IML
67577
[illegible]
[illegible]
312203021

① A learning algorithm, D a distribution, loss function in $[0,1]$

(a) $\forall \epsilon,\delta > 0 \ \exists m(\epsilon,\delta)$ s.t. $\forall m \geq m(\epsilon,\delta)$

$$\mathbb{P}_{S\sim D^m}[L_D(A(S))] \geq 1-\delta$$

(b) $\lim_{m\to\infty} \mathbb{E}_{S\sim D^m}[L_D(A(S))] = 0$

נניח: (b) $\Leftarrow$

$$\lim_{m\to\infty} \mathbb{E}_{S\sim D^m}[L_D(A(S))] = 0$$

לפי הגדרת הגבול קיים $\epsilon_1 > 0$ כך ש$\forall m \geq m(\epsilon_1)$:

$$\mathbb{E}_{S\sim D^m}[L_D(A(S))] < \epsilon_1$$

לכן, $\epsilon_1 = \delta\epsilon \geq 0$ אז לפי אי שוויון מרקוב:

$$\mathbb{P}_{S\sim D^m}[L_D(A(S)) > \epsilon] \leq \frac{\mathbb{E}_{S\sim D^m}(L_D(A(S)))}{\epsilon} = \frac{\delta\epsilon}{\epsilon} = \delta$$

$$\Downarrow$$

$$\mathbb{P}_{S\sim D^m}[L_D(A(S)) \leq \epsilon] \geq 1-\delta$$

$(a \Leftarrow b)$ $\quad$ $\Leftarrow$ מ.ש.ל

$\Rightarrow$ נניח (a)

$\forall \epsilon,\delta > 0 \ \exists m(\epsilon,\delta)$ s.t. $\forall m \geq m(\epsilon,\delta)$

$$\mathbb{P}_{S\sim D^m}[L_D(A(S))] \geq 1-\delta$$

$$\lim_{m\to\infty} \mathbb{E}_{S\sim D^m}[L_D(A(S))] \neq 0$$

נניח בשלילה: $\leftarrow$

אז:

$$\lim_{m\to\infty} \mathbb{E}_{S\sim D^m}[L_D(A(S))] > 0$$

כי נתון ש אי־שלילית $L_D(A(s)) \geq 0$.

לפי הגדרת הגבול קיים $\epsilon^* > 0$ כך ש$m \geq m(\epsilon^*)$:

$$\lim_{m\to\infty} \mathbb{E}_{S\sim D^m}[L_D(A(S))] > \epsilon^*$$

$$\Downarrow$$

$$\int_0^1 x\, \mathbb{P}_{S\sim D^m}[L_D(A(S)) = x]\,dx > \epsilon^* \qquad (*)$$

עבור $\epsilon = (\epsilon^*)^2,\ \delta = \frac{\epsilon^*}{2}$ אז:

$$\mathbb{P}_{S\sim D^m}[L_D(A(S)) \leq (\epsilon^*)^2] \geq 1 - \frac{\epsilon^*}{2} \qquad (**)$$

$$\Downarrow$$

ע"פ (*) מתקיים:

$$\epsilon^* < \int_0^1 x\,\mathbb{P}_{S\sim D^m}[L_D(A(S))]\,dx = \int_0^{\sqrt{\epsilon^*}} x\,\mathbb{P}_{S\sim D^m}[L_D(A(S))=x]\,dx +$$

$$+ \int_{\sqrt{\epsilon^*}}^1 x\,\mathbb{P}_{S\sim D^m}[L_D(A(S))=x]\,dx \leq$$

$$\leq \int_0^{\sqrt{\epsilon^*}} \sqrt{\epsilon^*}\,\mathbb{P}_{S\sim D^m}[L_D(A(S))=\sqrt{\epsilon^*}]\,dx + \int_{\sqrt{\epsilon^*}}^1 1\cdot\mathbb{P}_{S\sim D^m}[L_D(A(S))=1]\,dx \leq$$

$$< \int_0^{\sqrt{\epsilon^*}} \sqrt{\epsilon^*}\left(1-\frac{\epsilon^*}{2}\right)dx + \int_{\sqrt{\epsilon^*}}^1 \frac{\epsilon^*}{2}\,dx$$

