321123788 — תרגיל 4 - IML

אשלון זוממן

1) הגדרה נכונה לפי עקב:

$$\forall \varepsilon,\delta>0 \ \exists m(\varepsilon,\delta) \ \forall m\geq m(\varepsilon,\delta) \ \mathbb{P}_{S\sim D^m}\left(L_D(A(S))\leq\varepsilon\right)\geq 1-\delta$$

$$\Updownarrow$$

$$L_D(A(S)) \xrightarrow[m\to\infty]{} 0$$

$\Downarrow$: נניח בשלילה ש:

$$L_D(A(S)) \xrightarrow[m\to\infty]{} k \quad , \quad k>0$$

לכן בהינתן $0<\varepsilon'<k$ קיים $m'\in\mathbb{N}$ כך שעבור כל $m\geq m'$ מתקיים $L(A(S))>\varepsilon'$. לכן עבור $\varepsilon=\varepsilon'$ ו $\delta\not<1$ מתקיים:

$$\mathbb{P}_{S\sim D^{m}}\left(L_D(A(S))\leq\varepsilon'\right) \underset{m>m'}{\overset{\text{לכן}}{=}} 0 \geq 1-\delta \iff \delta\geq 1$$

בסתירה לכך ש $\delta\not<1$, לכן $L_D(A(S)) \xrightarrow[m\to\infty]{} 0$.

$\Uparrow$: אם $L_D(A(S)) \xrightarrow[m\to\infty]{} 0$ אז מהגדרת הגבול קיים לכל $\varepsilon>0$ $N\in\mathbb{N}$ כך שלכל $m>N$: $|L_D(A(S))|\leq\varepsilon \Leftarrow L_D(A(S))\leq\varepsilon$ (כי L_D פונקציה רק ערכים (~~חיוביים~~ אי-שליליים)). לכן

$$\mathbb{P}_{S\sim D^m}\left(L_D(A(S))\leq\varepsilon\right)=1\underset{0<\delta \text{ לכל}}{\geq} 1-\delta$$

כנדרש.

כעת נוכיח את התכונה:

$a\Rightarrow b$:

$$\mathbb{P}_{S\sim D^m}\left(L_D(A(S))\leq\varepsilon\right)\geq 1-\delta \iff L_D(A(S)) \xrightarrow[m\to\infty]{} 0$$

$$\Downarrow$$

$$\lim_{m\to\infty}\mathbb{E}_{S\sim D^m}\left(L_D(A(S))\right)=\lim_{m\to\infty}\left(\sum_{S\in D^m}L_D(A(S))\right)=\sum_{S\in D^m}\left(\lim_{m\to\infty}L_D(A(S))\right)=$$

$$=\sum_{S\in D^m}0=0$$

[illegible]:

$$\forall \varepsilon>0 \quad \frac{\mathbb{E}_{S\sim D^m}(L_D(A(S)))}{\varepsilon} \underset{\text{מרקוב}}{\geq} \mathbb{P}(L_D(A(S))\geq\varepsilon) = 1 - \mathbb{P}(L_D(A(S))<\varepsilon) \iff$$

$$\iff \mathbb{P}(L_D(A(S))<\varepsilon) \geq 1 - \frac{\mathbb{E}(L_D(A(S)))}{\varepsilon} \xrightarrow[m\to\infty]{} 1 - \frac{0}{\varepsilon} = 1 \implies$$

$$\implies L_D(A(S)) \xrightarrow[m\to\infty]{} 0$$ (שהרי ההסתברות זה מתקיים לכל $\varepsilon>0$) $\iff$ a (סעיף עבר)

ב) נציג את הגדרה ליטוי ה- threshold ~~[illegible]~~ מהתרגול ונוכיח את האלגוריתם הזה:

נניח: $S\subseteq X$ דגימה של נקודות על $\mathbb{R}^2$. האלגוריתם יחזיר $h_{r'}\in H$ כאשר מתקיים:

$$r' = \max\{\|x\| : x\in S \wedge h_R(x)=1\}$$

h_R היא הפונקציה ה״אמיתית״.

נשים לב כי עבור $x\in X$ כך ש- $\|x\|\le r'$ או $R<\|x\|$, תהיה הסכמה בין $h_{r'}$ ל- h_R (אין שגיאה), ושהמקום בו תהיה השגיאה הוא בקבוצת $x\in X$ כך ש- $r'<\|x\|\le R$, ושם $h_{r'}(x)=0$ אבל $h_R(x)=1$.

בדומה להוכחה בהרצאה, במידה ו- $D(\{x: 0\le\|x\|\le R\})<\varepsilon$ אז גם $D(\{x: r'<\|x\|\le R\})<\varepsilon$ ולכן הסיכוי לשגיאה תמיד יהיה קטן מ-ε, לכל גודל דגימה. אם כן נניח $\varepsilon\le D(\{x: 0\le\|x\|\le R\})$, ונגדיר

$\tilde r\ge 0$ כך ש- $D(\{x:\tilde r\le\|x\|\le R\})=\varepsilon$ ~~[illegible]~~

נשים לב כי ההסתברות כי מכיל את שיעור השגיאה בתוחלת, ולכן הסיכוי לשגיאה קטן או שווה מאפסילון. הסיכוי לקבל test sample שלא מכיל נק' שההסתברות הזו היא $(1-\varepsilon)^m$ ולסיכום מתקיים:

$$(1-\varepsilon)^m \le (e^{-\varepsilon})^m \le \delta$$

$$\Updownarrow$$

$$1 \le e^{\varepsilon m}\cdot\delta$$

$$\Updownarrow$$

$$\ln(1) \le \varepsilon\cdot m + \ln(\delta)$$

$$\Updownarrow$$

$$\ln(1/\delta) \le \varepsilon m$$

$$\Updownarrow$$

$$\frac{\ln(1/\delta)}{\varepsilon} \le m$$

3) כמו בתרגול, ניקח את הבסיס הסטנדרטי $B=\{e_1,\dots,e_d\}$ ונראה שהוא מנותץ:

יהי $y\in\{0,1\}^d$, ונגדיר בהתאם $I=\{i : y_i=0\}$ קבוצת האינדקסים בהן יש 0 בתוך y. כעת נגדיר גם $h:\{0,1\}^d\to\{0,1\}$ באופן הבא:

$$\forall x\in\{0,1\}^d \quad h(x)=\bigwedge_{i\in I}\bar{x}_i$$

כעת נבדיל בין שני מקרים:

$$y_i=1 \Rightarrow i\notin I \Rightarrow h(e_i)=\bigwedge_{j\in I}\bar{0}=1=y_i$$

$$y_i=0 \Rightarrow i\in I \Rightarrow h(e_i)=\Big(\bigwedge_{\substack{j\in I\\ j\neq i}}\bar{0}\Big)\wedge\bar{1}=0=y_i$$

מקבלים $h(e_i)=y_i$, ולכן מכיוון של-y יש 2^d אפשרויות נסיק ש-$|H|=2^d$ כנדרש.

