פתרון תרגיל מספר 4 ־ מערכות לומדות

שם: מיכאל גרינבאום, **ת.ז:** 211747639

25 במאי 2020

1. **צ"ל:** $\lim_{m\to\infty}\mathbb{E}_{S\sim D^m}\left[L_D\left(A\left(S\right)\right)\right]=0$ אמ"ם לכל $\varepsilon,\delta\in\left(0,1\right)$ קיים $m\left(\varepsilon,\delta\right)$ כך ש־ $\forall m\geq m\left(\varepsilon,\delta\right)$ מתקיים $\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\varepsilon\right)\geq1-\delta$

הוכחה:

$\Leftarrow$: נניח שלכל $\varepsilon,\delta\in\left(0,1\right)$ קיים $m\left(\varepsilon,\delta\right)$ כך ש־ $\forall m\geq m\left(\varepsilon,\delta\right)$ מתקיים $\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\varepsilon\right)\geq1-\delta$,
יהי $\varepsilon>0$, לכן מההנחה קיים $m\left(\frac{\varepsilon}{2},\frac{\varepsilon}{2}\right)$ כך ש־ $\forall m\geq m\left(\frac{\varepsilon}{2},\frac{\varepsilon}{2}\right)$ מתקיים $\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\frac{\varepsilon}{2}\right)\geq1-\frac{\varepsilon}{2}$,
נסמן

$$S_1=\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\frac{\varepsilon}{2}\right)\cdot\mathbb{E}\left[L_D\left(A\left(S\right)\right)\mid L_D\left(A\left(S\right)\right)\leq\frac{\varepsilon}{2}\right]$$
$$S_2=\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\frac{\varepsilon}{2}\right)\cdot\mathbb{E}\left[L_D\left(A\left(S\right)\right)\mid L_D\left(A\left(S\right)\right)>\frac{\varepsilon}{2}\right]$$

נשים לב כי

$$\begin{aligned}S_1&=\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\frac{\varepsilon}{2}\right)\cdot\mathbb{E}\left[L_D\left(A\left(S\right)\right)\mid L_D\left(A\left(S\right)\right)\leq\frac{\varepsilon}{2}\right]\\&\leq\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\frac{\varepsilon}{2}\right)\cdot\frac{\varepsilon}{2}\overset{\mathbb{P}_{S\sim D^m}\left(L_D(A(S))\leq\frac{\varepsilon}{2}\right)\leq1}{\leq}1\cdot\frac{\varepsilon}{2}=\frac{\varepsilon}{2}\end{aligned}$$

וגם נשים לב כי

$$\begin{aligned}S_2&=\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\frac{\varepsilon}{2}\right)\cdot\mathbb{E}\left[L_D\left(A\left(S\right)\right)\mid L_D\left(A\left(S\right)\right)>\frac{\varepsilon}{2}\right]\\&\overset{L_D(A(S))\in[0,1]}{=}\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\frac{\varepsilon}{2}\right)\cdot\mathbb{E}\left[L_D\left(A\left(S\right)\right)\mid\left(L_D\left(A\left(S\right)\right)>\frac{\varepsilon}{2}\right)\wedge\left(L_D\left(A\left(S\right)\right)\leq1\right)\right]\\&\leq\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\frac{\varepsilon}{2}\right)\cdot\mathbb{E}\left[L_D\left(A\left(S\right)\right)\mid L_D\left(A\left(S\right)\right)\leq1\right]\\&\leq\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\frac{\varepsilon}{2}\right)\cdot1=\left[1-\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\frac{\varepsilon}{2}\right)\right]\cdot1=\left[1-\left(1-\frac{\varepsilon}{2}\right)\right]\cdot1=\frac{\varepsilon}{2}\end{aligned}$$

עתה נשים לב כי לפי נוסחת ההסתברות השלמה

$$\mathbb{E}_{S\sim D^m}\left[L_D\left(A\left(S\right)\right)\right]=S_1+S_2\leq\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon$$

וגם נשים לב כי $0\leq L_D\left(A\left(S\right)\right)$ ולכן $-\varepsilon\leq0\leq\mathbb{E}_{S\sim D^m}\left[0\right]\leq\mathbb{E}_{S\sim D^m}\left[L_D\left(A\left(S\right)\right)\right]$,
כלומר הראנו שלכל $0<\varepsilon<1$ (זה מספיק), קיים $N=m\left(\frac{\varepsilon}{2},\frac{\varepsilon}{2}\right)$ כך ש־ $\forall m\geq N$ מתקיים $-\varepsilon\leq\mathbb{E}_{S\sim D^m}\left[L_D\left(A\left(S\right)\right)\right]\leq\varepsilon$
ולכן $\boxed{\lim_{m\to\infty}\mathbb{E}_{S\sim D^m}\left[L_D\left(A\left(S\right)\right)\right]=0}$, כנדרש.

$\Rightarrow$: נניח ש־ $\lim_{m\to\infty}\mathbb{E}_{S\sim D^m}\left[L_D\left(A\left(S\right)\right)\right]=0$, יהיו $\varepsilon,\delta\in\left(0,1\right)$, נשים לב שמאי שוויון מרקוב מתקיים

$$0\leq\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\varepsilon\right)\leq\frac{\mathbb{E}\left[L_D\left(A\left(S\right)\right)\right]}{\varepsilon}\overset{m\to\infty}{\to}0$$

לכן ממשפט סנדוויץ' מתקיים כי $\lim_{m\to\infty}\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\varepsilon\right)=0$,
כלומר קיים $N=m\left(\varepsilon,\delta\right)$ כך ש־ $\forall m\geq m\left(\varepsilon,\delta\right)$ מתקיים

$$\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\varepsilon\right)=\left|\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\varepsilon\right)-0\right|<\delta$$

לכן

$$\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\varepsilon\right)=1-\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\varepsilon\right)\geq 1-\delta$$

כלומר הראנו שלכל $\varepsilon,\delta\in(0,1)$ קיים $N=m\left(\varepsilon,\delta\right)$ כך ש־ $\forall m\geq m\left(\varepsilon,\delta\right)$ כך ש־ $\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\varepsilon\right)\geq 1-\delta$, כנדרש

מ.ש.ל.☺

2. **צ"ל:** $\mathcal{H}$ למידה PAC עם $m_{\mathcal{H}}\left(\varepsilon,\delta\right)\leq\frac{\log\left(\frac{1}{\delta}\right)}{\varepsilon}$
הוכחה:
יהיו $\varepsilon,\delta\in(0,1)$,
נגדיר אלגוריתם A שמקבל מדגם S ומוציא את חוק ההחלטה הבא:

(א) נמצא את הנקודה מהמחלקה 1 עם המרחק המקסימלי מהראשית ונסמנה r (כלומר המעגל הכי מגביל שצודק על כל המדגם, אם אין נקודות שהן 1 $r=(0,0)$)

(ב) נחזיר את חוק ההחלטה $h_r\left(x\right)=1_{\|x\|_2\leq\|r\|_2}$

תהי D התפלגות על ה־ $\mathbb{R}^2$ ו־ $h_{nature}\in\mathcal{H}$ פונקציה המקיימת $L_D\left(h_{nature}\right)=0$ (relizability assumption),
נסמן את הרדיוס של h_{nature} ב־ r_{nature}, ונסמן ב־ h_r את הניחוש של האלגוריתם A עם מדגם S, ואת הרדיוס נסמן ב־ r.
נשים לב ש־ $r\leq r_{nature}$ (אחרת אם $r>r_{nature}$ יש דגימה x במרחק r שקיבלה קטלוג 1 כאשר $nature$ היה אמור לקטלגה כ־ 0 כי $r>r_{nature}$)
עתה נראה שקיים $m\left(\varepsilon,\delta\right)$ כך ש־ $\forall m\geq m\left(\varepsilon,\delta\right)$ מתקיים $\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\varepsilon\right)\geq 1-\delta$,
נשים לב שאם A טועה על נקודה x אז $r<\|x\|_2\leq r_{nature}$ (אחרת h_{nature} וגם h_r יקטלגו את הנקודה לאותה מחלקה),
נחלק ל־ 2 מקרים:

