Optimization Methods Fall 2021

Homework 3

Instructor: Lijun Zhang *Name:* 方盛俊, *StudentId:* 201300035

Notice

- The submission email is: **zhangzhenyao@lamda.nju.edu.cn**.
- Please use the provided Latex file as a template.
- If you are not familiar with LaTeX, you can also use Word to generate a **PDF** file.

Problem 1: One inequality constraint

对于

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & f(x) \leqslant 0 \end{aligned}$$

我们有其 Lagrange 函数为 $L(x, \lambda) = c^T x + \lambda f(x)$, 其中 $\lambda \geqslant 0$

则其对偶函数为 $g(\lambda) = \inf_x (c^T x + \lambda f(x)) = -\lambda \sup_x (-\frac{c^T}{\lambda} x - f(x))$

由 $f^*(y) = \sup_x (y^T x - f(x))$ 可知

我们有 $g(\lambda) = -\lambda f^*(-\frac{c}{\lambda})$

即可转化为对偶问题

$$\begin{aligned} \max \quad & g(\lambda) = -\lambda f^*(-\frac{c}{\lambda}) \\ \text{s.t.} \quad & \lambda \geqslant 0 \end{aligned}$$

Problem 2: KKT conditions

(1)

Lagrange 函数为:

$L(x_1,x_2,\lambda_1,\lambda_2)=x_1^2+x_2^2+\lambda_1(x_1-1)^2+\lambda_1(x_2-1)^2-2\lambda_1+\lambda_2(x_1-1)^2+\lambda_2(x_2+1)^2-2\lambda_2$

(2)

有强对偶性, 我们可以找到点 $x=(1,0)$, 即 $x_1=1,x_2=0$ 满足

$(x_1-1)^2+(x_2-1)^2=1<2$ 和 $(x_1-1)^2+(x_2+1)^2=1<2$ 成立

并且原问题是一个凸问题

即我们有 Slater 条件成立, 则这个问题保持强对偶性.

(3)

KKT 条件为:

$$
\begin{aligned}
(x_1^*-1)^2+(x_2^*-1)^2 &\leqslant 2\\
(x_1^*-1)^2+(x_2^*+1)^2 &\leqslant 2\\
\lambda_1^* &\geqslant 0\\
\lambda_2^* &\geqslant 0\\
x_1^{*2}+\lambda_1^*(x_1^*-1)+\lambda_2^*(x_1^*-1) &= 0\\
x_2^{*2}+\lambda_1^*(x_2^*-1)+\lambda_2^*(x_2^*+1) &= 0
\end{aligned}
$$

Problem 3: Equality Constrained Least-squares

(1)

对应的 Lagrange 函数为

$L(x,v)=\frac{1}{2}\|Ax-b\|_2^2+v^T(Gx-h)=\frac{1}{2}x^TA^TAx+(v^TG-b^TA)x-v^Th+\frac{1}{2}b^Tb$

因为 $L(x,v)$ 是二次凸函数, 则求解

$\nabla_x L(x,v)=A^TAx-A^Tb+G^Tv=0$ 即有 $A^TAx=A^Tb-G^Tv$

再由 A 的 rank 为 n 可知 A^TA 的 rank 也为 n, 则可以求逆, 则 $x=(A^TA)^{-1}(A^Tb-G^Tv)$

代入可知对偶函数为 $g(v)=\frac{1}{2}(v^TG-b^TA)(A^TA)^{-1}(A^Tb-G^Tv)-v^Th+\frac{1}{2}b^Tb$

则转化为对偶问题

$$\max \quad g(v) = \frac{1}{2}(v^TG - b^TA)(A^TA)^{-1}(A^Tb - G^Tv) - v^Th + \frac{1}{2}b^Tb$$

(2)

使用 Lagrange 函数再对 v 求偏导得 $Gx - h = 0$ 即 $Gx = h$

那么我们有 $Gx = G(A^TA)^{-1}(b - G^Tv) = G(A^TA)^{-1}b - G(A^TA)^{-1}G^Tv = h$

即有 $G(A^TA)^{-1}G^Tv = G(A^TA)^{-1}b - h$

由于 $(A^TA)^{-1}$ 为 $n \times n$ 的满秩矩阵, 因此 $G(A^TA)^{-1}G^T$ 也为 $p \times p$ 的满秩矩阵, 存在逆

