Convex Optimization. Exercise 3: Convex Optimization Problems. Chao Ni

1. Local minima are global minima in Convex Optimization.

Assume. x^* is a local minimizer.

i.e. $\exists \delta > 0$. s.t. $\|x - x^*\| \le \delta$, $\forall x$. $f(x^*) \le f(x)$

Assume x^* is not global minimizer. then $\exists \tilde{x}$, such that $f(\tilde{x}) < f(x^*)$

Consider the point $x = \theta \tilde{x} + (1-\theta) x^*$, where $\theta = \min\{1, \frac{\delta}{\|\tilde{x} - x^*\|}\}$.

$\Rightarrow \quad \|x - x^*\| = \|\theta(\tilde{x} - x^*)\| \le \delta$.

$\Rightarrow \quad f(x^*) \le f(x)$

However. by Convex function definition

$$
\begin{aligned}
f(x) = f(\theta\tilde{x} + (1-\theta)x^*) &\le \theta f(\tilde{x}) + (1-\theta) f(x^*) \\
&< \theta f(x^*) + (1-\theta) f(x^*) \\
&= f(x^*)
\end{aligned}
$$

Which in contradict to the fact $f(x^*) \le f(x)$

Hence. x^* is global minimizer.

2. The Lasso Problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \lambda \|x\|_1 =: f(x).$$

$$
\begin{aligned}
f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 &\ge \frac{1}{2}\|Ax - b\|_2^2 + \|A^T b\|_\infty \|x\|_1 \\
&\ge \frac{1}{2}\|Ax - b\|_2^2 + |b^T A x|. \\
&= \frac{1}{2}\left(x^T A x - b^T A x - x^T A^T b + b^T b\right) + |b^T A x| =: g(x)
\end{aligned}
$$

$$\nabla_x g(x) = \frac{1}{2}\cdot(2Ax - 2b^T A) + b^T A \cdot \operatorname{sgn}(b^T A x). \quad , x \ne 0$$

Let $x = 0$. $g(x) = \frac{1}{2} b^T b = \frac{1}{2}\|b\|_2^2$.

We show $x = 0$ is the optimal solution by showing that it's a local optimum. and because Lasso is convex, it's indeed a global optimum.

$\forall \|x\| < \varepsilon$.

$$g(x) - \frac{1}{2} b^T b = \frac{1}{2} x^T A x - b^T A x + |b^T A x| \ge \frac{1}{2} x^T A x \ge 0 \qquad (*)$$

Hence $x^* = 0$ is the local optimum $\Rightarrow$ global optimum.

Actually. $(*)$ holds for $\forall x \in \mathbb{R}^n$,

$\Rightarrow x^* = 0$ is the global optimum, and the optimal value is $\frac{1}{2}\|b\|_2^2$. $\square$

3. Logistic Regression

(a). Because $y \in \{+1,-1\}$. we can write $P_y(x) = \frac{1}{1+e^{-y(W^Tx+b)}}$

Hence $l_W(x,y) = \prod_{i=1}^{N} \frac{1}{1+e^{-y_i(W^Tx_i+b)}}$

We have $l_W(x,y) > 0$, as $\log(\cdot)$ is monotonous.

$$\max l_W(x,y) \iff \max \log l_W(x,y).$$

$$\log l_W(x,y) = \sum_{i=1}^{N} \log \frac{1}{1+e^{-y_i(W^Tx_i+b)}} = -\sum_{i=1}^{N} \log(1+e^{-y_i(W^Tx_i+b)})$$

Hence $\max \log l_W(x,y) \iff \min \sum_{i=1}^{N} \log(1+e^{-y_i(W^Tx_i+b)}) =: g(W,b)$.

~~$\nabla g(W,b) =$~~ Consider $h(x) = \log(1+e^x)$.

$\nabla h(x) = \frac{e^x}{1+e^x}$, $\nabla^2 h(x) = \frac{e^x}{(1+e^x)^2} > 0$

Hence $h(x)$ is convex. $\Rightarrow$ $\log(1+e^{-y_i(W^Tx_i+b)})$ is convex. by composition rule.