כעת נניח בשלילה שקיימת קבוצה $V=\{v_1,\dots,v_{d+1}\}\subseteq X$ מנותצת. מכך שכל ההשמות אפשריות, נסיק שבין היתר ניתן לבחור קבוצת פונ' כך ש:

$$H=\{h_1,\dots,h_{d+1}\},\quad \forall i,j\in[d+1]\quad h_i(v_j)=\begin{cases}0 & i=j\\ 1 & i\neq j\end{cases}$$

נזכור כי לכל $i\in[d+1]$ הפונ' h_i מייצגת גימום של כמות מסוימת של ליטרלים, לכן מאופן הגדרתם לכל h_i קיים ליטרל (או משלימו) שערכו 0 עבור x_i ו-1 עבור x_j $i\neq j$, נסמן ליטרל זה ב l_i. אם כך נסיק מעקרון שובך היונים ומכך ש- V שקיימים שני ליטרלים l_k, l_m $k\neq m$ המתאימים לאותו האינדקס $i\in[d]$ בכל וקטור ב V. כעת נבחין:

__אם $l_k=l_m$__: לכן l_k מקבל ערך 1 עבור v_m ($k\neq m$ לכן זה נובע מהמסקנה על h_m), אבל אז גם l_m מקבל ערך 1 עבור v_m, בסתירה למסקנה.

__אם $l_k\neq l_m$__: אז עבור $n\neq k \wedge n\neq m$ l_k ו-l_m יקבלו ערכים שונים בוקטור v_n, בסתירה להגדרת h_n.

אם כך נסיק שלא ניתן להגדיר קבוצה H כזו ולכן לא קיימת קבוצה בגודל $d+1$ שניתן לנתץ.

לסיכום נסיק:

$$VC\ H_{con}\leq d \ \wedge\ VC\ H_{con}\geq d$$

$$\Downarrow$$

$$VC\ H_{con}=d$$

5) ההבדל בהגדרה בשאלה הוא הסדר של ההתפלגות D והגודל m_H.
בהגדרה agnostic PAC קיים m_H כך שעבור כל התפלגות D
ניתן ללמוד אלגו' A, אבל בשאלה לכל התפלגות D קיים גודל
m_H כך שניתן ללמוד אלגו' A.
אם כך האלגו' A בעצם יכול להיות תלוי בD, ולכן תמיד נוכל
לבנות את A כך שיקיים את תנאי השאלה. ~~[illegible]~~
~~[illegible]~~ היא נכון, כל מחלקת היפותזות מקיימת
תנאי השאלה, ולכן ניקח מחלקה שאינה למידה PAC
ונראה שמכך נגזר שהיא אינה למידה agnostic PAC.
תהי H מחלקת כל פונ' הסיווג הבינאריות, ראינו שאינה
למידה PAC. כעת נשים לב שאינה גם למידה agnostic PAC.
בהגדרה זו אומר שקיים אלג' למידה A ופונ' $m_H(\varepsilon,\delta)$ כך שעבור
כל התפלגות D מעל Z $(X \times \{\pm 1\})$, אם נריץ את A
על קבוצת דגימות בגודל $m \geq m_H(\varepsilon,\delta)$ נקבל בהסתברות לפחות
$1-\delta$ פונ' שמקיימת $A(S) \leq \min_{h \in H} h + \varepsilon$. בפרט דבר זה
יתקיים עבור התפלגות בה הלייבלים של הדגימות נקבעים
באופן דטרמיניסטי ע"י פונ' $f: X \to \{\pm 1\}$ כלשהי:

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

מכיוון שכל התפלגות כזאת ניתנת לבנייה ע"י בחירת התפלגות
D' מעל ל-X ופונ' $f \in H$, נסיק ש-H היא למידה
PAC, בסתירה להנחה.
לכן לסיכום H הנה אינה למידה agnostic PAC.

7) נניח בשלילה כי $VC\mathcal{H}_1 > VC\mathcal{H}_2$. לכן קיימת פונ' $h_1\in\mathcal{H}_1$
ו- $d\in\mathbb{N}$ כך שעבור קבוצת דגימות S בגודל d h_1
מנתצת אותה. מכך שהנחנו $VC\mathcal{H}_1 > VC\mathcal{H}_2$ נוכל לבחור d כזה
כך ש- $d > VC\mathcal{H}_2$.
כעת מהנתון $\mathcal{H}_1\subseteq\mathcal{H}_2$ נסיק ש- $h_1\in\mathcal{H}_1\subseteq\mathcal{H}_2$,
כלומר קיימת קבוצת דגימות בגודל d המנותצת ע"י פונ'
מתוך $\mathcal{H}_2$, כלומר $d\le VC\mathcal{H}_2$. וזו הינה סתירה
להנחה ולמסקנה $VC\mathcal{H}_2 < d$.

8 a) נשים לב כי $\tau_{\mathcal{H}}(m)$ זה מספר הסיווגים השונים המקסימלי שקבוצת דגימות $C\subseteq X$ בגודל m יכולה לקבל. במילים אחרות זה ~~מספר~~ גודל קבוצה הכי ~~מפוצצת~~ המקסימלי עבור קבוצת דגימות בגודל m.

b) בהגדרה:

$$VCdim(\mathcal{H})=\infty \iff \sup\{m\in\mathbb{N}: \exists C\subseteq X, |C|=m \text{ s.t. } \mathcal{H} \text{ shatters } C\}=\infty$$

או במילים אחרות: לכל קבוצת דגימות בגודל k יש את 2^k היפותזות. מכך שלא יתכן יותר מ- 2^k היפותזות עבור כל קבוצה שכזו נקבל:

$$\forall m\in\mathbb{N} \quad \tau_{\mathcal{H}}(m)=2^m$$

c) נשים לב כי עבור $m\le d$ קיימת $C\subseteq X$ כך שמספר ההיפותזות עליה הוא 2^m כאשר $|C|=m$. הדבר נכון לכל $m\le d$ מהגדרת $VCdim(\mathcal{H})$, לכן נקבל:

$$\forall m\in\mathbb{N}, m\le d \quad \tau_{\mathcal{H}}(m)=2^m$$

d) i) נוכיח באינדוקציה על $|C|$:

<u>$|C|=1$</u>: נשים לב כי במקרה זה $2\ge|\mathcal{H}_C|$, כעת נפרק למקרים:

* <u>$|\mathcal{H}_C|=1$</u>: לכן רק הקבוצה הריקה $C_0\subseteq C$ מנותצת, מכיוון שעבור $x\in C$, $x\ne\phi$ מתקיים שהוא מקבל בדיוק את אותו הסיווג עבור כל $h\in\mathcal{H}$. אם כך:

$$|\{B\subseteq C: \mathcal{H} \text{ shatters } B\}|=|\{\{\phi\}\}|=1$$

* <u>$|\mathcal{H}_C|=2$</u>: כלומר כל C מנותצת, ולכן גם כל תת קבוצה שלה מנותצת, לכן:

$$|\{B\subseteq C: \mathcal{H} \text{ shatters } B\}|=|\{\{\phi\},\{x\}\}|=2$$

והבסיס מתקיים.

כעת נניח נכונות עבור גודל קבוצה $k<m$.