因此 $v = [G(A^TA)^{-1}G^T]^{-1}[G(A^TA)^{-1}b - h]$

再代入式子可得 $x^* = (A^TA)^{-1}(b - G^T[G(A^TA)^{-1}G^T]^{-1}[G(A^TA)^{-1}b - h])$

对于对偶问题, 我们对 $g(v)$ 求偏导, 有

$\nabla_v g(v) = -G(A^TA)^{-1}(G^Tv - A^Tb) - h = 0$

则我们有 $G(A^TA)^{-1}(G^Tv - A^Tb) = -h$

则最后 $v^* = [G(A^TA)^{-1}G^T]^{-1}[G(A^TA)^{-1}b - h]$

Problem 4: Negative-entropy Regularization

对于 $x \in \Delta^n = \{x| \sum\limits_{i=1}^n x_i = 1, x_i \geqslant 0, i = 1, 2, \cdots, n\}$

则问题可以表达为

$$\begin{aligned} \arg\min \quad & b^Tx + c \cdot \sum_{i=1}^n x_i \ln x_i \\ \text{s.t.} \quad & -x_i \leqslant 0 \\ & \sum_{i=1}^n x_i - 1 = 0 \end{aligned}$$

对应的 Lagrange 函数为

$$L(x,\lambda,v)=b^Tx+c\cdot\sum_{i=1}^n x_i\ln x_i-\sum_{i=1}^n\lambda_i x_i+v\sum_{i=1}^n x_i-v$$

对应的对偶函数为

$$\begin{aligned}g(\lambda,v)&=\inf_x(b^Tx+c\cdot\sum_{i=1}^n x_i\ln x_i-\sum_{i=1}^n\lambda_i x_i+v\sum_{i=1}^n x_i-v)\\&=-c\sum_{i=1}^n\sup_x(\frac{1}{c}(\lambda_i-b_i-v)x_i-x_i\ln x_i)-v\\&=-ce^{-\frac{v}{c}-1}\sum_{i=1}^n e^{\frac{1}{c}(\lambda_i-b_i)}-v\end{aligned}$$

则我们转化为对偶问题

$$\begin{aligned}\arg\max\quad&-ce^{-\frac{v}{c}-1}\sum_{i=1}^n e^{\frac{1}{c}(\lambda_i-b_i)}-v\\\text{s.t.}\quad&\lambda_i\geqslant 0,\quad i=1,2,\cdots,n\end{aligned}$$

因为我们很容易找到点 $x_i=\dfrac{1}{n},i=1,2,\cdots,n$ 满足 $-x_i<0,\sum\limits_{i=1}^n x_i-1=0$

因此有强对偶性, 最优对偶间隙为零.

我们固定 λ, 对 v 求导数并等于零有

$$e^{-\frac{v}{c}-1}\sum_{i=1}^n e^{\frac{1}{c}(\lambda_i-b_i)}-1=0 \text{ 即 } v^*=c\ln(\sum_{i=1}^n e^{\frac{1}{c}(\lambda_i-b_i)})-c$$

将 v 的最优值代入对偶问题可得

$$\begin{aligned}\arg\max\quad&-c\ln(\sum_{i=1}^n e^{\frac{1}{c}(\lambda_i-b_i)})\\\text{s.t.}\quad&\lambda_i\geqslant 0,\quad i=1,2,\cdots,n\end{aligned}$$

这是一个非负约束的几何规划问题 (凸优化问题)

则我们求导可知该函数总是递减的,

因此我们可以求解出 $\lambda_i^*=0,i=1,2,\cdots,n$

可得 $v^* = c\ln(\sum_{i=1}^{n} e^{\frac{1}{c}(\lambda_i - b_i)}) - c = c\ln(\sum_{i=1}^{n} e^{-\frac{b_i}{c}}) - c$

进而根据 KKT 条件:

$$
\begin{aligned}
\sum_{i=1}^{n} x_i^* - 1 = 0 \\
x_i^* \geqslant 0, \quad i = 1, 2, \cdots, n \\
\lambda_i^* \geqslant 0, \quad i = 1, 2, \cdots, n \\
\lambda_i^* x_i^* = 0, \quad i = 1, 2, \cdots, n \\
b_i + c\ln x_i^* + c - \lambda_i^* + v^* = 0, \quad i = 1, 2, \cdots, n
\end{aligned}
$$

