THU-70250403, Convex Optimization (Fall 2020) **Homework: 5**

Dual Problems and Classificiation

Lecturer: Li Li li-li@tsinghua.edu.cn

Student:

Problem 1

Considering the following optimization problem

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{z}\in\mathbb{R}^n} \quad & \frac{1}{2}\left|\boldsymbol{x}\right|_2^2 + \frac{1}{2}\left|A\boldsymbol{z}-\boldsymbol{b}\right|_2^2 && (1)\\ \text{s.t.} \quad & \boldsymbol{x}-\boldsymbol{z}=\boldsymbol{c} && (2) \end{aligned}$$

where $A \in \mathbb{R}^{m\times n}$ are known constant matrix with $\text{rank}(A) = n$, $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{c} \in \mathbb{R}^n$ are known constant vectors.

Please derive the corresponding dual problem.

Solution

Let $\boldsymbol{y} = A\boldsymbol{z} - \boldsymbol{b}$, then the original problem becomes

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{y}} \quad & \frac{1}{2}|\boldsymbol{x}|_2^2 + \frac{1}{2}|\boldsymbol{y}|_2^2 && (3)\\ \text{s.t.} \quad & \boldsymbol{y} = A\boldsymbol{z}-\boldsymbol{b} && (4)\\ & \boldsymbol{x}-\boldsymbol{z}=\boldsymbol{c} && (5) \end{aligned}$$

Now we introduce $\boldsymbol{v_1} \in \mathbb{R}^n, \boldsymbol{v_2} \in \mathbb{R}^m$ and formulate the lagrange function as follows:

$$\begin{aligned} L(\boldsymbol{v_1},\boldsymbol{v_2}) &= \frac{1}{2}|\boldsymbol{x}|_2^2 + \frac{1}{2}|\boldsymbol{y}|_2^2 + \boldsymbol{v_1}^T(\boldsymbol{x}-\boldsymbol{z}-\boldsymbol{c}) + \boldsymbol{v_2}^T(A\boldsymbol{z}-\boldsymbol{b}-\boldsymbol{y}) && (6)\\ &= \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} + \boldsymbol{v_1}^T\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{v_2}^T\boldsymbol{y} + (\boldsymbol{v_2}^T A - \boldsymbol{v_1})\boldsymbol{z} - \boldsymbol{v_1}^T\boldsymbol{c} - \boldsymbol{v_2}^T\boldsymbol{b} && (7)\\ &\Rightarrow && (8)\\ g(\boldsymbol{v_1},\boldsymbol{v_2}) &= \inf_{\boldsymbol{x},\boldsymbol{y}}(\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} + \boldsymbol{v_1}^T\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{v_2}^T\boldsymbol{y} + (\boldsymbol{v_2}^T A - \boldsymbol{v_1})\boldsymbol{z} - \boldsymbol{v_1}^T\boldsymbol{c} - \boldsymbol{v_2}^T\boldsymbol{b}) && (9)\\ &= -\boldsymbol{v_1}^T\boldsymbol{c} - \boldsymbol{v_2}^T\boldsymbol{b} + \inf_{\boldsymbol{x}}(\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} + \boldsymbol{v_1}^T\boldsymbol{x}) + \inf_{\boldsymbol{y}}(\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{v_2}^T\boldsymbol{y}) + \inf_{\boldsymbol{z}}((\boldsymbol{v_2}^T A - \boldsymbol{v_1}^T)\boldsymbol{z}) && (10) \end{aligned}$$

Then we can get the following result,

$$g(\boldsymbol{v}) = \begin{cases} -\frac{1}{2}\boldsymbol{v_1}^T\boldsymbol{v_1} - \boldsymbol{v_1}^T\boldsymbol{c} - \frac{1}{2}\boldsymbol{v_2}^T\boldsymbol{v_2} - \boldsymbol{v_2}^T\boldsymbol{b} & \boldsymbol{v_2}^T A = \boldsymbol{v_1}^T \\ -\infty & \text{Otherwise} \end{cases} \quad (11)$$