(א) אם $D\left\{x\mid r<\|x\|_2\leq r_{nature}\right\}\leq\varepsilon$ אז

$$\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\varepsilon\right)=1\geq 1-\delta$$

בגלל שההסתברות לבחור נקודה בטווח שאנחנו טועים בו תמיד קטנה מ־ ε אז השגיאה קטנה תמיד מ־ ε

(ב) אחרת $D\left\{x\mid r<\|x\|_2\leq r_{nature}\right\}>\varepsilon$, נשים לב שדגמנו m דגימות ל־ S ואף דגימה לא הייתה בין r ל־ r_{nature} (לפי הגדרת r להיות מקסימלי),
מהיות המאורעות הן בת"ל ושווי התפלגות ההסתברות לא לקבל דגימה בין $(r,r_{nature}]$ הוא

$$\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\varepsilon\right)=\left(\mathbb{P}_{x\sim D}\left\{x\notin(r,r_{nature}]\right\}\right)^m=\left(1-\mathbb{P}_{x\sim D}\left\{x\in(r,r_{nature}]\right\}\right)^m$$
$$\leq(1-\varepsilon)^m\leq e^{-\varepsilon m}\overset{m\geq\frac{\ln\left(\frac{1}{\delta}\right)}{\varepsilon}}{\leq}e^{-\varepsilon\cdot\ln\left(\frac{1}{\delta}\right)}=e^{\ln(\delta)}=\delta$$

ולכן נקבל כי $\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\varepsilon\right)=1-\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)>\varepsilon\right)\geq 1-\delta$ עבור $m\geq\frac{\ln\left(\frac{1}{\delta}\right)}{\varepsilon}=m\left(\varepsilon,\delta\right)$

כלומר הראנו שלכל $\varepsilon,\delta\in(0,1)$ קיים אלגוריתם A ו־ $m:(0,1)^2\to\mathbb{N}$ שמוגדר על ידי $m\left(\varepsilon,\delta\right)=\frac{\ln\left(\frac{1}{\delta}\right)}{\varepsilon}$ כך שלכל התפלגות של D על $\mathbb{R}^2$ ולכל $h^*\in\mathcal{H}$ ו־ $\forall m\geq m\left(\varepsilon,\delta\right)$ מתקיים $\mathbb{P}_{S\sim D^m}\left(L_D\left(A\left(S\right)\right)\leq\varepsilon\right)\geq 1-\delta$.
כלומר $\mathcal{H}$ למידה PAC, ויותר מכך הראנו חסם עליון ללמידה שהוא $m\left(\varepsilon,\delta\right)=\frac{\ln\left(\frac{1}{\delta}\right)}{\varepsilon}$ ולכן $m_{\mathcal{H}}\left(\varepsilon,\delta\right)\leq m\left(\varepsilon,\delta\right)=\frac{\ln\left(\frac{1}{\delta}\right)}{\varepsilon}$, כנדרש

מ.ש.ל.☺

3. **צ"ל:** $\mathrm{VCdim}\left(\mathcal{H}_{con}\right)=d$

הוכחה:

תחילה נראה שקיימת C כך ש־ $|C|=d$ וגם $\mathcal{H}_{con}$ shutters C,

יהיו $e_1,\ldots,e_d\in\mathbb{F}_2^d$ הוקטורים הסטנדרטיים, כלומר $\left[e_i\right]_j=\begin{cases}1 & i=j\\ 0 & else\end{cases}$ לכל $i,j\in[d]$

נגדיר $C=\{e_1,\ldots,e_d\}$,

תהי $(y_1,\ldots,y_d)\in\{0,1\}^d$ השמה ל־ $(e_1,\ldots,e_d)$. נראה שקיימת $h\in\mathcal{H}_{con}$ כך ש־ $h(e_i)=y_i$ לכל $i\in[d]$.

נגדיר

$$h(x)=\bigwedge_{i\in[d],y_i=0}\overline{x_i}\in\mathcal{H}_{con}$$

יהי $i\in[d]$, נחלק ל2 מקרים:

(א) אם $y_i=0$ אז

$$\begin{aligned}h(e_i)&=\bigwedge_{j\in[d],y_j=0}\overline{x_j}=\left(\bigwedge_{j\in[d],j\neq i,y_j=0}\overline{x_j}\right)\wedge\left(\overline{x_i}\right)=\left(\bigwedge_{j\in[d],j\neq i,y_j=0}\overline{0}\right)\wedge\left(\overline{1}\right)\\&=\left(\bigwedge_{j\in[d],j\neq i,y_j=0}1\right)\wedge(0)=1\wedge0=0=y_i\end{aligned}$$

(ב) אם $y_i=1$ אז

$$h(e_i)=\bigwedge_{j\in[d],y_j=0}\overline{x_j}=\bigwedge_{j\in[d],j\neq i,y_j=0}\overline{x_j}=\bigwedge_{j\in[d],j\neq i,y_j=0}\overline{0}=\bigwedge_{j\in[d],j\neq i,y_j=0}1=1=y_i$$

כלומר קיבלנו ש־ $h(e_i)=y_i$ לכל $i\in[d]$.

כלומר לכל השמה $(y_1,\ldots,y_d)\in\{0,1\}^d$ ל־ $(e_1,\ldots,e_d)$ $h\in\mathcal{H}_{con}$ כך ש־ $h(e_i)=y_i$ לכל $i\in[d]$, כלומר $\mathcal{H}_{con}$ shutters C וגם $|C|=d$ ולכן $\boxed{\mathrm{VCdim}\left(\mathcal{H}_{con}\right)\geq d}$

עתה נראה כי $\mathrm{VCdim}\left(\mathcal{H}_{con}\right)<d+1$, תהי C כך ש־ $|C|=d+1$,

נניח בשלילה ש־ $\mathcal{H}_{con}$ shutters C , נסמן $C=\{s_1,\ldots,s_{d+1}\}$.

מהיות $\mathcal{H}_{con}$ shutters C קיימות $h_1,\ldots,h_{d+1}\in\mathcal{H}_{con}$ כך ש־ $h_i(s_j)=1_{i\neq j}$

נשים לב שלכל $i\in[d+1]$ ל ־ h_i יש לפחות משתנה אחד שלא מקבל 1 עבור s_i, אחרת יתקיים $0=1_{i\neq i}=h_i(s_i)=1$ בסתירה להגדרת h_i.)

לכן לכל $i\in[d+1]$ ל ־ h_i יש לפחות משתנה אחד שלא מקבל 1 עבור s_i שנסמנו ב־ l_i^k (כאשר ה־ k הוא האינדקס שלא מקבל 1),

ויש לכל היותר d משתנים (משתנה או שלילתו),

לכן משובך היונים, קיימים $i,j\in[d+1]$ כך של־ h_i,h_j יש משתנה משותף שלא מקבל 1 עבור s_i ו־ s_j שנסמנם ב־ l_i^k,l_j^k.

