Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Emilio Balda

Exercise 11
- Proposed Solution -

Friday, January 26, 2018

Solution of Problem 1

a) The first step is to formulate the Lagrangian function:

$$\Lambda(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}^T\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = (\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{x} - \boldsymbol{\lambda}^T\mathbf{b}$$

$\boldsymbol{\lambda}$ is Lagrange multiplier. Next step is to find Lagrange dual function:

$$g(\boldsymbol{\lambda}) = \inf_{\mathbf{x}\in\mathbb{R}^n} \Lambda(\mathbf{x}, \boldsymbol{\lambda}) = \begin{cases} -\boldsymbol{\lambda}^T\mathbf{b} & \mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda} = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Therefore the dual problem can be written as:

$$\begin{aligned} \max \quad & -\mathbf{b}^T\lambda \\ \text{s.t.} \quad & \mathbf{A}^T\boldsymbol{\lambda} + \mathbf{c} = 0 \\ & \boldsymbol{\lambda} \succeq 0 \end{aligned}$$

b) Again, the first step is to formulate the Lagrangian function:

$$\Lambda(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x}^T\mathbf{B}\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

Next step is to find Lagrange dual function:

$$g(\boldsymbol{\lambda}) = \inf_{\mathbf{x}\in\mathbb{R}^n} \Lambda(\mathbf{x}, \boldsymbol{\lambda})$$

Using derivation with repsect to $\mathbf{x}$, we have:

$$\frac{\partial}{\partial \mathbf{x}}\Lambda(\mathbf{x}, \boldsymbol{\lambda}) = 2\mathbf{B}x + \mathbf{A}^T\boldsymbol{\lambda} \implies \mathbf{x} = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{A}^T\boldsymbol{\lambda}.$$

Having this solution (verify that it is the minimum point indeed), the Lagrange dual function is given by:

$$g(\boldsymbol{\lambda}) = \frac{1}{4}\boldsymbol{\lambda}^T\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^T\boldsymbol{\lambda} + \boldsymbol{\lambda}^T(-\frac{1}{2}\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - \mathbf{b}) = -\frac{1}{4}\boldsymbol{\lambda}^T\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\mathbf{b}$$

Therefore the dual problem can be written as:

$$\begin{aligned} \max \quad & -\frac{1}{4}\boldsymbol{\lambda}^T\mathbf{A}\mathbf{B}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\mathbf{b} \\ \text{s.t.} \quad & \boldsymbol{\lambda} \succeq 0 \end{aligned}$$

c) Again, the first step is to formulate the Lagrangian function:

$$\Lambda(\mathbf{x}, \boldsymbol{\nu}) = \|\mathbf{x}\|_p + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

Next step is to find Lagrange dual function:

$$g(\boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathbb{R}^n} \Lambda(\mathbf{x}, \boldsymbol{\nu})$$

A related optimization problem can be written as:

$$\inf_{\mathbf{x}} \|\mathbf{x}\|_p + \mathbf{a}^T\mathbf{x}.$$

Consider the Hölder's inequality:

$$\|\mathbf{x}\|_p \|\mathbf{a}\|_q \geq |\mathbf{a}^T\mathbf{x}| \geq -\mathbf{a}^T\mathbf{x},$$

if $\frac{1}{p} + \frac{1}{q} = 1$. Note that if $\|\mathbf{a}\|_q \leq 1$, then $\|\mathbf{x}\|_p + \mathbf{a}^T\mathbf{x} \geq 0$ and therefore $\inf_{\mathbf{x}} \|\mathbf{x}\|_p + \mathbf{a}^T\mathbf{x} = 0$.

Now suppose that $\|\mathbf{a}\|_q > 1$. See that if $\mathbf{x} = -t(a_1|a_1|^{q-2}, \ldots, a_n|a_n|^{q-2})$ for some $t > 0$, then we have:

$$\|\mathbf{x}\|_p + \mathbf{a}^T\mathbf{x} = t\|\mathbf{a}\|_q^{q-1} - t\|\mathbf{a}\|_q^q = t(\|\mathbf{a}\|_q^{q-1} - \|\mathbf{a}\|_q^q).$$

Since $\|\mathbf{a}\|_q > 1$, the above expression is strictly negative and can be arbitrarily small when $t \to \infty$.

Applying the above result, the Lagrange dual function is given by:

$$g(\boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathbb{R}^n} \Lambda(\mathbf{x}, \boldsymbol{\nu}) = \begin{cases} -\boldsymbol{\nu}^T\mathbf{b} & \|\mathbf{A}^T\boldsymbol{\nu}\|_q \leq 1 \\ -\infty & \text{otherwise.} \end{cases}$$

Therefore the dual problem can be written as:

$$\begin{aligned} \max \quad & -\boldsymbol{\nu}^T\mathbf{b} \\ \text{s.t.} \quad & \|\mathbf{A}^T\boldsymbol{\nu}\|_q \leq 1 \end{aligned}$$

Solution of Problem 2

a) The first step is to formulate the Lagrangian function:

$$\Lambda(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{a}\|^2 + c\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \lambda_i[y_i(\mathbf{a}^T\mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^{n} \alpha_i \xi_i,$$

where $\lambda_i, \alpha_i \geq 0$ and we know that $y_i = 1$ or -1.

