מבוא למערכות לומדות

236756

Hw3_dry

מספר זהות

313511602

1

1.1

תהי A קבוצה בעלת 5 נקודות שונות במישור: $C = \{p_1 = (a_1, b_1), \dots, p_5 = (a_5, b_5)\}$.
נוכיח כי H אינה מנפצת את A. נבחר את התוויות של כל נקודה כך ש-H אכן אינה תנפץ את A.

נסמן את הנקודות כך:

- p_{left} – הנקודה בעלת שיעור ה-a הנמוך ביותר ואת שיעורי הנקודות ב-(a_{left}, b_{left}).
- p_{right} – הנקודה בעלת שיעור ה-a הגבוה ביותר ואת שיעורי הנקודות ב-(a_{right}, b_{right}).
- p_{down} – הנקודה בעלת שיעור ה-b הנמוך ביותר ואת שיעורי הנקודות ב-(a_{down}, b_{down}).
- p_{up} – הנקודה בעלת שיעור ה-b הגבוה ביותר ואת שיעורי הנקודות ב-(a_{up}, b_{up}).

כעת, נקבע את התוויות של כל נקודה $f: \{p_1, p_2, p_3, p_4, p_5\} \to \{0,1\}$ בצורה הבאה:

- במקרה בו כל הנקודות שסימנו שונות זו מזו אז נגדיר:
p_{middle} – הנקודה החמישית שלא בחרנו עד כה ואת שיעורי הנקודות ב-(a_{middle}, b_{middle}).

$$f(p_{left}) = f(p_{right}) = f(p_{up}) = f(p_{down}) = 1$$
$$f(p_{middle}) = 0$$

- אחרת, נגדיר את הסימון של הנקודות בצורה הבאה:
p_{middle} – נבחר את אחת הנקודות שלא בחרנו עד כה ונסמן את שיעוריה ב-(a_{middle}, b_{middle}). כמו כן נגדיר את התוויות כך:

$$\forall i \in \{1, \dots, 5\}: f(p_i) = \begin{cases} 1, p_i \in \{p_{down}, p_{left}, p_{right}, p_{up}\} \\ 0, \quad else \end{cases}$$

נשים לב שהסימונים הנוספים שהוספנו הן גם במקרה הזה (שסימון הנקודות המקורי אינו מכסה את כל הנקודות), וההבדל הוא שהבאנו את התווית לכל נקודה לפי הנקודה עצמה ולא לפי הסימון שהבאנו. בעצם המשך ההוכחה יהיה לפי הסימונים שהבאנו, אפילו במקרה הזה (כי בעצם כן יהיה לנו את הנקודות הקיצוניות ביותר שאנו צריכים להמשך ההוכחה).

על מנת לנתץ את C בהכרח נצטרך לבחור היפותזה שתכיל את כל הנקודות בעלות התווית 1. כלומר אם נסמן את 4 הקודקודים של המלבן שנבחר ב:

$$(x_{left}, y_{left}), (x_{right}, y_{right}), (x_{up}, y_{up}), (x_{down}, y_{down})$$

כאשר הבחירה היא על פי המיקום היחסי של הקודקודים בינם לבין עצמם (באופן דומה בדיוק לסימון הנקודות שעשינו לעיל) אזי בהכרח מתקיים:

$$x_{left} \leq a_{left}, a_{right}, a_{up}, a_{down} \leq x_{right}$$

$$y_{down} \leq b_{left}, b_{right}, b_{up}, b_{down} \leq y_{up}$$

לכן, על פי בחירת הנקודה החמישית - p_{middle}, בהכרח גם מתקיים:

$$x_{left} \leq a_{middle} \leq x_{right}$$

$$y_{down} \leq b_{middle} \leq y_{up}$$

ולכן, על פי הגדרת המלבן, נקבל כי $f(p_{middle}) = 1$, בניגוד לתווית שנתנו לו.

לסיכום, בחרנו קבוצה שרירותית של 5 נקודות במישור, והראנו שעבור תוויות מסוימות H אינה יכולה לנתץ את C, ולכן על פי הגדרה: $VCdim(H) < 5$.