(**) מ

$$\epsilon^* < (\sqrt{\epsilon^*})^2\left(1-\frac{\epsilon^*}{2}\right) + \frac{\epsilon^*}{2}\left(1-\frac{\epsilon^*}{2}\right)(1-\sqrt{\epsilon^*}) <$$

$$\epsilon^* < \epsilon^*\left(1-\frac{\epsilon^*}{2}\right) + \frac{\epsilon^*}{2}\left(1-\sqrt{\epsilon^*}\right) < \epsilon^*$$

$\Downarrow$

סתירה

מכאן שלילת ההנחה (b) היא הנכונה

לכן מש.ל. $\Rightarrow$ $(a \Rightarrow b)$

הוכחנו את שני הכיוונים ולכן מתקיימת שקילות.

(2) $X = \mathbb{R}^2$, $Y = \{0,1\}$ $\mathcal{H}$ – class of concentric circles: (ב)

$$\mathcal{H} = \{h_r : r \in \mathbb{R}^+\},\ h_r(x) = \mathbb{1}[\|x\|_2 \leq r]$$

נגדיר אלגוריתם ERM A שעבור קבוצת דגימות k היפותזה
h_r, S, כך ש h_r יהיה הכדור המוכל הקטן ביותר שמכיל את הדגימות
עם תיוג חיובי. עבור מדגם גדול מספיק r עד שנגיע מספיק סו
המוכר "רוויזכ"י, עבור $0 \leq \epsilon \leq 1$

נסמן את הטבעת:

$$G = \{x \in \mathbb{R}^2 \mid r' \leq |x| \leq r\}$$

$\Downarrow$

ההסתברות לדגימה בתוך G:

$$D(x \in \mathbb{R}^2 \mid 0 < |x| \leq r) \Rightarrow D(x \in G) \leq \epsilon$$

$\Downarrow$

$$\mathbb{P}_{S\sim D}(L_D(A(S)) \leq \epsilon) = \mathbb{P}(D(x \in G) \leq \epsilon) = 1 \geq 1-\delta$$

מכאן H למידה ע"י PAC.

אם מתקיים $D(x\in\mathbb{R}^2 \mid 0\le\|x\|\le r)\ge\epsilon$ אז נגדיר $r^*\le 0$ ...

אחרת נגדיר r^* מינימלי כך ש-$D(x\in\mathbb{R}^2 \mid r^*\le\|x\|\le r)\ge\epsilon$. אם $r^*<r$

$D(x\in\mathbb{R}^2 \mid r^*\le\|x\|\le r)\ge\epsilon$.

נגדיר r^* ... מצא נקודה בהסתברות ϵ ... ההסתברות כי $1-$

עבור $\forall x\in S$ $(\|x\|<r^*)\vee(\|x\|\ge r)$ מתקיים שאו ...

בכל ... לא נמצא ב-G בהסתברות $1-\epsilon$.

ניקח $m_H(\epsilon,\delta)=\frac{\log(1/\delta)}{\epsilon}$ אז לכל ...

$|S|\ge m_H(\epsilon,\delta)$ מ-ד'

$$(1-\epsilon)^{|S|}\le(1-\epsilon)^{m_H(\epsilon,\delta)}\le\exp\left(-\epsilon\cdot\frac{\log(1/\delta)}{\epsilon}\right)=\exp(\log\delta)=\delta$$

$$\forall x \qquad P(D(x\notin G))\le\delta$$

לכן ... שההסתברות שהדגימות ... ב-S ...

כלומר ההסתברות ... שגיאה קטנה מ-ϵ ... $1-\delta$

ולכן מספיק ... דגימות ... $\frac{\log(1/\delta)}{\epsilon}$.

3) $H=\{h_1,\dots,h_N\}$

מהגדרה: $\text{VCdim}(H)=\max\{|C| \mid H \text{ shatters } C\}\overset{\text{סימון}}{=}n$

כל ... 2^n היפותזות ... ב-H.

$|H|\ge 2^n \Longleftrightarrow n=\text{VCdim}(H)\le\log_2(|H|)$

4) $X=\{0,1\}^n$, $Y=\overline{\{0,1\}}^{\text{מ-ל-ס}}$, אם $I\subset[n]$: $h_I(x)=\left(\sum_{i\in I}x_i\right)\bmod 2$

$H_{parity}=\{h_I \mid I\subseteq[n]\}$.

