תרגיל 4 | IML (67329)

שם: רונאל חרדים | ת"ז:208917641

שאלה 1

נוכיח את הדרוש ע"י שימוש בא"ש מרקוב:מאי השוויון נובע כי

$$P_{S\sim D^m}L_D(A(S)\leq\varepsilon)=1-P_{S\sim D^m}L_D(A(S)\geq\varepsilon)\geq 1-\frac{E_{S\sim D^m}\left(L_D(A(S))\right)}{\varepsilon}$$

מכיוון ש

$$\lim_{m\to\infty}\frac{E_{S\sim D^m}\left(L_D(A(S))\right)}{\varepsilon}=0$$

אזי

$$\lim_{m\to\infty}P_{S\dots D^m}L_D(A(S)\leq\varepsilon)\geq 1$$

ולכן קיימת m כך ש

$$P_{S\sim D^m}L_D(A(S)\leq\varepsilon)$$

ששואפת ל 1 , וזה שווה ל

$$P_{S\sim D^m}\left(L_D(A(S)\leq\varepsilon\right)\geq 1-\delta,$$

כנדרש.

שאלה 2

נשתמש באלגוריתם הלומד הבא:

בהינתן סט $X\in R^2$ נחזיר מעגל כך שבמרכזו נמצאות כל הדגימות ששוות ל 1, ומה שמחוץ למעגל שוו ל 0 .

אם לא קיים מעגל כזה תחזיר קבוצה ריקה.

בנוסף בשביל cov נגדיר טבעת עם רדיוס פנימי r_1 ורדיוס חיצוני $R(r_1,r_2)=r_2$. עבור מעגל C הרדיוס שלו יסומן ב r_C.

מכיוון שהמעגל התואם הקרוב ביותר - C' תמיד מוכל במעגל C השגיאה יכולה להגיע רק משגיאות האמת ששייכות ל $R(r_{C'},r_C)$. ואם אנו יודעים כי המשקל של D הוא קטן מ ε, אזי המשקל של הטבעת הוא בדיוק ε הוא $R(r_{C'},r_{C''})$ עבור מעגל C'' כלשהו.

כעת נשים לב כי:

מתקיים כי משקל הטבעת $R(r_{C'}, r_C)$. הוא בדיוק ε אמ"מ $C' \subseteq C$.

וגם $C \subseteq C'$ אמ"מ אין נקודות מהסט S בתוך $R(r_{C'}, r_C)$.

אם נקודה p כלשהי נמצאת ב $R(r_{C'}, r_C)$. אזי האלגוריתן יכלול את p ב C'.

ההסתברות לפספס דגימה ב $R(r_{C'}, r_C)$. היא $1-\varepsilon$ ועבור m דגימות $(1-\varepsilon)^m$.

מכאן נובע כי אם נבחר m כזו כך ש $(1-\varepsilon)^m \leq \delta$ אזי בהסתברות $1-\delta$ עבור m דגימות רנדומליות, גודל הטעות יהיה לכל היותר ε.

כעת מכיוון ש $(1-\varepsilon)^m \leq e^{-m\varepsilon}$ נבחר $\delta \geq e^{-m\varepsilon}$ ויתקיים $m_H(\varepsilon, \delta) \leq \frac{\log(1/\delta)}{\varepsilon}$ כנדרש.

שאלה 3

נראה כי מימד VC של המחלקה שווה ל n.

קים תת סט $C \subseteq X$ בגודל n שמחולק לשתי מחלקות. נגדיר את C להיות $C = \{e_1, ...e_n\}$. בהינתן n תוויות $y \in \{0,1\}^n$, $h_{[n]}$ יכולה ליצור את כולן מכיוון שכשאנו סוכמים את תתי הסטים ב $\sum_{i\in[n]} x_i$ אנו מקבלים את כל הפרמוטציות של $\{0,1\}^n$. במילים אחרות מההגדרה של h_Iושימוש בסינגלטון $I = \{i\}$ עבור $i \in [n]$ מתקבל $\left\{ \begin{array}{c} 1, x_i = 1 \\ 0, x_i = 0 \end{array} \right\}$ $h_I =$ ומאיחוד סינגלטון נקבל כי $\bigcup_{i=1}^{k}\{i\} = [k]$. התווית שתחזור לנו היא פרמוטציה של k 1-ים ו $n-k$ אפסים. כך שהאינדקס $i = 1$ אם סינגלטון $\{i\}$ הוא חלק מהאיחוד:

