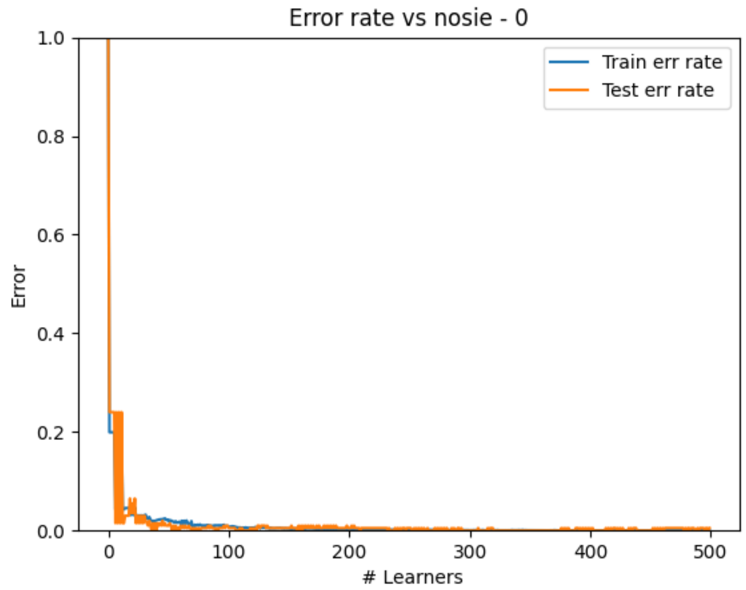
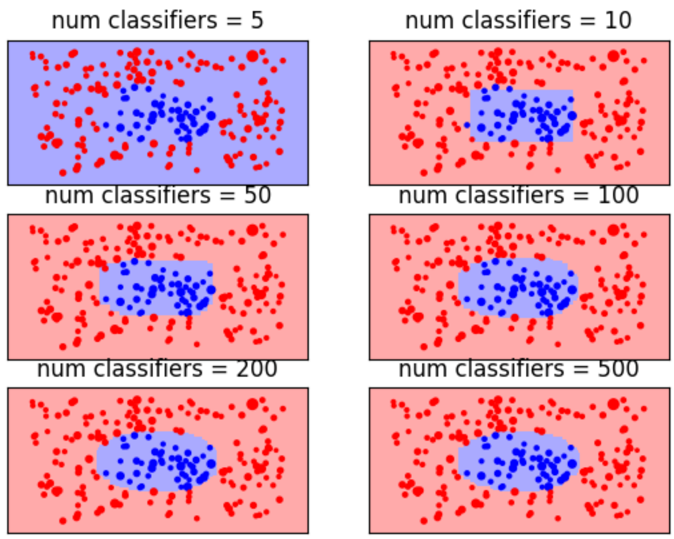
דורון ברודר
doronbruder 204917629

