Theoretical Part:

1.1.1) $f_1, \dots, f_m : C \to \mathbb{R}$ פונקציות קמורות

$y_1, \dots, y_m \in \mathbb{R}_+$

צ.ל: $g(u) = \sum_{i=1}^{m} y_i f_i(u)$ פונקציה קמורה.

הגדרת פונקציה קמורה:

לכל (u,v), $\alpha \in [0,1]$, מתקיים:

$$f(\alpha u + (1-\alpha)v) \le \alpha f(u) + (1-\alpha) f(v).$$

יהי (u,v):

הגדרת g: $g(u) = \sum_{i=1}^{m} y_i f_i(u)$

צ.ל: $g(\alpha u + (1-\alpha)v) \le \alpha g(u) + (1-\alpha) g(v)$

נתחיל מאגף שמאל:

$$g(\alpha u + (1-\alpha)v) \overset{\text{הגדרת } g}{=} \sum_{i=1}^{m} y_i f_i(\alpha u + (1-\alpha)v)$$

$$\overset{\text{כל } f \text{ קמורות}}{\le} \sum_{i=1}^{m} y_i (\alpha f_i(u) + (1-\alpha) f(v)) \overset{\text{הגדרת } g}{=} \alpha g(u) + (1-\alpha) g(v)$$

כנדרש $\boxtimes$

(1.1.2)

ניתן דוגמה נגדית:

$$f(x) = e^{-x} \qquad g(x) = x^2$$

שתיהן פונקציות קמורות, אבל מדק שני

$$f \circ g = e^{-x^2}$$

היא פונקציה שאינה קמורה, נראה לפי ע"י נגזרת שנייה:

$$f \circ g' = -2xe^{-x^2}$$

$$f \circ g'' = e^{-x^2}(4x^2-2)$$

ונאות פונקציה שיש לה ערכים חיוביים וגם שליליים

$$\boxed{\begin{aligned} f \circ g(0)'' &= e^0(-2) = -2 < 0 \\ f \circ g(1)'' &= e(2) = 2e > 0 \end{aligned}}$$

1.2.3) $x \in \mathbb{R}^d$ $\quad y \in \{\pm 1\}$

$$f(w,b) := \max\left(0, 1 - y(x^T w + b)\right)$$

יהי: $(w_1, b_1) = z_1$, $(w_2, b_2) = z_2$

צ.ל: $f(\alpha(z_1) + (1-\alpha)z_2) \leq \alpha f(z_1) + (1-\alpha) f(z_2)$

$$f(\alpha(z_1) + (1-\alpha)z_2) =$$

$$f\left(\alpha(w_1, b_1) + (1-\alpha)(w_2, b_2)\right)$$

$$= \max\left(0, 1 - y\left(x^T(\alpha w_1 + (1-\alpha) w_2) + \alpha b_1 + (1-\alpha) b_2\right)\right)$$

$$= \max\left(0, \alpha(1 - y(x^T w_1 + b_1)) + (1-\alpha)(1 - y(x^T w_2 + b_2))\right)$$

$$= \alpha f(z_1) + (1-\alpha) f(z_2)$$

כנדרש $\boxtimes$

1.2.4)

הגדרת סאב גרדיאנט:

$f(x) \geq f(x_0) + g^T(x - x_0)$ לכל $x \in \mathbb{R}^n$

נסמן: $1 - y(x^T w + b) = m$

נחלק למקרים:

$\underline{m > 0}$

$\Updownarrow$

$1 > y(x^T w + b)$

נגזור לפי b, w:

$\frac{\partial m}{\partial w} = -yx$, $\frac{\partial m}{\partial b} = -y$

$\Downarrow$

$\boxed{g=(-yx,-y)}$

$\underline{m \le 0}$

$y(x^T w + b) \le 0$

$\Downarrow$

מאחר ואנחנו בנקודת מינימום נקבל: $\boxed{g(0,0)}$

1.2.5)

$f_1,\dots,f_m:\mathbb{R}^d\to\mathbb{R}$

$g_k \in \partial f_k(x),\ k\in[m]$ סאב גרדיאנטים

מהגדרת סאב גרדיאנטים:

לכל z: $f_k(z)\ge f_k(x)+g_k^T(z-x)$

ולכן:

$$f(x) = \sum_{k=1}^{m} f_k(z) \geq \sum_{k=1}^{m} \left[f_k(x) + g_k^T (z-x) \right]$$

$$= \sum_{k=1}^{m} f_k(x) + \sum_{k=1}^{m} g_k^T (z-x) = f(x) + \left(\sum_{k=1}^{m} g_k \right)^T (z-x)$$

$$\Downarrow$$

$$\boxed{\sum_{k=1}^{m} g_k \in \partial f(x)}$$

כנדרש מהגדרת סאב גרדיאנט

1.2.6) בסעיפים קודמים ראינו שהנגזרת של $\ell^{hinge}_{x_i,y_i}(w,b)$

$$\overset{\ell^{hinge}_{x_i,y_i}}{(0,0)} \ , \ \overset{\ell^{hinge}_{x_i,y_i}}{(-y_i x_i, -y_i)}$$

$$t_i = \begin{cases} (0,0) & \ell = 0 \\ (-y_i x_i, -y_i) & \ell \neq 0 \end{cases}$$

ולכן מסעיף 5:

$$\partial \left(\frac{1}{m} \sum_{i=1}^{m} \ell^{hinge}_{x_i,y_i}(w,b) \right) \overset{\text{מכלל השרשרת}}{=} \frac{1}{m} \sum_{i} g_i$$

נגזור את $\frac{\lambda}{2}\|w\|^2$ מכלל השרשרת סאב גרדיאנט

שלו (בכיוון) $\lambda(w,0)$ ולכן כשמחברים את שני

חלקי הפונקציה נקבל בסך הכל

$$\boxed{\underbrace{\frac{1}{m} \sum_{i} t_i + \lambda(w,0)}_{\text{סאב גרדיאנט של הסכום.}} \in \partial \frac{1}{m}} \sum_{i=1}^{m} \ell^{hinge}_{x_i,y_i}(w,b) + \frac{\lambda}{2}\|w\|^2$$