$|C|=m$:

נסמן ב- הקבוצה בגודל m כלשהי $C=\{c_1,\dots,c_m\}$:

$$C'=\{c_2,\dots,c_m\}:$$

כעת נגדיר שתי קבוצות נוספות:

$$Y_0=\{(y_2,\dots,y_m):(-1,y_2,\dots,y_m)\in\mathcal{H}_C \vee (1,y_2,\dots,y_m)\in\mathcal{H}_C\}$$

$$Y_1=\{(y_2,\dots,y_m):(-1,y_2,\dots,y_m)\in\mathcal{H}_C \wedge (1,y_2,\dots,y_m)\in\mathcal{H}_C\}$$

כעת נשים לב ש- $\mathcal{H}_C=Y_0\cup Y_1$, ולכן:

$$|\mathcal{H}_C|=|Y_0\cup Y_1|\le|Y_0|+|Y_1|$$

כמו כן נשים לב ש:

$$\forall y\in Y_0\ \exists h\in\mathcal{H}\quad y(x)=h(x)\ ,x\in C'$$

לכן $Y_0=\mathcal{H}_{C'}$ ומהנחת האינדוקציה מתקיים:

$$|Y_0|=|\mathcal{H}_{C'}|\le|\{B\subseteq C':\mathcal{H}\text{ shatters }B\}|$$

כעת נגדיר קבוצה חדשה של פונקציות:

$$\mathcal{H}'=\{h\in\mathcal{H}:\exists h'\in\mathcal{H}\ s.t\ (-h'(c_1),\dots,h'(c_m))=(h(c_1),\dots,h(c_m))\}$$

הקבוצה הזו מורכבת מפונ' אשר קיימת פונ' אחרת שמסכימה איתן על כל הערכים ב-C' ולא על c_1. נבחין כי מההגדרה הפונ' האחרת גם כן מוכלת ב-$\mathcal{H}'$. גם כך נסיק שבמידה ו-$\mathcal{H}'$ מנתצת $B\subseteq C'$ st, אז לכל השמה של ערכים עבור $c_2,\dots,c_m$ יהיו שתי פונקציות שישנו אותה, רק של פונקציה כזו תיתן ערך שונה ל-c_1, ולכן $\mathcal{H}'$ מנתצת גם את $B\cup\{c_1\}$. הגדרה בכיוון ההפוך זהה.

כעת נשים לב שמהגדרת Y_1 מתקיים $Y_1 = \mathcal{H}'_{C'}$.

בשילוב עם הטענה A עבור $B \cup \{c_1\}$ נקבל:

$$|Y_1| = |\mathcal{H}'_{C'}| \underset{\text{הנחת האינדוקציה}}{\le} |\{B \subseteq C' : \mathcal{H}' \text{ shatters } B\}| =$$

$$= |\{B \subseteq C' : \mathcal{H}' \text{ shatters } B \cup \{c_1\}\}| \le |\{B \subseteq C' : \mathcal{H} \text{ shatters } B \cup \{c_1\}\}|$$

ולסיכום:

$$|\mathcal{H}_C| \le |Y_0| + |Y_1| \le |\{B \subseteq C' : \mathcal{H} \text{ shatters } B\}| + |\{B \subseteq C' : \mathcal{H} \text{ shatters } B \cup \{c_1\}\}| =$$

$$= |\{B \subseteq C : c_1 \notin B \wedge \mathcal{H} \text{ shatters } B\}| + |\{B \subseteq C : c_1 \in B \wedge \mathcal{H} \text{ shatters } B\}| =$$

$$= |\{B \subseteq C : \mathcal{H} \text{ shatters } B\}|$$

כנדרש.

ii) משמעות הלמה אומרת כי כמות הפונקציות השונות המתקבלות בצמצום $\mathcal{H}$ לתת קבוצה C גדלה לכל היותר כמספר תתי הקבוצות של C המנותצות ע"י $\mathcal{H}$. ~~[illegible]~~

זה טוב מכיוון שאומר לנו שהמשמעות לדגום קבוצה גדולה יותר ~~[illegible]~~ תביא גדילה עד פרופורציונלית לפונקציות שהולכות ויורדות לאפס.

~~iii) נשים לב כי הביטוי $\sum_{k=0}^{d}\binom{m}{k}$ מבטא את סכום [illegible] בגודל m [illegible] הקבוצות שבגודל d לכל היותר~~

~~$\sum_{k=0}^{d}\binom{m}{k} = 2^d$~~

iii) מכך ש- $VCdim(\mathcal{H}) = d$ נסיק שאף קבוצה מגודל גדול יותר מ-d אינה מנותצת. לכן כל תתי הקבוצות המנותצות ~~[illegible]~~ נלקחות בחשבון בביטוי $\sum_{k=0}^{d}\binom{m}{k}$, המייצג את כמות כל הקבוצות מגודל קטן או שווה ל-d. אם כך נסיק:

$$|\{B \subseteq C : \mathcal{H} \text{ shatters } B\}| \le \sum_{k=0}^{d}\binom{m}{k}$$

Сканировано с CamScanner

iv) מהטענות שהוכחנו ב-i וב-iii נקבל כי:

$$|\mathcal{H}_C| \le \sum_{k=0}^{d} \binom{m}{k}$$

כעת מכך ש:

$$\sum_{k=0}^{d} \binom{m}{k} \le \left(\frac{em}{d}\right)^d$$

$$\Downarrow$$

$$|\mathcal{H}_C| \le \left(\frac{em}{d}\right)^d$$

בהגדרה:

$$\tau_{\mathcal{H}}(m) = \max\{|\mathcal{H}_C| : C \subseteq X, |C| = m\}$$

מכך שהטענות הקודמות תקפות לכל $\mathcal{H}_C$ ולכן בפרט גם ל-מקסימלי, נסיק לסיכום:

$$\tau_{\mathcal{H}}(m) \le \left(\frac{em}{d}\right)^d$$

e) לא שוויון שכן עדיין מתקיים ומתקיים $\tau_{\mathcal{H}}(m) \le e^d$.
אם נבחר החסם שלנו הדוק, כי ב-iii נקבל!

$$|\{B \subseteq C \mid \mathcal{H} \text{ shatters } B\}| \le \sum_{k=0}^{d} \binom{d}{k}$$

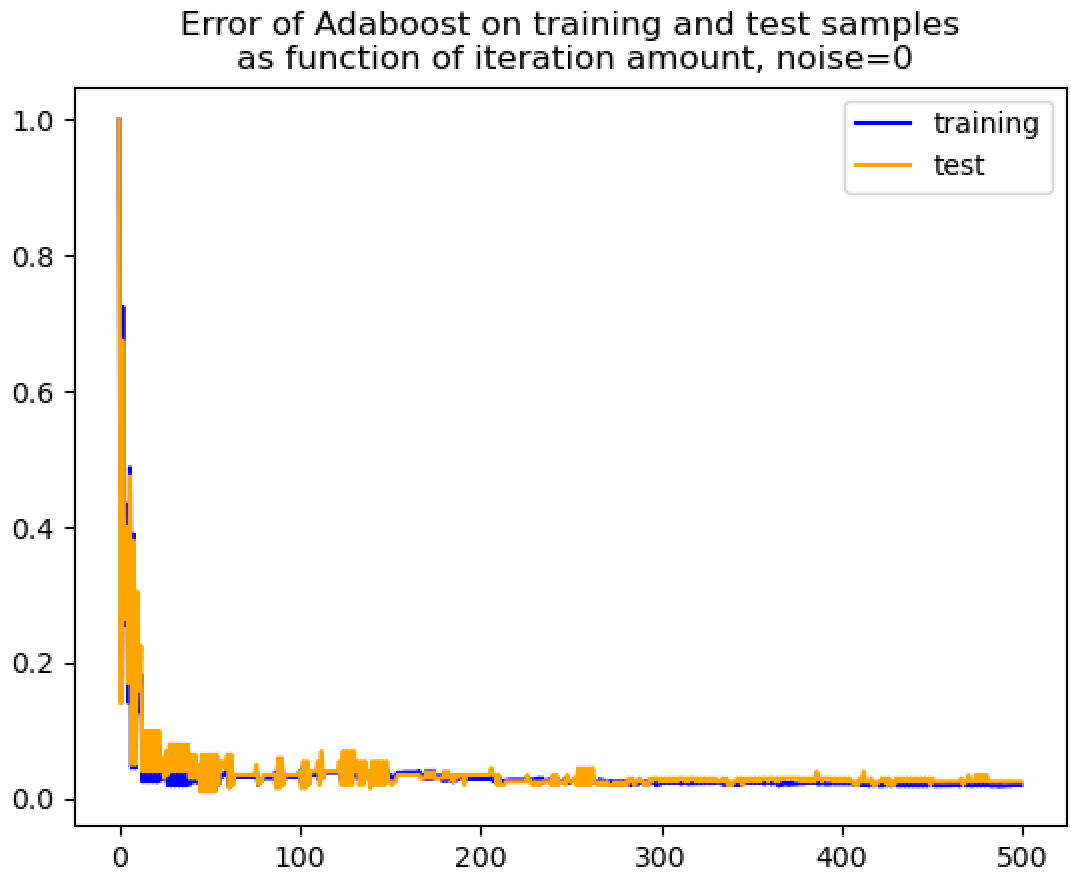
כאשר הביטוי מימין מייצג את $\boxed{\text{כל}}$ תתי הקבוצות של קבוצה בגודל d, ולכן:

