IML, תרגיל 4
מגיש: אבי כוגן, ת.ז: 205417710

שאלה 1

מתקיים

$$\begin{aligned}
&P_{S\sim D^m}(L_D(A(S) \leq \epsilon) \geq 1-\delta \Leftrightarrow \\
&P_{S\sim D^m}(L_D(A(S) > \epsilon) \leq \delta \Leftrightarrow \\
&\lim_{m\to\infty} P_{S\sim D^m}(L_D(A(S) > \epsilon) - \delta \leq \lim_{m\to\infty} 0 \overset{*}{\Leftrightarrow} \\
&\lim_{m\to\infty} \frac{E_{S\sim D^m}[L_D(A(S)]}{\epsilon} - \delta \leq \lim_{m\to\infty} 0 \overset{\epsilon>0}{\Leftrightarrow} \\
&\lim_{m\to\infty} E_{S\sim D^m}(L_D(A(S)) \leq \lim_{m\to\infty} \delta\epsilon \overset{**}{\leq} \lim_{m\to\infty} \frac{|H|}{e^{m\epsilon}}\epsilon = \lim_{m\to\infty} \cancelto{0}{\frac{|H|}{e^m}} \underbrace{\cancelto{0}{\frac{\epsilon}{e^{\epsilon}}}}_{\epsilon>0} = 0 \overset{\triangle}{\Leftrightarrow} \\
&\lim_{m\to\infty} E_{S\sim D^m}(L_D(A(S)) = 0
\end{aligned}$$

דם

* נתון $L_D(A(S) \in [0,1]$, לכן מכיוון שהוא אי שלילי וגם $\epsilon > 0$ ניתן להשתמש בא"ש מרקוב.
** ראינו בהרצאה שתחת התנאים בשאלה מתקיים $\delta \leq \frac{|H|}{e^{m\epsilon}}$.
$\triangle$: $L_D(A(S) \in [0,1] \Rightarrow 0 \leq E_{S\sim D^m}(L_D(A(S))$, קיבלנו ש- $0 \leq \lim_{m\to\infty} E_{S\sim D^m}(L_D(A(S)) \leq 0$

שאלה 2

נניח ראליזביליות, כלומר שקיימת f המתייגת את הנקודות נכון לפי $\mathbb{I}[||x||_2 \leq r]$ עבור $r \in R_+$ כלשהו.
נראה ש-H היא למידה PAC בעזרת אלג' הלמידה A הבא:

- $max\_x = 0$
- לכל $x \in S$ כך ש-y המתאים הוא 1:
  - אם $||x||_2 > max\_x$ אז $||x||_2 = max\_x$.
- נחזיר את כלל ההחלטה $\mathbb{I}[||x||_2 \leq max\_x]$.

מהנחת הראליזביליות כל הנקודות שהתיוג שלהן הוא 1 מקיימות שהנורמה שלהן $\geq max\_x$,
בנוסף מההנחה כל הנקודות שהתיוג שלהן הוא 0 מקיימות שהנורמה שלהן $\leq max\_x$.

טענה: שלכל r אמיתי שלפיו f מתייגת, לכל התפלגות D של הנקודות ולכל $\epsilon, \delta \in (0,1)$, אם נבחר את m יתקיים $m_H(\epsilon, \delta) \leq m$
$P_{S \sim D^m}[L_D(A(S)) \leq \epsilon] \geq 1 - \delta$.
הוכחה:
נסמן את המעגל שנתקבל ע"י max_x ע"י r'. נשים לב שתמיד מתקיים $r' \leq r$ עבור r האמיתי לפיו f מתייגת.
לכן הטעות של האלג' יכולה להיות בנקודות $x \in R^2$ שמקיימות $r' < ||x||_2 \leq r$ מכיוון שהאלג' יתייג אותן 0, אבל f תייגה אותן 1.
נסמן $T' = \{x \mid r' < ||x||_2 \leq r\}$ - השטח בין היקף המעגל האמיתי לבין היקף המעגל שקבע A.
נסמן T את השטח החל מהיקף המעגל הנקבע לפי r ועד היקף המעגל $\tilde{r}$ עבור $0 \leq \tilde{r} \leq r$ כך שהשטח ביניהם מקבל הסתברות ϵ תחת ההתפלגות D, כלומר שמתקיים
$P_D(x \in \pi(r^2 - \tilde{r}^2)) = \epsilon$. כלומר T הוא השטח השטח בין היקף המעגל האמיתי לבין היקף מעגל פנימי כך שההסתברות שנקודה תמצא בשטח זה תחת D היא ϵ.