# IML – EX 4

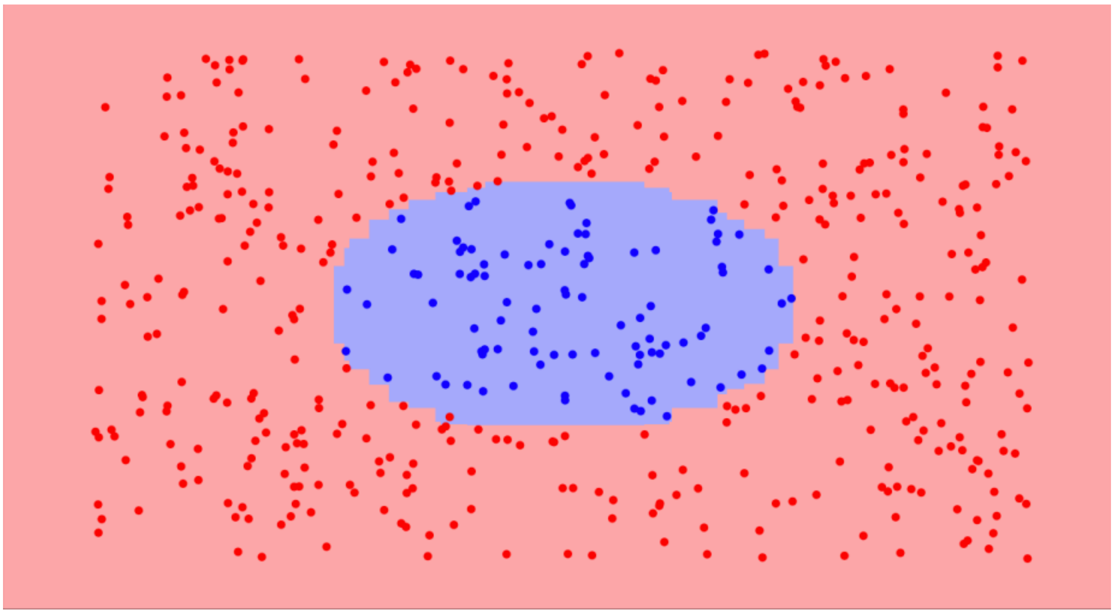
.13




.14

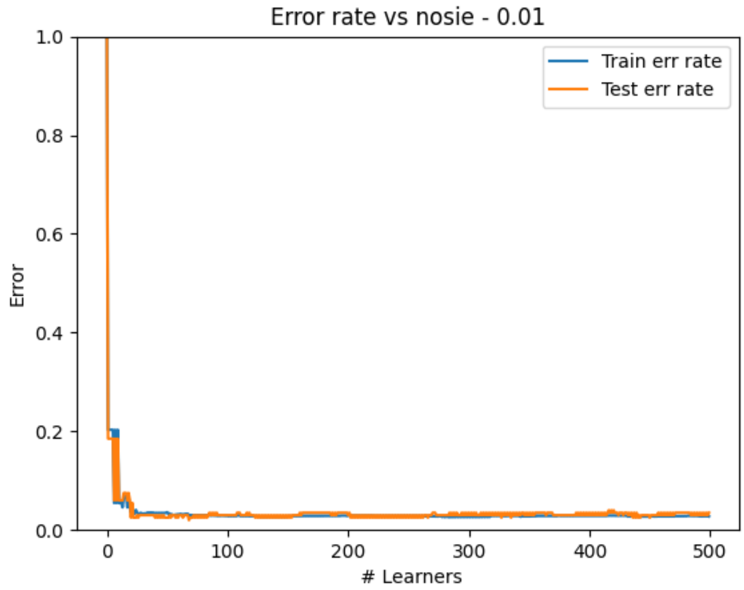


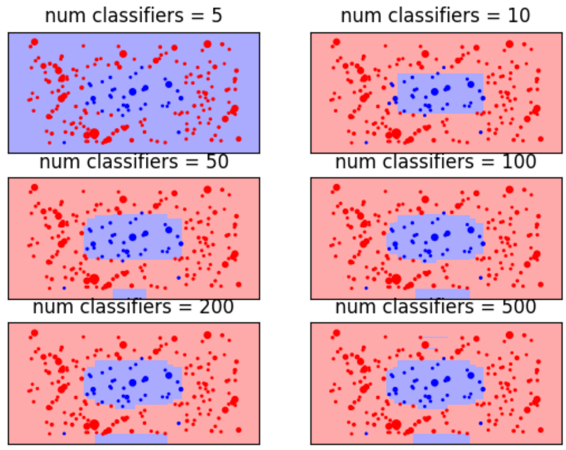
.15
המספר האופטימלי היה כ –200

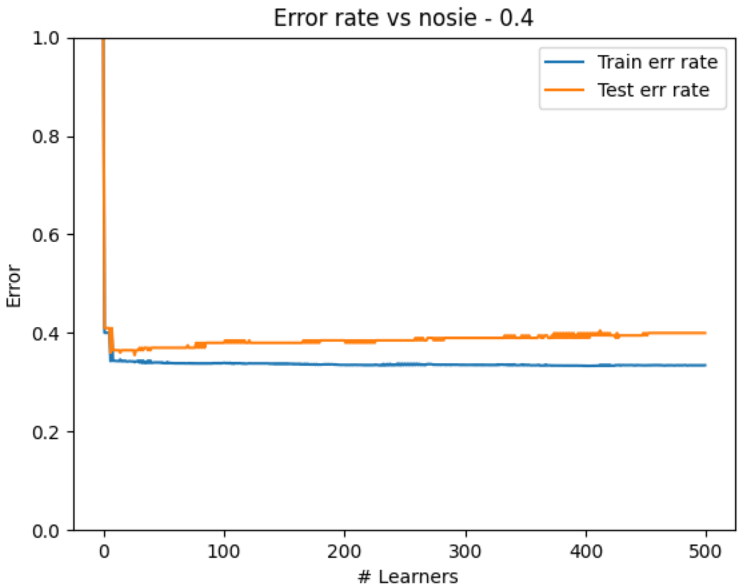


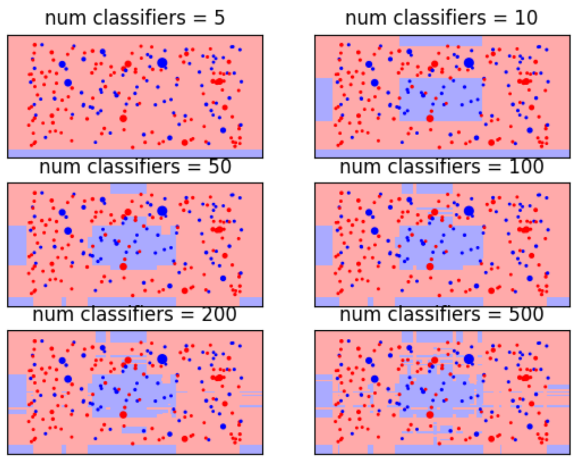
.16
הנקודות שיותר קרובות לגבול  המפריד (BOUNDERY)  קשות יותר לסיווג ולכן המסווג שלנו טועה בהם יחסית הרבה פעמים בתהליך הלמידה ולכן יש להם משקל גדול יותר . הנק' שבקצוות לעומת זאת קלות לזיהוי ולכן בעלות משקל נמוך

17.




- ניתן לראות שבמקרה שהרעש הוא 0.01 המספר האופטימלי הוא 100 או 200
- וכאשר הרעש הוא 0.4 המספר האופטימלי הוא 10

הסבר ל13 +15 :

ההסבר לשניהם נעוץ בעובדה שכאשר יש רעש ככל שT גדל יותר נוצר OVERFIT על הדטה וה-estimation error גדל ושגיאת ההכללה גדלה גם היא , ולכן נקבל ביצועים מיטביים דווקא עבור T שהוא הכי גדול .

# תרגיל 4 : IML

## שאלה 1:

3.ב: $\forall \epsilon,\delta>0\ \exists m(\epsilon,\delta)\ s.t\ \forall m>m(\epsilon,\delta)$

