기계인공지능 HW #2 Sol

1. (a)

$$\min_x \left[(2x_1 - 1)^2 + (-x_1 + x_2)^2 + (2x_2 + 1)^2 \right]$$

Let $f = (2x_1 - 1)^2 + (-x_1 + x_2)^2 + (2x_2 + 1)^2$

To find x s.t minimize f, differentiate f

$$\frac{\partial f}{\partial x_1} = 2(2x_1 - 1) \cdot 2 + 2(-x_1 + x_2) \cdot (-1) = 0$$

$$\frac{\partial f}{\partial x_2} = 2(-x_1 + x_2) + 2(2x_2 + 1) \cdot 2 = 0$$

we have two unknow & two constraint

Rewrite above equations as following

$$\begin{cases} 10x_1 - 2x_2 - 4 = 0 \\ -2x_1 + 10x_2 + 4 = 0 \end{cases}$$

$$\begin{bmatrix} 10 & -2 \\ -2 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -4 \\ 4 \end{bmatrix} = 0$$

So, $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 10 & -2 \\ -2 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ -4 \end{bmatrix} = \frac{1}{96} \begin{bmatrix} 10 & +2 \\ +2 & 10 \end{bmatrix} \begin{bmatrix} 4 \\ -4 \end{bmatrix} = \frac{1}{96} \begin{bmatrix} 32 \\ -32 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

Thus, $x_1 = \frac{1}{3}$ and $x_2 = -\frac{1}{3}$

(b)

$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$

Then, $\|Ax - b\|_2^2 = f$

Since $\min_x \|Ax - b\|_2^2 = \min_x \|Ax - b\|$

cvxpy $\Rightarrow$ solution $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}$

(c) Zero. this function has no suboptimal point since it is convex.
we need to prove the function is convex function

$$\frac{\partial f}{\partial x_1} = 2(2x_1 - 1)\cdot 2 + 2(-x_1 + x_2)\cdot(-1) = 0$$

$$\frac{\partial f}{\partial x_2} = 2(-x_1 + x_2) + 2(2x_2 + 1)\cdot 2 = 0$$

$$\frac{\partial^2 f}{\partial x_1^2} = 2\cdot 2\cdot 2 + 2 = 10 \qquad \frac{\partial^2 f}{\partial x_1 \partial x_2} = -2$$

$$\frac{\partial^2 f}{\partial x_2^2} = 2 + 2\cdot 2\cdot 2 = 10 \qquad \frac{\partial^2 f}{\partial x_1 \partial x_2} = -2$$

$f : \mathbb{R}^2 \to \mathbb{R}$ & it is twice differentible over an open domain.
Then

$$\frac{\partial^2 f}{\partial x^2} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 10 & -2 \\ -2 & 10 \end{bmatrix}$$

(Use Hessian determinent)

$$\det\left(\frac{\partial^2 f}{\partial x^2}\right) = 100 - 4 = 96 \geq 0 \quad \text{for all } x \in \text{domain}(f)$$

Thus, f is a convex function // done

2. with python ~

3.

(a) $|y_i| = \max(-y_i, y_i)$

(i) if $y_i < 0$ $(-y_i > 0)$

L.H.S $= |y_i| = -y_i$

& R.H.S $= \max(-y_i, y_i)$ since we know $y_i < 0 < -y_i$

so, RHS $= -y_i$ $\therefore$ LHS = RHS, done

(ii) if $y_i \geq 0$, $(-y_i \leqslant 0)$

L.H.S $= |y_i| = y_i$

& R.H.S $= \max(-y_i, y_i)$, since we know $-y_i \leqslant 0 \leqslant y_i$

So, RHS $= y_i$ $\therefore$ LHS = RHS, done $\boxtimes$

(b) Let define set T and G as following

$T = \{ t_i \mid t_i \geq \max(-y_i, y_i) \ \forall i \}$, $G = \{ g_i \mid g_i = \max(-y_i, y_i) \ \forall i \}$

If there is j s.t

$\max(-y_j, y_j) \leqslant \max(-y_i, y_i) \ \forall i$, then we can say

$\max(-y_j, y_j)$ is a minimum value of set G.

That means, $\max(-y_j, y_j) = \min[\max(-y_i, y_i)]$

By the definition of t_i,

$\max(-y_j, y_j) \leqslant \max(-y_i, y_i) \leqslant t_i$

So, $\max(-y_j, y_j) \leqslant t_i$ and $\min t_i = \max(-y_j, y_j)$

$= \min[\max(-y_i, y_i)]$ // done

(c)

$\|t\|_1 = t_1 + t_2 + t_3 + \cdots + t_m$

$t \geq \max[-r, r]$

from previous problem 3-(b)

$\min t = \min[\max(-r,r)]$

<u>Note</u>

we need to prove

if $t \geq \max[-r,r]$, $\Rightarrow \|t\|_1 \geq \|\max(-r,r)\|$

since $t \geq \max(-r,r) \geq 0$, this inequality is valid

Thus, $\min \|t\|_1 = \min \|\max(-r,r)\|_1$

Since we can ignore sign of r when it is in the norm,

we can say $\|\max(-r,r)\|_1 = \|-r\|_1 = \|r\|_1 = \|A\theta - b\|_1$

Hence, $\min \|t\|_1 = \min \|A\theta - b\|_1$

// done

(d)

$$\|t\|_1 = t_1 + \cdots + t_m = [\underbrace{0 \cdots\cdots 0}_{n} \ \underbrace{1 \cdots\cdots 1}_{m}] \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \\ t_1 \\ \vdots \\ t_m \end{bmatrix}$$

$$\left.\begin{aligned} A\theta - b &\leq t \\ -(A\theta - b) &\leq t \end{aligned}\right\} \Rightarrow \left.\begin{aligned} A\theta - t &\leq b \\ -A\theta - t &\leq -b \end{aligned}\right\} \xRightarrow{\text{matrix form}} \begin{bmatrix} A & -I \\ -A & -I \end{bmatrix} \begin{bmatrix} \theta \\ t \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix}$$

4.

(a) Let $\|A\theta - y\|_2^2 = (A\theta - y)^T(A\theta - y)$

and $\lambda\|\theta\|_2^2 = \lambda\theta^T\theta$

to find $\hat{\theta}$ such that minimize $[\|A\theta - y\|_2^2 + \lambda\|\theta\|_2^2]$,

differentiate the objective function. w.r.t θ

let $\|A\theta - y\|_2^2 + \lambda\|\theta\|_2^2 = f$

$$f = \theta^T A^T A\theta - y^T A\theta - \theta^T A^T y + y^T y + \lambda\theta^T\theta$$

$$\frac{\partial f}{\partial \theta} = 2A^T A\theta - (y^T A)^T - A^T y + 2\lambda\theta$$

for $\hat{\theta}$ s.t

$$2A^T A\hat{\theta} - 2A^T y + 2\lambda\hat{\theta} = 0$$

, f has minimum val

so, $\cancel{2}A^T A\hat{\theta} - \cancel{2}A^T y + \cancel{2}\lambda\hat{\theta} = 0$

$$\Rightarrow [A^T A + \lambda I_n]\hat{\theta} = A^T y$$

$$\Rightarrow \hat{\theta} = [A^T A + \lambda I_n]^{-1} A^T y$$