Linear and Convex Optimization Homework 13

Qiu Yihang, 2021/12/27-2022/1/2

1.(a) ***Solution:***

The Lagrangian function is

$$\mathcal{L}(x,\mu) = \log(1+e^x) - \mu x$$

By the optimality condition,

$$\begin{cases} \nabla_x \mathcal{L} = \dfrac{e^x}{1+e^x} - \mu = 0 \\ -\mu x \geq 0 \\ \mu \geq 0 \\ x \geq 0 \end{cases}$$

Since $\mu = \frac{e^x}{1+e^x} > 0,$ we know $x = 0.$

Thus, the optimal solution is $x^\star = 0$ and the optimal value is $f^\star = f(0) = \log 2.$ ■

(b) ***Solution:***

Since $\mathcal{L}(x,\mu)$ is convex in $x,$ x's minimum satisfies

$$\nabla_x \mathcal{L}(x,\mu) = \frac{e^x}{1+e^x} - \mu = 0 \Rightarrow x = \log\frac{\mu}{1-\mu}$$

Note that this requires $\mu \in (0,1).$ Thus, the dual function is

$$\phi(\mu) = \inf_x \mathcal{L}(x,\mu) = \begin{cases} -\infty, & \mu \leq 0 \\ \log\dfrac{1}{1-\mu} - \mu\log\dfrac{\mu}{1-\mu}, & \mu \in (0,1). \\ -\infty, & \mu \geq 1 \end{cases}$$

The dual problem is

$$\max_\mu \ \phi(\mu)$$

$$\text{s.t.} \ \ \mu \geq 0$$ ■

(c) ***Solution:***

Considering

$$\phi^\star = \max_{\mu \geq 0} \ \phi(\mu) = \max_{\mu \in (0,1)} \ \log\frac{1}{1-\mu} - \mu\log\frac{\mu}{1-\mu}$$

and when $\mu \in (0,1),$

$$\nabla_\mu \phi(\mu^\star) = \frac{1}{1-\mu^\star} - \log\frac{\mu^\star}{1-\mu^\star} - \frac{1}{1-\mu^\star} = -\log\frac{\mu^\star}{1-\mu^\star} = 0 \Leftrightarrow \frac{\mu^\star}{1-\mu^\star} = 1 \Leftrightarrow \mu^\star = \frac{1}{2},$$

we have

$$\phi^\star = \phi\left(\frac{1}{2}\right) = \log 2.$$

Thus, the dual optimal solution is $\mu^\star = 1/2$ and the dual optimal value is $\log 2$. ■

Since $f^\star = \phi^\star = \log 2$, the strong duality holds. ■

2.(a) ***Solution:***

Note that $X = \{(1,0)\}$.

The Lagrange function is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}) = x_1^2 + x_2^2 + \mu_1((x_1-1)^2 + (x_2-1)^2 - 1) + \mu_2((x_1-1)^2 + (x_2+1)^2 - 1)$$

Since $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu})$ is convex in $\boldsymbol{x}$, its minimum satisfies

$$\begin{cases} \nabla\mathcal{L}_{x_1}(\boldsymbol{x}, \boldsymbol{\mu}) = (2 + 2\mu_1 + 2\mu_2)x_1 - 2(\mu_1 + \mu_2) = 0 \\ \nabla\mathcal{L}_{x_2}(\boldsymbol{x}, \boldsymbol{\mu}) = (2 + 2\mu_1 + 2\mu_2)x_2 - 2(\mu_1 - \mu_2) = 0 \end{cases} \Rightarrow \begin{cases} x_1 = \dfrac{\mu_1 + \mu_2}{\mu_1 + \mu_2 + 1} \\ x_2 = \dfrac{\mu_1 - \mu_2}{\mu_1 + \mu_2 + 1} \end{cases}$$

Thus, the Lagrangian dual function is

$$\begin{aligned} \phi(\mu_1, \mu_2) &= \inf_x \mathcal{L}\left(\frac{\mu_1 + \mu_2}{\mu_1 + \mu_2 + 1}, \frac{\mu_1 - \mu_2}{\mu_1 + \mu_2 + 1}, \mu_1, \mu_2\right) \\ &= \begin{cases} 1 - \dfrac{(\mu_1 - \mu_2)^2 + 1}{\mu_1 + \mu_2 + 1}, & \mu_1 + \mu_2 + 1 \neq 0 \\ -\infty, & \mu_1 + \mu_2 + 1 = 0 \end{cases} \end{aligned}$$

The dual problem is

$$\max_{\mu} \ \phi(\mu_1, \mu_2)$$

$$\text{s.t. } \mu_1, \mu_2 \geq 0$$ ■

(b) ***Solution:***

We consider the case when $\mu_1, \mu_2 \geq 0$. This yields $\mu_1 + \mu_2 + 1 > 0 \Rightarrow \frac{(\mu_1 - \mu_2)^2 + 1}{\mu_1 + \mu_2 + 1} > 0$.

