1/7
311142905 מור מלכה
205487846 שני כפיר

למידה חישובית - תרגיל 6:

1. a. נוכיח כי $VC(H) = 100$

עבור כל תת קבוצה של X בגודל 100, נסמנה ב- $A=(z_1,\dots,z_{100})$ נבחר היפותזה h כך ש- $h(z_i) = -1$ אם"ם $y_i=-1$ לכל $1\le i\le 100$ ולכל השאר h תחזיר כפלט 1.

נשים לב כי $h\in H$ שכן יש לכל היותר 100 instances שעבורם h מחזירה 1. ולכן מתקבל כי עבור כל השמה של 100 הנקודות $h(z_i)=y_i$, כלומר A מנופצת קבוצה של 100 נקודות, ולכן $VC(H)\ge 100$.

כעת, נניח בשלילה ש- H מנפצת קבוצה בגודל 101, ולכן עבור כל השמה קיימת היפותזה h' כך ש- $h'(m_i)=y_i$ לכל $1\le i\le 101$, אבל ההשמה שעבור כל i $y_i=-1$, קיימות 101 נקודות שעבורן $h'(m_i)=1$ סתירה! $h'\notin H$ ולמעשה לא קיימת היפותזה ב-H שיכולה להסכים עם השמה של מעל 100 נקודות עם $y_i=1$.

$\Downarrow$

לכן מתקבל $VC(H) = 100$

b. נגדיר X = מרחב instances אינסופי

$$H = \{h: X\to\{-1,1\}, |x: h(x)=-1| \le 2019\}$$

בדומה לסעיף a, נבחר תת קבוצה בגודל 2019 ובאותה הצורה נמצא היפותזה שמקיימת את כל ההשמות

לכן $VC(H) = 2019$.

Scanned with CamScanner

① .c

$$H = \left\{ h : \exists w_1, w_2, b \in \mathbb{R} \ s.t \ h(x_1,x_2) = \begin{cases} 1 & w_1x_1 + w_2x_2 + b > 0 \\ -1 & w_1x_1 + w_2x_2 + b \le 0 \end{cases} \right\}$$

.1 נבחר 3 נקודות:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

נביט במטריצה:

$$m = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

מהיות המטריצה m הפיכה (עמודות בת"ל) אז מובטח שיש פתרון למשוואה עבור כל אחת מ-8 הקומבינציות האפשריות של (y_1, y_2, y_3).

.2 ראשית, נוכיח את הטענה:

נניח בה"כ כי $h(z) = h(z') = 1$, כלומר

$$\text{sign}(\langle w, z\rangle + b) = \text{sign}(\langle w, z'\rangle + b) = 1$$

נחשב:

$$h((1-\alpha)z + \alpha z') =$$

$$= \text{sign}(\langle w, (1-\alpha)z + \alpha z'\rangle + b)$$

$$= \text{sign}((1-\alpha)\langle w, z\rangle + \alpha\langle w, z'\rangle + b)$$

$$= \text{sign}((1-\alpha)\langle w, z\rangle + \alpha\langle w, z'\rangle + b + \alpha b - \alpha b)$$

$$= \text{sign}((1-\alpha)\langle w, z\rangle + \alpha\langle w, z'\rangle + (1-\alpha)b + \alpha b)$$

$$= \text{sign}(\underbrace{(1-\alpha)}_{\substack{** \\ >0}}\underbrace{(\langle w, z\rangle + b)}_{\substack{* \\ >0}} + \underbrace{\alpha}_{\substack{** \\ >0}}\underbrace{(\langle w, z'\rangle + b)}_{\substack{* \\ >0}}) = 1 = y$$

* לפי ההנחה מהיות $h(z) = h(z') = 1$ אז הביטוי גדול ממש מ-0

** $0 < \alpha < 1$ ולכן 2 הביטויים גדולים ממש מ-0

באותו האופן מתקיים אם $h(z) = h(z') = -1$.

① **2.c. המשך** כעת, נראה כי $VC(H)<4$. נפצל למקרים:

a. הנקודות נפרשות לקו ישר.

כאשר מדובר ב-4 נקודות אז השמה אפשרית היא

$- - + +$ (על) קו ישר.

יהיו 4 נקודות על הישר $z_1=+1, z_2=+1, z_3=-1, z_4=-1$

נשים לב כי מהיות הנקודות על הישר אזי z_3 הוא

כפל בסקלר של z_1 ו- z_2 ולכן לפי סגירות הצבר

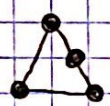
מתקבל -

$$h(z_3)= h((1-\alpha)z_1+\alpha z_2)=1$$

סתירה, $h(z_3)=-1$.

לכן לא ניתן לנפץ קבוצה בגודל 4 על קו ישר.

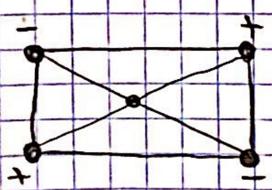
b. הנקודות יוצרות משולש.



* אם יש 3 נקודות על אותו ישר אז לפי הסעיף הקודם לא ניתן לנפץ.