$|H_{parity}|=2^n$ כיוון שכל ... מוגדרת לכל תת קבוצה של $[n]$.

$\text{VCdim}(H_{parity})\le\log(|H_{parity}|)=\log_2(2^n)=n$ ... 3

עבור קבוצה של אינטרוולים על $\mathbb{R}$ $([a_i,b_i])_{i=1}^{k}$, $k\in\mathbb{Z}$

$$\left(h_A(x)=\begin{cases}0 & x\notin A\\ 1 & x\in A\end{cases}\right) \subseteq H_{k\text{-intervals}}, \quad A=\bigcup_{i=1}^{k}[a_i,b_i]$$

ניקח קבוצה בגודל ונמיין את האיברים את הנקודות הבאות:

$X=\{x_1,\dots,x_{2k}\}$ $x_1<x_2<\dots<x_{2k}$.

עם תיוג כלשהו מבין 2^{2k} התיוגים האפשריים

נראה כי ניתן להגיע לתיוג ע״י אינטרוולים עבור כל תיוג

(x_i,y_i) עבור $y_i=0$ הנק' לא תיכלל באינטרוולים, לכן נ״ב רק

$y_i=1$ עבור y_{i-1} בנוסף ל y_i. אם כולם שווים לא

אז x_{i-1},x_i בהכרח יהיו עם אותו קטע. נבחר את הקטע $[x_{i-1},x_i]$.

אם הם שונים $y_{i-1}\neq y_i$, x_{i-1},x_i עם אותו קטע לכן נבחר את

הקטע ש x_{i-1} סוגר לפני x_i.

במקרה הכי גרוע התיוג יהיה $\{1,0,1,\dots,0,1\}$ כלומר עבור

כל נקודה בגודל k, לכן שכל זוג תיוגים והיא קבוצה בגודל

$2k$ של $H_{k\text{-intervals}}$ מנותצת ע״י

נראה שכל קבוצה בגודל $2k+1$ אינה מנותצת

תהי קבוצה: עם איברים ממוינים כמקודם:

$$X=\{x_1,\dots,x_{2k+1}\}$$

אם ניקח תיוג שמשנה כל איבר $\{1,0,1,\dots\}$ נקבל בסה״כ $k+1$

אחדות. לכן לא ניתן להגיע לתיוג כזה עם k קטעים עבור התיוג

$$\mathrm{VCdim}(H_{k\text{-interval}})=2k$$

כאשר k הולך לאינסוף:

$$\lim_{k\to\infty}\mathrm{VCdim}=\lim_{k\to\infty}2k=\infty$$

6

Let $\mathcal{X}=\{0,1\}^d$, $\mathcal{Y}=\{0,1\}$, $d\geq 2$

נראה כי $VCdim(H_{con})=d$ עבור הבסיס הקנוני נקבל: נתון כי

$E=\{e_1\dots e_d\}$

נבחר פסוקית x_i לכל i כך שהיא מסמנת לכל אפשרות התיוג:

$$h(x_i)=\begin{cases} x_i & y_i=1 \\ \overline{x_i} & o.w. \end{cases}$$

מצאנו קבוצה בגודל d ; $VCdim(H_{con})\geq d$

נניח בשלילה כי יש קבוצה מנותצת בגודל $d+1$ ונסמן

נניח בשלילה שקיימת מנותצת כזו $C=\{x_1,\dots,x_{d+1}\}$ נקבל:

$h_i(x_j)=\delta_{ij}$

משפט היונים עם הפסוקים שיהיו 2

אינדקסים לפחות המתאימים לאותו קורדינטה x, ומכאן שני פסוקים

באותו תיוג, ומכך מופיע הפסוק, כלומר לא מפה את התיוג לא

מכך מתקבלת סתירה, לכן $VCdim < d+1$ אז d

$\Downarrow$

$VCdim(H_{con})=d$

מ.ש.ל

7

$m_H^{uc}:(0,1)^2\rightarrow\mathbb{N}$ אם H מתכנס באופן אחיד עם הסיבוכיות

אז H למידה ע"י Agnostic-PAC עם סיבוכיות מדגם

$m_H(\epsilon,\delta)\leq m^{uc}(\epsilon/2,\delta)$

אם מתקיים ש:

$\forall\epsilon,\delta>0,\ \forall D,\ \forall m\geq m_H^{uc}(\frac{\epsilon}{2},\delta)$

$\mathbb{P}_{S\sim D^m}[L_D(\mathcal{A}(S))\leq \min_{h\in H}\{L_D(A(h))+\frac{\epsilon}{2}\}]$ מתקיים!

