Short HW4 – Optimization, Regression, and Boosting

.1

1.1. הפונקציה f קמורה.

1.2. נגדיר את קבוצת ה- $sub-gradients$ של $f: V \to \mathbb{R}$ באופן הבא על פי ההגדרה מהשאלה:

$$\partial f(u) \triangleq \left\{ q \in V | \forall v \in V: f(v) \geqslant f(u) + q^T(v-u) \right\}$$

נציע $sub-gradients$ ל-$f(x)$ שנסמנה $g(x)$, כלומר $g(x) \in \partial f$ ונראה כי לכל $u \in \mathbb{R}$ מתקיים $g(u) \in \partial f(u)$:

עבור הפונקציה הבאה:

$$g(x) = \begin{cases} 2x & x < 0 \\ 2 & x \geqslant 0 \end{cases}$$

יהא $u \in \mathbb{R}$, נראה שלכל $v \in \mathbb{R}$ מתקיים $f(v) \geqslant f(u) + q^T(v-u)$ עבור $q = g(u)$:

נפריד למקרים:

<u>עבור $u < 0$ מתקיים:</u>

$$f(x) = u^2,\ g(u) = 2u$$

עבור $v \geqslant 0$ נקבל $f(v) = 2v$ כלומר בוודאות מתקיים $v > u$ ולכן:

$$f(v) = 2v \geqslant 0 > \underbrace{u}_{\text{שלילי}} \underbrace{(2v-u)}_{\text{חיובי}}$$

$$\Longrightarrow 2v \geqslant 2uv - u^2$$
$$\Longrightarrow 2v \geqslant u^2 + 2u(v-u)$$
$$\Longrightarrow f(v) \geqslant f(u) + g(u)(v-u)$$

עבור $v < 0$ נקבל $f(v) = v^2$ כלומר:

$$v^2 + u^2 \geqslant 2uv \qquad (1)$$
$$\Longrightarrow v^2 + 2u^2 \geqslant u^2 + 2uv$$
$$\Longrightarrow v^2 \geqslant u^2 + 2uv - 2u^2$$
$$\Longrightarrow v^2 \geqslant u^2 + 2(v-u)$$
$$\Longrightarrow f(v) \geqslant f(u) + g(u)(v-u)$$

נימוק ל-(1): לפי אי שוויון הממוצעים מתקיים $\frac{u^2+v^2}{2} \geqslant \sqrt{u^2v^2} = uv$

עבור $u \geqslant 0$ מתקיים:

$$f(u) = 2u,\ g(u) = 2$$

עבור $v \geqslant 0$ נקבל $f(v) = 2v$ כלומר:

$$f(v) = 2v \geqslant 2u + 2v - 2u$$
$$\Longrightarrow 2v \geqslant 2u + 2(v-u)$$
$$\Longrightarrow 2v \geqslant u^2 + 2u(v-u)$$
$$\Longrightarrow f(v) \geqslant f(u) + g(u)(v-u)$$

עבור $v < 0$ נקבל $f(v) = v^2$ כלומר:

$$v^2 \geqslant 2v$$

$$\Longrightarrow v^2 \geqslant 2v + 2u - 2u$$
$$\Longrightarrow v^2 \geqslant 2u + 2(v-u)$$
$$\Longrightarrow f(v) \geqslant f(u) + g(u)(v-u)$$

כלומר הראנו כי לכל $u \in \mathbb{R}$ מתקיים כי $g(u) \in \partial f(u)$, כלומר $g \in \partial f$ כמבוקש.

1.3. נציג את הטבלה כמבוקש:

i	x_i	$f(x_i)$	$\frac{\partial}{\partial x} f(x_i) = g(x_i)$
0	$x_0 = -1.5$	$f(x_0) = \frac{9}{4}$	$g(x_0) = -3$
1	$x_1 = x_0 - \frac{1}{4} \cdot (-3) = -\frac{3}{4}$	$f(x_1) = \frac{9}{16}$	$g(x_1) = -\frac{3}{2}$
2	$x_2 = x_1 - \frac{1}{4} \cdot \left(-\frac{3}{2}\right) = -\frac{3}{8}$	$f(x_2) = \frac{9}{64}$	$g(x_2) = -\frac{3}{4}$
3	$x_3 = x_2 - \frac{1}{4} \cdot \left(-\frac{3}{4}\right) = -\frac{3}{16}$	$f(x_3) = \frac{9}{256}$	$g(x_3) = -\frac{3}{8}$
4	$x_4 = x_3 - \frac{1}{4} \cdot \left(-\frac{3}{8}\right) = -\frac{3}{32}$	$f(x_4) = \frac{9}{1024}$	$g(x_4) = -\frac{3}{16}$

נבחין כי האלגוריתם יתכנס בסופו של דבר ל-0 (ברור כי ערכי $x_i \to 0$, שזוהי הנקודה שבה $f(x)$ מקבלת מינימום), כלומר בסופו של דבר לאחר מספיק איטרציות האלגוריתם יתכנס למינימום.