מתקיים ש $T \subset T' \Leftrightarrow P_D(x \in T') > \epsilon$.
בנוסף $T \subseteq T' \Leftrightarrow$ לא קיים $s \in S$ כך ש-$f(s) = 1$ וגם $s \in T$.
מכיוון שאם קיימת נק' s כזו, מהגדרת האלג' A הוא היה מתייג אותה ב-1 ולכן היה מתקיים $s \in T'$ והיינו מקבלים ש$T \not\subseteq T'$ מכיוון שהיקף המעגל שקובע את T' היה בתוך השטח של T.

נקבל שמהגדרת T הסיכוי להגריל m נקודות $i.i.d$ תחת D שאינן ב-T הוא $(1-\epsilon)^m$, לכן אם נבחר m כך שיתקיים $(1-\epsilon)^m \leq \delta$ נקבל $P_{S \sim D^m}(L_D(A(S)) \leq \epsilon) \geq 1 - \delta$.
לכן נקבל שעבור $(1-\epsilon)^m \leq e^{-m\epsilon}$ אם נבחר mכך ש- $e^{-m\epsilon} \leq \delta$ נקבל שהתנאי מתקיים.
$e^{-m\epsilon} \leq \delta \Leftrightarrow -m\epsilon \leq ln(\delta) \overset{\epsilon > 0}{\Leftrightarrow} m \geq \frac{ln(\delta^{-1})}{\epsilon}$.

קיבלנו שלכל $\epsilon, \delta \in (0,1)$ ולכל D, f, תחת הנחת הראליזביליות, אם נגריל $m \geq \frac{ln(\delta^{-1})}{\epsilon}$ דגימות $i.i.d$ מ-D אז אלג' A יהיה עם שגיאת הכללה לכל היותר ϵ בהסתרות $1 - \delta$.

שאלה 3

נניח שמימד VC של H הוא d. מהגדרה קיימת תת קבוצה C שמנתצת את H. נקבל מכך שב-Hיש לפחות 2^d היפותזות, כלומר $|H| \geq 2^d \Leftrightarrow d \leq \lfloor log_2|H| \rfloor$.

שאלה 4

נראה שמימד VC הוא לכל הפחות n. עבור הקבוצה $C = \{e_1, .., e_n\}$ וקטורי היחידה ב-R^n.
עבור תיוג y כלשהו של f נגדיר $I = \{i | y_i = 1\}$, ונקבל $h_I(e_i) = y_i$. קיבלנו שעבור כל תיוג y קיימת היפותזה ולכן C מנתצת את H.
מצד שני מתקיים $|H| = 2^n$, מכיוון שכל וקטור ניתן לתייג 0 או 1.
נקבל מכיוון ש-H סופית בעזרת שאלה 3 שמתקיים $VC(H) \leq log_2(2^n) = n$.
קיבלנו $VC(H) = n$.

שאלה 5

נראה שמימד VC הוא לכל הפחות $2k$.

נראה שהקבוצה $C = \{x_1, .., x_{2k}\}$ כאשר $x_i = i$ עבור $i \in [2k]$ מנתצת את H.

יהי $y \in \{0,1\}^{2k}$ תיוג של הנקודות ב-C אזי נגדיר $E = \{j \mid y_j = 1 \wedge (j = 2k \vee y_{j+1} = 0)\}, S = \{j \mid y_j = 1 \wedge (j = 1 \vee y_{j-1} = 0)\}$.

S קבוצת כל האינדקסים הראשונים באינטרוול כלשהו.

E קבוצת כל האינדקסים האחרונים באינטרוול כלשהו.

כאשר נזכיר שנקודות בתוך אינטרוול מתוייגות 1.

נמיין את S, E לפי הסדר.

נראה ש-$|S| = |E|$, תהי $j \in S$ אזי אם $j \in E$ סיימנו.