$$X = \{0,1\}^n$$

$$C = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\} \subset X$$

$$I \subseteq \{1, 2, \ldots, n\} := [n]$$

$h_I(x \in C) = \left(\sum_{i\in I} x_i\right) \bmod 2$ מייצרת את התוויות הבאות:

$$I = \emptyset : h_\emptyset \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = h_\emptyset \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \cdots h_\emptyset \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = 0 \rightarrow \text{ label } 00\ldots0$$

$$I = \{1\} : h_{\{1\}} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 1, h_{\{1\}} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} = 0 \ldots h_\emptyset \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = 0 \rightarrow \text{ label } 10\ldots0$$

$$I = \{n\} : h_{\{1\}} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0, h_{\{1\}} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} = 0 \ldots h_\emptyset \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = 1 \rightarrow \text{ label } 00\ldots1$$

ועבור כל i שאינו סינגלטון הוא קומבינציה של סטים של תוויות של סינגלטונס.

שאלה 4

נראה כי מימד VC של המחלקה שווה ל $2k$ ואם k אינו מוגבל אזי המימד אינסופי.

תחילה נראה כי $VCdim\left(H_{k\ \textbf{intervals}}\right) \geq 2k$: כל תוית מגודל $2k$ יכולה להיווצר תוך k איטרציות, המקרה הגרוע עבור תווית עבור מספר מרווחים שבהם אנו צריכים להשתמש,

זה כאשר אין לנו 1. מכיוון שעבור כל חזרה של 1 אנו יכולים להשתמש במרווח בכדי לכלול את כולם.א"כ עבור k ־1ים לא סמוכים צריכים להיות לנו k מרווחים, יתרה מזאת תווית בגודל $2k$ יש לכל היותר k ־1ים לא בסמוכים, כי בין כל שני אחדים יהיה לפחות 0 אחד.

נראה כי $VCdim\left(H_{k\ \textbf{intervals}}\right) \leq 2k$: נניח בשלילה כי קיימת קבוצה בגודל $2k+1$ שיכולה לנתץ את המחלקה. א"כ אנו יכולים לייצר תווית עם אחדים לא צמודים ואפסים לא צמודים. בסתירה לכך ש $k+1$ אחדים לא צמודים יכולים להיווצר רק ע"י $k+1$מרווחים.

הוכחנו שני צרדיים כל א"ש ולכן מתקיים השוויון הנדרש.

שאלה 5

נוכיח כי $m_H\left(\varepsilon_1, \delta\right) \geq m_H\left(\varepsilon_2, \delta\right)$ $\varepsilon_1 \leq \varepsilon_2 \rightarrow$: נניח בשלילה כי $\varepsilon_1 \leq \varepsilon_2$ אבל $m_H\left(\varepsilon_1, \delta\right) < m_H\left(\varepsilon_2, \delta\right)$, מההגדרה $m_H\left(\varepsilon_2, \delta\right)$ היא המספר המינימלי של הדגימות בכדי שנוכל להשיג את $D^m\left(\left\{S : L_{D,f}\left(h_S\right) \leq \varepsilon\right\}\right) \geq 1 - \delta_2$ אך מההנחה קיבלנו ש $D^m\left(\left\{S : L_{D,f}\left(h_S\right) \leq \varepsilon\right\}\right) \geq 1 - \delta_1 \geq 1 - \delta_2$ נשתמש ב $m_H\left(\varepsilon, \delta_1\right) < m_H\left(\varepsilon, \delta_2\right)$ במילים אחרות ־ השתמשנו בפחות דגימות בכדי להדיג $accuracy$ טובה יותר בסתירה.