即有 $\frac{b_i}{c} + \ln x_i^* + \ln(\sum_{i=1}^{n} e^{-\frac{b_i}{c}}) = 0$

最后有 $x_i = \frac{e^{-\frac{b_i}{c}}}{\sum_{i=1}^{n} e^{-\frac{b_i}{c}}}, i = 1, 2, \cdots, n$

Problem 5: Support Vector Machines

(1)

引入 u_i 后变为

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n} l(u_i) + \frac{\lambda}{2}\|w\|_2^2 \\
\text{s.t.} \quad & u_i = y_i(w^T x_i + b) \qquad i = 1, 2, \cdots, n
\end{aligned}
$$

(2)

其 Lagrange 函数为

$$
\begin{aligned}
L(u,w,b,v) &= \sum_{i=1}^{n} l(u_i) + \frac{\lambda}{2}\|w\|_2^2 + \sum_{i=1}^{n} v_i(u_i - y_i(w^T x_i + b)) \\
&= \sum_{i=1}^{n} l(u_i) + \sum_{i=1}^{n} \frac{\lambda}{2} w_i^2 + \sum_{i=1}^{n} v_i(u_i - y_i(w^T x_i + b)) \\
&= \sum_{i=1}^{n} [l(u_i) + \frac{\lambda}{2} w_i^2 + v_i u_i - v_i y_i w^T x_i - v_i y_i b] \\
&= \sum_{i=1}^{n} [l(u_i) + v_i u_i] + \frac{\lambda}{2}\|w\|_2^2 - \sum_{i=1}^{n} v_i y_i x_i^T w - b y^T v
\end{aligned}
$$

因此有对偶函数

$$
\begin{aligned}
g(v) &= \inf_{u,w,b} [\sum_{i=1}^{n} [l(u_i) + v_i u_i] + \frac{\lambda}{2}\|w\|_2^2 - \sum_{i=1}^{n} v_i y_i x_i^T w - b y^T v] \\
&= -\sum_{i=1}^{n} \sup_{u_i} [-v_i u_i - l(u_i)] + \inf_{w} [\frac{\lambda}{2}\|w\|_2^2 - \sum_{i=1}^{n} v_i y_i x_i^T w] - \sup_{b} b y^T v \\
&= -\sum_{i=1}^{n} l^*(-v_i) + \inf_{w} [\frac{\lambda}{2}\|w\|_2^2 - \sum_{i=1}^{n} v_i y_i x_i^T w] - \sup_{b} b y^T v \\
&= \sum_{i=1}^{n} v_i + \frac{1}{2\lambda} \sum_{i=1}^{n} \sum_{j=1}^{n} v_i v_j y_i y_j x_i^T x_j
\end{aligned}
$$

其中要满足 $0 \leqslant v_i \leqslant 1, y^T v = 0$

因此可以转化为对偶问题:

$$
\begin{aligned}
\max \quad & g(v) = \sum_{i=1}^{n} v_i + \frac{1}{2\lambda} \sum_{i=1}^{n} \sum_{j=1}^{n} v_i v_j y_i y_j x_i^T x_j \\
\text{s.t.} \quad & y^T v = 0 \\
& 0 \leqslant v_i \leqslant 1 \qquad\qquad\qquad\qquad i = 1, 2, \cdots, n
\end{aligned}
$$

(3)

对应的 KKT 条件为:

$$
\begin{aligned}
u_i = y_i(w^T x_i + b), \quad i = 1, 2, \cdots, n \\
\sum_{i=1}^{n} \nabla_u l(u_i) + \sum_{i=1}^{n} v_i = 0 \\
\lambda w - \sum_{i=1}^{n} v_i y_i x_i = 0 \\
\sum_{i=1}^{n} v_i y_i = 0 \\
\sum_{i=1}^{n} (u_i - y_i(w^T x_i + b)) = 0
\end{aligned}
$$

其中 $\nabla_u l(u_i) = \begin{cases} -1, & u_i < 1 \\ 0, & u_i > 1 \end{cases}$