$$\sum_{k=0}^{d} \binom{d}{k} = 2^d$$

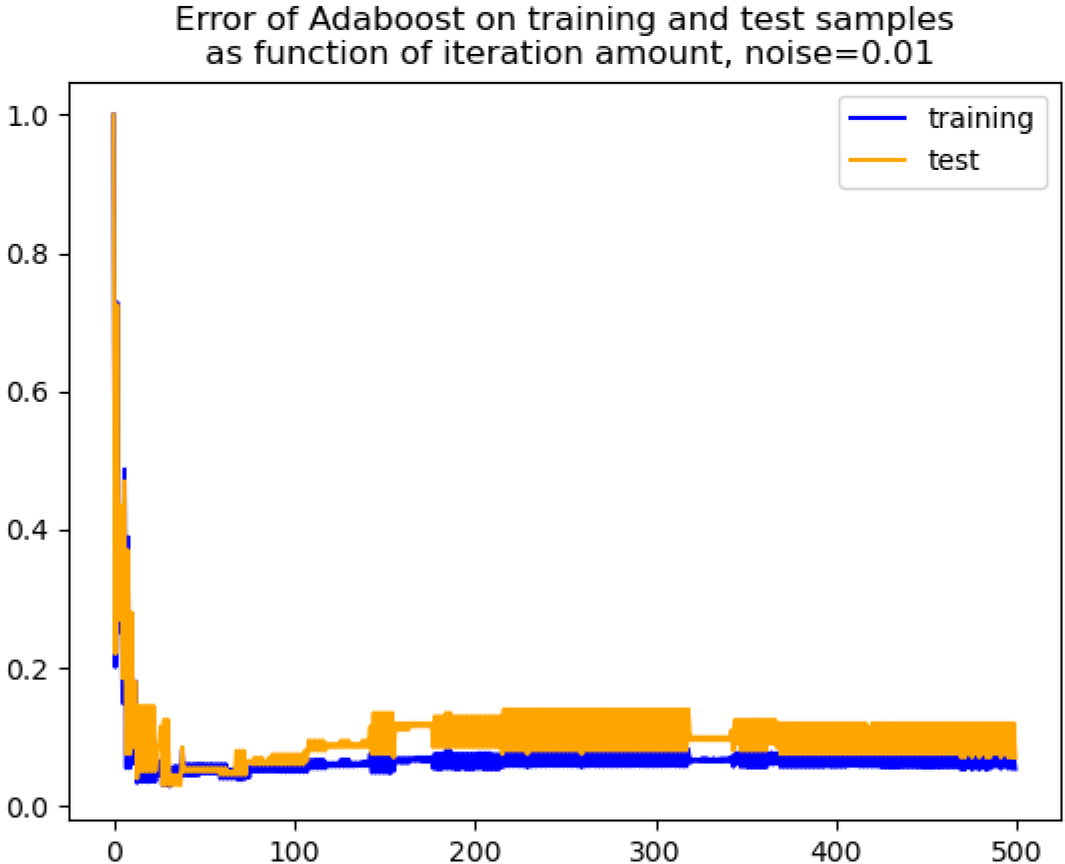
ובמיוחד $2^d < e^d$.

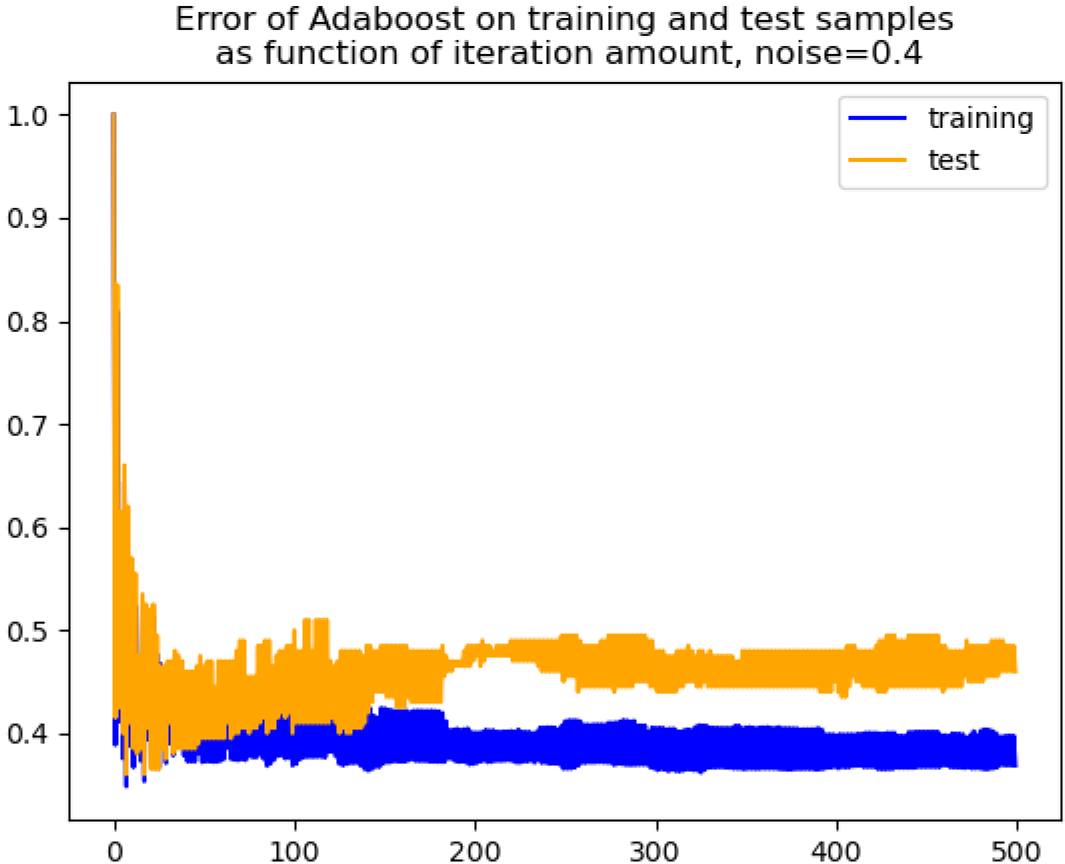
f) $\underline{m \le d}$: מהגדרת $\tau_{\mathcal{H}}(m)$ ו- $VCdim(\mathcal{H}) = d$ עבור $m \le d$ קיימת קבוצה שמתנפצת בגודל m ולכן $\tau_{\mathcal{H}}(m) = 2^m$:
$\underline{d < m}$: מההגדרה: ניתן לראות ש- $\tau_{\mathcal{H}}(m)$ גדלה באופן פולינומיאלי עבור $d < m$.