* אם 3 נקודות על המשולש ונקודה אחת בפנים אז לפי סגירות הצבר כל מה שבתוך המשולש (ניתן להגיע אליו ע"י קו ישר מאחד הקודקודים) גם מסווג לאותה השמה כמו של הנקודות שיוצרות את המשולש, לכן ההשמה $-,+,+,+$ לא אפשרית.

c. הנקודות יוצרות מרובע.



השמה בה יש $+,+$ על קודקודי של אלכסון אחד ו- $-,-$ על קודקודי של אלכסון שני אינה אפשרית כיוון שלפי הלמה 5 מה: שגם האלכסון שעובר בין 2 קודקודי ה+ חייב להיות + ולהפך ישנה נקודת חיתוך שמסווגת גם כ- + וגם כ- −.

4/7

1 .ד

$$H = \left\{ h : \exists \bar{w} \in \mathbb{R}^d, b \in \mathbb{R} \text{ s.t } h(\bar{x}) = \begin{cases} +1 & \bar{w}\bar{x} + b > 0 \\ -1 & \bar{w}\bar{x} + b \le 0 \end{cases} \right\}$$

נבחר $d+1$ וקטורים:

$$A = \left\{ \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\}$$

נביט במטריצה:

$$\underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}}_{m} \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_d \\ b \end{bmatrix}}_{w} = \begin{bmatrix} y_1 \\ \vdots \\ y_d \\ y_{d+1} \end{bmatrix}$$

נשים לב כי המטריצה m הפיכה ולכן מובטח שיש
פתרון למשוואה ובפרט פתרון לכל קומבינציה אפשרית
של y, כלומר לכל 2^{d+1} השמות של y קיים וקטור
w כך ש- $mw = y$.
הוקטור w הוא $w = m^{-1}y$ (m הפיכה ולכן
נכפול ב-m^{-1} ונקבל את הוקטור).
כלומר התקבל כי $VC(H) \ge d+1$.

כעת, נראה כי $VC(H) \le d+1$:
יהיו $d+2$ נקודות $z_1, \dots, z_{d+1}, z_{d+2}$,
מכיוון שהמימד הוא גדול מ-d אזי בהכרח ישנו לפחות
וקטור אחד שתלוי לינארית באחרים ולכן קיים $1 \le j \le d+2$
כך ש-

$$z_j = \sum_{i \ne j} a_i z_i$$

ולא כל $a_i = 0$.
מתקבל:

$$z_j = \sum_{i \ne j} a_i z_i \Rightarrow \bar{w}^T z_j = \sum_{i \ne j} a_i w^T z_i$$

נביט בדיכוטומיה הבאה:
x_i אם a_i שאינו 0 מקבל $y_i = \text{sign}(a_i)$
ו- x_j מקבל $y_j = -1$.

1 d. המשך

כלומר אם $y_i = sign(w^T x_i) = sign(a_i)$ אז

כיוון שאם $a_i w^T x_i > 0$ אז $sign(a_i) = 1$

אז a_i גדול מ-0 ומכיון ש $sign(w^T x_i) = 1$ אז

גם $w^T x_i > 0$.

(בצורה הפוכה ניתן להגיע אם $sign(a_i) = -1$).

ולכן בהכרח לפי המשוואה מתקבל-

$$w^T x_j = \sum_{i \neq j} \underbrace{a_i w^T x_i}_{>0} > 0$$

סתירה! עבור x_j מתקיים $y_j = -1$, לכן

לא ניתן לנפץ $d+2$ נקודות, כלומר $VC(H) \leq d+1$

$\Downarrow$

מתקבל- $VC(H) = d+1$

2 נתחיל עם כל $2n$ הליטרלים - $x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n$

עבור כל ה- instances שמקבלים טג חיובי, כלומר $c(\bar{x}) = true$

נוריד את כל הליטרלים אשר מקיימים אם $x_i = 1$ אז

נוריד את $\neg x_i$, אם $x_i = 0$ נוריד את x_i (כלומר נשאיר

רק את הליטרלים שהם 1).

נשים לב שעבור כל ה- instances שמקבלים טג חיובי,

גם לאחר ההורדה נישאר עם טג חיובי כיוון שהשארנו רק

את הליטרלים שהם 1.

עבור ה- instances שמקבלים טג שלילי, מכיון שהאלגוריתם

מוריד את הליטרלים שהם 0 עבור החיוביים אז חייב

להיות לפחות ליטרל אחד של השליליים שנמצא

במשוואה שקיבלנו.

כלומר האלגוריתם קונסיסטנטי עם הקונספט.