Thus, $\phi(\mu_1, \mu_2) \leq 1$, **i.e.** $\phi^\star = 1$.

Meanwhile, since $X = \{(1,0)\}, f^\star = f(1,0) = 1 = \phi^\star$. Therefore, strong duality holds. ■

(c) ***Solution:***

Since $X = \{(1,0)\}$, there exists no $x \in \textbf{int}\, D = \mathbb{R}$ that is strictly feasible, **i.e.**

$$(x_1 - 1)^2 + (x_2 - 1)^2 < 1,$$

$$(x_1 - 1)^2 + (x_2 + 1)^2 < 1.$$

Thus, Slater's condition **does not hold**. ■

Considering the strong duality still holds in this case (proved in 2(b)), we conclude that Slater's condition **is not necessary** for strong duality. ■

(d) ***Solution:***

The dual optimal value $\phi^\star$ is **not attained** by any dual feasible point.

This is expected since the optimal point, $(1,0)$, does not satisfy KKT conditions and is not a regular point. ■

3.(a) ***Solution:***

Let $\mathcal{L}(\boldsymbol{x},\mu) = f(\boldsymbol{x}) + \mu(1 - x_1 - x_2) = x_1^3 + x_2^3 + \mu(1 - x_1 - x_2)$.

We know

$$\begin{cases} \nabla\mathcal{L}_{x_1}(\boldsymbol{x},\boldsymbol{\mu}) = 3x_1^2 - \mu = 0 \\ \nabla\mathcal{L}_{x_2}(\boldsymbol{x},\boldsymbol{\mu}) = 3x_2^2 - \mu = 0 \end{cases} \Rightarrow x_1 = x_2 = \sqrt{\frac{\mu}{3}}.$$

This requires $\mu \geq 0$. When $\mu < 0$, the optimal condition is $x_1 = x_2 = 0$.

Thus, the explicit expression of $\phi(\mu)$ is

$$\phi(\mu) = \inf_{\boldsymbol{x}\geq\boldsymbol{0}} \mathcal{L}(\boldsymbol{x},\mu) = \begin{cases} \mu, & \mu < 0 \\ \mu - \frac{4}{3\sqrt{3}}\mu^{\frac{3}{2}}, & \mu \geq 0 \end{cases}.$$ ■

(b) ***Solution:***

The dual problem is

$$\max_{\mu\in\mathbb{R}} \ \phi(\mu)$$

When $\mu < 0$, $\phi(\mu) < 0$.

When $\mu \geq 0$, $\phi'(\mu) = 1 - \frac{2}{\sqrt{3}}\mu^{\frac{1}{2}} = 0 \Leftrightarrow \mu^\star = \frac{3}{4}$. In this case, $\phi^\star = \phi(\mu^\star) = \frac{1}{4} > 0$.

Thus, the optimal solution is $\mu^\star = \frac{3}{4}$. ■

(c) ***Solution:***

By weak duality, we know $f^\star \geq \phi^\star = \frac{1}{4}$.

Meanwhile, we notice that $f\left(\frac{1}{2},\frac{1}{2}\right) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$. Thus, $f^\star \leq \frac{1}{4}$.

Therefore, $f^\star = \frac{1}{4}$, **i.e.** the primal optimal value is $\frac{1}{4}$. ■

(d) ***Solution:***

The dual problem of (P2) is

$$\max_{\boldsymbol{\mu}\in\mathbb{R}^3} \ \tilde{\phi}(\boldsymbol{\mu})$$

where $\tilde{\phi}(\boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \ \tilde{\mathcal{L}}(\boldsymbol{x},\boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \ x_1^3 + x_2^3 + \mu_1(1 - x_1 - x_2) - \mu_2 x_1 - \mu_3 x_2$.

