חלק תיאורטי

1. Let $f_1, \ldots, f_m : C \to \mathbb{R}$ be a set of convex functions and $\gamma_1, \ldots, \gamma_m \in \mathbb{R}_+$. Prove from definition that $g(\mathbf{u}) = \sum_{i=1}^{m} \gamma_i f_i(\mathbf{u})$ is a convex function.

נרצה להוכיח כי $g(u)$ היא פונקציה קמורה כאשר לכל

$i \in [m]$ f_i פונ' קמורה ו- γ_i מספר ממשי חיובי לכל $i \in [m]$

נבחין שמכיון ש $\gamma_i > 0$ לכל $i \in [m]$ אז גם $\gamma_i f_i(u)$ קמורה

שהרי מתקיים $f(\alpha v + (1-\alpha)u) \leq \alpha f(v) + (1-\alpha) f(u)$

ואם נכפול בקבוע האי"ש ישמר ולכן

$$g(\alpha v + (1-\alpha)u) \leq \sum_{i=1}^{m} \gamma_i f_i(\alpha v + (1-\alpha)u) \leq$$

$$\leq \sum_{i=1}^{m} \alpha \gamma_i f_i(v) + (1-\alpha) f(u) = \alpha \sum_{i=1}^{m} \gamma_i f_i(v) + (1-\alpha) \sum_{i=1}^{m} \gamma_i f_i(u) =$$

$$= \alpha g(v) + (1-\alpha) g(u)$$

2. Give a counterexample for the following claim: Given two functions $f, g : \mathbb{R} \to \mathbb{R}$, define a new function $h : \mathbb{R} \to \mathbb{R}$ by $h = f \circ g$. If f and g are convex than h is convex as well.

דוגמא נגדית: $g(x) = x^2$, $f(x) = -x$ שתי הפונקציות קמורות

אבל $h(x) = -x^2$ שהיא ההרכבה אינה קמורה.

3. Let $f : C \to \mathbb{R}$ be a function defined over a convex set C. Prove that f is convex iff its *epigraph* is a convex set, where $\text{epi(f)} = \{(u,t) : f(u) \leq t\}$.

נניח ש-f קמורה ונראה ש $epi(f)$ סט קמור

נניח בשלילה שקיימות 2 נקודות שונות $(a,b),(c,d)\in epi(f)$

וגם $\lambda\in(0,1)$ ו- $\lambda(a,b)+(1-\lambda)(c,d)\in epi(f)$

נקבל ש: $f(a)\le b$ וגם $f(c)\le d$

וגם $f(\lambda a+(1-\lambda)f(c))\le\lambda f(a)+(1-\lambda)f(c)\le$

$\le\lambda b+(1-\lambda)d$ סתירה.

בכיוון השני: נניח ש $epi(f)$ כס קמור אזי לכל

$(x,y),(z,t)\in epi(f)$ ולכל $a\in[0,1]$ מתקיים:

$$\alpha(x,y)+(1-\alpha)(z,t)=(\alpha x+(1-\alpha)z,\alpha y+(1-\alpha)t)\in epi(f)$$

כלומר: $f(\alpha x+(1-\alpha)z)\le\alpha y+(1-\alpha)z$

נבחר $y=f(x)$, $t=f(z)$ ונקבל

$$f(\alpha x+(1-\alpha)z)\le\alpha f(x)+(1-\alpha)f(z)$$

ולכן f קמורה.

4. Let $f_i : V\to\mathbb{R}$, $i\in I$. Let $f:V\to\mathbb{R}$ given by

$$f(u)=\sup_{i\in I} f_i(u).$$

If f_i are convex for every $i\in I$, then f is also convex.