1.2

נניח ש $\mathcal{H}1 \subseteq \mathcal{H}2$ אזי לכל $h \in \mathcal{H}1$ מתקיים $h \in \mathcal{H}2$.

נניח בשלילה ש $VCdim(\mathcal{H}1) > VCdim(\mathcal{H}2)$ אזי קיים סט נקודות C כך ש $\mathcal{H}1$ מנתץ את C אבל $\mathcal{H}2$ לא מנתץ את C

<= קיים מיתוג ערכים וקיים $h1 \in \mathcal{H}1$ כך ש $h1(x_i) = y_i,\ \forall i \in m$ אבל לכל $h2 \in \mathcal{H}2$ מתקיים $h2(x_i) \neq y_i$.

<= וזה סתירה כי הנחנו שלכל $h \in \mathcal{H}1$ מתקיים $h \in \mathcal{H}2$ ובפרט ל $h1 \in \mathcal{H}1$ מתקיים $h1 \in \mathcal{H}2$

<= בהכרח $VCdim(\mathcal{H}1) \leq VCdim(\mathcal{H}2)$.

1.3

נשים לב שעץ בעומק 4 יכול לייצר 4 חוצצים במקביל לצירים $a1, a2, b1, b2$.

אז נרצה לומר ש $\mathcal{H}rect \subseteq \mathcal{H}ds$

ואז נקבל לפי הסעיף הקודם ש $4 = VCdim(\mathcal{H}rect) \leq VCdim(\mathcal{H}ds)$.

נוכיח ש $\mathcal{H}rect \subseteq \mathcal{H}ds$:

יהי $h \in \mathcal{H}rect$ אזי קיים מלבן לפי $a1, a2, b1, b2$ שכל מה שבפנים ערכו 1 וכל מה שבחוץ ערכו 1-.

נגדיר פיצול לפי אותם תנאים של המלבן דהיינו

פיצול 1: $\begin{matrix} -1 & x < a1 \\ 1 & x > a1 \end{matrix}$

פיצול 2: $\begin{matrix} 1 & x < a2 \\ -1 & x > a2 \end{matrix}$

פיצול 3: $\begin{matrix} -1 & x < b1 \\ 1 & x > b1 \end{matrix}$

פיצול 4: $\begin{matrix} 1 & x < b2 \\ -1 & x > b2 \end{matrix}$

קיבלנו ש $h \in \mathcal{H}ds$

הוכחנו שלכל $h \in \mathcal{H}rect$ מתקיים $h \in \mathcal{H}ds$ כדרוש.

2

נגדיר את הקרנל החדש בצורה הבאה:

נבחר $n_3 = n_1 + n_2$,ונגדיר $\phi^3 : \chi \to R^{n_1+n_2}$ כך:

$$\phi_3(u) = \begin{pmatrix} 2[\phi_1(u)] \\ 3[\phi_2(u)] \end{pmatrix}$$

נוכיח כי אכן זהו קרנל חוקי וכן שהוא מתאים לדרישה שלנו:

$$k_3(u, v) = 4k_1(u, v) + 9k_2(u, v) = 4\phi_1(u) * \phi_2(v) + 9\phi_1(u) * \phi_2(v)$$

$$= (2[\phi_1(u)] \quad 3[\phi_2(u)]) \cdot \begin{pmatrix} 2[\phi_1(u)] \\ 3[\phi_2(u)] \end{pmatrix} = \phi_3(u)^T \cdot \phi_3(v) = \langle \phi_3(u), \phi_3(v) \rangle$$

והוכחנו שעבור n_3 שבחרנו ועבור ϕ^3 שהגדרנו, אכן מתקיים הדרוש.

3

3.1

יהיו $x_1, x_2 \in C$ ויהי $t \in [0,1]$, צריך להוכיח: $q(tx_1 + (1-t)x_2) \leq tq(x_1) + (1-t)q(x_2)$.