מה שנדרש לקיים הוא ש- $\{$ מייצגת ומתכנסת

$\mathbb{P}_{S\sim D^m}(L_D(A(S)))\leq \min_{h\in H}\{L_D(A(h))+\epsilon\}$

נסמן $\epsilon/2=\xi$ ומההתכנסות האחידה נקבל:

$\mathbb{P}_{S\sim D^m}[L_D(A(S))\leq \min_{h\in H}(L_D(A(h))+\epsilon/2)]\geq D^m(\{S\in X\times Y\mid S$ is $\epsilon/2$ represents $\}) \geq 1-\delta$

מכאן מורכבות המדגם מוכחת

8) $Z = X \times \{\pm 1\}$ ופונקציית loss function 0-1, נניח שקיימת פונ' m_H
שעבורה לכל התפלגות מעל Z קיים אלגוריתם A שמחזיר:
כאשר מקבלים את A עם m_H דוגמאות מ-D
מובטחת ההצלחה בהסתברות של לפחות $1-\delta$ (הגדרה):

$$h(S): X \to \{\pm 1\} \qquad L_D(h(S)) \le \min_{h \in H} L_D(h) + \epsilon$$

נכון שבהנחה זו מוגדר לפי agnostic PAC. נניח בשלילה שהוא לא כן,
אז קיים אלגוריתם A ופונ' $m_H(\epsilon,\delta)$ כך שלכל
$1 \ge \epsilon,\delta > 0$ ולכל התפלגות D מעל Z לכל הנ"ל
$f \in H$ שמתקיים במקרה הדטרמיניסטי:

$$D(y|x) = \begin{cases} 1 & y = f(x) \\ 0 & \text{אחרת} \end{cases}$$

נשים לב כי כל התפלגות מהצורה הזו, ניתן לייצר ע"י בחירת D_x
מעל X ובחירת $f \in H$, המשמעות היא ש-H היא
למידה לפי PAC הסטנדרטי. לכן הטענה לא נכונה מ.ש.ל

9) לא מתקיימת הסתירה עבור קומפלקסיות המשלים:
H היא PAC learnable $\forall \epsilon \in (0,1)$ $\quad 0 < \epsilon_1 \le \epsilon_2 < 1$
נניח בשלילה שמתקיים כאשר $m_H(\epsilon_2,\delta) < m_H(\epsilon_1,\delta)$, ו-$m_H$ מתפרק.
אם המספר המינימלי m שעבורו H היא למידה PAC.
$\epsilon_2 \ge \epsilon_1$, וגם $m_H(\epsilon_1,\delta) \le m_H(\epsilon_2,\delta)$ לא ניתן לקבל דיוק יותר
גבוה עם יותר דוגמאות, כי זה סותר להגדרה לכך ש-m_H
מוגדר כמס' מינימלי של דוגמאות שעבורו דיוק זה.
לכל $\epsilon \in (0,1)$ נניח בשלילה שקיימים δ כך ש-$0 < \delta_1 \le \delta_2 < 1$
ומתקיים $m_H(\epsilon,\delta_1) < m_H(\epsilon,\delta_2)$ בסתירה דומה לקודם סותרת
למינימליות ההגדרה.

10) $H_1 \subseteq H_2$ לכן $VCdim(H_1) \le VCdim(H_2)$
מינימליות והגדרה הנתונה $VCdim(H_1) \le VCdim(H_2)$. נגדיר קבוצה
של נקודות $C \subseteq X$, $|C| \le VCdim(H_1)$ מתקיים ש-H_1 מנפצת
את C לכל תיוג שלה. מהגדרה גם H_2 מנפצת, ומכאן מתקיים
$VCdim(H_1) \le VCdim(H_2)$ מ.ש.ל