$$(*)\quad \left\{ P_{S\sim D^m}[L_D(A(S))\le\epsilon]\ge 1-\delta \right.$$

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

הוכחה: נשתמש במרקוב ונקבל:

$$P_{S\sim D^m}(L_D(A(S))\le\epsilon)=1-P_{S\sim D^m}(L_D(A(S))\ge\epsilon)$$

$$\ge 1-\frac{\mathbb{E}_{S\sim D^m}[L_D(A(S))]}{\epsilon}$$

כעת נשים לב שאם (*) מתקיים לכל $\epsilon,\delta$ ולכל $m>m(\epsilon,\delta)$

מתקיים: $\frac{\mathbb{E}[L_D(A(S))]}{\epsilon}\le\delta$

$$\Rightarrow \mathbb{E}[L_D(A(S))]<\delta\epsilon$$

מכיוון שזה נכון לכל $0<\epsilon,\delta$ נקבל $\lim_{m\to\infty}\mathbb{E}[L_D(A(S))]=0$

אם (**) מתקיים אז קיים $m'$ כך שלכל $m>m'$

$$\mathbb{E}[L_D(A(S))]<\delta\epsilon$$

ולכן (*) מתקיים.

## תרגיל ב:

נגדיר את האלגוריתם $A$ באופן הבא:

בהינתן קבוצת נק' $S$, $A$ תחזיר את המעגל שמכיל את כל הנק' שהן '1' עם הרדיוס המינימלי שמקיים זאת.