נחלק ל2 מקרים:

(א) אם ב־ h_i,h_j המשתנה x_k מופיע באותה צורה (כלומר בשניהם x_k או בשניהם $\overline{x_k}$)
אז נשים לב שהמשתנה של x_k ב־ h_i,h_j לא מקבל 1 על ידי s_i וגם s_j

$$0=h_i(s_i)=h_i(s_j)=1_{i\neq j}=1$$

ולכן קיבלנו סתירה

(ב) אחרת, ב־ h_i,h_j המשתנה x_k מופיע בצורות שונות (כלומר ב־ h_i מופיע x_k וב־ h_j מופיע $\overline{x_k}$ וא הפוך)
נשים לב ש־ $d+1\geq3$ ולכן קיים $i,j\neq n\in[d+1]$,
נניח בלי הגבלת הכלליות כי h_i מופיע x_k וב־ h_j מופיע $\overline{x_k}$

נשים לב ש־ s_n לא מקבל 1 במשתנה x_k או ש־ s_n לא מקבל 0 במשתנה x_k,
ולכן $0 = h_i(s_n) = 1_{i \neq n} = 1$ או ש־ $0 = h_j(s_n) = 1_{i \neq n} = 1$,
כלומר $0 = 1$ או $0 = 1$, סתירה

כלומר הראנו שבכל מקרה אפשרי, נקבל סתירה ל־ $\mathcal{H}_{con}$ shutters C , ולכן אין קבוצה C כך ש־ $|C| = d+1$ וגם $\mathcal{H}_{con}$ shutters C, לכן נקבל כי $\boxed{\text{VCdim}(\mathcal{H}_{con}) < d+1}$,
לכן מ2 האי שוויונות שקיבלנו, נסיק כי $\boxed{\text{VCdim}(\mathcal{H}_{con}) = d}$

מ.ש.ל.☺

4. **צ"ל:** agnostic PAC עם $m_{UC}^{\mathcal{H}}\left(\frac{\varepsilon}{2}, \delta\right)$
הוכחה:
תהי S קבוצה $\frac{\varepsilon}{2}$ מייצגת,
נסמן ב־ h_S את ההיפותזה הנלמדה על ידי ERM,
תהי $h \in \mathcal{H}$, נשים לב כי

$$\begin{aligned} L_D(h_S) &\overset{\text{S is } \frac{\varepsilon}{2} \text{ representative}}{\leq} L_S(h_S) + \frac{\varepsilon}{2} \\ &\overset{h_S \text{ minimizes the error on S}}{\leq} L_S(h) + \frac{\varepsilon}{2} \\ &\overset{\text{S is } \frac{\varepsilon}{2} \text{ representative}}{\leq} L_D(h) + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = L_D(h) + \varepsilon \end{aligned}$$

כלומר קיבלנו כי $L_D(h_S) \leq L_D(h) + \varepsilon$ לכל $h \in \mathcal{H}$ ולכן $L_D(h_S) \leq \min_{h \in \mathcal{H}} L_D(h) + \varepsilon$,
כלומר הראנו שאם S קבוצה $\frac{\varepsilon}{2}$ מייצגת אז $L_D(h_S) \leq \min_{h \in \mathcal{H}} L_D(h) + \varepsilon$ כאשר h_S היפותזה הנלמדה על ידי ERM.
עתה נוכיח את מה שהתבקש,
יהיו $\varepsilon, \delta \in (0,1)$, נשים לב שעבור $m(\varepsilon, \delta) = m_{UC}^{\mathcal{H}}\left(\frac{\varepsilon}{2}, \delta\right)$, לכל $m \geq m(\varepsilon, \delta) = m_{UC}\left(\frac{\varepsilon}{2}, \delta\right)$ מתקיים

$$D^m \left\{ S \in (\mathcal{X} \times \mathcal{Y})^m \mid \text{S is } \frac{\varepsilon}{2} \text{ representative} \right\} \geq 1 - \delta$$

וגם ראינו שכאשר S הוא קבוצה $\frac{\varepsilon}{2}$ מייצגת אז $L_D(h_S) \leq \min_{h \in \mathcal{H}} L_D(h) + \varepsilon$ כאשר h_S היפותזה הנלמדה על ידי ERM,
לכן

$$D^m \left\{ S \in (\mathcal{X} \times \mathcal{Y})^m \mid \text{S is } \frac{\varepsilon}{2} \text{ representative} \right\} \leq \mathbb{P}_{S \sim D^m} \left\{ L_D(h_S) \leq \min_{h \in \mathcal{H}} L_D(h) + \varepsilon \right\}$$

כלומר קיבלנו כי

$$1 - \delta \leq D^m \left\{ S \in (\mathcal{X} \times \mathcal{Y})^m \mid \text{S is } \frac{\varepsilon}{2} \text{ representative} \right\} \leq \mathbb{P}_{S \sim D^m} \left\{ L_D(h_S) \leq \min_{h \in \mathcal{H}} L_D(h) + \varepsilon \right\}$$

לכן הראנו שקיימת פונקציה $m : (0,1)^2 \to \mathbb{N}$ (המוגדרת על ידי $m(\varepsilon, \delta) = m_{UC}^{\mathcal{H}}\left(\frac{\varepsilon}{2}, \delta\right)$ לכל $\varepsilon, \delta \in (0,1)$) כך שלכל ε, δ והתפלגות D מתקיים שלכל $m \geq m(\varepsilon, \delta)$ מתקיים

$$\mathbb{P}_{S \sim D^m} \mid L_D(h_S) \leq \min_{h \in \mathcal{H}} L_D(h) + \varepsilon \geq 1 - \delta$$

ולכן $\mathcal{H}$ היא למידה agnostic PAC עם $m_{\mathcal{H}}(\varepsilon, \delta) \leq m(\varepsilon, \delta) = m_{UC}^{\mathcal{H}}\left(\frac{\varepsilon}{2}, \delta\right)$ לכל $\varepsilon, \delta \in (0,1)$, כנדרש

מ.ש.ל.☺

5. **צ"ל:** לא בהכרח agnostic PAC
הוכחה:
$\mathcal{H}$ לא בהכרח agnostic PAC,
נבחר $\mathcal{X}=\mathbb{N}$, $\mathcal{Y}=\{0,1\}$, ו- $\mathcal{H}=\mathcal{Y}^{\mathcal{X}}$.
תחילה נוכיח כי $\text{VCdim}(\mathcal{H})=\infty$,
יהי $n\in\mathbb{N}$, תהי $C=\{1,2,\ldots,n\}$, נשים לב כי

$$\begin{aligned}\mathcal{H}_C&=\{h_c:C\to\mathcal{Y}\mid\exists h\in\mathcal{H}\text{ s.t. }\forall x\in C\Rightarrow h_c(x)=h(x)\}\\&=\left\{h_c:C\to\mathcal{Y}\mid\exists h\in\mathcal{Y}^{\mathcal{X}}\text{ s.t. }\forall x\in C\Rightarrow h_c(x)=h(x)\right\}\\&\overset{h_c:C\to\mathcal{Y}\Rightarrow h_c\in\mathcal{Y}^{\mathcal{X}}}{=}\{h_c:C\to\mathcal{Y}\}=\mathcal{Y}^C\end{aligned}$$

ולכן $|\mathcal{H}_C|=\left|\mathcal{Y}^C\right|=2^{|C|}=2^n$,
ולכן $\text{VCdim}(\mathcal{H})\geq n$ לכל $n\in\mathbb{N}$, ולכן $\text{VCdim}(\mathcal{H})=\infty$,
ולכן מהמשפט המרכזי של IML ש- $\mathcal{H}$ למידה agnostic PAC אם"ם $\text{VCdim}(\mathcal{H})\neq\infty$, נקבל ש- $\mathcal{H}$ לא למידה agnostic PAC.
נשאר להראות ש- $\mathcal{H}$ מקיימת את ההנחות בשאלה.
נשים לב שבהנחות האלגוריתם A מודע להתפלגות $\mathcal{D}$ ולכן יוכל להיות בפרט bias optimal classifier שעבורו מתקיים $L_{\mathcal{D}}(A(S,\mathcal{D}))\leq L_{\mathcal{D}}(h)$ לכל $h\in\mathcal{H}$.
ולכן $L_{\mathcal{D}}(A(S,\mathcal{D}))\leq\min_{h\in\mathcal{H}}L_{\mathcal{D}}(h)\leq\min_{h\in\mathcal{H}}L_{\mathcal{D}}(h)+\varepsilon$
ולכן

$$\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}(A(S,\mathcal{D}))\leq\min_{h\in\mathcal{H}}L_{\mathcal{D}}(h)+\varepsilon\right)=1\geq 1-\delta$$