When $x_1, x_2 \to -\infty$, we have $\tilde{\mathcal{L}}(\boldsymbol{x},\boldsymbol{\mu}) \to -\infty$, **i.e.** the dual function

$$\tilde{\phi}(\boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \ \tilde{\mathcal{L}}(\boldsymbol{x},\boldsymbol{\mu}) = -\infty.$$ ■

Therefore,

$$\tilde{\phi}^\star = -\infty \neq f^\star = \frac{1}{4},$$

i.e. the strong duality <u>**does not hold**</u>. ■

4.(a) ***Solution:***

The Lagrangian function of the primal problem is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \sum_{i=1}^{n} \mu_i \left(1 - y_i\left(\boldsymbol{x}_i^T \boldsymbol{w}_i + b_i\right)\right)$$

Consider the dual problem. The domain is $\boldsymbol{\mu} \in \mathbb{R}$. By the assumption that the SVM problem is linearly separatable, we know the feasible set is not empty, **i.e.** exists at least one solution $\boldsymbol{\mu}$. Obviously $\boldsymbol{\mu} \in \textbf{int}\, \mathbb{R}$. Thus, the Slater's condition holds.

By Slater's Theorem, strong duality holds, **i.e.** $f^\star = \phi^\star$.

By the definition given in the problem, we know $f^\star = f(\boldsymbol{w}^\star, b^\star) = \phi^\star = \phi(\boldsymbol{\mu}^\star)$

By the complementary slackness in the KKT condition, we have

$$\begin{cases} \mu_i \left(1 - y_i\left(\boldsymbol{x}_i^T \boldsymbol{w}^\star + b^\star\right)\right) = 0 \\ \mu_i \geq 0 \\ 1 - y_i\left(\boldsymbol{x}_i^T \boldsymbol{w}^\star + b^\star\right) \leq 0 \end{cases}$$

Thus, for any i **s.t.** $\mu_i^\star > 0$,

$$1 - y_i\left(\boldsymbol{x}_i^T \boldsymbol{w}^\star + b^\star\right) = 0, \text{ i.e. } y_i\left(\boldsymbol{x}_i^T \boldsymbol{w}^\star + b^\star\right) = 1.$$ ■

CASE 01. When $y_i = 1$, $\boldsymbol{x}_i^T \boldsymbol{w}^\star + b^\star = 1 \Rightarrow b^\star = 1 - \boldsymbol{x}_i^T \boldsymbol{w}^\star = y_i - \boldsymbol{x}_i^T \boldsymbol{w}^\star$.

CASE 02. When $y_i = -1$, $\boldsymbol{x}_i^T \boldsymbol{w}^\star + b^\star = -1 \Rightarrow b^\star = -1 - \boldsymbol{x}_i^T \boldsymbol{w}^\star = y_i - \boldsymbol{x}_i^T \boldsymbol{w}^\star$.

Thus, $b^\star = y_i - \boldsymbol{x}_i^T \boldsymbol{w}^\star$. ■

(b) ***Solution:***

Complete <u>`svm.py`</u>. The results of the code are as follows.

```
In [1]: runfile('D:/Textbooks/2021-2022-1/Linear and Convex Optimization/hw13/
p4.py', wdir='D:/Textbooks/2021-2022-1/Linear and Convex Optimization/hw13')
primal optimal:
    w = [-1.09090908  1.45454545]
    b = [-0.09090911]

dual optimal:
    mu = [1.65289255e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
 0.00000000e+00 0.00000000e+00 0.00000000e+00 1.65289254e+00
 0.00000000e+00 0.00000000e+00 7.11813933e-09 0.00000000e+00
 0.00000000e+00]
```

Fig.01. The Results of Program (4)

The visualization of the result of hard-margin SVM is as follows (on the next page).

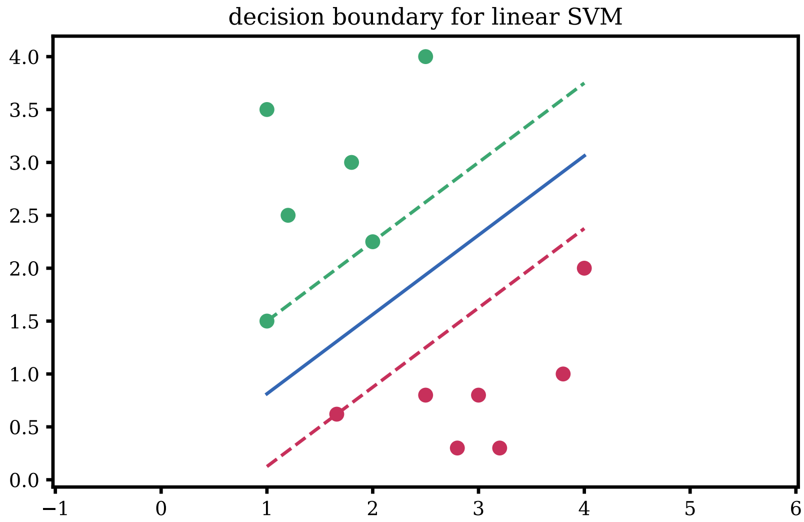

Fig.02. The Visualization of the Result of Hard-Margin SVM