Therefore, the dual problem of the original problem can be formulated as follows:

$$\begin{aligned} \max_{\boldsymbol{v_1},\boldsymbol{v_2}} \quad & -\frac{1}{2}\boldsymbol{v_1}^T\boldsymbol{v_1} - \boldsymbol{v_1}^T\boldsymbol{c} - \frac{1}{2}\boldsymbol{v_2}^T\boldsymbol{v_2} - \boldsymbol{v_2}^T\boldsymbol{b} && (12)\\ \text{s.t.} \quad & \boldsymbol{v_2}^T A = \boldsymbol{v_1}^T && (13) \end{aligned}$$

Also, we can use the subject condition to eliminate $\boldsymbol{v_1}$ and get,

$$\max_{\boldsymbol{v_2}} \quad -\frac{1}{2}\boldsymbol{v_2}^T AA^T\boldsymbol{v_2} - \boldsymbol{v_2}^T Ac - \frac{1}{2}v_2{}^T v_2 - \boldsymbol{v_2}^T\boldsymbol{b} \tag{14}$$

Problem 2

The sum of the largest elements of a vector.

Define $f : \mathbb{R}^n \to \mathbb{R}$ as

$$f(\boldsymbol{x}) = \sum_{i=1}^{r} \boldsymbol{x}_{[i]},$$

where r is an integer between 1 and n , and $\boldsymbol{x}_{[1]} \geqslant \boldsymbol{x}_{[2]} \geqslant \cdots \geqslant \boldsymbol{x}_{[r]}$ are the components of x sorted in decreasing order. In other words, $f(x)$ is the sum of the r largest elements of x . In this problem we study the constraint

$$f(\boldsymbol{x}) \leqslant \alpha.$$

This is a convex constraint, and equivalent to a set of $n!/(r!(n-r)!)$ linear inequalities

$$\boldsymbol{x}_{i_1} + \cdots + \boldsymbol{x}_{i_r} \leqslant \alpha, \quad 1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n.$$

The purpose of this problem is to derive a more compact representation.

1. Given a vector $\boldsymbol{x} \in \mathbb{R}^n$, show that $f(\boldsymbol{x})$ is equal to the optimal value of the LP

   $$\begin{array}{ll} \text{maximize} & \boldsymbol{x}^T\boldsymbol{y} \\ \text{subject to} & \boldsymbol{0} \preceq \boldsymbol{y} \preceq \boldsymbol{1} \\ & \boldsymbol{1}^T\boldsymbol{y} = r \end{array}$$

   with $\boldsymbol{y} \in \mathbb{R}^n$ as variable.

2. Derive the dual of the LP in part (a). Show that it can be written as

   $$\begin{array}{ll} \text{minimize} & rt + \boldsymbol{1}^T\boldsymbol{u} \\ \text{subject to} & t\boldsymbol{1} + \boldsymbol{u} \succeq \boldsymbol{x} \\ & \boldsymbol{u} \succeq \boldsymbol{0}, \end{array}$$

   where the variables are $t \in \mathbb{R}$, $\boldsymbol{u} \in \mathbb{R}^n$. By duality this LP has the same optimal value as the LP in (a), *i.e.*, $f(x)$. We therefore have the following result: $\boldsymbol{x}$ satisfies $f(\boldsymbol{x}) \leqslant \alpha$ if and only if there exist $t \in \mathbb{R}$, $\boldsymbol{u} \in \mathbb{R}^n$ such that

   $$rt + \boldsymbol{1}^T\boldsymbol{u} \leqslant \alpha, \qquad t\boldsymbol{1} + \boldsymbol{u} \succeq \boldsymbol{x}, \qquad \boldsymbol{u} \succeq \boldsymbol{0}.$$

   These conditions form a set of $2n+1$ linear inequalities in the $2n+1$ variables $\boldsymbol{x}, \boldsymbol{u}, t$.

Solution

1. Obviously, the optimal solution of the problem is to assign weight 1 to r largest elements of $\boldsymbol{x}$, which gives

   $$\max_{\boldsymbol{y}} \quad \boldsymbol{x}^T\boldsymbol{y} = \sum_{i=1}^{r} \boldsymbol{x}_{[i]} \tag{15}$$

   $$\text{s.t.} \quad \boldsymbol{0} \preceq \boldsymbol{y} \preceq \boldsymbol{1} \tag{16}$$

   $$\boldsymbol{1}^T\boldsymbol{y} = r \tag{17}$$

We can prove this by contradiction. Suppose an element $\boldsymbol{x}_{[j]},\ j > r$ is assigned weight of $a,\ 0 < a < 1$. Assume this weight originally belongs to $\boldsymbol{x}_{[k]},\ 1 \leq k \leq r$. Then the object value becomes $\sum_{i=1}^{r} \boldsymbol{x}_{[i]} + a(\boldsymbol{x}_{[j]} - \boldsymbol{x}_{[k]}) < \sum_{i=1}^{r} \boldsymbol{x}_{[i]}$ since $\boldsymbol{x}_{[k]} > \boldsymbol{x}_{[j]}$ according to the sorting result. So the original assignment of weights is optimal as shown above.

