מבוא למערכות לומדות - תרגיל קצר 4

מגיש - עומר שמחי, 316572593

6 ביוני 2021

1. ניזכר בנוסחאת ה־$Soft-SVM$ -

$$argmin_{w\in\mathbb{R}^d}\frac{1}{m}\sum_{i=1}^{m}\max\left\{0,1-y_iwx_i\right\}+\lambda\left\|w\right\|_2^2$$

(א) תהיינה $f,g:C\to\mathbb{R}$ שתי פונקציות קמורות. הוכיחו כי $q\left(z\right):=\max\left\{f\left(z\right),g\left(z\right)\right\}$ קמורה.

הוכחה - נוכיח עפ"י ההגדרה. יהיו $z_1,z_2\in\mathbb{R}$ וכן $t\in\left[0,1\right]$. נרשום:

$$\boldsymbol{q\left(tz_1+\left(1-t\right)z_2\right)}=\max\left\{f\left(tz_1+\left(1-t\right)z_2\right),g\left(tz_1+\left(1-t\right)z_2\right)\right\}\underbrace{\leq}_{f,g\,convax}$$

$$\leq\max\left\{tf\left(z_1\right)+\left(1-t\right)f\left(z_2\right),tg\left(z_1\right)+\left(1-t\right)g\left(z_2\right)\right\}\leq$$

$$\leq\max\left\{tf\left(z_1\right),tg\left(z_1\right)\right\}+\max\left\{\left(1-t\right)f\left(z_2\right),\left(1-t\right)g\left(z_2\right)\right\}=$$

$$=t\cdot\max\left\{f\left(z_1\right),g\left(z_1\right)\right\}+\left(1-t\right)\cdot\max\left\{f\left(z_2\right),g\left(z_2\right)\right\}=$$

$$\underbrace{=}_{by\,def.}\boldsymbol{tq\left(z_1\right)+\left(1-t\right)q\left(z_2\right)}$$

ולכן מהגדרה q קמורה.

(ב) הסק כי פונקציית ה־$hing$ - $l_{hinge}\left(z\right)=\max\left\{0,1-z\right\}$ קמורה.

הוכחה - בהסתמך על הסעיף הקודם מספיק להוכיח כי $0,1-z$ קמורות. אולם זה מיידי היות שפונקציות ליניאריות וקבועות הן קמורות (מתקיים עבורן שיוויון ממש בהגדרת הקמירות, הוכח בתרגול).

(ג) הוכיחו כי הפונקציה $\max\left\{0,1-y_iw^Tx_i\right\}$ קמורה כפונקציה של w.

הוכחה - בהתסמך על הסעיפים הקודמים, מספיק להראות כי $1-y_iw^Tx_i$ קמורה במשתנה w. מההרצאה $f_w\left(x\right):=w^Tx$ ליניארית ב־w. מכאן שגם $y_if_w\left(x_i\right)$ ליניארית, וסה"כ $1-y_iw^Tx_i=1-y_if_w\left(x_i\right)$ ליניארית ב־w. הוכחנו בתרגול שפונקציה ליניארית בפרט קמורה ולכן סיימנו.

(ד) תהי $f:C\to\mathbb{R}$ פונקציה קמורה ויהי $\alpha\in\mathbb{R}_{\geq 0}$. הוכיחו כי $u(z):=\alpha f(z)$ קמורה.

הוכחה - נוכיח עפ"י ההגדרה. יהיו $z_1,z_2\in\mathbb{R}$ וכן $t\in[0,1]$. נרשום:

$$u(tz_1+(1-t)z_2)=\alpha f(tz_1+(1-t)z_2)\underbrace{\leq}_{\alpha\geq 0,f\,convax}$$
$$\leq\alpha\cdot(tf(z_1)+(1-t)f(z_2))=t\alpha f(z_1)+(1-t)\alpha f(z_2)=$$
$$=tu(z_1)+(1-t)u(z_2)$$

ולכן u קמורה כנדרש.

(ה) הוכיחו כי נוסחאת ה־$Soft-SVM$ קמורה.

הוכחה - בתרגול הוכחנו כי $f(w):=\|w\|_2^2$ קמורה. מסעיף קודם, היות ש־$\lambda\geq 0$ אזי גם $\lambda\|w\|_2^2$ קמורה. מסעיף 3, נקבל כי לכל i $\max\{0,1-y_iw^Tx_i\}$ קמורה. בתרגול הוכחנו כי לכל f,g קמורות, גם $f+g$ קמורות. לכן באנדוקציה, היות ש־$f_i(w):=\max\{0,1-y_iw^Tx_i\}$ קמורות, אזי שגם $\sum_{i=1}^m\max\{0,1-y_iwx_i\}$ ושוב ע"י הסעיף הקודם היות ש־$m>0$ נסיק כי $\frac{1}{m}\sum_{i=1}^m\max\{0,1-y_iwx_i\}$ קמורה ולכן שוב ע"י שימוש בקמירות סכום של פונקציות, נקבל כי $\frac{1}{m}\sum_{i=1}^m\max\{0,1-y_iwx_i\}+\lambda\|w\|_2^2$ קמורה ובזאת סיימנו (בביטוי יש גם את $argmin_{w\in\mathbb{R}^d}$, אולם זה כמובן לא משנה את עובדת הקמירות היות שעבור כל w הוכחנו קמירות, ובפרט לזה שמספק את הביטוי המינימאלי).