ולכן הנחות השאלה מתקיימות וגם $\mathcal{H}$ לא למידה agnostic PAC, כנדרש

מ.ש.ל.☺

6. **צ"ל:** $m_{\mathcal{H}}$ מונוטונית לא עולה ב- ε,δ
הוכחה:
יהיו $\varepsilon_1,\varepsilon_2,\delta\in(0,1)$ כך ש- $\varepsilon_1<\varepsilon_2$.
נשים לב ש- $\forall m\geq m_{\mathcal{H}}(\varepsilon_1,\delta)$ מתקיים

$$\begin{aligned}1-\delta&\leq\mathbb{P}_{S\sim D^m}(L_D(h_S)\leq\varepsilon_1)\\&\overset{0\leq\mathbb{P}(\text{something})}{\leq}\mathbb{P}_{S\sim D^m}(L_D(h_S)\leq\varepsilon_1)+\mathbb{P}_{S\sim D^m}(\varepsilon_1<L_D(h_S)\leq\varepsilon_2)\\&\overset{\text{disjoint events}}{=}\mathbb{P}_{S\sim D^m}((L_D(h_S)\leq\varepsilon_1)\vee(\varepsilon_1<L_D(h_S)\leq\varepsilon_2))=\mathbb{P}_{S\sim D^m}(L_D(h_S)\leq\varepsilon_2)\\&\Rightarrow\boxed{1-\delta\leq\mathbb{P}_{S\sim D^m}(L_D(h_S)\leq\varepsilon_2)}\end{aligned}$$

ולכן מהגדרה מתקיים כי $m_{\mathcal{H}}(\varepsilon_2,\delta)$ הוא המינימלי המקיים $1-\delta\leq\mathbb{P}_{S\sim D^m}(L_D(h_S)\leq\varepsilon_2)$ וראינו ש- $m_{\mathcal{H}}(\varepsilon_1,\delta)$ מקיים את התנאי ולכן $\boxed{m_{\mathcal{H}}(\varepsilon_2,\delta)\leq m_{\mathcal{H}}(\varepsilon_1,\delta)}$
יהיו $\varepsilon,\delta_1,\delta_2\in(0,1)$ כך ש- $\delta_1<\delta_2$.
נשים לב ש- $\forall m\geq m_{\mathcal{H}}(\varepsilon,\delta_1)$ מתקיים

$$1-\delta_2\overset{\delta_1<\delta_2}{\leq}1-\delta_1\leq\mathbb{P}_{S\sim D^m}(L_D(h_S)\leq\varepsilon_1)$$

ולכן מהגדרה מתקיים כי $m_{\mathcal{H}}(\varepsilon_1,\delta_2)$ הוא המינימלי המקיים $1-\delta_2\leq\mathbb{P}_{S\sim D^m}(L_D(h_S)\leq\varepsilon)$ וראינו ש- $m_{\mathcal{H}}(\varepsilon,\delta_1)$ מקיים את התנאי ולכן $\boxed{m_{\mathcal{H}}(\varepsilon,\delta_2)\leq m_{\mathcal{H}}(\varepsilon,\delta_1)}$,
כלומר הראנו ש- $m_{\mathcal{H}}$ היא מונוטונית לא עולה בכל אחד מקלטיה, כנדרש

מ.ש.ל.☺

7. **צ"ל:** $\text{VCdim}\left(\mathcal{H}_1\right) \le \text{VCdim}\left(\mathcal{H}_2\right)$
הוכחה:
תהי $C \subseteq \mathcal{X}$, נשים לב כי

$$\begin{aligned} \mathcal{H}_1^C &= \left\{h_c : C \to \mathcal{Y} \mid \exists h \in \mathcal{H}_1 \text{ s.t. } \forall x \in C \Rightarrow h\left(x\right) = h_c\left(x\right)\right\} \\ &\overset{\exists h \in \mathcal{H}_1 \subseteq \mathcal{H}_2}{\subseteq} \left\{h_c : C \to \mathcal{Y} \mid \exists h \in \mathcal{H}_2 \text{ s.t. } \forall x \in C \Rightarrow h\left(x\right) = h_c\left(x\right)\right\} \\ &= \mathcal{H}_2^C \end{aligned}$$

עתה תהי C כך ש־ C shutters $\mathcal{H}_1$, אזי $\left|\mathcal{H}_1^C\right| = 2^{|C|}$, נשים לב כי $2^{|C|} = \left|\mathcal{H}_1^C\right| \le \left|\mathcal{H}_2^C\right| = 2^{|C|}$ ולכן $\left|\mathcal{H}_2^C\right| = 2^{|C|}$,
כלומר C shutters $\mathcal{H}_2$,
לכן נקבל כי

$$\begin{aligned} \text{VCdim}\left(\mathcal{H}_1\right) &= \max\left\{C \subseteq \mathcal{X} \mid C \text{ shutters } \mathcal{H}_1\right\} \\ &\overset{C \text{ shutters } \mathcal{H}_1 \Rightarrow C \text{ shutters } \mathcal{H}_2}{\le} \max\left\{C \subseteq \mathcal{X} \mid C \text{ shutters } \mathcal{H}_2\right\} = \text{VCdim}\left(\mathcal{H}_2\right) \end{aligned}$$

כלומר קיבלנו כי $\boxed{\text{VCdim}\left(\mathcal{H}_1\right) \le \text{VCdim}\left(\mathcal{H}_2\right)}$

מ.ש.ל.☺

8. **פתרון:**

(א) **צ"ל:** מה $\tau_{\mathcal{H}}\left(m\right)$ אומר?
הוכחה:
מה ש־ $\tau_{\mathcal{H}}\left(m\right)$ אומר זה כמות הסיווגים השונים המקסימלית שקבוצה בגודל m יכולה לקבל, בהנחה וניתן לסווג כל קלט לכל פלט, נקבל שיש 2^m אפשרויות כמו שיוסבר בסעיפים הבאים.