אחרת אם $j \notin E$ מתקיים $y_j = 1 \wedge (j = 1 \vee y_{j-1} = 0)$ אבל לא $y_j = 1 \wedge (j = 2k \vee y_{j+1} = 0)$ מהגדרת S, E.

יהי $j < p \leq 2k$ האינדקס הראשון כך ש-$y_p = 1 \wedge (p = 2k \vee y_{p+1} = 0)$, אזי מתקיים $p \in E$ וגם $y_{p-1} \neq 1$ לכן $p \notin S$.

ולכן $j < l < p$ מתקיים $l \notin E \wedge l \notin S$ מהגדרת S, E. לכן נקבל שלכל $j \in S$ כך ש-$j \notin E$ קיימת $p \in E$ כך ש-$p \notin S$, לכן נקבל $|S| = |E|$.

בנוסף מאופן הגדרת S, E מתקיים (לאחר מיון S, E) $S[i] \leq E[i]$ לכל $i \in [|S|]$.

לכן אם נגדיר $A = \bigcup_{i \in [|L|]} [S[i], E[i]]$, נקבל ש- $y_j = 1 \Leftrightarrow j \in [S[i], E[i]]\ for\ i \in [|L|]$

לכן $h_A(x_j) = y_j$ ונקבל ש-C מנתצת את H.

נסמן ב-d את מימד VC של H. נראה ש-$d < 2k + 1$.

תהי $C = \{x_1, .., x_{2k+1}\}$, ויהי התיוג h^* שמתייג כך - $y_j = j\ mod 2$, מספר הנק' שיתויגו 1 באופן זה הוא $k + 1$.

לכן לכל $A = \cup k - intervals$ נקבל שקיים אינטרוול שמכיל לפחות 2 נק'.

מבניית צורת התיוג בין כל 2 נק' שתוייגו 1 יש נק' שתויגה 0 (מכיוון שתבצע $mod 2$).

לכן נקבל שבאיחוד יש נקודה שמתוייגת לא נכון, הנקודה בין 2 הנק' שבאותו האינטרוול.

לכן ההיפותזה h^* לא יכולה להתקבל ע"י $H_{k-intervals}$, לכן C אינה מנתצת את H.

מכיוון שזו קבוצה כללית מגודל $2k + 1$ נסיק שאף קבוצה מגודל $2k + 1$ אינה מנתצת את H.

קיבלנו שמימד VC של $H_{k-intervals}$ הוא $2k$.

עבור $H_{intervals}$ נקבל מההוכחה של $H_{k-intervals}$ שעבור כל איחוד של k אינטרוולים קיימת קבוצה מגודל $2k$ שמנתצת אותה, לכן אם k אינו מוגבל מתקיים גם שגודל הקבוצה המנתצת הגדולה ביותר אינו מוגבל ולכן $VC(H_{intervals}) = \infty$.

שאלה 6

נסמן את מימד VC של H_{con} ב-p.

נראה ש-$p \geq d$ בעזרת הקבוצה $C = \{e_i\}$ עבור $i \in [d]$ וקטורי יחידה ב-R^d שנראה שהקבוצה C מנתצת את H_{con}.

יהי $y \in \{0,1\}^d$ תיוג כלשהו, נגדיר $I = \{i \mid y_i = 0\}$ ובהתאם $h = \bigwedge_{i \in I} \overline{x_i}$. נקבל $h(e_i) = y_i$ לכל $i \in [d]$, לכן $p \geq d$.

נראה ש-$p < d + 1$.
נניח בשלילה שקיימת $C = \{x_1, .., x_{d+1}\}$ שמנתצת את H_{con}, אזי קיימים $h_1, .., h_{d+1} \in$ H_{con} כך ש-$\forall i, j \in [d + 1]$ ניתן להגדיר:

$$h_i(x_j) = \begin{cases} 0 & i = j \\ 1 & o.w \end{cases}$$

נקבל שכל h_i מחזיק ליטרל שנסמנו l_i שהוא 1 על x_j לכל $j \neq i$ והוא 0 עבור x_i. משובך היונים (כאשר נתון $d \geq 2$) קיימים לפחות 2 ליטרלים מבין $d + 1$ הליטרלים האילו שעונים באותו האופן עבור x_k כלשהו.
נניח בה"כ שl_1, l_2 הם הליטרלים שעונים באותו האופן על משתנה מסויים. נקבל שאם $l_1 = l_2$ אזי $l_1 = 1$ על x_1 וגם $l_1 = l_2 = 1$ על x_2, בסתירה להגדרת h_1.
אם $l_1 \neq l_2$ נקבל שלא יתכן ששניהם יהיו 1 על x_k עבור $k \notin \{1, 2\}$ ומההנחה הם עונים שונים על x_k $k \in \{1, 2\}$, בסתירה לכך שיש משתנה שהם עונים עליו באותו האופן.
לכן קיבלנו שמימד VC הוא d.