ניעזר בהוכחה שלנו ב-2 למות:

- למה 1: יהיו $a, b, c, d \in R$ אזי מתקיים:
  $$\max\{a + b, c + d\} \leq \max\{a, c\} + \max\{b, d\}$$
  הוכחה: נשים לב שמתקיים:
  $$a \leq \max\{a, c\}, b \leq \max\{b, d\}$$
  $$\Rightarrow a + b \leq \max\{a, c\} + \max\{b, d\}$$

באופן דומה ניתן להוכיח כי:

$$c + d \leq \max\{a, c\} + \max\{b, d\}$$

ולכן סה"כ נקבל שאם גם $a+b$ וגם $c+d$ חסומים מלמעלה על ידי $\max\{a,c\}+\max\{b,d\}$, ולכן:

$$\max\{a+b,c+d\}\le\max\{a,c\}+\max\{b,d\}$$

- למה 2: עבור $\alpha\ge 0$ ועבור $x,y\in R$ מתקיים: $\max\{\alpha x,\alpha y\}=\alpha\max\{x,y\}$.

$$q(tx_1+(1-t)x_2)=\max\{f(tx_1+(1-t)x_2),g(tx_1+(1-t)x_2)\}\underbrace{\le}_{f,g\ are\ convex}$$

$$\le\max\{tf(x_1)+(1-t)f(x_2),tg(x_1)+(1-t)g(x_2)\}\underbrace{\le}_{lemma\ 1}$$

$$\max\{tf(x_1),tg(x_1)\}+\max\{(1-t)f(x_2),(1-t)g(x_2)\}\underbrace{=}_{lemma\ 2}$$

$$=t\max\{f(x_1),g(x_1)\}+(1-t)\max\{f(x_2),g(x_2)\}=$$

$$tq(x_1)+(1-t)q(x_2)$$

(3.2

על פי טענה שראינו בתרגול 7, כל פונקציה לינארית היא קמורה, ולכן $1-y_iw^Tx_i$ גם היא קמורה (ניתן בנוסף להוכיח את זה באמצעות המשפט שראינו בתרגול שפונקציה גזירה פעמיים היא קמורה אמ"מ $\nabla^2(1-y_iw^Tx_i)\succcurlyeq 0$, והרי מטריצת ההסיאן של הפונקציה הזו היא מטריצת האפס שהיא *PSD)*. גם פונקציית ה-0 היא קמורה מאותה הטענה. ולכן על פי הסעיף הקודם:

$\max\{0,1-y_iw^Tx_i\}$ היא פונקציה קמורה.

(3.3

הוכחנו בסעיף הקודם כי $\max\{0,1-y_iw^Tx_i\}$ פונקציה קמורה. על פי טענה שראינו בתרגול, סכום שופי של פונקציות קמורות הוא קמור, ולכן: $\sum_{i=1}^{m}\max\{0,1-y_iw^Tx_i\}$ הוא פונקציה קמורה.

על פי הטענה שנתונה לנו בתחילת השאלה, מכיוון ש-$0<\frac{1}{m}$, אז $\frac{1}{m}\sum_{i=1}^{m}\max\{0,1-y_iw^Tx_i\}$ זה גם פונקציה קמורה.

כמו כן, ראינו בתרגול כי $||w||_2^2$ היא פונקציה קמורה, וכמו כן מהטענה הנתונה בתחילת השאלה עבור $\lambda\ge 0$, נקבל כי $\lambda||w||_2^2$ הוא פונקציה קמורה.

סך הכל, $\frac{1}{m}\sum_{i=1}^{m}\max\{0,1-y_iw^Tx_i\}$ וכן $\lambda||w||_2^2$ הן קמורות וסכום של שתי פונקציות קמורות הוא קמור:

$$\underset{w\in R^d}{argmin}\frac{1}{m}\sum_{i=1}^{m}\max\{0,1-y_iw^Tx_i\}+\lambda||w||_2^2$$

ולכן סך הכל *soft-SVM* היא בעיה קמורה.