שאלה 6

נגדיר $\text{VCdim}\,(H_i) = d_i \text{ for } i \in \{1,2\}$. נניח בשלילה כי $d_1 > d_2$, לכן קיימת קבוצה $C = \{c_1, c_{d_1}\}$ שמנותצת ע"י H_1 אך לא ע"י H_2'כלומר קיים $h_1 \in H_2$ כך שבהינתן C מייצרת כל תווית. מכיוון ש $H_1 \subseteq H_2$אזי $h_1 \in H_2$ ולכן גם H_2 מנתצת את C . בסתירה לכך שמימד VC שלה קטן ממימד של H_1. לכן קיבלנו

$$H_1 \subseteq H_2 \Rightarrow \text{VCdim}\,(H_1) \leq \text{VCdim}\,(H_2)$$

שאלה 7

H מתכנסת באופן אחיד עם

$$m_H^{UC} : (0,1)^2 \to \mathbb{N} \Rightarrow \forall(\varepsilon, \delta) \in (0,1) \forall D \text{ on } X \times Y$$

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

מכיוון שטענה זו נכונה לכל $1 < \varepsilon < 1$ נבחר $\varepsilon = \frac{\varepsilon}{2}$ ולכן

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

כעת, מכיוון ש S הוא $\varepsilon/2$ representative אנו יודעים ש

$$L_D\left(h_S\right) \leq \min_{h \in H} L_D(h) + \varepsilon, \text{ where } h_S = \text{argmin}\, L_{S \in H}(h)$$

ומתקיים

$$D^m\left(\left\{S : L_D\left(h_S\right) \leq \min_{h' \in H} L_d(h) + \varepsilon\right\}\right) \geq 1 - \delta$$

שזאת ההגדרה של $Agnostic - PAC$ כנדרש.

שאלה 8

טענה: H **אינה למידה** $Agnostic - PAC$.

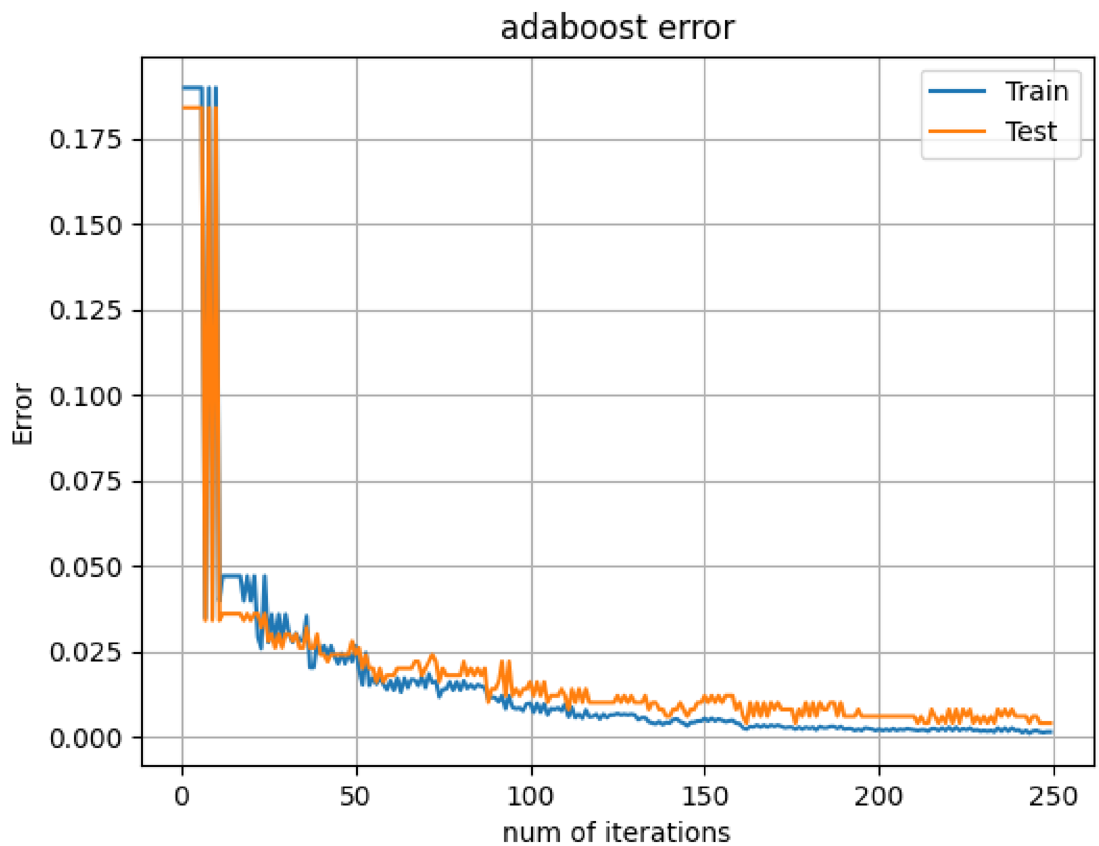
לפי הטענה מתקיים כי A תלויה ב D, למרות שזה לא לפי הגדרת למידות PAC . התלות הזו מציעה כי כל מחלקת היפותזות מספקת את התנאי.