נסמן את הכדור שנוצר בין המעגל ה"אמיתי" לבין המעגל שהחזרתי כ-$T'$. כעת נגדיר מעגל נוסף אשר הכדור שנוצר בינו לבין המעגל "האמיתי" הוא בעל משקל $\varepsilon$ ביחס להתפלגות $D$. נסמנה ב-$T$,

נשים לב כי האלגוריתם $A$ יטעה אך ורק בנקודות של $T'$ אם $T'$ מוכל ב-$T$ אז ההסתברות לשגיאת הכללה ביחס ל-$D$ (לטעות) היא $\varepsilon$ ונסיים. לכן מעתה נניח כי $T \subsetneq T'$ ואז:

$$D(x \notin T) = 1-\varepsilon$$

$$\Rightarrow \underbrace{D(x_1 .. x_m \notin T)}_{\substack{= \\ \text{סיכוי טעות} \\ \text{(שגיאת הכללה)}}} = (1-\varepsilon)^m \le e^{-\varepsilon m} \underset{\text{(נדרש)}}{<} \delta$$

$$\Rightarrow \underbrace{1 - D(x_1 .. x_m \notin T')}_{\text{זה שגיאת הכללה } \varepsilon \text{ (לכל היותר)}} > 1-\delta$$

$-\varepsilon m < \ln\delta$

$$\Longleftrightarrow e^{-\varepsilon m} < \delta$$

$$\Rightarrow -\varepsilon m > \log\delta$$

$$\varepsilon m > \ln\frac{1}{\delta}$$

$$\boxed{m > \frac{\ln(\frac{1}{\delta})}{\varepsilon}}$$

**תר' 3:**

יהי $H=\{H_1,\dots,H_N\}$ קבוצה סופית

צ"ל: $\text{vc-dim}(H)\le\lfloor\log_2(|H|)\rfloor$

**הוכחה:**

לפי משפט מההרצאה מכיוון ש-H סופית היא למידה -PAC

ומכיוון לפי המשפט היסודי $\text{vc-dim}(H)$ סופי

(נסמן ב-d) כלומר קיימת קבוצה $C\subset X$

כך ש: $|H|\overset{(def)}{\ge}|H_C|=2^d$

ומכיוון שיש לפחות $2^d$ היפותזות שונות ב-H (שכן אם היו פחות לא היה אפשר לנתץ את C.)

**תר' 4:** יהי $X=\{0,1\}^n$ ו- $Y=\{0,1\}$

לכל $I\subseteq[n]$ נגדיר: $\log(2^n)$ $n$

$$h_I(x)=\left(\sum_{i\in I}x_i\right)\bmod 2$$

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

ראשית נשים לב ש-H סופית (בגלל התנאי המספר הפונקציות.)

(I יכול להיות כל תת קבוצה של $[n]$ ולכן יש $2^n$ היפותזות.)

מטענה 3 נקבל כי: $\text{vc-dim}(H)\le\log 2^n=n$

כעת נראה כי $\text{vc dim}(H)\ge n$ ונסיים.

נראה זאת ע"י כך שנמצא קבוצה בגודל n ש-H מנתצת.

נקח קבוצה: $C=\{\underline{e}_1,\dots,\underline{e}_n\}$ איבר הבסיס הסטנדרטי ב- $\mathbb{R}^n$

~~מספיק להראות~~ נראה שלכל וקטור תיוגים $y=[y_1,\dots,y_n]$ כלשהו

קיימת היפותזה $h\in H$ כך ש: $h(\underline{e}_i)=y_i \quad \forall i\in[n]$

ה-h המתאימה תהיה $h_I$ כך ש: $I=\{i\in[n]: y_i=1\}$

נשים לב שעבור $y_i=1$ נקבל: $h_I(e_i)=0+\dots+\underset{i\text{-th}}{1}+\dots 0\,\%\,2=1$

ועבור $y_i=0$ נקבל: $h_I(e_i)=(\sum 0)\%2=0 \bmod 2=0$

הערה:

תר' 5:

1 א. רווחים נסגרים על $\mathbb{R}$ המחלקה הנתונה: $H=\{x\in\mathbb{R} \mid a\le x\le b,\ a,b\in\mathbb{R}\}$

פונקציה שמקבלת $x$ מ-$H$ מכילה את כל הנקודות שבתוך $[a,b]$

ובהינתן $x$ מחזירה $1$ אם $x\in[a,b]$ (נסמן פונקציה כזו ע"י $h_{a,b}$) ואחרת מחזירה $0$.

