Short HW 3

Ori Zohar - 205960750

VC-dimension exercises . 1

1.1) יהי סט של 5 נק' ב- R^2 נסמנם $C=\{P_1,P_2,P_3,P_4,P_5\}$

$C \subseteq R^2$ לכן לכל נקודה P_i יש 2 קורדינטות (X_1,X_2).

נסמן $X_1^{max}, X_1^{min}, X_2^{min}, X_2^{max}$ את הערכים המינימלים והמקסימלים

של ערכי X_1 ו- X_2 בקבוצה C.

נבחר P_i שמק"ית $X_1(P_i)=X_1^{min}$ ונסמן y_i המתאים ל-P_i הוא 1.

נבחר P_i שמק"ית $X_1(P_i)=X_1^{max}$ ונסמן y_i המתאים ל-P_i הוא 1.

נבחר P_i שמק"ית $X_2(P_i)=X_2^{min}$ ונסמן y_i המתאים ל-P_i הוא 1.

נבחר P_i שמק"ית $X_2(P_i)=X_2^{max}$ ונסמן y_i המתאים ל-P_i הוא 1.

* אם בחרנו P_i שכבר נבחרה לא נעשה דבר.

* בבחירה בכל אחד מה-4 הבחירות יש P_i מתאים

כי ערכי $X_i^{min/max}$ לקוחים מתוך הקבוצה.

על מנת לסווג נכונה את הנקודות שנבחרו כל הקבוצה

$h_\Theta \in H_{rect}$ תקיים כי הריבוע המתקבל מערכי Θ

מכיל את הריבוע המתקבל מערכי $x_1^{max}, x_1^{min}, x_2^{min}, x_2^{max}$.

בה"כ בחרנו לכל היותר 4 מ-C ולכן קיימת נקודה p'

שלא נבחרה ומקיימת ש- $x_1^{min} \le p'(x_1) \le x_1^{max}$ וגם

$x_2^{min} \le p'(x_2) \le x_2^{max}$. נקבע $y'=-l$ ונקבל שלא קיימת

היפותזה המסווגת אותה נכונה.

הראנו ש- $VC(H_{rect})<5$ ובתרגול ראינו ש- $VC(H_{rect})\geq 4$

לכן $VC(H_{rect})=4$.

ב.2) נתונות 2 מחלקות היפותזות H_1, H_2 כך ש-

$H_1 \subseteq H_2$. נסמן $VCDIM(H_1)=m$.

מהגדרת $VCDim$ קיים סט C של m נקודות כך

שלא משנה איזה לייבל ניתן להם קיימת היפותזה ב-H_1

המסווגת אותן נכונה. בגלל ש- $\mathcal{H}_1 \subseteq \mathcal{H}_2$ אז גם קיימת

היפותזה ב- $\mathcal{H}_2$ המסווגת אותן נכונה (אותה אחת שסווגה נכונה

ב- $\mathcal{H}_1$). לפי מהגדרת VCDim:

$$VCDim(\mathcal{H}_2) \geq |C| = m = VCDim(\mathcal{H}_1)$$

1.3) נראה ש- $|\mathcal{H}_{DS(4)}| \leq 4$ ע"י כך ש-

נראה ש- $\mathcal{H}_{DS(4)} \subseteq \mathcal{H}_{rect}$ מטענה 1 נקבל ש- $VC(\mathcal{H}_{DS(4)}) \leq VC(\mathcal{H}_{rect})$

ומטענה 2 ראינו ש- $VC(\mathcal{H}_{rect}) = 4$ ולכן מתקבל ש- $|\mathcal{H}_{DS(4)}| \leq 4$