Next step is to find Lagrange dual function:

$$g(\boldsymbol{\lambda}, \boldsymbol{\alpha}) = \inf_{\boldsymbol{\xi}, \mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}} \Lambda(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$$

Taking the derivation with respect to these parameters lead to:

$$\frac{\partial}{\partial \mathbf{a}} \Lambda(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = 0 \implies \mathbf{a} = \sum_{i=1}^{n} \lambda_i y_i \mathbf{x}_i$$

$$\frac{\partial}{\partial b}\Lambda(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = 0 \implies 0 = \sum_{i=1}^{n} \lambda_i y_i$$

$$\frac{\partial}{\partial \boldsymbol{\xi}}\Lambda(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = 0 \implies c = \lambda_i + \alpha_i \quad i = 1, \dots, n$$

Using these assumptions, the dual problem can be written as:

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\sum_{i,j} y_i y_j \lambda_i \lambda_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{s.t.} \quad & 0 \le \lambda_i \le c \\ & \sum_{i=1}^{n} \lambda_i y_i = 0. \end{aligned}$$

b) Take two support vector $\mathbf{x}_k$ and $\mathbf{x}_l$ with $y_k = 1$ and $y_l = -1$. Those support vectors are those with $0 < \lambda < c$. Then since for support vectors we have $y_i(\mathbf{a}^T\mathbf{x}_i + b) = 1$, we have:

$$(\mathbf{a}^{\star T}\mathbf{x}_k + b^\star) = -(\mathbf{a}^{\star T}\mathbf{x}_l + b^\star) \implies b^\star = \frac{-1}{2}\mathbf{a}^{\star T}(\mathbf{x}_k + \mathbf{x}_l) \tag{1}$$

In litterature, somethimes an average is calculated over all support vectors for better robustness.

Solution of Problem 3

a) b is given as $-\frac{1}{2}\mathbf{a}^T(\mathbf{x}_1 + \mathbf{x}_2) = 3$.

b) Supporting vectors are those with $\lambda_i \neq 0$, which are $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$.

$\mathbf{x}_i$	y_i	λ_i	$\mathbf{x}_i$	y_i	λ_i
$\mathbf{x}_1 = \begin{pmatrix}1\\1\end{pmatrix}$	$y_1 = -1$	$\lambda_1 = 0$	$\mathbf{x}_4 = \begin{pmatrix}0.5\\-0.5\end{pmatrix}$	$y_4 = 1$	$\lambda_4 = 4.73$
$\mathbf{x}_2 = \begin{pmatrix}2\\0\end{pmatrix}$	$y_2 = -1$	$\lambda_2 = 0.67$	$\mathbf{x}_5 = \begin{pmatrix}-2\\1\end{pmatrix}$	$y_5 = 1$	$\lambda_5 = 0.94$
$\mathbf{x}_3 = \begin{pmatrix}0\\0\end{pmatrix}$	$y_3 = -1$	$\lambda_3 = 5$	$\mathbf{x}_6 = \begin{pmatrix}0\\-1\end{pmatrix}$	$y_6 = 1$	$\lambda_1 = 0$

c) For those vectors, the normal vector of the hyperplane is obtained as:

$$\mathbf{a} = \sum_{i=1}^{6} \lambda_i y_i \mathbf{x}_i = \lambda_2 y_2 \mathbf{x}_2 + \lambda_3 y_3 \mathbf{x}_3 + \lambda_4 y_4 \mathbf{x}_4 + \lambda_5 y_5 \mathbf{x}_5$$

$$\mathbf{a} = -0.67\begin{pmatrix}2\\0\end{pmatrix} - 5\begin{pmatrix}0\\0\end{pmatrix} + 4.73\begin{pmatrix}0.5\\-0.5\end{pmatrix} + 0.94\begin{pmatrix}-2\\1\end{pmatrix} = \begin{pmatrix}-0.86\\-1.43\end{pmatrix}$$

To find b, take two support vectors $\mathbf{x}_k$ and $\mathbf{x}_l$ with $y_k = 1$ and $y_l = -1$ with $0 < \lambda < 5$. For these support vectors, we have $y_i(\mathbf{a}^T\mathbf{x}_i + b) = 1$. Hence:

$$b^\star = \frac{-1}{2}\mathbf{a}^{\star T}(\mathbf{x}_k + \mathbf{x}_l) = -\frac{1}{2}\begin{pmatrix}-0.86 & -1.43\end{pmatrix}(\begin{pmatrix}2\\0\end{pmatrix} + \begin{pmatrix}-2\\1\end{pmatrix}) = \frac{1.43}{2} = 0.715. \tag{2}$$