מ.ש.ל.א.☺

(ב) **צ"ל:** נוסחא ל־ $\tau_{\mathcal{H}}\left(m\right)$ כאשר $\text{VCdim}\left(\mathcal{H}\right) = \infty$
הוכחה:
נטען ש־ $\tau_{\mathcal{H}}\left(m\right) = 2^m$.
תחילה נשים לב שקיימת קבוצה סופית C כך ש־ $|C| \ge m$ וגם $\mathcal{H}$ shutters C (אחרת $\text{VCdim}\left(\mathcal{H}\right) < m$),
נסמן $C = \left\{c_1, \ldots, c_n\right\}$ ונגדיר $B = \left\{c_1, \ldots, c_m\right\}$. נטען ש־ $\mathcal{H}$ shutters B.
תהי $\left(y_1, \ldots, y_m\right) \in \left\{0,1\right\}^m$ השמה ל־ $\left(c_1, \ldots, c_m\right)$, מהיות $\mathcal{H}$ shutters C קיימת $h \in \mathcal{H}$ כך ש־ $h\left(c_i\right) = \begin{cases} y_i & i \le m \\ 0 & else \end{cases}$ לכל $i \in [n]$, נשים לב שהצמצום של h ל־ B הוא $h\left(c_i\right) = y_i$ לכל $i \in [m]$,
כלומר קיימת $h \in \mathcal{H}$ כך ש־ $h\left(c_i\right) = y_i$ לכל $i \in [m]$, כלומר $\mathcal{H}$ shutters B ולכן בפרט

$$\begin{aligned} &\tau_{\mathcal{H}}\left(m\right) = \max\left\{\left|\mathcal{H}_G\right| \mid G \subseteq \mathcal{X} \mid |G| = m\right\} \ge \left|\mathcal{H}_B\right| = 2^m \\ &\tau_{\mathcal{H}}\left(m\right) = \max\left\{\left|\mathcal{H}_G\right| \mid G \subseteq \mathcal{X} \mid |G| = m\right\} \le \max\left\{\left|h : G \to \mathcal{Y} \mid G \subseteq \mathcal{X} \mid |G| = m\right|\right\} \le 2^m \end{aligned}$$

ולכן נקבל כי $\boxed{\tau_{\mathcal{H}}\left(m\right) = 2^m}$

מ.ש.ל.ב.☺

(ג) **צ"ל:** נוסחא ל־ $\tau_{\mathcal{H}}\left(m\right)$ כאשר $m \le d$ ו־ $\text{VCdim}\left(\mathcal{H}\right) = d$
הוכחה:
נטען ש־ $\tau_{\mathcal{H}}\left(m\right) = 2^m$.
תחילה נשים לב שקיימת קבוצה C כך ש־ $|C| = d$ וגם $\mathcal{H}$ shutters C מהגדרת $\text{VCdim}\left(\mathcal{H}\right) = d$
נסמן $C = \left\{c_1, \ldots, c_n\right\}$ ונגדיר $B = \left\{c_1, \ldots, c_m\right\}$. נטען ש־ $\mathcal{H}$ shutters B.
תהי $\left(y_1, \ldots, y_m\right) \in \left\{0,1\right\}^m$ השמה ל־ $\left(c_1, \ldots, c_m\right)$, מהיות $\mathcal{H}$ shutters C קיימת $h \in \mathcal{H}$ כך ש־ $h\left(c_i\right) =$

$\begin{cases} y_i & i \le m \\ 0 & else \end{cases}$ לכל $i \in [n]$, נשים לב שהצמצום של h ל- B הוא $h(c_i) = y_i$ לכל $i \in [m]$,
כלומר קיימת $h \in \mathcal{H}$ כך ש- $h(c_i) = y_i$ לכל $i \in [m]$, כלומר $\mathcal{H}$ shutters B ולכן בפרט

$$\tau_{\mathcal{H}}(m) = \max\{|\mathcal{H}_G| \mid G \subseteq \mathcal{X} \mid |G| = m\} \ge |\mathcal{H}_B| = 2^m$$
$$\tau_{\mathcal{H}}(m) = \max\{|\mathcal{H}_G| \mid G \subseteq \mathcal{X} \mid |G| = m\} \le \max\{|h : G \to \mathcal{Y} \mid G \subseteq \mathcal{X} \mid |G| = m|\} \le 2^m$$

ולכן נקבל כי $\boxed{\tau_{\mathcal{H}}(m) = 2^m}$

מ.ש.ל.ג.☺

(ד) **צ"ל:** $\tau_{\mathcal{H}}(m) \le \left(\frac{em}{d}\right)^d$
הוכחה:

i. תחילה נוכיח באינדוקציה על גודל הקבוצה C כי $|\mathcal{H}_C| \le |\{B \subseteq C \mid B \text{ shutters } \mathcal{H}\}|$
<u>בסיס:</u> $n = 1$, נשים לב שאם $\mathcal{H}$ shutters C אז

$$|\mathcal{H}_C| = 2 \le 2 = |\{\emptyset, C\}| = |\{B \subseteq C \mid B \text{ shutters } \mathcal{H}\}|$$

אחרת $\mathcal{H}$ doesn't shutter C אז

$$|\mathcal{H}_C| = 1 \le 1 = |\{\emptyset\}| = |\{B \subseteq C \mid B \text{ shutters } \mathcal{H}\}|$$

<u>צעד:</u> נניח שהטענה נכונה ל- $n-1$ ונוכיח ל- n
נסמן $C = \{c_1, \dots, c_n\}$, נגדיר $C' = \{c_2, \dots, c_n\}$.
נגדיר קבוצת עזר $H = \{(h_1, \dots, h_m) \mid \exists h \in \mathcal{H}_C \text{ s.t. } \forall i \in [m] \Rightarrow h(c_i) = h_i\}$, נשים לב ש- $|H| = |\mathcal{H}_C|$ (ייצוג חח"ע ועל בין הקבוצות)
עתה נגדיר קבוצות עזר

$$H_0 = \{(h_2, \dots, h_m) \mid (0, h_2, \dots, h_m) \in H \vee (1, h_2, \dots, h_m) \in H\}$$
$$H_1 = \{(h_2, \dots, h_m) \mid (0, h_2, \dots, h_m) \in H \wedge (1, h_2, \dots, h_m) \in H\}$$

נשים לב ש- H_0 סופר את כל האפשרויות שיכולות להתקבל על ידי צמצום $\mathcal{H}_C$ ל- $\mathcal{H}_{C'}$.
נשים לב שאם $(h_2, \dots, h_m)$ הופיע פעמיים (פעם בתור $(0, h_2, \dots, h_m)$ ופעם בתור $(1, h_2, \dots, h_m)$) ספרנו אותו רק פעם אחת, אז קבוצה H_2, באה לספור את המופעים שאמורים להספר פעמיים ונספרו רק פעם אחת.
לכן $\boxed{|\mathcal{H}_C| = |H| = |H_0| + |H_1|}$,
נשים לב כי $H_0, H_1 \subseteq \mathcal{H}_{C'}$ וגם עבור $|C'| = m - 1$ מתקיימת הנחת האינדוקציה ש-

$$|\mathcal{H}_{C'}| \le |\{B \subseteq C' \mid B \text{ shutters } \mathcal{H}\}|$$

ולכן נקבל כי $|H_0| \le |\mathcal{H}_{C'}| \le |\{B \subseteq C' \mid B \text{ shutters } \mathcal{H}\}|$
נגדיר

$$\mathcal{H}_{helper} = \{h \in \mathcal{H} \mid \exists h_{helper} \in \mathcal{H} \text{ s.t. } \forall 1 \ne i \in [m] \Rightarrow h(c_i) = h_{helper}(c_i) \wedge h(c_1) \ne h_{helper}(c_1)\}$$

כלומר כל הפונקציות ב- $\mathcal{H}$ שיכולות לסווג את c_1 ב- 2 אפשרויות.
נשים לב כי $|H_1| = |\mathcal{H}_{helper}|$ (ייצוג חח"ע ועל בין הקבוצות מההגדרה),
עתה תהי $B \subseteq C'$ כך ש- $\mathcal{H}_{helper}$ shutters B אז $\left|\mathcal{H}^B_{helper}\right| = 2^{|B|}$, ולפי הגדרה יש 2 אפשרויות לc_1 לכל איבר בקבוצה ולכן $\left|\mathcal{H}_{B \cup \{c_1\}}\right| = 2^{|B|} \cdot 2 = 2^{|B \cup \{c_1\}}$ ולכן $\mathcal{H}$ shutters $B \cup \{c_1\}$.
עתה נשתמש שוב בהנחת האינדוקציה ונקבל כי

$$|H_1| \le \left|\mathcal{H}^B_{helper}\right| \le |\{B \subseteq C' \mid B \text{ shutters } \mathcal{H}_{helper}\}| \le |\{B \subseteq C \mid B \text{ shutters } \mathcal{H} \wedge c_1 \in B\}|$$