$\Rightarrow \sum_{i=1}^{N} \log(1+e^{-y_i(W^Tx_i+b)})$ is convex

However, the final optimization problem is not strongly convex.

Strong convex requires $\nabla^2 h(x) \succeq \sigma I$. $\exists \sigma > 0$, $\forall x$.

But. $\nabla^2 h(x) = \frac{e^x}{(1+e^x)^2}$

and $\lim_{x \to \infty} \frac{e^x}{(1+e^x)^2} = 0$.

Hence it's not strongly convex

(b) If the class are labeled with $\{0,1\}$ instead of $\{-1,1\}$

we can write $P_y(x) = \frac{1}{1+e^{(-2y+1)(W^Tx+b)}}$

$\Rightarrow \max l_W(x,y) \iff \min \sum_{i=1}^{N} \log(1+e^{(-2y+1)(W^Tx+b)}) := g(W,b)$.

$$g(W,b) = \sum_{i=1}^{N} \log(1+e^{(-2y+1)(W^Tx+b)}) \overset{\iff}{=} \sum_{y=0} \log(1+e^{W^Tx+b}) + \sum_{y=1} \log\left(\frac{1+e^{W^Tx+b}}{e^{W^Tx+b}}\right)$$

$$= \sum_{i=1}^{N} \log(1+e^{W^Tx+b}) - \sum_{y=1} (W^Tx_i+b)$$

$$= \sum_{i=1}^{N} \log(1+e^{W^Tx+b}) - y_i(W^Tx_i+b)). \quad \square$$

(c). $\sum_{i=1}^{N} \log(1+e^{-y_i(w^T x_i+b)}) := f_i(w,b)$

First, we know $f_i(w,b) > 0$ and Hence $\sum_{i=1}^{N} f_i(w,b) > 0$. Strictly.

We will show that $\inf_{w,b} \sum_{i=1}^{N} f_i(w,b) = 0$.

Hence the optimal value is 0 but cannot be reached.

Because there exists $(\bar{w},\bar{b})$ such that.

$\bar{w}^T x_i + \bar{b} > 0 \quad \forall y_i = 1.$

$\bar{w}^T x_i + \bar{b} < 0 \quad \forall y_i = -1.$

We can construct $(w_n, b_n) = n\cdot(\bar{w},\bar{b})$,

then $f_i(w_n,b_n) = \log(1+e^{-y_i(w_n^T x_i+b_n)})$

$= \log(1+e^{n\cdot(-y_i(\bar{w}^T x_i+\bar{b}))})$,

Hence. $\lim_{n\to\infty} f_i(w_n,b_n) = 0$.

$\lim_{n\to\infty} \sum_{i=1}^{N} f_i(w_n,b_n) = 0$.

Hence, $\inf_{w,b} \sum_{i=1}^{N} f_i(w,b) = 0$

And we know this can't be reached. i.e. there exist no such w,b. that $\sum_{i=1}^{N} f_i(w,b) = 0$. $\square$

4. ℓ_1-, ℓ_∞-, and ℓ_4-norm Approximation Problems.

(a). (1) Min $\|Ax-b\|_\infty$

$\iff \min \; t$

s.t. $a_i^T x - b_i \le t$, $i=1,2,\dots m$. $A=[a_1^T\, a_2^T \cdots a_m^T]^T$.

$-a_i^T x + b_i \le t$

Denote by x^* the original optimal solution, y^* the optimal solution of the corresponding LP. (same for the rest problems)

$x^* \in \mathbb{R}^n$. $y^* \in \mathbb{R}^{n+1}$; ~~$y^*=[t,x$~~

$y = [t. x^T]^T$, denote y the new variables in the corresponding LP.

$\Rightarrow x^* = [0 \; I_n]\cdot y^*$.

(2) $\min \ \|Ax-b\|_1 = \sum_{i=1}^{m} |a_i^T x - b_i|$.

$\Leftrightarrow$. $\min \ 1^T s$

s.t. $Ax - b \le s$

$-Ax + b \le s$.

New optimization variable

$y = [s^T, x^T]^T$, $s \in \mathbb{R}^m$

$x^* = [0_{n\times m} \ I_{n\times n}] \cdot y^*$.

