-\sum_{i=1} nx_iy_i+\sum_{i=1}^n\frac{x_ie^{\beta_0+\beta^Tx_i}}{1+e^{\beta_0+\beta^Tx_i}}$

$\frac{\partial^2 R}{\partial \beta^2}=\sum_{i=1}^n\frac{x_ie^{\beta_0+\beta^Tx_i}}{(1+e^{\beta_0+\beta^Tx_i})^2}$

$\frac{\partial^2 R}{\partial \beta_0\partial\beta}=\frac{\partial^2 R}{\partial \beta_0\partial\beta_0}=\sum_{i=1}^n\frac{x_ie^{\beta_0+\beta^Tx_i}}{(1+e^{\beta_0+\beta^Tx_i})^2}$

设 $t_i=\frac{e^{\beta_0+\beta^Tx_i}}{(1+e^{\beta_0+\beta^Tx_i})^2}\ge 0$，则逻辑岭回归 Hessian 矩阵为

$H=\begin{pmatrix}\sum_{i=1}^n t_i, & \sum_{i=1}^n x_it_i\\ \sum_{i=1}^n x_it_i, & \sum_{i=1}^n x_i^2t_i\end{pmatrix}$

$y^THy=\sum_{i=1}^n(x_i+1)^2t_iy^Tt\ge 0$

即 Hessian 矩阵半正定，故凸函数得证。

12 Self-Concordant 函数性质与实例

12.1 自和谐函数性质

已知为 $f(x)$ 自和谐函数，则做仿射变换 $x = ay+b$，令 $f(y) = f(ay+b)$

$f''(y) = \alpha^2 f''(x), f'''(y) = \alpha^3 f'''(x)$

$f(x)$ 为自和谐函数，有

$|f'''(x)| \leq 2(f''(x)^{\frac{3}{2}})$

两边同时乘 $\alpha^3(\alpha > 0)$，得

$|\alpha^3 f'''(x)| \leq 2(\alpha^2 f''(x)^{\frac{3}{2}})$

$|f'''(x)| \leq |2(f''(y))^{\frac{3}{2}}|$

自和谐函数仿射变换之后仍为自和谐函数，affine function 成立

12.2 自和谐函数实例

12.2.1 第一个

内生函数是关于 x 的仿射变换，为自和谐函数，外层函数也为自和谐函数，故整个函数为自和谐函数。

12.2.2 第二个

令 $g(t) = f(X + tV)$ 经过分解

$g(t) = -logdet(X + tV)$

$= -logdetX - logdet(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}})$

$= -logdetX - \sum_{i=1}^{n} log(1 + t\lambda_i)$，$\lambda_i$ 为 $X^{-\frac{1}{2}}VX^{-\frac{1}{2}}$ 的特征值

$g(t)$ 内外层函数均为自和谐函数，故 $g(t)$ 为自和谐函数，故原函数为自和谐函数。

12.2.3 第三个

$f(x) = -log(y^2 - x^T x)$

$= -log(y^2 - x^T x) - logx - log\frac{1}{x}$

$= -log(\frac{y^2}{x}) - logx$

令 $g(x) = \frac{y^2}{x} - x$

$g^{'}(x) = -1 - \frac{y^2}{x^2}$

$g^{''}(x) = \frac{2y^T}{x^3}$

$g^{'''}(x) = -\frac{6y^2}{x^4}$

有 $|g^{'''}(x)| \leq 3\frac{g^{''}(x)}{x}$

由此 $f(x)$ 为自和谐函数