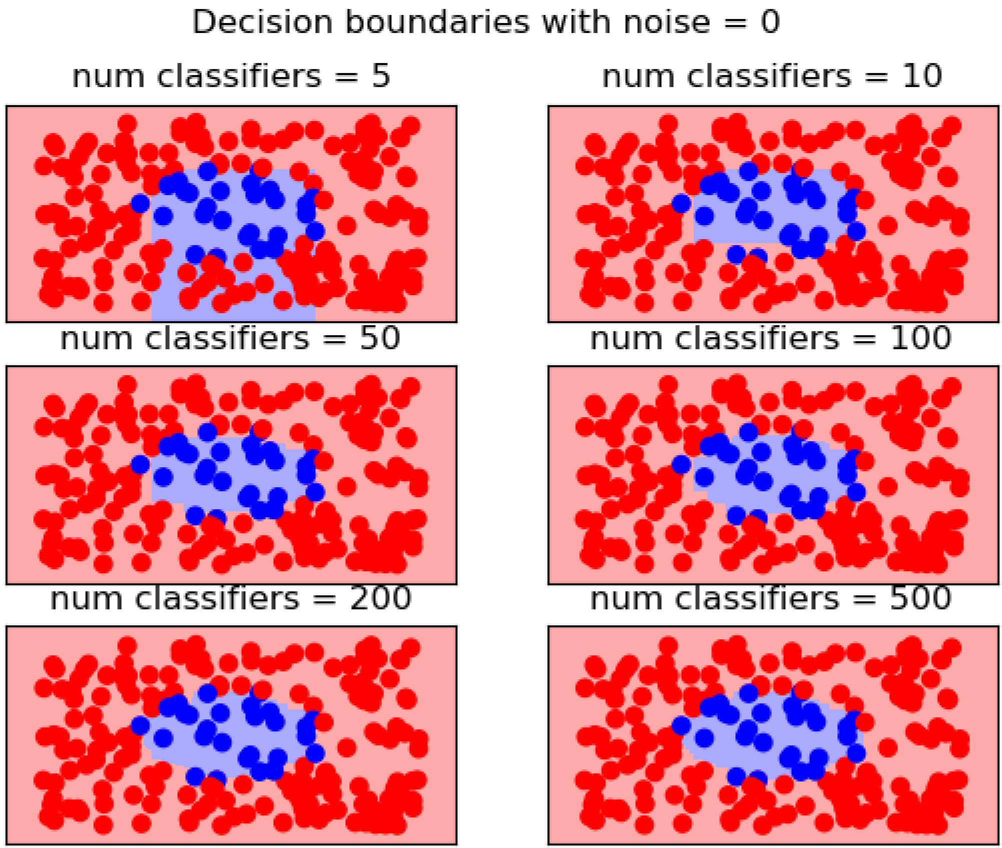
גם כך ניתן להשתמש בתוצאות הללו כדי לומר ש-
$VCdim(\mathcal{H}) = d \iff \tau_{\mathcal{H}}(m)$ מתנהג כפולינום למעלה.

