Exercise 06 of Machine Learning [IN 2064]

Name: Yiman Li
Matr-Nr: 03724352
cooperate with Kejia Chen(03729686)

Problem 1

The visualization is shown in Figure 1. We can see the set $\mathcal{X}$ is convex, so in order to find the maximum, we can just try in 4 different vertices and will find that the maximizer $\boldsymbol{\theta}_{max}$ is located at the point $(9, 3)$.
In contrast, in order to find the minimizer, we just need to find the maximizer of $-f(\boldsymbol{\theta})$. Since $f(\boldsymbol{\theta})$is a linear function, then both $f(\boldsymbol{\theta})$ and $-f(\boldsymbol{\theta})$ are convex functions, so still doing experiments in different vertices in order to find the maximizer of $-f(\boldsymbol{\theta})$, which is located at the point $(5, 7)$. So we

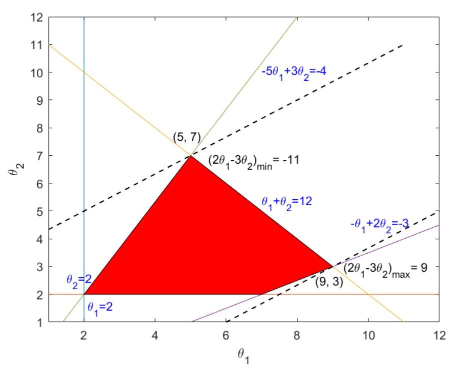

Figure 1: Visulization 1

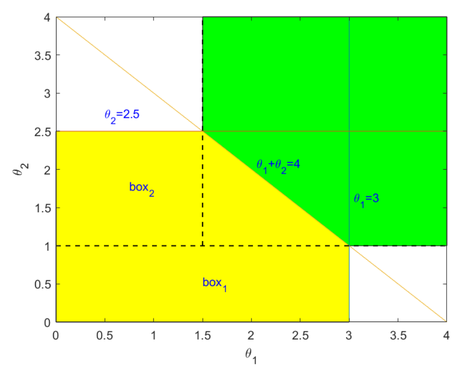

Figure 2: Visulization 2

have:

$$f(\boldsymbol{\theta})_{max} = f(9, 3) = 9 \tag{1}$$
$$f(\boldsymbol{\theta})_{min} = f(5, 7) = -11 \tag{2}$$

Problem 2

a)The Visulization is shown in Figure 2. When doing the projection, the points outside the green area are just project onto box, so

$$(\pi_{\mathcal{X}}(\boldsymbol{p}))_i = min(max(l_i, p_i), u_i) \tag{3}$$

But for points $\boldsymbol{p} = (x, y)$ in green region, the hyperplane is defined by points $\boldsymbol{a} = (1.5, 2.5)$, $\boldsymbol{b} = (3, 1)$, so we have

$$\begin{aligned}\pi_{\mathcal{X}_{a,b}}(\boldsymbol{p}) &= (1.5, 2.5) + \frac{(x - 1.5)\cdot 1.5 + (y - 2.5)\cdot -1.5}{4.5}\cdot(1.5, -1.5)\\ &= (\frac{x - y + 4}{2}, \frac{y - x + 4}{2})\end{aligned} \tag{4}$$

So in summary, we have the closed form projection for point $\boldsymbol{p} = (p_x, p_y)$

$$\pi_{\mathcal{X}}(\boldsymbol{p}) = \begin{cases} (min(max(0,3), p_x),\ min(max(0,1), p_y)), & \text{if } p_y \le 1; \\ (x = min(max(0,1.5), p_x),\ y = min(max(1,2.5), p_y)), & \text{if } p_x < 1.5, p_y > 1; \\ (\frac{p_x - p_y + 4}{2}, \frac{p_y - p_x + 4}{2}), & \text{if p in region green.} \end{cases} \tag{5}$$

b)Firstly, the derivative of the function is

$$f'(\boldsymbol{\theta}) = f'(\theta_1, \theta_2) = (2(\theta_1 - 2), 4(2\theta_2 - 7)) \tag{6}$$