שאלה 7

נראה לכל D, ϵ, δ, בהינתן $m \geq m^{UC}(\frac{\epsilon}{2}, \delta)$ מתקיים $P_{S\sim D^m}[L_D(h_S) \leq \min_{h' \in H} L_D(h') + \epsilon] \geq 1 - \delta$ ולכן נסיק ש-H היא למידה $Agnostic - Pac$.
הוכחה:
יהי $m \geq m^{UC}(\frac{\epsilon}{2}, \delta)$ ראינו בתרגול 8 (תרגול PAC 2) שתקיים כאשר S היא $\frac{\epsilon}{2}$-מייצגת אזי מתקיים $L_D(h_S) \leq \min_{h \in H} L_D(h) + \epsilon$.
מתכונת UC נקבל:

$$P_{S\sim D^m}[L_D(h_S) \leq \min_{h' \in H} L_D(h') + \epsilon] \overset{*}{\geq} D^m(\{S \in (X \times Y)^m | S \ is \ \frac{\epsilon}{2} - representative) \overset{UC}{\geq} 1 - \delta$$

* מכיוון שברור שההסתברות להגריל מדגם שהוא $\epsilon - representative$ (הביטוי בצד שמאל) $\leq$ מההסתברות להגריל מדגם $\frac{\epsilon}{2} - representative$ (הביטוי באמצע).

שאלה 8

תהי H **לא** למידה PAC נניח בשלילה ש-H למידה $Agnostic - PAC$ אזי מהגדרה קיים אלג' למידה A עבור H ופונק' סיבוכיות מדגם - $m_H(\epsilon, \delta)$ כך שלכל $\epsilon, \delta \in (0, 1)$ ולכל התפלגות D מעל Z אם המדגם S מקיים $S \sim D^m$ כאשר $m \geq m_H(\epsilon, \delta)$ אזי $P[A(S) \leq min_{h \in H} h + \epsilon] \geq 1 - \delta$.
כאשר הטענה מתקיימת לכל D מעל Z כך שהתיוג מתקבל בעזרת פונק' תיוג $f \in H$ דטרמיניסטית לדוגמה: $D(y|x) = \begin{cases} 1 & y = f(x) \\ 0 & o.w \end{cases}$.
קיבלנו שניתן ליצור כל התפלגות D בדרך זו מעל X ואז לבחור פונק' תיוג $f \in H$ ולכן נקבל ש-H למידה PAC, בסתירה להנחה. לכן H לא למידה $Agnostic - PAC$.

שאלה 9

נתון H למידה PAC, נקבל:
עבור $0 < \epsilon_1 \leq \epsilon_2 < 1$ ו-$\delta \in (0,1)$, בהינתן $m \geq m_H(\epsilon_1, \delta)$ דגימות iid מ-D כלשהי, מתקיים $L_{D,f}(h) \leq \epsilon_1 \leq \epsilon_2$,
לכן נסיק מהמינימליות של $m_H(\epsilon_2, \delta)$ שמתקיים $m_H(\epsilon_2, \delta) \leq m_H(\epsilon_1, \delta)$.

עבור $0 < \delta_1 \leq \delta_2 < 1$ ו-$\epsilon \in (0,1)$,
בהינתן $m_1 \geq m_H(\epsilon, \delta_1)$ דגימות iid מ-D כלשהי, מתקיים $P_{S\sim D^{m_1}}(L_{D,f}(h) \leq \epsilon) \geq 1-\delta_1 \geq 1-\delta_2$,
בהינתן $m_2 \geq m_H(\epsilon, \delta_2)$ דגימות iid מ-D כלשהי, מתקיים $P_{S\sim D^{m_2}}(L_{D,f}(h) \leq \epsilon) \geq 1-\delta_2$,
לכן נסיק מהמינימליות של $m_H(\epsilon, \delta_2)$ שמתקיים $m_H(\epsilon, \delta_2) \leq m_H(\epsilon_1, \delta_1)$.