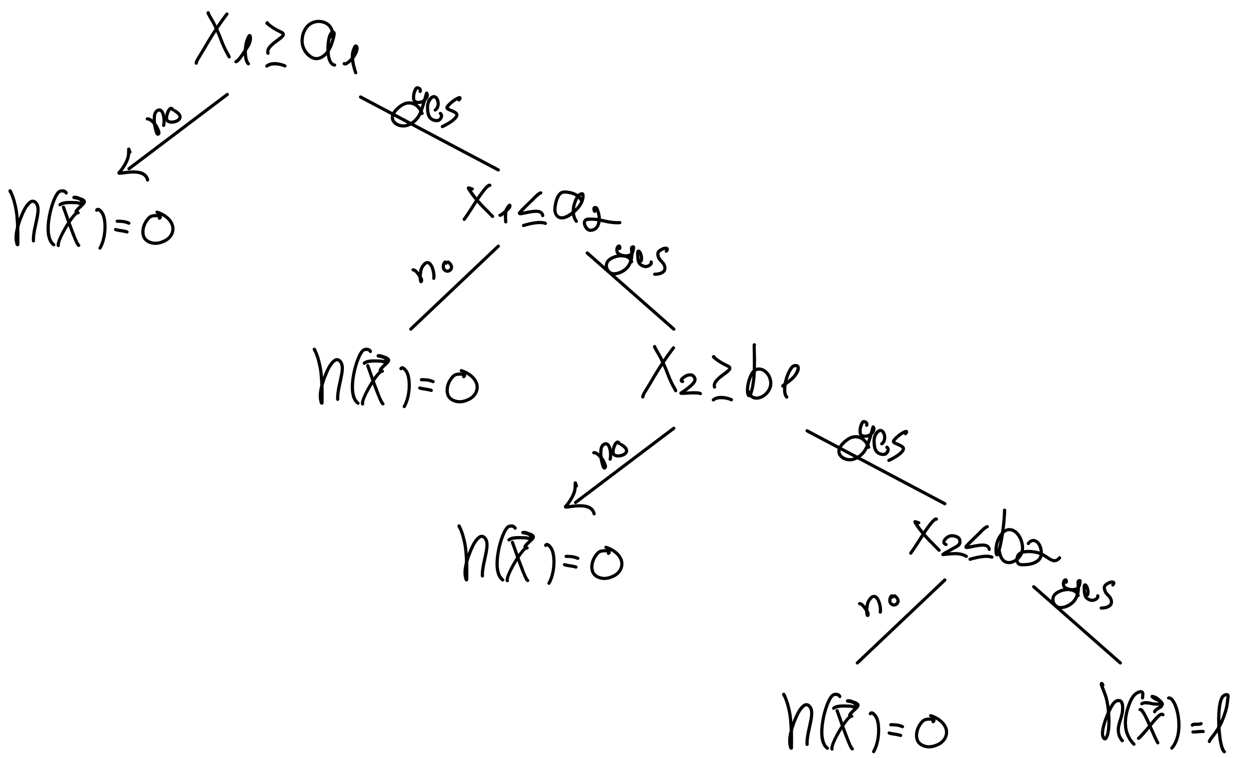
כנדרש.

יהי $h \in \mathcal{H}_{rect}$ ולכן מהגדרת $\mathcal{H}_{rect}$ קיימים a_1, a_2, b_1, b_2

כך שלכל דוגמה $\vec{X} = \binom{x_1}{x_2}$ מתקיים.

$$h(\vec{X}) = \begin{cases} 1 & a_1 \leq x_1 \leq a_2,\ b_1 \leq x_2 \leq b_2 \\ 0 & \text{אחרת} \end{cases}$$

את h ניתן לייצג באמצעות עץ ההחלטה הבא:



זהו עץ החלטה בעומק 4 ולכן $h \in H_{DT(4)}$ כנדרש.

Kernel 2

$$K_3(u,v) = 4 \cdot K_1(u,v) + 9 \cdot K_2(u,v)$$

$$= 4 \cdot (\phi_1(u)^T \cdot \phi_1(v)) + 9(\phi_2(u)^T \cdot \phi_2(v))$$

$$\overset{*}{=} (2\phi_1(u)^T, 3\phi_2(u)^T) \begin{pmatrix} 2\phi_1(v) \\ 3\phi_2(v) \end{pmatrix}$$

$$\overset{*}{=} \begin{pmatrix} 2\phi_1(u) \\ 3\phi_2(u) \end{pmatrix}^T \cdot \begin{pmatrix} 2\phi_1(v) \\ 3\phi_2(v) \end{pmatrix}$$

$$= \left\langle \begin{pmatrix} 2\phi_1(u) \\ 3\phi_2(u) \end{pmatrix}, \begin{pmatrix} 2\phi_1(v) \\ 3\phi_2(v) \end{pmatrix} \right\rangle$$

$$\overset{*}{=} \langle \phi_3(u), \phi_3(v) \rangle$$

ובסה"כ ראינו ש- $K_3(u,v) = \langle \phi_3(u), \phi_3(v) \rangle$

לכן K_3 קרנל חוקי.

* $\phi_1(x)$ הוא וקטור באורך n_1 ו- $\phi_2(x)$ הוא וקטור באורך n_2 לכן בסימון $(2\phi_1(u)^T, 3\phi_2(u)^T)$ הכוונה לוקטור שורה באורך n_1+n_2 כך ש- $2\phi_1(u)^T$ ו- $3\phi_2(u)^T$ הם בלוקי שורה באורכים n_1 ו- n_2 בהתאמה ובאופן דומה $\begin{pmatrix}2\phi_1(u)\\3\phi_2(v)\end{pmatrix}$ זהו וקטור עמודה.