ממד VC של H: $VC\text{-}\dim H = 2$

עבור כל 2 נק' $x_1<x_2\in\mathbb{R}$ נוכל לקבל $x_1\to0,\ x_2\to0$

אם $x_2<a,b$ נוכל לקבל $x_1\to1,\ x_2\to1$ אם

$a\le x_1$ $x_2\le b$ נוכל לקבל $x_1\to1,\ x_2\to0$ אם

$a\le x_1,\ b<x_2$ נוכל לקבל $x_1\to0,\ x_2\to1$ אם

$x_1<a\le x_2$ $x_2\le b$.

עבור כל קבוצה $C$ כך ש-$|C|=3$ לא נוכל לנתץ

למשל: $x_1\to1,\ x_2\to0,\ x_3\to1$ (לא שייך לרווח)

כלומר $VC\dim H=2$.

המקרה הכללי: לאיחוד $n$ קטעים בקבוצת $2n$ נק' $x_1,\dots,x_{2n}$

(נניח ש- $x_1<\dots<x_{2n}$) נוכל לבחור $n$ קטעים

$[a_1,b_1],\dots,[a_n,b_n]$ לכל אחת מ-$2n$ הנק' כך שנבחר

לכל $[a_i,b_i]$ שמכיל את הזוג $[x_{2i-1},x_{2i}]$

אין אפשרות $2n+1$ נק' כי לא נוכל לקבל את

ההשמה: $(\underbrace{1,-1,\dots,1,-1,1}_{2n+1})$ כלומר המימד Vcdim הוא $2n$.

(כי נצטרך יותר - פיצול שעושה לקבל פרדה לכל נקודה.)

Scanned with CamScanner

שאלה 5!

ניתן להראות שכאשר ה-k שלנו יהי האינסופי-

H תהיה קבוצת ערכין לתשובה של ה-k בעיה:

$$H_{unlimited\ k} = \bigcup_{k \in \mathbb{N}} H_A$$

$$A = \bigcup_{i=1}^{k} [a_i, b_i]$$

עבור כל A נוכל לנתץ קבוצה בגודל 2k

וכיון ש-A יכול להיות כל מספר טבעי

אזי נוכל לנתץ $\aleph_0$ נק', (כמות)

כלומר $VC\dim(H_{unlimited\ k}) = \infty$

## שאלה 6:

נראה ש- $d = VCdim(H_{con})$ אזי:

(1) קיימת קבוצה בגודל $d$ כך ש-$H$ מנתצת את הקבוצה

(2) שכל קבוצה בגודל $\geq d+1$ לא תוכל להיות מנותצת ע"י $H_{con}$

(1): הקבוצה ש-$H$ מנתצת היא: $C = \{e_j\}_{j=1}^{d}$ תהי $J \subseteq [d]$.

ת"ע קבוצה של האינדקסים ונראה שקיימת השערה $h \in H$

שמתייגת '1' עבור $\{e_j\}_{j \in J}$ (ו-'0' אחרת) ונסווג

(סימון שצריך כו ניתן להגדיר לכל פונ' מ- $Y \to C$) $\theta(C \to Y)$

נחלק למקרים:

- אם $J = [d]$ אזי נגדיר את $\phi$ כלומר נוסחה ריקה (שנוסחה ריקה) שמחזירה '1' עבור כל הדגימות

- אם $J = \phi$ אזי נגדיר: $x_1 \wedge \bar{x}_1$ כלומר מחזירה '0' עבור כל הדגימות

- אם $\phi \subsetneq J \subsetneq [d]$ אזי נגדיר: $\wedge_{j \in J} x_j$

וכן מתקיים כי $h(e_j) = 1 \iff j \in J$

(2): נניח בשלילה שקיימת קבוצה בגודל $d+1$:

$(x_1 \ldots x_{d+1}) \subset X$

שמנותצת ע"י $H$

ולכן מפרק הדוגמה שיש,

ישנם 2 ליטרלים ששייכים לאותו משתנה (נסמן $l_1, l_2$).