כדוגמה נגדית נגדיר מחלקת היפותזות H להיות מחלקת כל הפונקציות ממרחב X אינסופי, ראינו בכיצה כי היא לא למידת PAC. יתרה מזאת ראינו בכיה כי H למידה PAC אמ"מ היא למידה $Agnostic - PAC$, מכיוון שהמחלקה הזאת אינה למידה PAC היא גם לא למידה $Agnostic - PAC$ אפילו שהיא מספקת את הטענה.

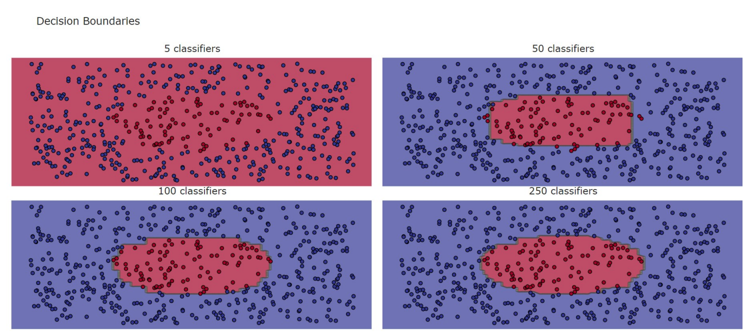
חלק I
$Practical$:

שאלה 1

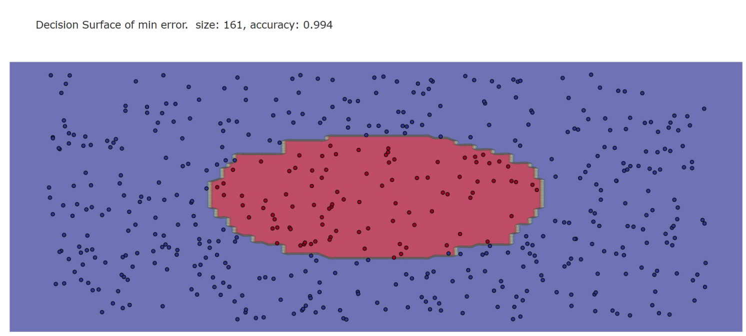
ניתן לראות כי ככל שאנו לוקחים בחשבון יותר מסווגים התוצאה שלנו משיגה $loss$ קטן יותר.בנוסף סט האימון משיג תוצאה טובה יותר כי הלומד ראה כבר את הנקודות האלו וסיווג לפיהן.


adaboost error
Train
Test
Error
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
50
100
150
200
250
num of iterations