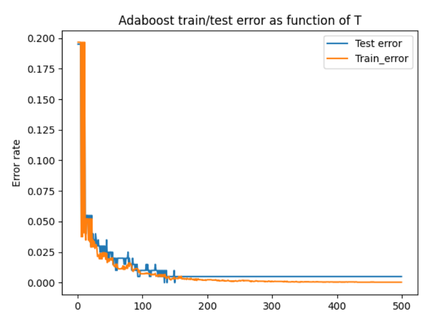
שאלה 10

נתון $H_1 \subseteq H_2$, לכן לכל $C \subseteq X$ מתקיים בהכרח שאם C מנתצת את H_1 אזי היא גם מנתצת את H_2, ניתן בעזרת ההיפותזות של H_1 לנתץ את C גם תחת H_2.
לכן נקבל $VC(H_1) \leq VC(H_2)$.

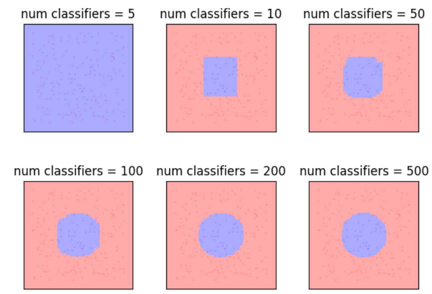
שאלה 12

בקוד.

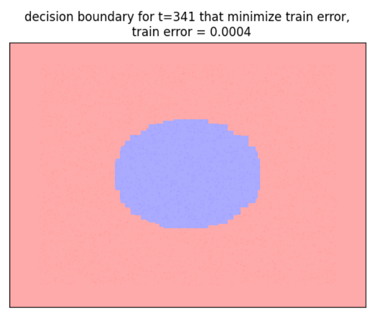
שאלה 13



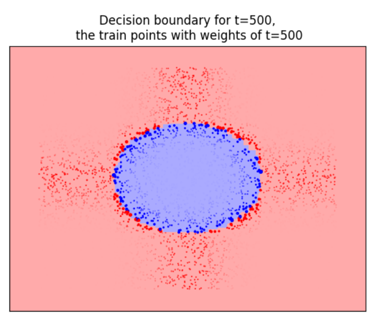
שאלה 14



שאלה 15



שאלה 16

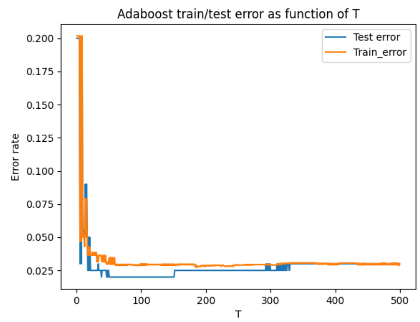


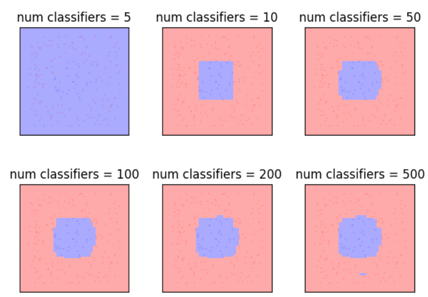
ניתן לראות שהמשקולות של הנקודות שקרובות לכלל ההחלטה הן בעלות משקל גבוהה יותר בהתפלגות בהשוואה לנקודות רחוקות יותר ממנו באופן הגיוני, כיוון שבמהלך האיטרציות האלג' טעה בנקודות אילו יותר ולכן העלה את המשקל בהן כדי להתאים את עצמו טוב יותר אליהם בהיפותזות שנוצרו במהלך האיטרציות והוריד את ההסתברות מנקודות רחוקות מכיוון שעליהן צדק פעמים רבות.