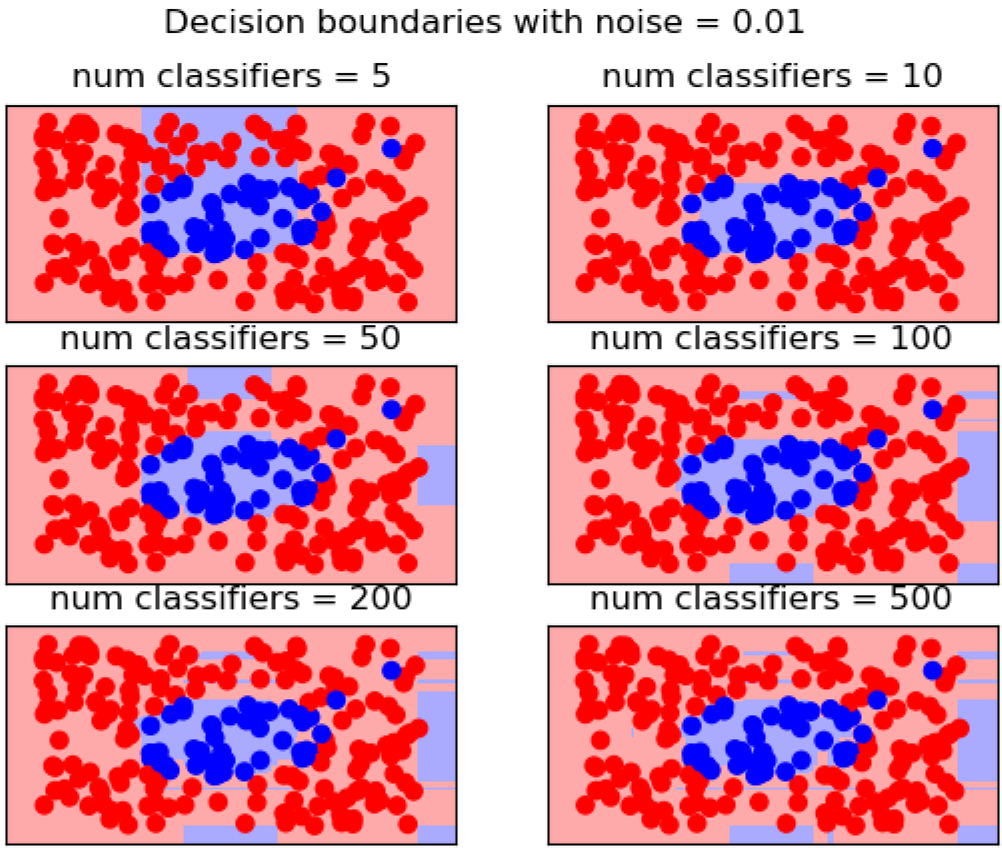


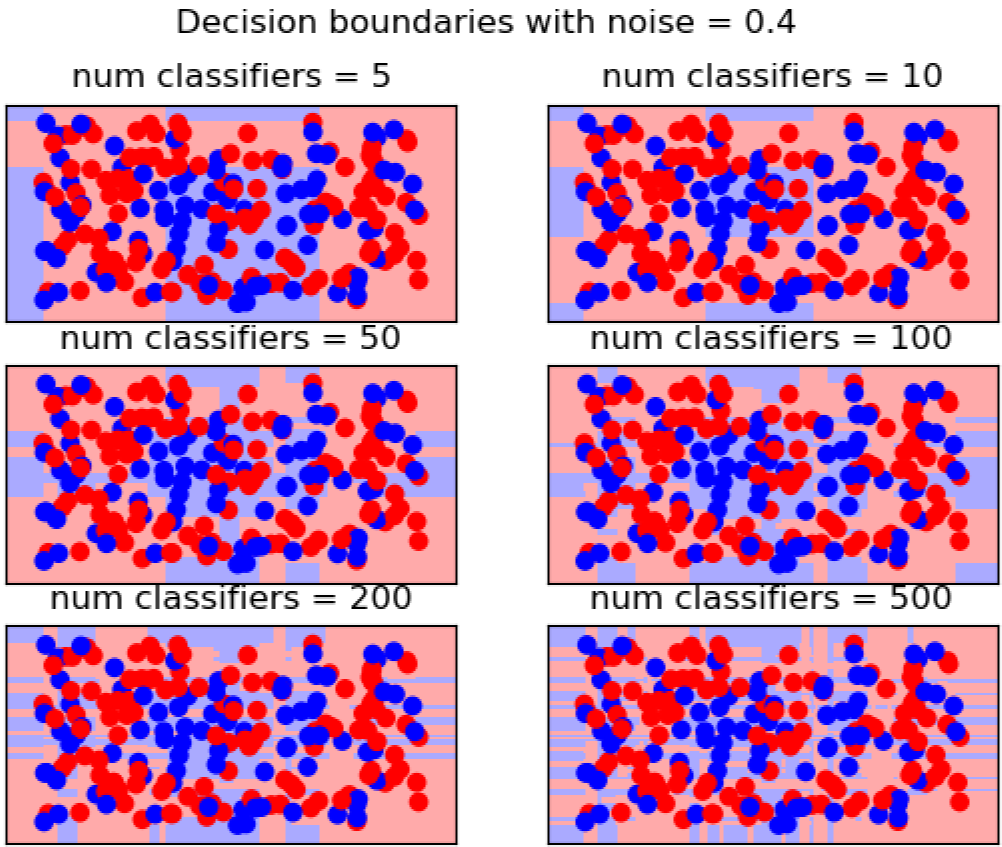
Error of Adaboost on training and test samples
as function of iteration amount, noise=0
training
test
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
400
500


Error of Adaboost on training and test samples
as function of iteration amount, noise=0.01
training
test
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
400
500


Error of Adaboost on training and test samples
as function of iteration amount, noise=0.4
training
test
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
100
200
300
400
500


Decision boundaries with noise = 0
num classifiers = 5
num classifiers = 10
num classifiers = 50
num classifiers = 100
num classifiers = 200
num classifiers = 500


Decision boundaries with noise = 0.01
num classifiers = 5
num classifiers = 10
num classifiers = 50
num classifiers = 100
num classifiers = 200
num classifiers = 500

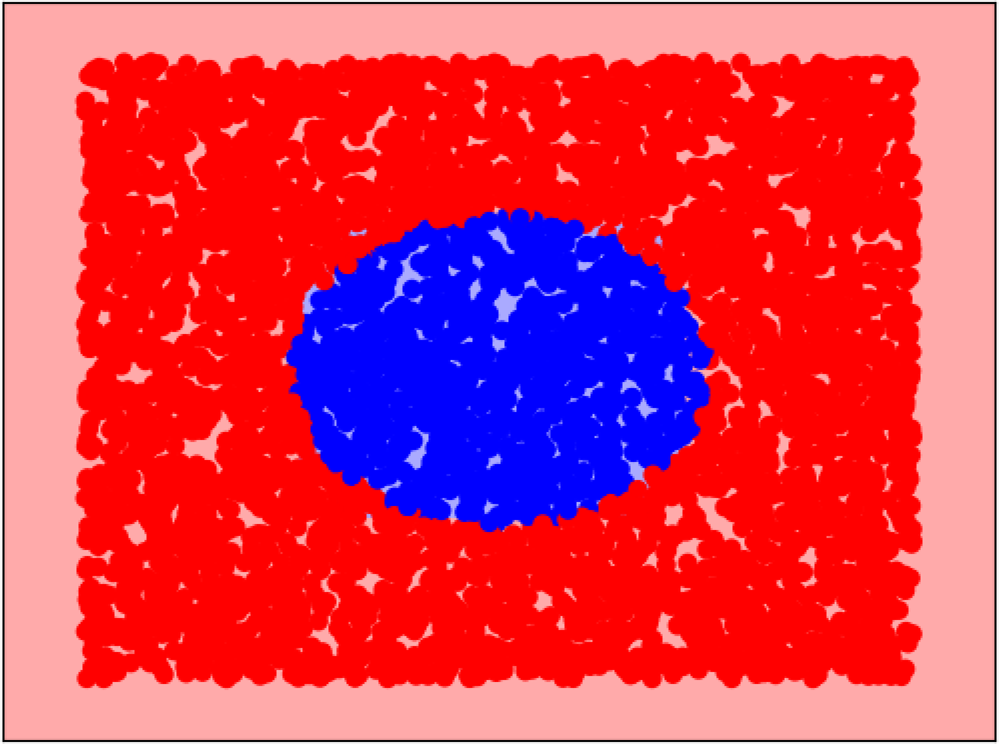

Decision boundaries with noise = 0.4
num classifiers = 5
num classifiers = 10
num classifiers = 50
num classifiers = 100
num classifiers = 200
num classifiers = 500