שאלה 2


Decision Boundaries
5 classifiers
50 classifiers
100 classifiers
250 classifiers

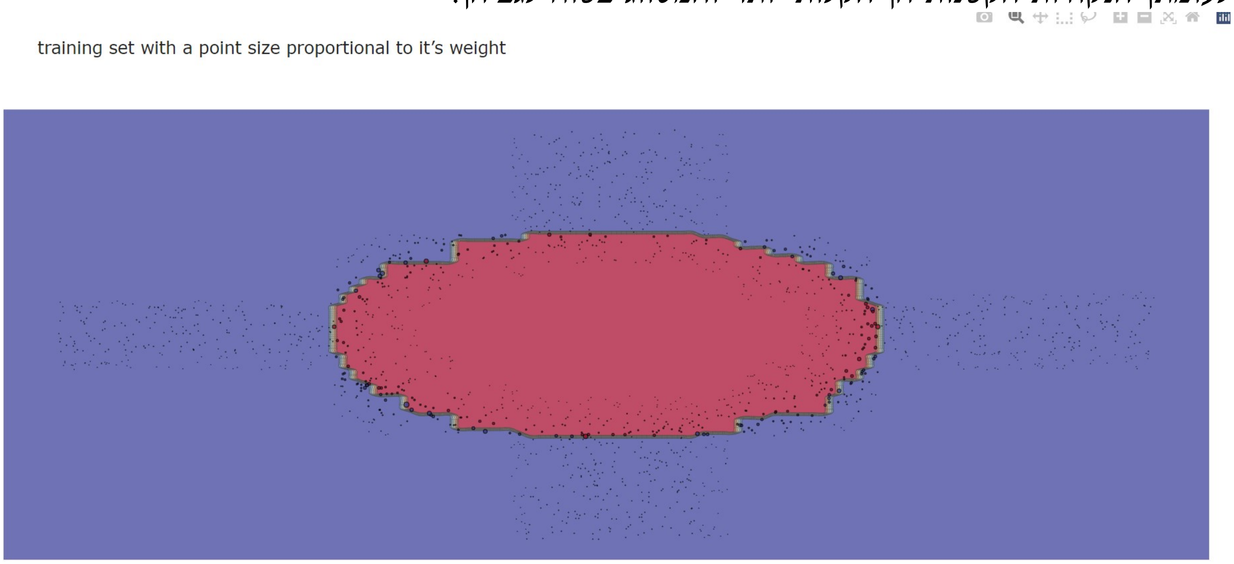
שאלה 3


Decision Surface of min error. size: 161, accuracy: 0.994

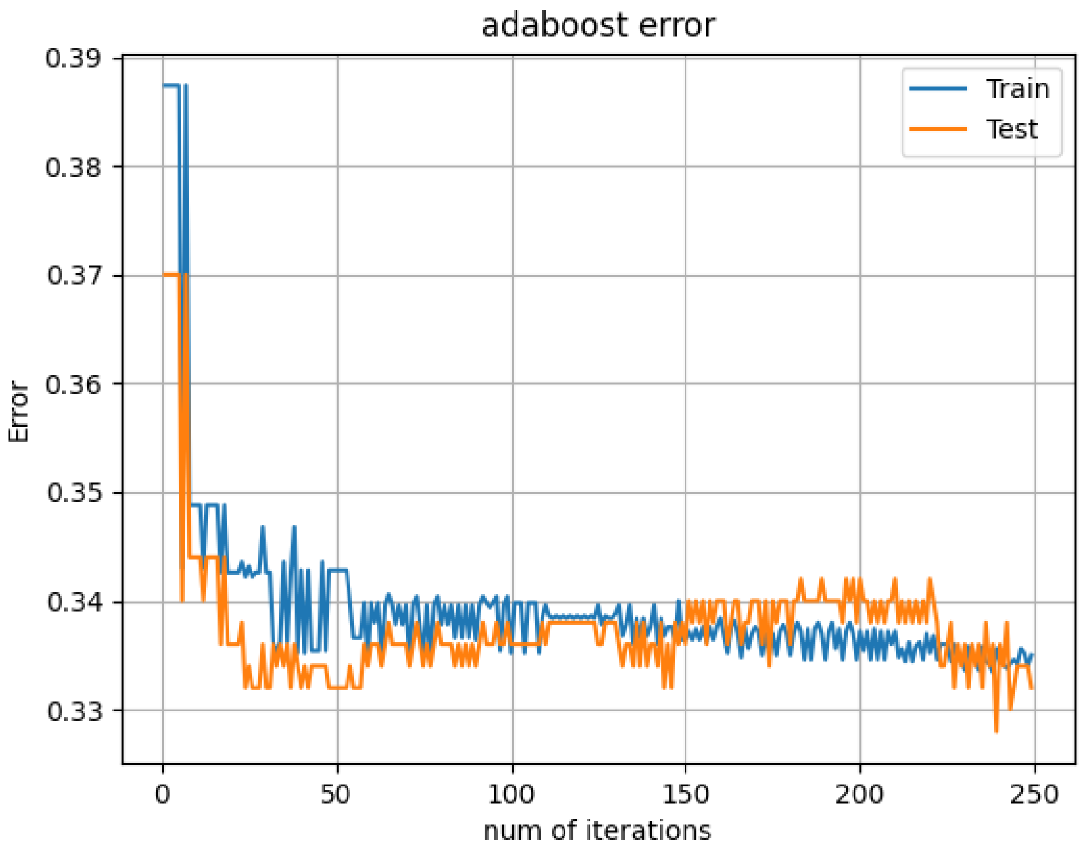
הטעות המינימלית מתקבלת כאשר מספר האיטרציות שווה ל 161 והיא שווה 0.04