מרחב ההיפותזות הוא 3^n (נקיט ליטרל, לא נקיט או

לא משנה), כמו שראינו בכיתה מתקבל:

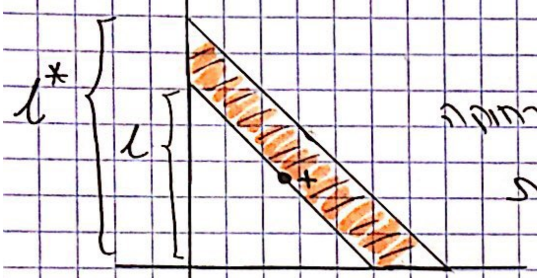
② המשך

$$m \geq \frac{1}{\varepsilon} \ln \frac{|H|}{\delta} = \frac{1}{\varepsilon}\left(\ln|H| + \ln\frac{1}{\delta}\right)$$

$$= \frac{1}{\varepsilon}\left(\ln 3^n + \ln\frac{1}{\delta}\right) = \frac{1}{\varepsilon}\left(n\ln 3 + \ln\frac{1}{\delta}\right)$$

כך שזמן האלגוריתם תלוי ב-$n, \frac{1}{\varepsilon}, \frac{1}{\delta}$ - בדיקה של instance עם התחשבות ב- $O(n)$ ליטרלים שנמצאים במשוואה.

ולכן מתקבל כי "וגם" של n ליטרלים הוא PAC learnable.



③ אלגוריתם: ניקח את נקודת ה- + הרחוקה ביותר מראשית הצירים. נמצא את הישר שעובר בנקודה זו ובנקודות $(x,0)$, $(0,x)$, כלומר במידה ונעביר ישר כזה יווצר לנו משולש ישר זווית ושווה צלעות וההיפותזה שלנו היא המשולש שנוצר.

טענה: C הוא PAC-Learnable ע"י L עם H

הוכחה:

יהי l צלע המשולש ישר הזווית שמתקבל לאחר ביצוע האלגוריתם, כלומר ההיפותזה, בנוסף מתקיים כי $h \in H$.

יהי l^* צלע המשולש שהקונספט מחזיר.

נשים לב כי $l^* \geq l$ תמיד, וכי מרחב הטעות הוא השטח הצבוע בציור, כלומר ההיפותזה מסווגת כ- - והקונספט כ- +.

נגדיר: $l_\varepsilon = \arg\inf_l \Pr[(x_1, x_2) \in A] \leq \varepsilon$

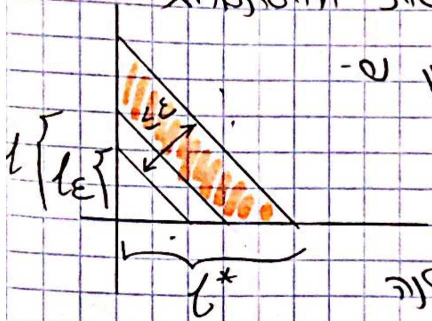
כלומר l_ε היא הצלע הגדולה ביותר שהסתברות לטעות קטנה או שווה ל- ε.

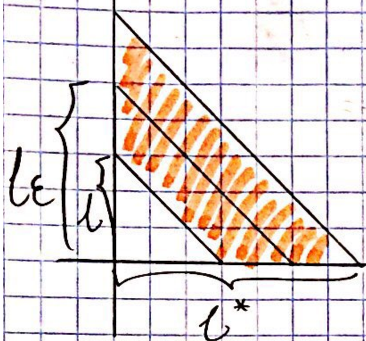
③ המשך

נחלק למקרים:

① $l_\varepsilon = l$: מכיוון ש- l_ε מקיים טעות קטנה מ-ε SC בפרט גם l.

② $l < l_\varepsilon$: במקרה זה אני מקטינים את ההסתברות לנקודה ליפול בשטח הצבוע ולכן מכיוון ש- l_ε מקיים שההסתברות לטעות קטנה מ-ε SC למעשה l (אשר יותר גדול מ-l_ε) יתן הסתברות לטעות שאף קטנה מזו של l_ε והפרט קטנה מ-ε.





③ $l > l_\varepsilon$: מכיוון שההסתברות לטעות גדולה מ-ε נרצה להראות שהאורך חסום ע"י δ. נרצה לחשב את כמות המופעים שנהיה חייבים על מנת להבטיח שהאורך של m המשתנים לא יפלו בשטח הצבוע יהיה קטן מ-δ. ההסתברות שמשתנה לא יפול בשטח הצבוע הוא $1-\varepsilon$ כיוון שההסתברות של השטח היא ε ולכן ההסתברות של m המשתנים לא יפלו בשטח הצבוע היא $(1-\varepsilon)^m$.

מתקיים- $(1-\varepsilon)^m \le \exp(-\varepsilon m)$

עבור מספר דגימות $m \ge \frac{\ln \frac{1}{\delta}}{\varepsilon}$ נקבל:

$$(1-\varepsilon)^m \le \exp(-\varepsilon m) \le \exp\left(-\varepsilon \cdot \frac{\ln\left(\frac{1}{\delta}\right)}{\varepsilon}\right)$$

$$= \exp\left(-\ln \frac{1}{\delta}\right) = \exp(\ln \delta) = \delta$$

כלומר האורך של m המשתנים יפלו מחוץ לשטח הצבוע חסומה ע"י δ ולכן הטעות חסומה ע"י δ.