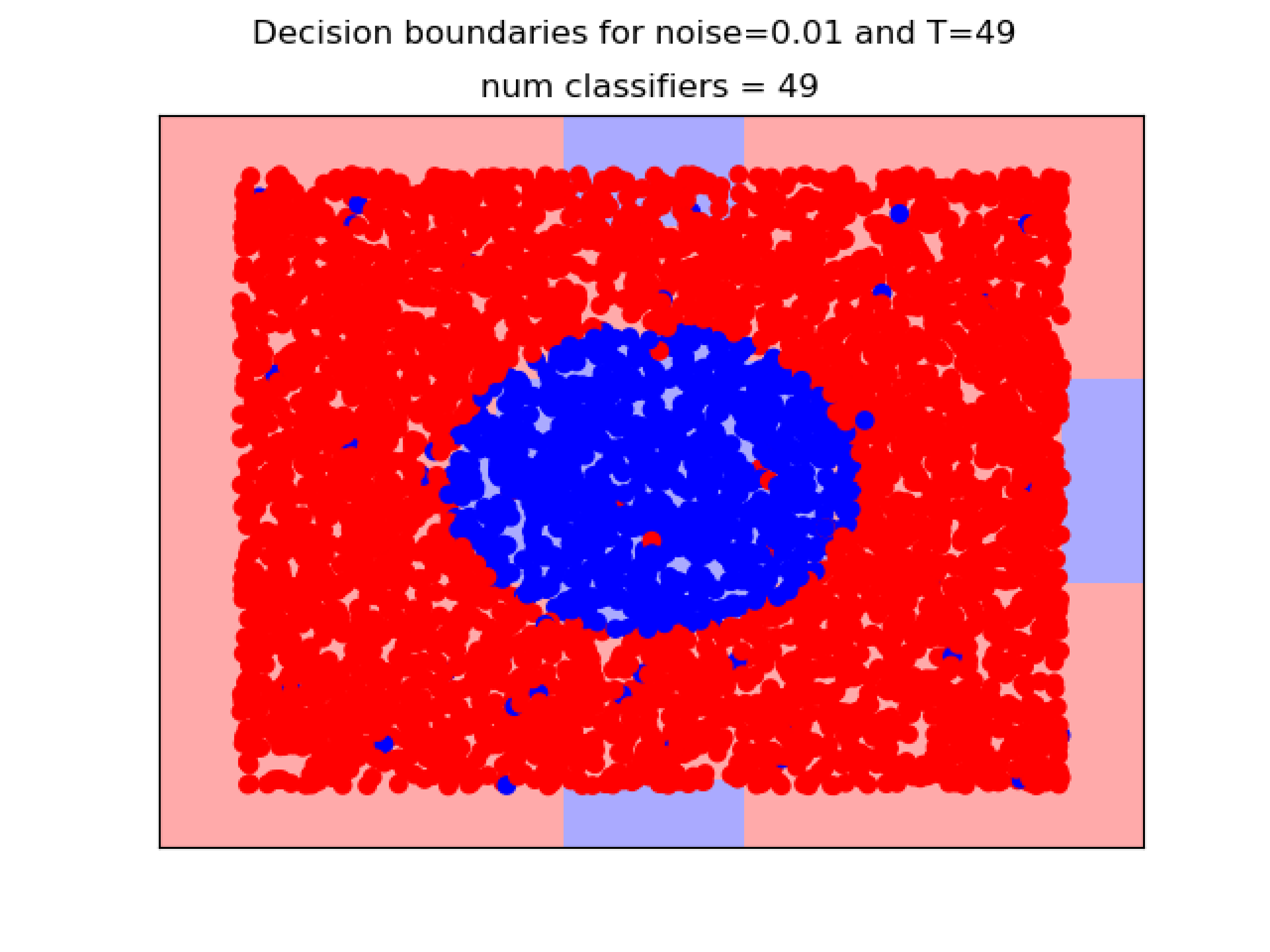
<u>שאלה 12</u>

עבור T noise = 0-כובע הוא בסביבות 15-20. השגיאה עלולה לרדת גם אחרי ערך זה אבל באופן זניח בלבד.

Decision boundaries for noise=0 and T=15

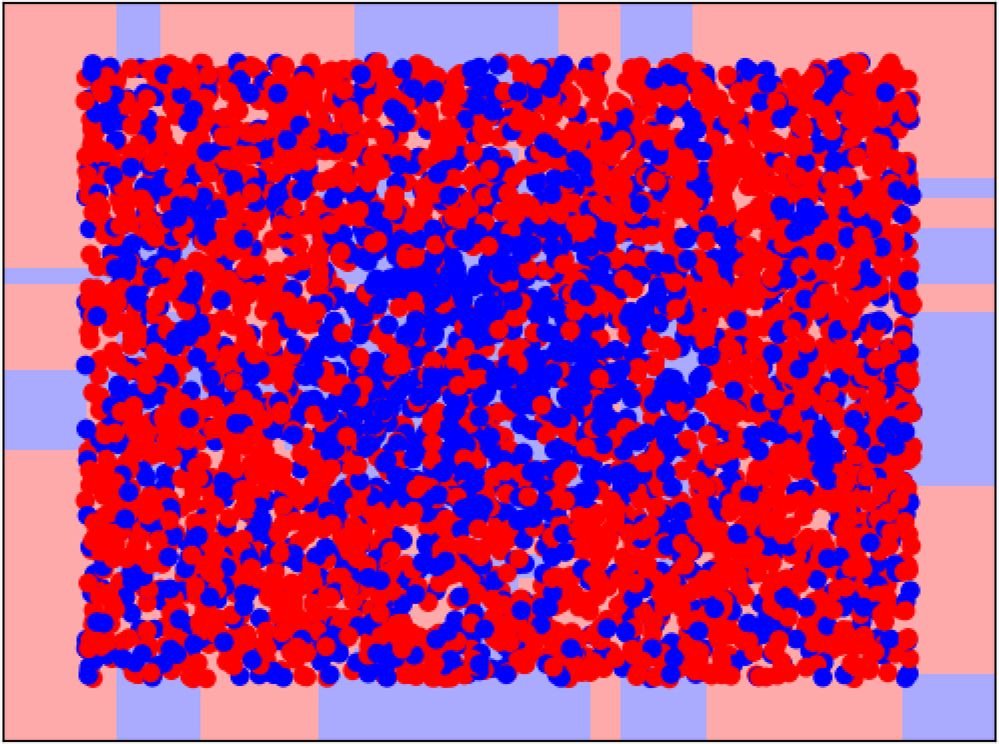
num classifiers = 15




Decision boundaries for noise=0.01 and T=49
num classifiers = 49

Decision boundaries for noise=0.4 and T=100

num classifiers = 100

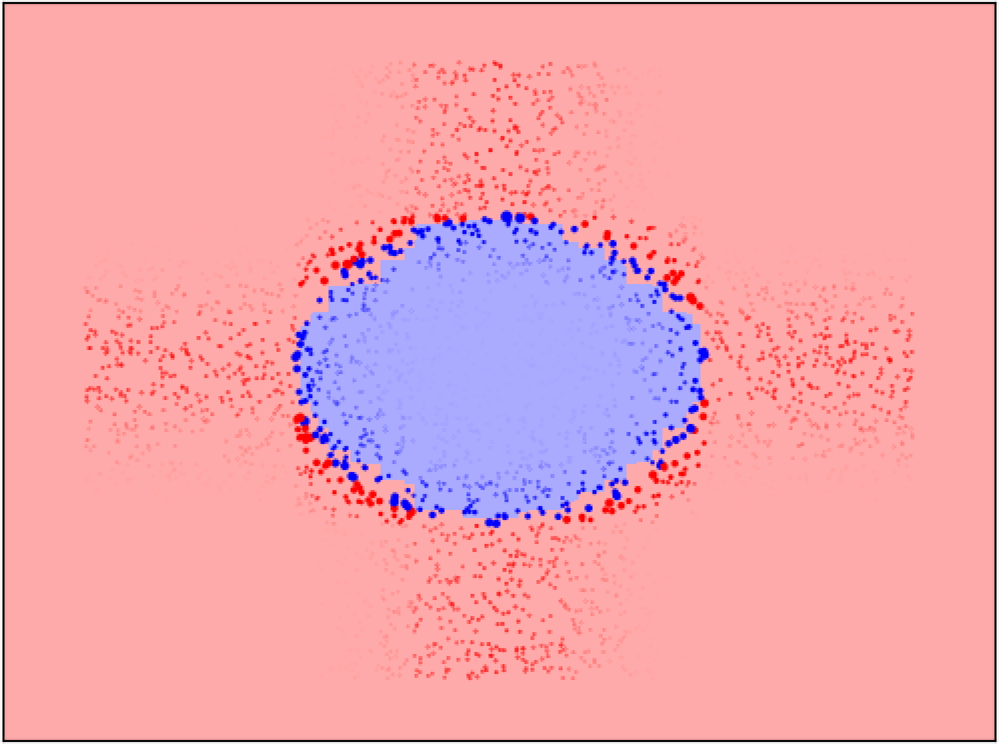


שאלה 13

לאחר נרמול המשקולות ניתן לראות בבירור את ההבדל בין נקודות בעלות משקל נמוך יותר לנקודות בעלות משקל גבוה יותר. הסיבה לכך היא פשוט שהמשקל לפני הנרמול היה קטן מדי בלי קשר ליחס בין המשקלים של נקודות שונות. הנרמול בסך הכל הזיז את המשקלים לטווח שניתן לראות באופן ברור יותר על גרף בלי להפר את היחס בין הנקודות.