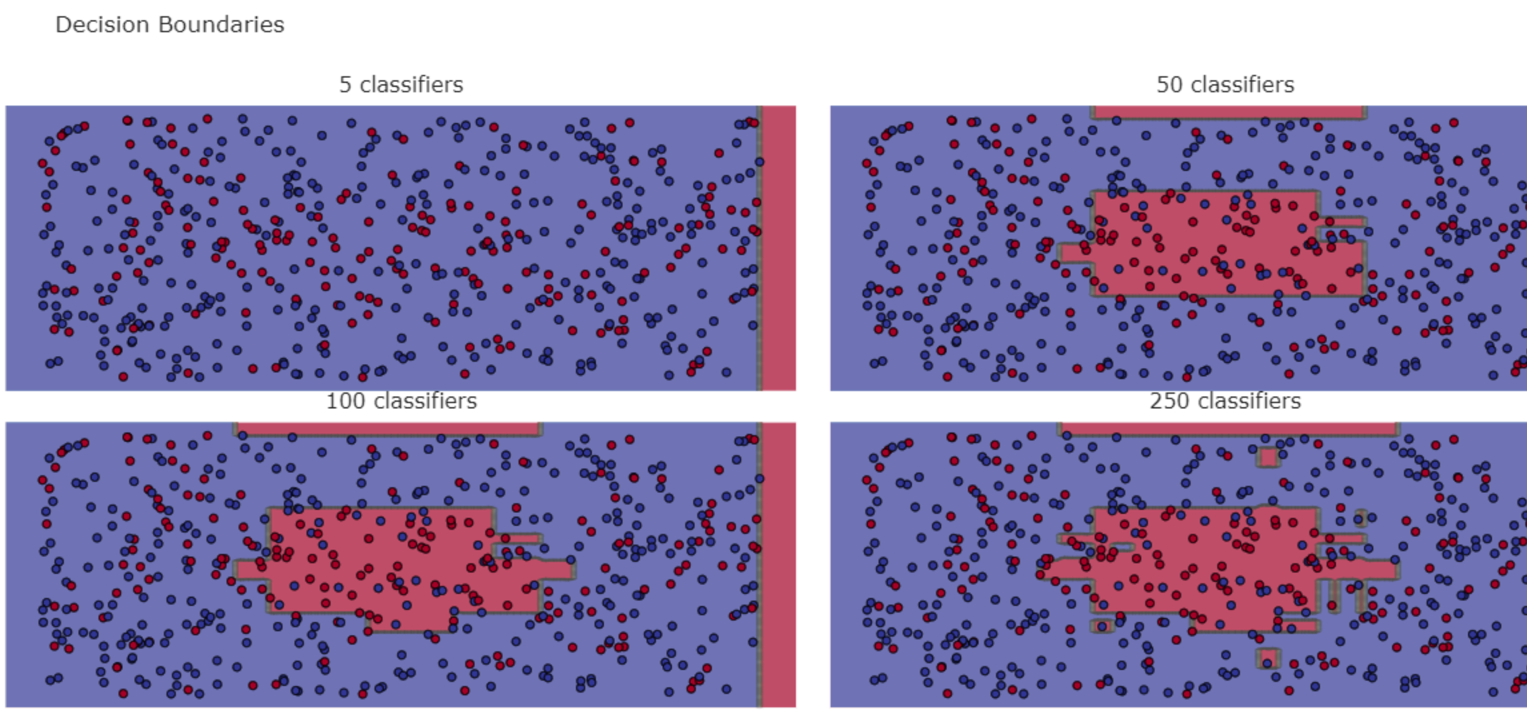
שאלה 4

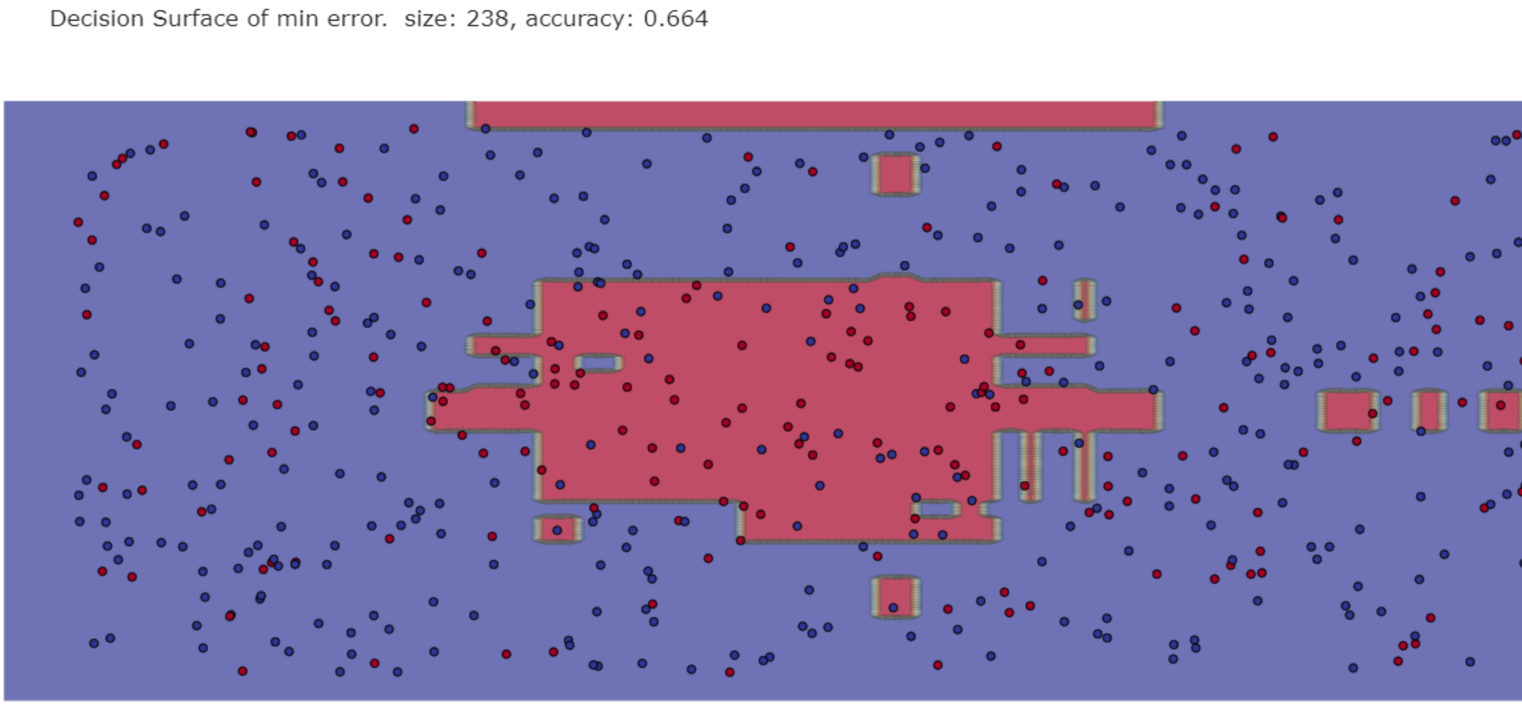
אנו רואים כי הנקודות שטעינו עליהן קיבלו משקל גבוה יותר לפעם הבאה ולכן הנקודות הגדולות הן הקשות יותר לסיווג, לעומתן הנקודות הקטנות הן הקלות יותר והמסווג בטוח לגביהן.