2. The original problem is equivalent to:

$$\begin{aligned} \min_{\boldsymbol{y}} \quad & -\boldsymbol{x}^T\boldsymbol{y} && (18)\\ \text{s.t.} \quad & \mathbf{0} \preceq \boldsymbol{y} \preceq \mathbf{1} && (19)\\ & \mathbf{1}^T\boldsymbol{y} = r && (20) \end{aligned}$$

Now we introduce $\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2} \in \mathbb{R}^n, v \in \mathbb{R}$ and formulate the lagrange function as follows:

$$\begin{aligned} L(\boldsymbol{y}, \boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, v) &= -\boldsymbol{x}^T\boldsymbol{y} - \boldsymbol{\lambda_1}^T\boldsymbol{y} + \boldsymbol{\lambda_2}^T(\boldsymbol{y} - \mathbf{1}) + v(\mathbf{1}^T\boldsymbol{y} - r) && (21)\\ &= (-\boldsymbol{x}^T - \boldsymbol{\lambda_1}^T + \boldsymbol{\lambda_2}^T + v\mathbf{1}^T)\boldsymbol{y} - \boldsymbol{\lambda_2}^T\mathbf{1} - vr && (22)\\ &\Rightarrow && (23)\\ g(\boldsymbol{v_1}, \boldsymbol{v_2}) &= \inf_{\boldsymbol{y}}((-\boldsymbol{x}^T - \boldsymbol{\lambda_1}^T + \boldsymbol{\lambda_2}^T + v\mathbf{1}^T)\boldsymbol{y} - \boldsymbol{\lambda_2}^T\mathbf{1} - vr) && (24) \end{aligned}$$

Then we can get the following result,

$$g(\boldsymbol{v}) = \begin{cases} -\boldsymbol{\lambda_2}^T\mathbf{1} - vr & \boldsymbol{\lambda_2}^T + v\mathbf{1}^T = \boldsymbol{x}^T + \boldsymbol{\lambda_1}^T \\ -\infty & \text{Otherwise} \end{cases} \tag{25}$$

Therefore, the dual problem of the original problem can be formulated as follows:

$$\begin{aligned} \max_{\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, v} \quad & -\boldsymbol{\lambda_2}^T\mathbf{1} - vr && (26)\\ \text{s.t.} \quad & \boldsymbol{\lambda_2}^T + v\mathbf{1}^T = \boldsymbol{x}^T + \boldsymbol{\lambda_1}^T && (27)\\ & \boldsymbol{\lambda_1} \succeq \mathbf{0} && (28)\\ & \boldsymbol{\lambda_2} \succeq \mathbf{0} && (29) \end{aligned}$$

By eliminating $\boldsymbol{\lambda_1}$ and substitute $\boldsymbol{\lambda_2},\ v$ with $\boldsymbol{u},\ t$ respectively, we can get

$$\begin{aligned} \min_{\boldsymbol{u}, t} \quad & rt + \mathbf{1}^T\boldsymbol{u} && (30)\\ \text{s.t.} \quad & t\mathbf{1}^T + \boldsymbol{u} \succeq \boldsymbol{x} && (31)\\ & \boldsymbol{u} \succeq \mathbf{0} && (32) \end{aligned}$$

which is the same with the problem statement.

Problem 3

The so called *Support Vector Data Description* (SVDD) [1] is a one-class classification method. Given a set of data $(\boldsymbol{x}_i)$, $i = 1, \ldots, l$, $\boldsymbol{x}_i \in \mathbb{R}^n$, SVDD tries to find a closed boundary (indeed an hypersphere centered at $\boldsymbol{a} \in \mathbb{R}^n$ and has a radius $r \in \mathbb{R}$) around the data by solving the following problem

$$\begin{aligned} \min_{\boldsymbol{a}, r} \quad & r^2 + C\sum_{i=1}^{L} \xi_i && (33)\\ \text{s.t.} \quad & |\boldsymbol{x}_i - \boldsymbol{a}|_2^2 \leq r^2 + \xi_i,\ i = 1, \ldots, L && (34)\\ & \xi_i \geq 0,\ i = 1, \ldots, L && (35) \end{aligned}$$

where $C > 0$ is a penalty parameter.