(3) $\min \ \|Ax-b\|_1$

s.t. $\|x\|_\infty \le 1$

$\Leftrightarrow$ $\min \ 1^T s$

s.t. $Ax - b \le s$

$-Ax + b \le s$

$e_i^T x \le 1$, $i = 1, \dots n$.

$-e_i^T x \le 1$.

where $e_i = [0, \dots 0, \underset{\substack{\uparrow \\ i}}{1}, 0 \dots 0]^T$

New Optimization Variable

$y = [s^T, x^T]^T$

$x^* = [0_{n\times m} \ I_{n\times n}] \cdot y^*$.

(4) $\min \ \|x\|_1$

s.t. $\|Ax - b\|_\infty \le 1$

$\Leftrightarrow$. $\min \ 1^T s$

s.t. $x \le s$

$-x \le s$

$a_i^T x - b \le 1$ $i = 1, \dots m$.

$-a_i^T x + b \le 1$

New Optimization variable

$y = [s^T, x^T]^T$

$x^* = [0_{n\times m} \ I_{n\times n}] \cdot y^*$

(5) $\min \ \|Ax-b\|_1 + \|x\|_\infty$

$\Leftrightarrow$ $\min \ 1^T s + t$

s.t. $Ax - b \le s$

$-Ax + b \le s$

$e_i^T x \le t$, $i = 1, \dots n$.

$-e_i^T x \le t$

New Variable

$y = [s^T, t, x^T]^T \in \mathbb{R}^{m+n+1}$

$x^* = [0_{n\times(m+1)} \ I_{n\times n}] \cdot y^*$.

(b) Formulate $\min \|Ax-b\|_4$ as a QCQP

$$\|Ax-b\|_4 = \left(\sum_{i=1}^{m} |a_i^T x - b_i|^4\right)^{1/4}$$

$$\min \|Ax-b\|_4 \iff \min \sum_{i=1}^{m} |a_i^T x - b_i|^4 := g(x)$$

$$g(x) = \sum_{i=1}^{m} |a_i^T x - b_i|^4 = (|a_1^T x - b_1|^2, \cdots, |a_m^T x - b_m|^2) \cdot \begin{bmatrix} 1 & & \\ & 1 & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \cdot \begin{bmatrix} |a_1^T x - b_1|^2 \\ \vdots \\ |a_m^T x - b_m|^2 \end{bmatrix}$$

Let $s_i = |a_i^T x - b_i|^2$. $i \in [m]$.

$$g(s) = s^T s$$

$$s_i = |a_i^T x - b_i|^2 \iff [e_i^T(Ax-b)]^T [e_i^T(Ax-b)] = s_i$$

$$\iff (Ax-b)^T e_i e_i^T (Ax-b) = s_i$$

Let $t = Ax-b \in \mathbb{R}^m$.

then we have $t^T e_i e_i^T t = s_i$. $i \in [m]$.

Hence. formulate $\min \|Ax-b\|_4$ to.

$$\min \ \tfrac{1}{2} s^T s \qquad (*)$$

Subject to

$$t^T e_i e_i^T t = s_i \ , \ i \in [m]$$

$$Ax - b = t \longrightarrow [A \ \ {-I}] \begin{bmatrix} x \\ t \end{bmatrix} = b.$$

New optimization variable $y = [s^T, t^T, x^T]^T \in \mathbb{R}^{2m+n}$

Obviously $e_i e_i^T \succeq 0$.

Hence. $(*)$ is a QCQP $\square$

Standard form.

$$\min \quad y^T \begin{bmatrix} I_m & & \\ & O_m & \\ & & O_n \end{bmatrix} y$$

Subject to.

$$y^T \begin{bmatrix} O_m & & \\ & e_i e_i^T & \\ & & O_n \end{bmatrix} y - e_i^T [I_m \ O_m \ O_{m\times n}] \cdot y = 0 \qquad i \in [m]$$

$$\cancel{[A \ {-I_m}]} \ [O_m \ {-I_m} \ A] \cdot y = b.$$