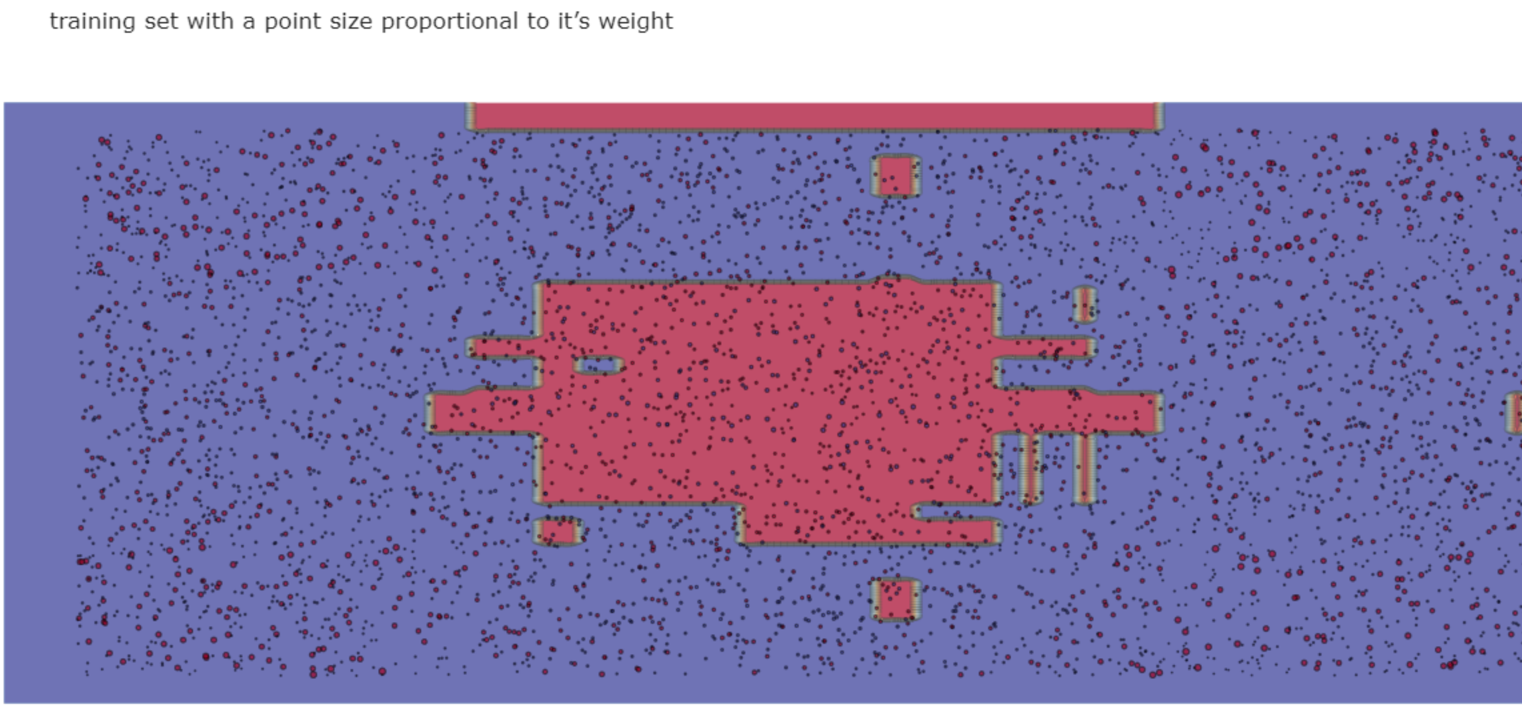


שאלה 5


adaboost error
Error
num of iterations
Train
Test


Decision Boundaries
5 classifiers
50 classifiers
100 classifiers
250 classifiers





נשים לב כי מכיוון שיש רעש הלומד "רודף" אחרי הדאטה נגרם $overfit$ ולכן בשלב מסויים ככל שאנו מוספים מסווגים ה $loss$ רק עולה.