אם $l_1 = l_2$ אזי נסמן את $l_1$ ולא את $l_2$ ...

ס- $l_2 \neq l_1$ נקבל שהנוסחה לא תקיים $\theta$ (וכו' ...)

שאלה 7:

לפי ~~[illegible]~~ מההרצאה (הגדרה של VC)

מכיוון ש-H מקיים VC אז קיים $m^{VC}(\varepsilon,\delta)$ כך

שלכל $m \geq m^{VC}$ ולכל D מעל X מתקיים:

$$P[S\in(X,Y)^m \text{ over } D \mid S \text{ is } \varepsilon \text{ representative}] \geq 1-\delta$$

כלומר: $L_D(h_S) \leq \min_{h\in H} L_D(h^*) + \frac{\varepsilon}{2}$ : $\forall \varepsilon,\delta>0$

(כאשר נסמן ב-$h^*_D$ את המפיקה ~~[illegible]~~ ממזערת את $L_D$) ו-$h_S$ ממזער את ה-ERM.

נוכיח:

$$(*)\ \begin{cases} L_D(h_S) \overset{\text{הנחה}}{\leq} L_S(h_S)+\frac{\varepsilon}{2} \overset{h^*\text{ ממזער ERM}}{\leq} L_S(h^*)+\frac{\varepsilon}{2} \\ \overset{\text{הנחה}}{\leq} (L_D(h^*)+\frac{\varepsilon}{2})+\frac{\varepsilon}{2} = L_D(h^*)+\varepsilon \end{cases}$$

כלומר קיבלנו ש: ה-$m^{VC}$ מתאים

~~[illegible]~~

ל הנו $\frac{\varepsilon}{2}$ מינימאלי עבורם זה את ה-$m_H$ מהסעיף

ה- sample complexity של H - מינון ש(*) מתקיים

לכל $m \geq m^{VC}$ ! -(*) היא הגדרה ל-Agnostic-PAC learnable

כלומר $m_H \leq m^{VC}$

כאשר $m_H$ הוא ה- sample complexity של H.

Scanned with CamScanner

## שאלה 8:

H הוא לא אגנוסטית - למיד מכיוון שההגדרה המקורית PAC

האלגוריתם A יכול לא להיות תלוי בהתפלגות D.

למעשה G מחלקת היפותזות מקיימת את ההגדרה הבאה:

כאשר נבחר את $A_n$ להיות האלגוריתם שמחזיר $h^*$

תמיד לא משנה מה הקלט כאשר $h^* = \underset{h\in H}{\arg\min}\, L_D(h)$

## שאלה 7:

תהי D התפלגות על X ותהי $f\in H$ הפונקציה

שמגדירה. נסמן ב-A את האלגוריתם שלומד את H בסיבוכיות מדגם

של $m_H$ ונניח $\delta\in(0,1)$ ותהיין $\epsilon_1,\epsilon_2\in(0,1)$

כך ש- $\epsilon_1\le\epsilon_2$.

נסמן: $m_1=m_H(\epsilon_1,\delta)$ $m_2=m_H(\epsilon_2,\delta)$

נקח את $m_1$ מתקיים:

$$L_{D,f}(h)\le\epsilon_1\le\epsilon_2$$

ומהמינימליות של $m_2$ נקבל ש- $m_2\le m_1$

(כאשר δ קבוע ולכן עבור $\delta_1,\delta_2$.)

שאלה 10:

$H_1 \subseteq H_2$ ולכן לכל $X \supseteq C=\{c_1 \dots c_m\}$ שס

$C$ מנותצת ע"י $H_1$ היא מנותצת גם ע"י $H_2$

ולכן עפ"י ההגדרה: $VC(H_2) \geq VC(H_1)$