2. נגדיר עבור פונקציה $f:V\to\mathbb{R}$ את הקבוצה הבאה:

$$\partial f(u)=\left\{q\in V\,|\,\forall v\in V:\ f(v)\geq f(u)+q^T\cdot(v-u)\right\}$$

תהי $f(x)=|2x-1|$.

(א) האם f קמורה?

פתרון - **כן.** נשים לב כי $f(x)=|2x-1|=\max\{0,2x-1\}$ ולכן קמורה כפי שהראינו בשאלה 1.

(ב) הוכיחו כי $0\in\partial f(0.5)$.

הוכחה - ראשית נרשום את הקבוצה $\partial f(0.5)$:

$$\partial f(0.5)=\left\{q\in\mathbb{R}\,|\,\forall v\in\mathbb{R}:\ f(v)\geq f(0.5)+q^T\cdot(v-0.5)\right\}=$$
$$\underbrace{=}_{f(0.5)=0}\{q\in\mathbb{R}\,|\,\forall v\in\mathbb{R}:\ f(v)\geq q\cdot(v-0.5)\}$$

נשים לב כי מהגדרתה תמיד מתקיים $f(v)\geq 0$ לכל $v\in\mathbb{R}$ ולכן עבור $q=0$ נקבל:

$$f(v)\geq 0=q\cdot(v-0.5)$$

ולכן מהגדרה $0\in\partial f(0.5)$.

(ג) ראשית:

$$\frac{\partial f}{\partial x}=\begin{cases}2 & x>0.5\\ -2 & x<0.5\\ 0 & x=0.5\end{cases}$$

נמלא את הטבלה בדומה לתרגול עבור $x_0=-1$ ו־$\eta=0.25$:

i	x_i	$\frac{\partial f}{\partial x}(x_i)$
0	-1	-2
1	-0.5	-2
2	0	-2
3	0.5	0

חישובי עזר:

$$\begin{aligned} x_1 &= x_0-\eta\frac{\partial f}{\partial x}(x_0)=-1-0.25\cdot(-2)=-0.5\\ x_2 &= x_1-\eta\frac{\partial f}{\partial x}(x_1)=-0.5-0.25\cdot(-2)=0\\ x_3 &= x_2-\eta\frac{\partial f}{\partial x}(x_2)=0-0.25\cdot(-2)=0.5\end{aligned}$$

והגענו לנקודת המינימום הגלובלי של f, $x=0.5$. מכאן האלגוריתם אכן מתכנס למינימום.

(ד) נראה כי קריטריון ההתכנסות הוא $\eta<1$. ראשית, עבור $\eta=1$ נקבל את הריצה הבאה:

i	x_i	$\frac{\partial f}{\partial x}(x_i)$
k	x_k	-2
$k+1$	$x_k-(-2)\eta=x_{k+1}$	2
$k+2$	$x_k-2\eta+2\eta=x_k$	-2
$k+3$	$x_k-(-2)\eta=x_{k+1}$	2

ולכן האלגוריתם נתקע בלולאה בין שתי נקודות קבועות וכמובן לא יתכנס. לכן, כדי שנקבל התכנסות בהכרח צריך להתקיים $\frac{\partial f}{\partial x}(x_k)=-2$ לכל איטרציה k עד אשר נתכנס. עבור איטרציה i שבה עוד לא הגענו להתכנסות מתקיים:

$$x_i=-1-\sum_{j=1}^{i-1}\eta\cdot\frac{\partial f}{\partial x}(x_j)=-1`-\eta\cdot\sum_{j=1}^{i-1}-2=-1-2\eta(i-1)$$

ולכן עבור η קטנה מספיק, נקבל כי באיטרציה ה־i נגיע ל־0, כלומר צריך להתקיים:

$$-1-2\eta(i-1)=0.5\Longrightarrow\eta=\frac{3}{4\cdot(i-1)}$$

ומכאן נקבל כי האלגוריתם יתכנס אמ"מ $\eta\in\left\{\frac{3}{4\cdot k}\,|\,k\in\mathbb{N}\right\}$. נשים לב כי הערכים הללו דיסקרטיים, ולכן לא לכל η קטנה מספיק יש התכנסות, אלא רק ל־ηות המסוימות הנ"ל.

(ה) **הטענה לא נכונה.** ניקח את f מהסעיף הקודם שאיננה גזירה בכל תחום הגדרתה (תמיד יהיה "שפיץ"). ראינו כי האלגוריתם יתכנס אמ"מ $\eta \in \left\{\frac{3}{4\cdot k} \mid k \in \mathbb{N}\right\}$.
כעת נבחר η קטנה כרצונינו שאיננה מתוך הקבוצה הנ"ל. ראינו כי בתנאי זה אין התכנסות ולכן במקרה זה האלגוריתם לא יתכנס.