ולכן נקבל כי

$$\begin{aligned}
|\mathcal{H}_C| &\leq |H_0| + |H_1| \\
&\leq |\{B \subseteq C' \mid B \text{ shutters } \mathcal{H}\}| + |\{B \subseteq C \mid B \text{ shutters } \mathcal{H} \wedge c_1 \in B\}| \\
&= |\{B \subseteq C \setminus \{c_1\} \mid B \text{ shutters } \mathcal{H}\}| + |\{B \subseteq C \mid B \text{ shutters } \mathcal{H} \wedge c_1 \in B\}| \\
&= |\{B \subseteq C \mid B \text{ shutters } \mathcal{H} \wedge c_1 \notin B\}| + |\{B \subseteq C \mid B \text{ shutters } \mathcal{H} \wedge c_1 \in B\}| \\
&= |\{B \subseteq C \mid B \text{ shutters } \mathcal{H}\}|
\end{aligned}$$

כלומר קיבלנו כי $\boxed{|\mathcal{H}_C| \leq |\{B \subseteq C \mid B \text{ shutters } \mathcal{H}\}|}$, כמו שרצינו

ii. האי שוויון שקיבלנו אומר שמספר האפשרויות השונות לסווג של איברים ב- $\mathcal{H}_C$ הוא לכל היותר מספר הקבוצות שאפשר לסווג בכל אופן אפשרי.

iii. עתה נשים לב שאם $\text{VCdim}(\mathcal{H}) \leq d$ וגם B shutters $\mathcal{H}$ אז מתקיים כי $|B| \leq d$ ולכן נקבל כי

$$\begin{aligned}
|\mathcal{H}_C| &\leq |\{B \subseteq C \mid B \text{ shutters } \mathcal{H}\}| \leq |\{B \subseteq C \mid |B| \leq d\}| \\
&= \left|\bigcup_{k=0}^{d} \{B \subseteq C \mid |B| = k\}\right| \leq \sum_{k=0}^{d} |\{B \subseteq C \mid |B| = k\}| = \sum_{k=0}^{d} \binom{m}{k}
\end{aligned}$$

iv. עתה נשים לב כי

$$\begin{aligned}
\sum_{k=0}^{d} \binom{m}{k} &= \sum_{k=0}^{d} \frac{m \cdot (m-1) \cdot \ldots \cdot (m-d+1)}{k!} \leq \sum_{k=0}^{d} \frac{m^k}{k!} = \sum_{k=0}^{d} \frac{\left(\frac{m \cdot d}{d}\right)^k}{k!} \\
&= \sum_{k=0}^{d} \frac{\left(\frac{m}{d}\right)^k \cdot d^k}{k!} \overset{m \geq d}{\leq} \left(\frac{m}{d}\right)^d \cdot \sum_{k=0}^{d} \frac{d^k}{k!} \leq \left(\frac{m}{d}\right)^d \cdot e^d = \left(\frac{m \cdot e}{d}\right)^d
\end{aligned}$$

כלומר קיבלנו כי

$$|\mathcal{H}_C| \leq \sum_{k=0}^{d} \binom{m}{k} \leq \left(\frac{m \cdot e}{d}\right)^d$$

ולכן נסיק כי עבור $m \geq d$ מתקיים

$$\tau_{\mathcal{H}}(m) = \max\{|\mathcal{H}_C| \mid C \subseteq \mathcal{X} \wedge C = |m|\} \leq \max\left\{\left(\frac{m \cdot e}{d}\right)^d \mid C \subseteq \mathcal{X} \wedge C = |m|\right\} = \left(\frac{m \cdot e}{d}\right)^d$$

כלומר קיבלנו ש- $\boxed{\tau_{\mathcal{H}}(m) \leq \left(\frac{m \cdot e}{d}\right)^d}$

מ.ש.ל.ד.☺

(ה) **צ"ל:** האם האי שוויון חלש $\tau_{\mathcal{H}}(m) \leq \left(\frac{em}{d}\right)^d$

הוכחה:

נשים לב כי $\tau_{\mathcal{H}}(d) = 2^d$ (הוכח בסעיף ג) ונשים לב כי $\left(\frac{ed}{d}\right)^d = e^d$,
לכן נקבל כי $2^d < e^d$ ולכן האי שוויון הוא לא חלש אלא חזק (ניתן לראות במעבר של ההוכחה בטור טיילור, ניתן היה לעשות אי שוויון חזק), אלא אם $d = 0$

מ.ש.ל.ה.☺

(ו) **צ"ל:** הגדרה נוספת ל- VCdim

הוכחה:

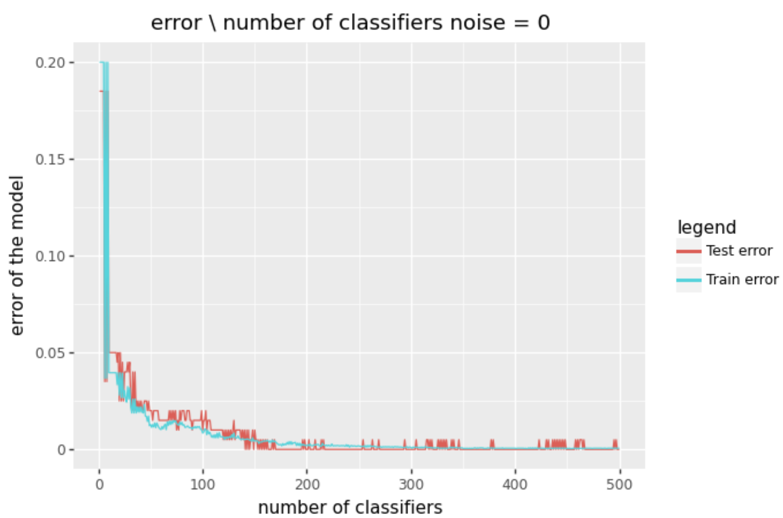
נשים לב ש- $\tau_{\mathcal{H}}(m) = 2^m$ רק כאשר $m \leq d$ ולאחר מכן $\tau_{\mathcal{H}}(m) \leq \left(\frac{em}{d}\right)^d$, ובפרט פולינומי ומתקיים $\tau_{\mathcal{H}}(m) \neq 2^m$

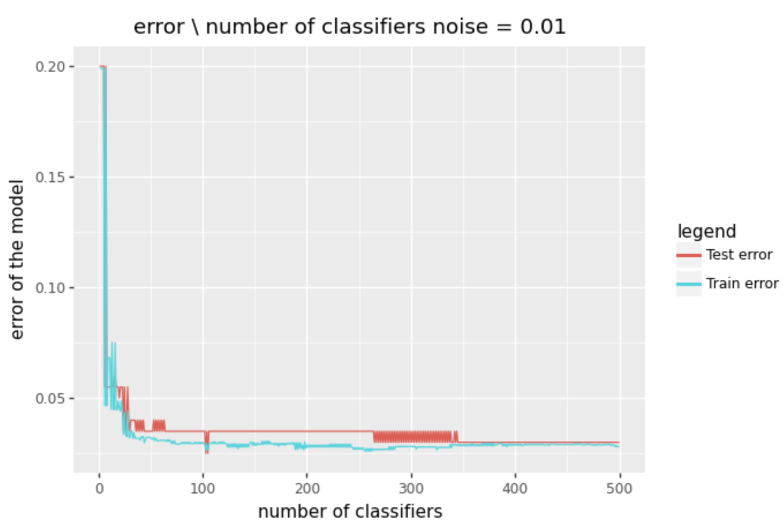
לכל $m \geq d+1$ (אחרת הקבוצה בגודל m הייתה מנתצת את $\mathcal{H}$ בסתירה לכך ש- d הוא הגודל המקסימלי המנתץ את $\mathcal{H}$)
כלומר ההתנהגות היא אקפוננציאלית עד ל- d ומ - d היא פולינומית בחזקה d לכל היותר.
הגדרה נוספת ל- $\mathrm{VCdim}\left(\mathcal{H}\right)$ יכולה להיות $\mathrm{VCdim}\left(\mathcal{H}\right) = \max_{C\subseteq\mathcal{X}}\left\{|C| \mid |\mathcal{H}_C| = 2^{|C|}\right\}$ והגדרה נוספת שניתן להגדיר זאת הנקודה שבה הגדילה של $\tau_{\mathcal{H}}$ משתנה מאקפוננציאלית לפולינומית.