Please explain the geometry meaning of ξ_i above and derive the dual problem.

Solution

The existence of ξ_i makes the boundary of the hypersphere soft. That is, data points are allowed to have small disturbance from the boundary with tolerance equal to ξ_i in terms of square root of distance from $\boldsymbol{a}$. So the geometry meaning of ξ_i is that it makes the boundary from a 'line' to a 'circular ring'.

Now we introduce $\boldsymbol{\alpha} \in \mathbb{R}^L,\ \boldsymbol{\beta} \in \mathbb{R}^L$ and formulate the lagrange function as follows:

$$L(r, \boldsymbol{a}, \xi_i, \boldsymbol{\alpha}, \boldsymbol{\beta}) \quad = \quad r^2 + C\sum_{i=1}^{L}\xi_i + \sum_{i=1}^{L}\alpha_i(|\boldsymbol{x}_i - \boldsymbol{a}|_2^2 - r^2 - \xi_i) + \sum_{i=1}^{L}\beta_i(-\xi_i) \tag{36}$$

Letting its partial derivatives with respect to $r, \boldsymbol{a}, \xi_i$ be zero, we have,

$$\frac{\partial L(r, \boldsymbol{a}, \xi_i, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial r} = 2r(1 - \sum_{i=1}^{L}\alpha_i) = 0 \quad \Rightarrow \quad \sum_{i=1}^{L}\alpha_i = 1 \tag{37}$$

$$\frac{\partial L(r, \boldsymbol{a}, \xi_i, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{a}} = \sum_{i=1}^{L}\alpha_i(\boldsymbol{a} - \boldsymbol{x}_i) = 0 \quad \Rightarrow \quad \boldsymbol{a} = \sum_{i=1}^{L}\alpha_i\boldsymbol{x}_i \tag{38}$$

$$\frac{\partial L(r, \boldsymbol{a}, \xi_i, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \quad \Rightarrow \quad \alpha_i + \beta_i = C \tag{39}$$

Eliminating the partial decision variables $r, \boldsymbol{a}, \xi_i$, we have the objective of the Lagrange dual problem as

$$g(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{L}\alpha_i|\boldsymbol{x}_i - \sum_{j=1}^{L}\alpha_j\boldsymbol{x}_j|_2^2 \tag{40}$$

Therefore, the dual problem of the original problem can be formulated as follows:

$$\begin{aligned}
\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \quad & \sum_{i=1}^{L}\alpha_i|\boldsymbol{x}_i - \sum_{j=1}^{L}\alpha_j\boldsymbol{x}_j|_2^2 && (41)\\
\text{s.t.} \quad & \boldsymbol{\alpha} \succeq \mathbf{0} && (42)\\
& \boldsymbol{\beta} \succeq \mathbf{0} && (43)\\
& \boldsymbol{\alpha} + \boldsymbol{\beta} = C\mathbf{1} && (44)\\
& \sum_{i=1}^{L}\alpha_i = 1 && (45)
\end{aligned}$$

Note that $\boldsymbol{\beta}$ is irrelevant to the objective funtion, so we can eliminate it from the subject conditions.

$$\begin{aligned}
\max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^{L}\alpha_i|\boldsymbol{x}_i - \sum_{j=1}^{L}\alpha_j\boldsymbol{x}_j|_2^2 && (46)\\
\text{s.t.} \quad & C\mathbf{1} \succeq \boldsymbol{\alpha} \succeq \mathbf{0} && (47)\\
& \sum_{i=1}^{L}\alpha_i = 1 && (48)
\end{aligned}$$

Acknowledgment

We would like to thank Mr. Shan Jiang with the Department of Civil Engineering for providing answers to this homework.

References

[1] D. M. J. Tax, R. P. W. Duin, "Support Vector Data Description," *Machine Learning*, vol. 54, no. 1, pp. 45-66, 2004.