נראה ש $epi(f)=epi(\sup f_i(u))$ כס קמור ומשאלה 3

זה מוכיח ש- $\sup f_i(\omega)$ פונקציה קמורה

מתקיים $(a,b) \in epi(\sup f_i(\omega))$ אמ"מ $f(a) \le b$

$(a,b) \in epi(f_i) \Leftarrow \forall i \in I \quad f_i(\omega) \le b \quad \Leftarrow$

אם כן נקבל ש $epi(f) = \bigcap_{i \in I} epi(f_i)$ קמור כי חיתוך

של סט'ים קמורים הוא קמור

5. Given $\mathbf{x} \in \mathbb{R}^d$ and $y \in \{\pm 1\}$. Show that the hinge loss is convex in $\mathbf{w}, b$. That is, define

$$f(\mathbf{w}, b) := \ell^{hinge}_{\mathbf{x},y}(\mathbf{w}, b) = \max\left(0, 1 - y(\mathbf{x}^\top \mathbf{w} + b)\right)$$

and show that f is convex in $\mathbf{w}, b$.

ראינו ש- Pointwise-max קמורה כלומר בהנחה

שהפונקציות $1-y(w^t x+b)$ ו- 0 קמורות

$\max(0, 1-y(w^t x+b))$ קמורה

ונבחין שהפונקציה 0 קמורה כי היא ליניארית

ו- $1-y(w^t x+b)$ קמורה כי היא הרכבה על פונקציה אפינית

6. Deduce some sub-gradient of the hinge loss function $g \in \partial \ell^{hinge}_{\mathbf{x},y}(\mathbf{w}, b)$.

נבחין כי $\ell^{hinge}_{x,y}$ דיפרנציאבילית לכל w,b שלא מקיימים

$$y(x^t w + b) = 1$$

נבחין כי לכל (w,b) המקיימים $y(x^t w+b) \geq 1$ מתקיים

ש- $0 \in \partial l^{hinge}_{x,y}(w,b)$ ולכל $y(x^t w+b)<1$ מתקיים:

$$l^{hinge}_{x,y}(w,b) = 1 - y(x^t w+b) = 1 - yx^t w - yb \Rightarrow$$

$$\Rightarrow \partial^{hinge}_{x,y}(w,b) = \left(\frac{\partial}{\partial w}, \frac{\partial}{\partial b}\right) = (-yx, -y)$$

נבחר כי:

$$g = \begin{cases} 0 & y(x^t w+b) \geq 1 \\ (-yx, -y) & y(x^t w+b) < 1 \end{cases}$$

7. Let $f_1, \ldots, f_m : \mathbb{R}^d \to \mathbb{R}$ be a set of convex functions and $\mathbf{g}_k \in \partial f_k(\mathbf{x})$ for all $k \in [m]$ be sub-gradients of these functions. Define $f : \mathbb{R}^d \to \mathbb{R}$ by $f(\mathbf{x}) = \sum_{i=1}^m f_i(\mathbf{x})$. Show that $\sum_k \mathbf{g}_k \in \partial \sum_k f_k(\mathbf{x})$.

מהגדרת סאבגרדיאנט מתקיים לכל $i \in [m]$ ולכל $u \in \mathbb{R}^d$:

$$f_i(u) \geq f_i(x) + \langle g_i, u-x \rangle$$

$$\Rightarrow \sum_{i=1}^m f_i(u) \geq \sum_{i=1}^m f_i(x) + \langle g_i, u-x \rangle$$

$$\sum_k g_k \in \partial \sum f_k(x)$$ ← מההגדרה

8. Let $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^m \subseteq \mathbb{R}^d \times \{\pm 1\}$ be a sample and define $f : \mathbb{R}^d \to \mathbb{R}$ by:

$$f(\mathbf{w}, b) = \frac{1}{m} \sum_{i=1}^m \ell^{hinge}_{\mathbf{x}_i, y_i}(\mathbf{w}, b) + \frac{\lambda}{2} ||\mathbf{w}||^2$$

Find a sub-gradient of f for any $\mathbf{w}$.

בדומה לשאלה 6 נגדיר

$$g = \begin{cases} 0 & y_i(x_i^t w + b) \geq 1 \\ (-y_i x_i, -y_i) & y_i(x_i^t w + b) < 1 \end{cases}$$

ונקבל ש- $g_i \in \partial l^{hinge}_{x_i,y_i}(w,b)$

ומשאלה 7 נקבל:

$$\frac{1}{m}\sum_{i=1}^{m} g_i + \lambda(w,0) \in \partial \frac{1}{m}\sum_{i=1}^{m} l^{hinge}_{x_i,y_i}(w,b) + \frac{\lambda}{2}\|w\|^2$$