1.4. נציג את הטבלה כמבוקש:

i	x_i	$f(x_i)$	$\frac{\partial}{\partial x} f(x_i) = g(x_i)$
0	$x_0 = -1.5$	$f(x_0) = \frac{9}{4}$	$g(x_0) = -3$
1	$x_1 = x_0 + 3 = 1.5$	$f(x_1) = 3$	$g(x_1) = 2$
2	$x_2 = x_1 - 2 = -\frac{1}{2}$	$f(x_2) = \frac{1}{4}$	$g(x_2) = -1$
3	$x_3 = x_2 + 1 = \frac{1}{2}$	$f(x_3) = 1$	$g(x_3) = 2$
4	$x_4 = x_2 + 1 = -1.5$	$f(x_4) = \frac{9}{4}$	$g(x_3) = -3$
5	$x_5 = x_2 + 1 = 1.5$	$f(x_5) = 3$	$g(x_3) = 2$
6	$x_6 = x_2 + 1 = -\frac{1}{2}$	$f(x_6) = \frac{1}{4}$	$g(x_3) = -1$

נבחין כי במקרה זה האלגוריתם לא יתכנס בגלל גודל הצעד, נקבל כי שלבים 2-5 יחזרו על עצמם בלולאה אינסופית ולכן האלוריתם לא יתכנס למינימום אלא יעבור בין כמה ערכים קבועים.

.2

a. על מנת להראות כי מטריצה A היא PD יש להראות שני תנאים:

1. המטריצה סימטרית
2. המטריצה מקיימת:

$$\forall u \neq 0 : u^T A u > 0$$

נראה את שני התנאים עבור המטריצה המבוקשת $X^T X + m\lambda I$:

1. המטריצה $X^T X$ סימטרית מהיותה מטריצת $Gram$:

$$\left(X^T X\right)^T = X^T X^{T^T} = X^T X$$

כל מטריצה סקלרית היא סימטרית ובפרט $m\lambda I$, כאשר לפי תכונות חיבור מטריצות סכום של מטריצות סימטריות הוא סימטרי, לכן המטריצה $X^T X + m\lambda I$ היא סימטרית.

2. יהי וקטור $u \neq 0$ אשר תואם למימדי המטריצה ומאפשר כפל בה, כעת נכפול:

$$u^T \left(X^T X + m\lambda I\right) u = u^T X^T X u + u^T m\lambda I u$$

ידוע כי מטריצת $Gram$ היא לכל הפחות PSD ולכן לפי הגדרת מטריצה חיובית חצי מוגדרת נקבל כי $u^T X^T X u \geqslant 0$, בנוסף $u^T m\lambda I u = m\lambda u^T u = m\lambda ||u||^2$ ועבור וקטור u שאינו וקטור האפס נקבל כי $||u||^2 > 0$, כאשר m מייצג את מספר הדוגמאות הנדגמות ולכן תמיד חיובי (אחרת הבעיה אינה מוגדרת), וגם $\lambda > 0$.

כלומר סך הכל מתקיימים שני התנאים ולכן $X^T X + m\lambda I$ היא מטריצה PD.

b. הפונקציה $\frac{1}{m}||Xw - y||_2^2 + \lambda||w||_2^2$ היא קמורה (בגלל שהיא סכום של פונקציות קמורות - הוכח בתרגול), ניתן למצוא את נקודת המינימום שלה על ידי גזירה לפי w והשוואה ל-0.

נראה שלאחר הגזירה נקבל נקודת קיצון יחידה שתהיה נקודת המינימום:

$$\nabla_w \left(\frac{1}{m}||Xw - y||_2^2 + \lambda||w||_2^2\right) = \frac{2}{m} X^T (Xw - y) + 2\lambda w$$

נשווה ל-0 על מנת למצוא את נקודת הקיצון:

$$\frac{2}{m} X^T (Xw - y) + 2\lambda w = 0$$
$$\Longrightarrow X^T X w - X^T y + \lambda m w = 0$$
$$\Longrightarrow \left(X^T X + \lambda m I\right) w = X^T y$$

לפי סעיף א ידוע כי $X^T X + \lambda m I$ היא PD ולכן הפיכה, כלומר $\left(X^T X + \lambda m I\right)^{-1}$ היא מטריצה מוגדרת היטב, לכן קיים פתרון יחיד עבור $w = \left(X^T X + \lambda m I\right)^{-1} X^T y$, ולפי תחילת הסעיף ידוע כי לפונקציה נקודת מינימום יחידה (לפי תכונות של פונקציה קמורה), נקבל:

$$\hat{w} = \left(X^T X + \lambda m I\right)^{-1} X^T y$$

כנדרש.

c. תחילה נרשום את פירוק ה-SVD של X:

$$X^TX = \left(U\Sigma V^T\right)^T U\Sigma V^T = V\Sigma^T U^T U\Sigma V^T \underbrace{=}_{\substack{\text{עבור מטריצות} \\ \text{אורתונורמליות} \\ U^TU=I}} V\Sigma^T\Sigma V^T$$

וכעת נציב בביטוי שמצאנו בסעיף b:

$$\begin{aligned}
\hat{w} &= \left(X^TX + \lambda mI\right)^{-1} X^Ty \\
&= \left(V\Sigma^T\Sigma V^T + \lambda mI\right)^{-1} V\Sigma^T U^T y \\
&= \left(V\Sigma^T\Sigma V^T + \lambda m \underbrace{VV^T}_{=I}\right)^{-1} V\Sigma^T U^T y = \\
&= \left(V\left(\Sigma^T\Sigma + \lambda mI\right)V^T\right)^{-1} V\Sigma^T U^T y = \\
&= \underbrace{\left(V^T\right)^{-1}}_{=V} \left(\Sigma^T\Sigma + \lambda mI\right)^{-1} \underbrace{V^{-1}V}_{=I}\Sigma^T U^T y \\
&= V\left(\Sigma^T\Sigma + \lambda mI\right)^{-1} \Sigma^T U^T y
\end{aligned}$$

כלומר הראנו את המבוקש.

3. הדוגמאות אשר אפשר לקבל אחרי איטרציה אחת של $AdaBoost$ ומסווג חלש עבור כן אחת מהן:

$(a), (c)$

אציין כי המלבן הכתום מייצג תיוג שלילי והסיווג הכחול תיוג חיובי.