* בגלל שאלו וקטורי שורה ועמודה יש שקילות בסימונים.

* נגדיר $\phi_3(v) = \begin{pmatrix}2\phi_1(v)\\3\phi_2(u)\end{pmatrix}$ ומתקיים ש- $n_3=n_1+n_2$.

3. Convexity

3.1) יהיו c_1,c_2 ו- $t\in(0,1)$:

$$g(t\cdot c_1+(1-t)\cdot c_2)$$
$$=\max\left(f(t\cdot c_1+(1-t)\cdot c_2),\, g(t\cdot c_1+(1-t)\cdot c_2)\right)$$
$$\overset{(*)}{\leq}\max\left(t\cdot f(c_1)+(1-t)\cdot f(c_2),\, t\cdot g(c_1)+(1-t)\cdot g(c_2)\right)$$
$$\overset{(*)}{\leq}\max\left(t f(c_1),\, t g(c_1)\right)+\max\left((1-t)f(c_2),\,(1-t)g(c_2)\right)$$
$$=t\cdot\max\left(f(c_1),g(c_1)\right)+(1-t)\max\left(f(c_2),g(c_2)\right)$$
$$=t\cdot g(c_1)+(1-t)g(c_2)$$

(*) f,g פונ' קמורות ולכן לפי ההגדרה

לכל c_1,c_2 ולכל $t\in(0,1)$ מתקיים :

$$f(t\cdot c_1+(1-t)\cdot c_2)\le t\cdot f(c_1)+(1-t)\cdot f(c_2)$$

$$g(t\cdot c_1+(1-t)\cdot c_2)\le t\cdot g(c_1)+(1-t)\cdot g(c_2)$$

⊛ לפני המעבר ערך הביטוי שיכול להתקבל הוא:

$t\cdot g(c_1)+(1-t)\cdot g(c_2)$ או $t\cdot f(c_1)+(1-t)\cdot f(c_2)$

אך כעת בנוסף ל-2 האפשרויות הנ"ל נוכל לקבל

גם את:

$t\cdot g(c_1)+(1-t)\cdot f(c_2)$ או $t\cdot f(c_1)+(1-t)\cdot g(c_2)$

לכן בהכרח נגדיל את המקסימום.

3.2) בתרגול 7 ראינו שכל פונ' לינארית היא פונ' קמורה

לכן $f(w)=0$ ו- $g(w)=1-y_iw^Tx_i$ הם פונ' קמורות

ומ- 1.a נובע ש- $\max\{f(w),g(w)\}$ גם פונ' קמורה.

3.3) נראה שהבע"ה קמורה משני שלבים:

1. $\frac{1}{m}\sum_{i=1}\max\{0,1-y_iw^Tx_i\}+\lambda\|w\|_2$ פונ' קמורה:

ראינו ש- $\max\{0,1-y_iw^Tx_i\}$ היא פונ' קמורה ולכן

גם $\sum_{i=1}^{m}\max\{0,1-y_iw^Tx_i\}$ פונ' קמורה כסכום של פונ' קמורות

(לפי למה). m הוא מספר הדגימות ולכן חיובי ולכן

גם $\frac{1}{m}$ חיובי ולכן גם $\frac{1}{m}\sum_{i=1}^{m}\max\{0,1-y_iw^Tx_i\}$ קמורה

כהכפלה בסקלר חיובי של פונ' קמורה.

בתרגול 7 ראינו ש- $\|w\|_2$ היא פונ' קמורה ולפי

הגדרת ההייפר פרמטר λ ב- Soft SVM הוא חיובי

ולכן גם $\lambda\|w\|_2$ קמורה.

כסכום של שתי פונ' קמורות מתקבל ש-

$\frac{1}{m}\sum_{i=1}^{m}\max\{0,1-y_i w^T x_i\}+\lambda\|w\|_2$ פונ' קמורה.

<u>2. התחום של הבעיה $\arg\min_{w\in\mathbb{R}^d}$ קמור:</u>

בתרגול ראינו שהתחום אותו ניתן לכתוב כ'

$\{w\in\mathbb{R}^d \mid \forall i\in[m]: y_i w^T x_i \ge 1\}$ הוא תחום.

<u>בסה"כ</u> ראינו שבעיית Soft-SVM היא בעיה קמורה.