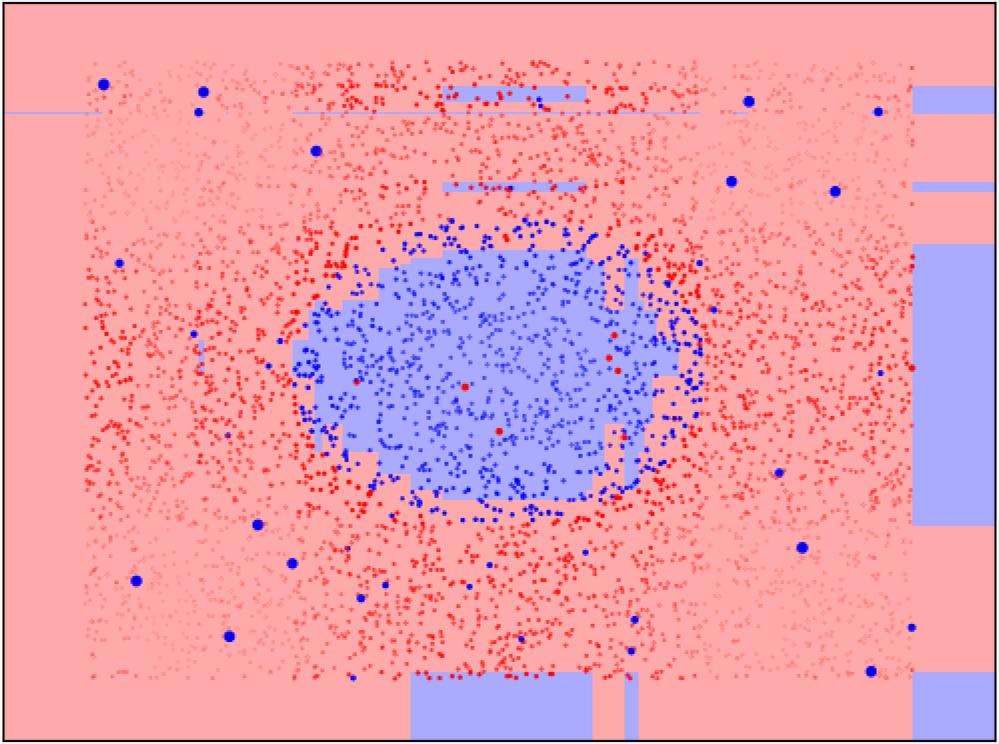
Size proprtional to weights with noise=0

num classifiers = 500



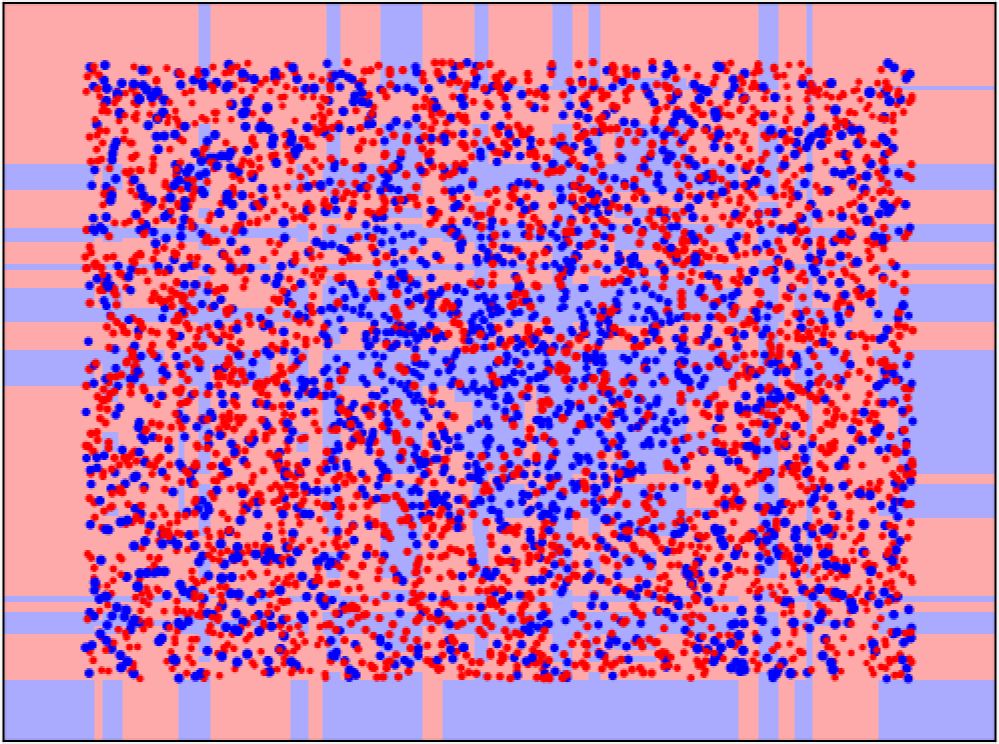
Size proprtional to weights with noise=0.01

num classifiers = 500



Size proprtional to weights with noise=0.4

num classifiers = 500



<u>שאלה 14</u>

* בשאלה 10 ניתן לראות שככל שהרעש עולה כך גם עולה הערך אליו מתכנסים ערכי השגיאה. עבור רעש 0.4 ניתן לראות כי פרקטית כלל אין אף תועלת כי השגיאה מתכנסת סביב 0.5, ככה שהיתרון זניח לעומת הטלת מטבע. לגבי ה**<u>Bias/Variance tradeoff</u>**: ניתן לראות שהשגיאה עולה ככל שהרעש עולה, כניתן לצפות. למעשה ה Boost לא עזר עם ויסות הרעש ומידל אותו יחד עם הנתונים שבאמת רצינו ללמוד, ומכך נובעת השגיאה. בנוסף נבחין כי השגיאה התחילה בערך נמוך יותר (עבור T=30\40), לכן ניתן לומר שה Variance בערכים אלו היה נמוך יותר, אבל ה Bias גבוה יותר.
* בשאלה 11 פשוט רואים את החלוקות שמבצעים הלומדים על המרחב עבור כמויות שונות של T. כמצופה ככל שהערך של T עולה הלומדים נעשים מדוייקים יותר (עד ל T-כובע משאלה 12) ולכן נעשת חלוקה ליותר אזורים, פרט למקרה בו הרעש שווה ל-0, שם הלומד מגיע לתוצאות טובות עבור T נמוכים יחסית.
* בשאלה 12 ניתן לראות כי הערך של T-כובע עולה ככל שהרעש עולה. בנוסף אם נתבונן בגרפים נראה שהלומדים מחלקים את המרחב ליותר אזורים עם עליית הרעש, זאת בגלל עליית ה Variance.
* בשאלה 13 עם עליית הרעש ניתן לראות יותר נקודות על הגרף, כלומר ככלל משקלן של הנקודות גבוה יותר. זאת מכיוון שהלומד לא מצליח לסווג באופן נכון יותר ויותר נקודות מכיוון שהוא לא מצליח להפריד את הרעש מהנתונים אותם אנחנו רוצים ללמוד. לכן כמות הנקודות הנמצאות בסיווג הלא נכון גדלה ולכן משקל הנקודות ככלל גדל.