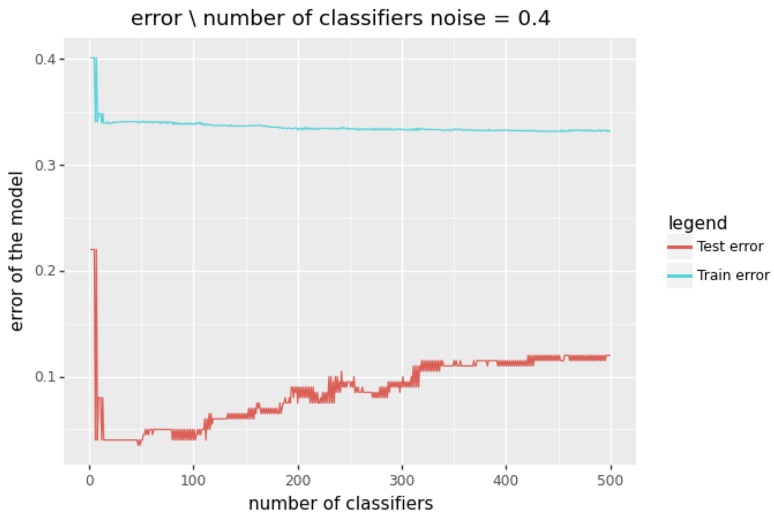
מ.ש.ל.ו.☺

9. **בוצע**

10. **צ"ל:** גרפים ומסקנות
הוכחה:



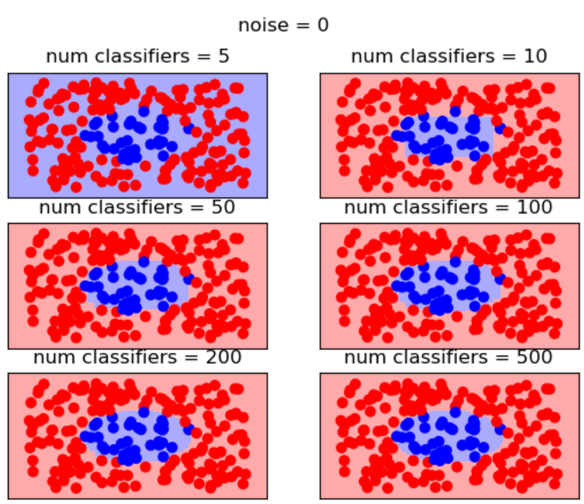


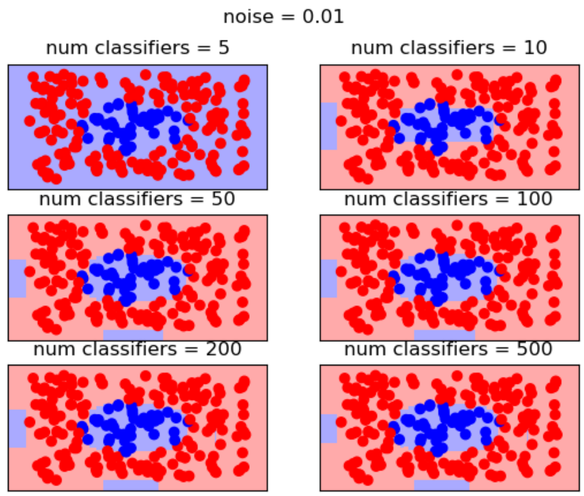


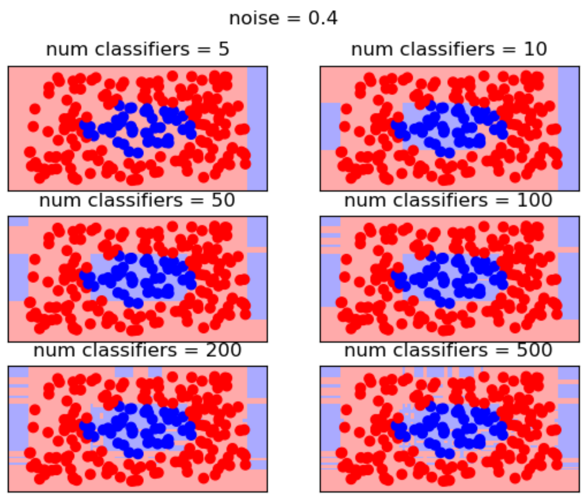
אפשר לשים לב שבכל הגרפים, השגיאה של ה־ $train$ דועכת ממש מהר ומתייצבת, שמתאים למה שהוכחנו בהרצאה שהדעיכה היא אקפוננציאלית.
נשים לב שעם רעש, הנחת הריזביליות לא מתקיימת יותר, וזה מתכנס לדעתי לשגיאה המינימלית $\min_{h\in\mathcal{H}} L_D(h) + \varepsilon$ (כמו שראינו ב־ agnostic PAC).
אפשר לשים לב, שהשגיאה של ה־ $test$ נמוכה גם היא אפילו שיש רעש גדול, שמראה את הכוח של $Adaboost$ להתמודדות עם רעש.
ניתן לראות שעם רעש גדול מספיק ומספיק מודלים, מתחיל להיות $overfit$ וה־ $Adaboost$ מתחיל ללמוד את הרעש ולא את המודל שמוביל לעלייה בשגיאה של ה־ $test$.
ניתן לראות את הסיבה להצלחה של ה־ $Adaboost$ כמו הנאמר בהרצאה לגבי ה־ Bias Variance tradeoff שה־ $Adaboost$ מוריד את שגיאת ה־ $Bias$ וכמעט לא משנה את שגיאת ה־ $Variance$ (טיפה מעלה) שמוביל לשגיאה נמוכה יותר בגרפים בו ה־ $Variance$ נמוך מאוד, ניתן לראות בגרף עם $0.0, 0.01$ רעש.