Starting from $\boldsymbol{\theta}^{(0)} = (2.5, 1)$, we have

$$\begin{aligned} \boldsymbol{\theta}^{(1)} &= \boldsymbol{\theta}^{(0)} - \tau \cdot f_0'(\theta_1, \theta_2) \\ &= (2.5, 1) - 0.05 \cdot (1, -20) \\ &= (2.45, 2) \end{aligned} \tag{7}$$

which is out of the feasible region, so we do projection and obtain that

$$\begin{aligned} \boldsymbol{\theta}^{(1)} &= (\frac{2.45 - 2 + 4}{2}, \frac{2 - 2.45 + 4}{2}) \\ &= (2.225, 1.775) \end{aligned} \tag{8}$$

Then

$$\begin{aligned} \boldsymbol{\theta}^{(2)} &= \boldsymbol{\theta}^{(1)} - \tau \cdot f_1'(\theta_1, \theta_2) \\ &= (2.225, 1.775) - 0.05 \cdot (0.45, -13.8) \\ &= (2.2025, 2.465) \end{aligned} \tag{9}$$

which is again not in the feasible areas, so we need to do the projection job, and will find out the true value:

$$\begin{aligned} \boldsymbol{\theta}^{(2)} &= (\frac{x - y + 4}{2}, \frac{y - x + 4}{2}) \\ &= (1.86875, 2.13125) \end{aligned} \tag{10}$$

Problem 3

Let $\boldsymbol{\theta}^*$ be a minimizer of the primal problem and $\boldsymbol{\alpha}^*$ a maximizer of the dual problem, then the dual function is given by

$$\begin{aligned} g(\boldsymbol{\alpha}) &= \arg\min_{\theta}\ \ L(\boldsymbol{\theta}, \boldsymbol{\alpha}) \\ &= \arg\min_{\theta}\ \ (\theta_1 - \sqrt{3}\theta_2) + \alpha(\theta_1^2 + \theta_2^2 - 4) \end{aligned} \tag{11}$$

find the minimizing $\boldsymbol{\theta}$ from the optimality condition

$$\nabla_x\ L(\boldsymbol{\theta}, \boldsymbol{\alpha}) = (2\alpha\theta_1 + 1) + (2\alpha\theta_2 - \sqrt{3}) = 0 \tag{12}$$

which yields $\boldsymbol{\theta}^* = (-\frac{1}{2\alpha}, \frac{\sqrt{3}}{2\alpha})$.Therefore the dual function is

$$\begin{aligned} g(\alpha) &= (\theta_1 - \sqrt{3}\theta_2) + \alpha(\theta_1^2 + \theta_2^2 - 4) \\ &= -(\frac{1}{\alpha} + 4\alpha) \end{aligned} \tag{13}$$

Since $\alpha \ge 0$, so we have $\boldsymbol{\alpha}^* = \frac{1}{2}$. Then according to reference [1], when Slater's condition holds and the primal problem is convex, which is exactly the problem 3's condition, we will have strong duality here. So using $\alpha = \frac{1}{2}$, we get $\boldsymbol{\theta}^* = (\theta_1, \theta_2) = (-1, \sqrt{3})$

Problem 4

We can rewrite the problem in the following form:

$$\begin{aligned} &\text{minimize}_{\boldsymbol{w},b} \;\; f_0(\boldsymbol{w}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \\ &\text{subject to} \;\; f_i(\boldsymbol{w}) = 1 - y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \leq 0 \end{aligned} \tag{14}$$

- Firstly, we get the derivatives of the function as below:

  $$\nabla_{\boldsymbol{w}} f_0 = \boldsymbol{w} \tag{15}$$
  $$\nabla^2_{\boldsymbol{w}} f_0 = 1 \tag{16}$$

  We see the second order derivative is positive, so the object function is convex.

- Then since all the constraint functions are affine, so their second order derivatives are all 0, and we can say that they are all convex.

- If the data is linearly separable, there exists a hyperplane to separate the data, so that $\boldsymbol{w}^T\boldsymbol{x}_i + b \geq 1$ for $y_i = 1$, and $\boldsymbol{w}^T\boldsymbol{x}_i + b \leq -1$ for $y_i = -1$. So there exists a feasible $\boldsymbol{w}$ such that all the constraints are satisfied.

Under the 3 conditions mentioned above, the Slater's constraint qualification is fulfilled, so the strong duality holds here.

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.