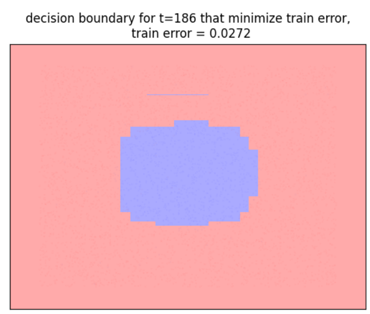
ניתן לראות בשאלה 14 שכבר לאחר 10 איטרציות נוצר כלל החלטה (משוקלל עד איטרציה זו) שמסווג נכון את הנקודות שהן במשקל נמוך בתרשים של שאלה 16, לכן בהמשך על מנת ליצור את כלל ההחלטה בצורת המעגל הועלתה ההסתברות לנקודות הרלוונטיות וירדה לפחות רלוונטיות.

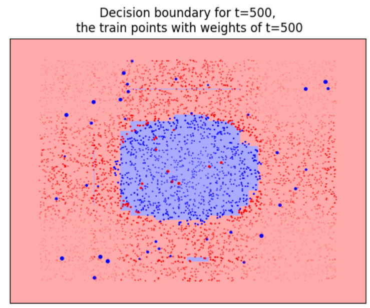
שאלה 17

תרשימים עבור רעש 0.01

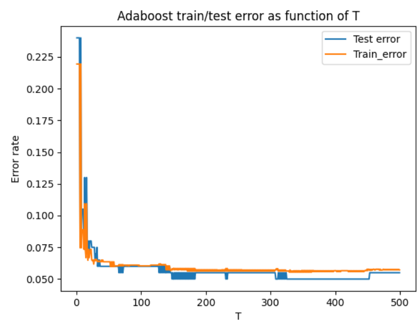


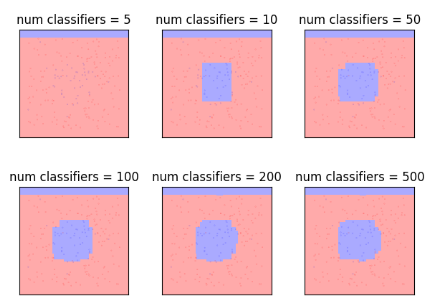


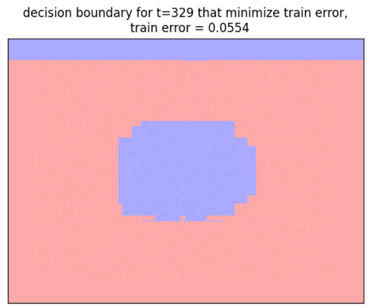


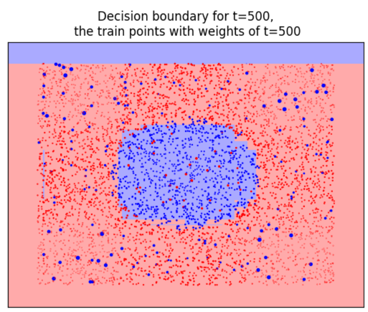


תרשימים עבור רעש 0.04









- השיויים: ניתן לראות שכללי ההחלטה בתרשימים בשאלה 14 השתנו החל מ-$T = 5$ בין כללי ההחלטה של כל דאטה (עם רעש שונה). ניתן גם לראות שבתרשים של שאלה 16 יש יותר נקודות עם משקל גבוה יחסית בדאטה עם הרעש 0.04 בהשוואה לתרשים המתאים בדאטה עם רעש 0.01.
- בתרשימים של שאלה 13 ניתן לראות שבמונחים של "$bias - complexity\ tradeoff$":
  - $\epsilon-approximation$: ככל ש-T גדל $\epsilon-approximation$ יורד מכיוון שניתן להראות ששגיאת האימון יורדת.
  - $\epsilon - estimation$: ניתן לראות ששגיאת האימון המינימלית גבוה יותר עבור הדאטה המורעש יותר, כלומר האלג' רגיש יותר לרעש בדאטה.
- השוני בשאלה 15 הוא ששגיאת האימון המינימלית בדאטה עם רעש 0.04הוא 0.05 ומתקבל עבור $T = 329$, כאשר בדאטה עם רעש 0.01הוא 0.02 ומתקבל עבור $T = 186$. כלומר ניתן להסיק שבדאטה הפחות מורעש ניתן לקבל שגיאת אימון נמוכה יותר ותוך פחות איטרציות.