מ.ש.ל.☺

11. **צ"ל:** גרפים ומסקנות
הוכחה:



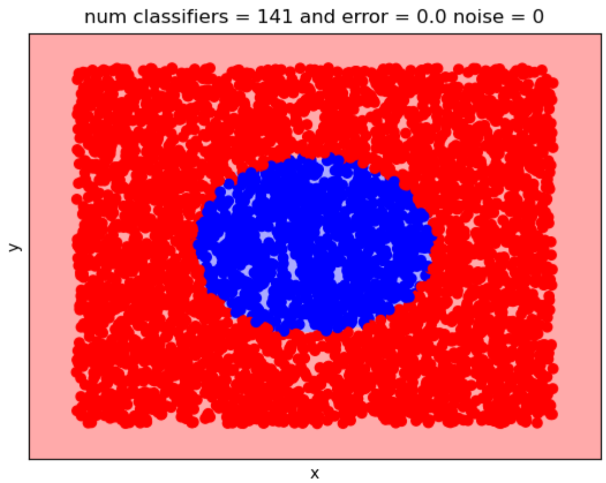


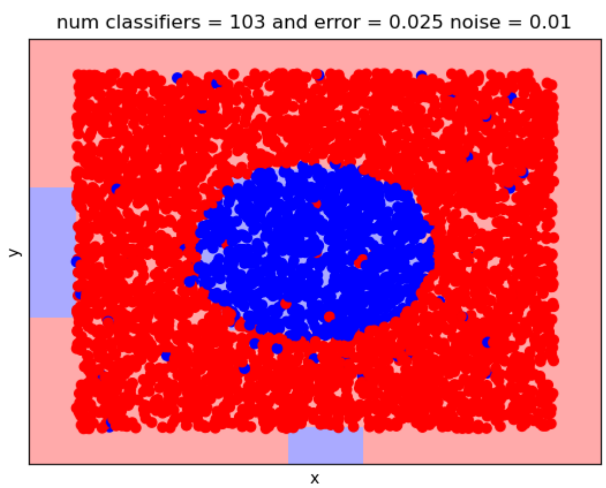


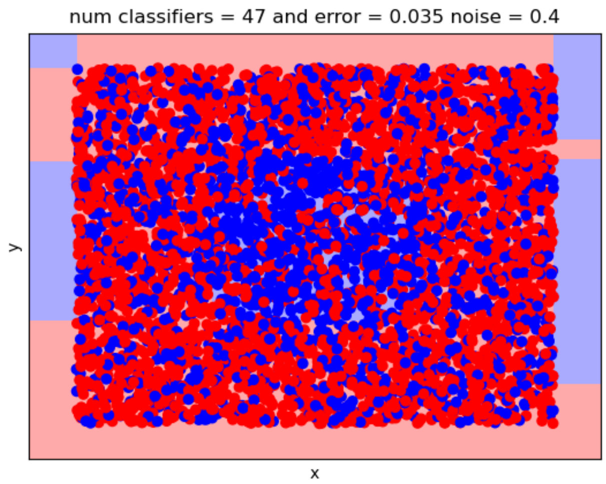
אפשר לראות שאפילו עם $T = 10$ ה־ $Adaboost$ כבר מצליח ללמוד את הצורה של העיגול שאמורה להווצר ומנסה לשפר אותה ככל ש־ T עולה.
ניתן לראות שעם רעש גדול, ו־ T מספיק גדול, הוא מתחיל ללמוד טיפה את הרעש ולא רק את מה שרצינו ללמוד.

מ.ש.ל.☺

12. **צ"ל:** גרפים ומסקנות
הוכחה:


num classifiers = 141 and error = 0.0 noise = 0
y
x

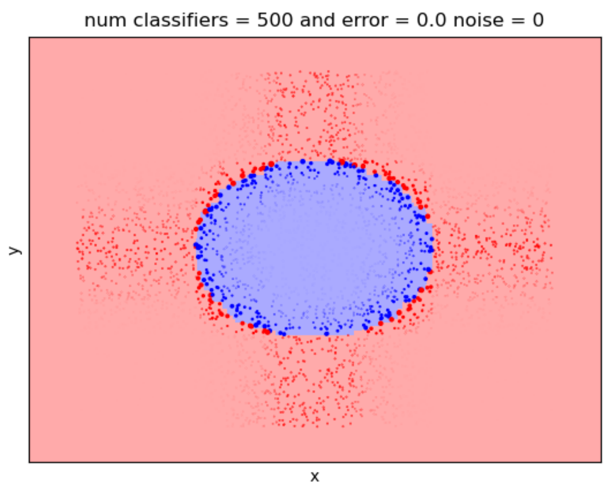

num classifiers = 103 and error = 0.025 noise = 0.01
y
x

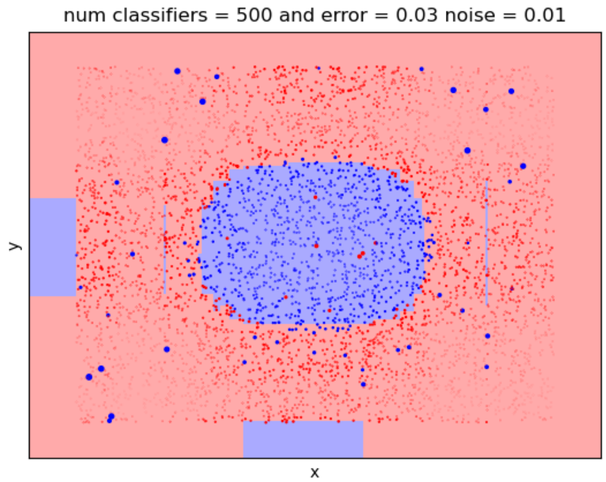


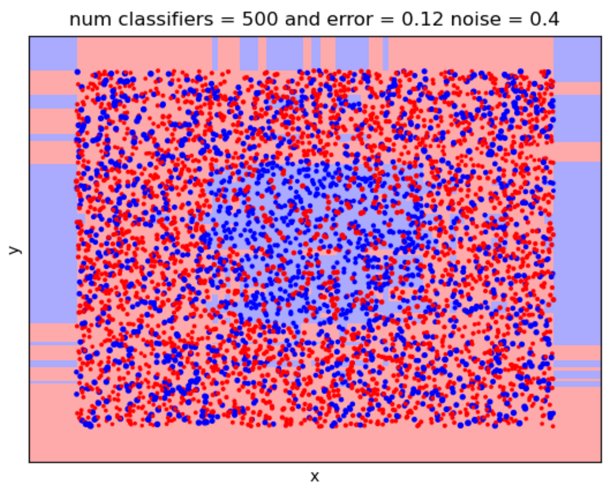
כמו שראינו בגרפים הקודמים, ה־ $Bias$ יורד ממש מהר (אקפוננציאלית לפי הנאמר בהרצאה) בזמן שה־ $variance$ מתחיל להשפיע ככל שמסתכלים על יותר קלסיפיירים ולכן כש־ $variance$ נמוך מאוד, נראה שהשגיאה המינימלית מתקבלת עם T גדול.
נשים לב שכשהוספנו רעש, בגלל שה־ $Bias$ יורד מהר לנקודת התכנסות שלו וה־ $Variance$ מתחיל להשפיע רק ב־ T גדולים, ולכן ככל ש־ T ייגדל מתישהו השגיאה תתחיל לעלות כי הוא מתחיל ללמוד את הרעש.
ולכן כשיש רעש, נשתמש נקבל שהשגיאה יותר נמוכה עם $T = 40,100$ מאשר עם $T = 141,500$ ששם הוא התחיל ללמוד את הרעש שהוספנו.

מ.ש.ל.☺

13. **צ"ל:** גרפים ומסקנות
הוכחה:







אפשר לראות שללא רעש, הנקודות עם המשקל הכי גדול הן דווקא על שפת המעגל, בדיוק כמו שרצינו כי זה מכיל את המידע הכי חשוב ושאותו הכי מומלץ ללמוד.
אפשר לשים לב שעם טיפה רעש, הנקודות עם המשקל הרב ביותר הן אלה שהוא טועה בהם וכצפוי, הוא טועה דווקא ברעש ומצליח לזהות שהצורה היא עיגול.
אפשר לשים לב שעם הרבה רעש, רוב הנקודות בעלות אותו משקל כי בכל ריצה הוא טועה בהרבה כי אין באמת משהו שניתן ללמוד עם כמות כזאת גדולה של רעש, אפשר לראות שאפילו בני אדם לא יכולים לראות את הצורה שאמורה להפריד בין הנקודות האדומות לכחולות בתמונה השלישית עם רעש 0.4

מ.ש.ל.☺