Theoretical part

2.1 ① נסתכל על אלגוריתם A - ERM אשר בהינתן סט אימון מחזיר את המעגל h_S הכי "הדוק" אשר מכיל את כל הדוגמאות החיוביות. כלומר את הרדיוס של h_S כ-r_S. נסמן רדיוס ונגדיר h^* להיות המעגל המטרה ס מרכזו, ורדיוס להיות r^*.

יהי $\varepsilon>0$, $r<r^*$. $D_x(\{x: r\le\|x\|\le r^*\})=\varepsilon$

נגדיר $E=\{x\in\mathbb{R}^2 \mid r\le\|x\|\le r^*\}$.

ההסתברות ש- $L_D(h_S)\ge\varepsilon$ חסומה ע"י ההסתברות שאף נקודה מ-S שייכת ל-E

ולכן ההסתברות לאירוע הנ"ל היא:

$$P_{S\sim D^m}[L_D(h_S)\ge\varepsilon]\le(1-\varepsilon)^m\le e^{-\varepsilon m}$$

כעת נבחר m כך ש $e^{-\varepsilon m}\le\delta$ ונקבל $P_{S\sim D^m}[L_D(h_S)<\varepsilon]\ge 1-\delta$

כנדרש.

2.2 ② נניח ש- $H_1\subseteq H_2$ לכל קבוצה כך $C=\{c_1,\dots,c_m\}$ $C\subseteq X$

אם C מנופצת ע"י H_1 אזי היא גם מנופצת ע"י H_2 ולכן

$VC\text{-}Dim(H_1)\le VC\text{-}Dim(H_2)$

③ נראה שהמימד של VC-Dim עבור H_{con}

יהי d, מספיק למצוא קבוצה בגודל d ש-H_{con} מנפצת אותה

נציג אלגוריתם לבנייה

נבחר את הוקטור הבסיסי ה-i שמסווג כחיובי אם ורק אם x_i מופיע בו

ביטוי $x_i \wedge \bar{x}_i$ = false

נשים לב כי כאשר $|X|<d$ אז ... לכל $1\le i\le d$ אם $x_i=0$ נבחר $\bar{x}_i$ ואם $x_i=1$ נבחר x_i

ובנוסף ... 0 נבחר ... $VC\text{-}dim=d$

4) נרצה להראות שסיבוכיות של הדגימות מונוטונית יורדת בתלות ב-ϵ

יהיו D פונקציית התפלגות על X ו- $f\in H$ הפונקציה המטרה

נבחר A כלשהו לומד עם סיבוכיות הדגימות $m_H(\cdot,\cdot)$ ויהיו $\delta\in(0,1)$

נניח כי $0<\epsilon_1\le\epsilon_2\le 1$

נניח התפלגות אחידה על ה train set בגודל $m_1\ge m_2$ נקבל כי בהסתברות $1-\delta$

לומד A יחזיר לנו $h\in H$ כך ש $L_{D,f}(h)\le\epsilon_1\le\epsilon_2$ ולכן בפרט $m_1\ge m_2$

ולכן הסיבוכיות $m_1=m_H(\epsilon_1,\delta)\ge m_2=m_H(\epsilon_2,\delta)$

כלומר מונוטונית ~~יורדת~~ לפי δ ההוכחה זהה עם הבדל כי נקבע ϵ במקום δ

5) H היא לא אגנוסטית PAC למידה

ע״פ ההגדרה צריך אלגוריתם לומד של H לכל התפלגות D של דגימות H

שיוכל לבצע למידה אגנוסטית PAC.

א' $A_D=\arg\min L_D(h)$

נסתכל על $H=\{h \mid \text{כל הפונקציות הבינאריות ב-}X\}$

הראינו בתרגול כי היא לא למידה PAC

נראה כי לא PAC ← לא אגנוסטית PAC

לכן, קיים אלגוריתם לומד A עבור H וסיבוכיות דגימה $m_H(\epsilon,\delta)$ כך שלכל

$\epsilon,\delta\in(0,1)$ ולכל התפלגות D על Z אם $S\sim D^m$ וגם $m\ge m_H(\epsilon,\delta)$

אז $P\left(A(S)\le\min_{h\in H}L_D(h)+\epsilon\right)\ge 1-\delta$

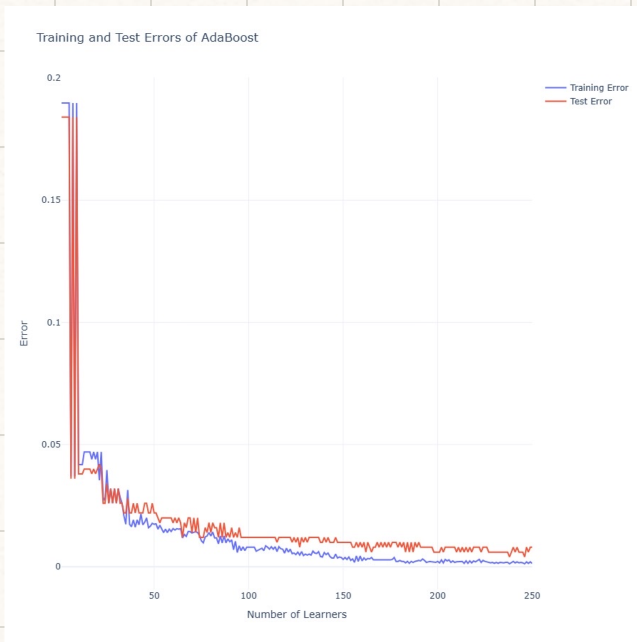
נראה כי כל התפלגות D כוללת אגנוסטית לפי H מתאימה ויכולה להיות מוגדרת

כלומר D_x על X וכל בחירה של $f\in H$ ולכן נקבל כי H למידה PAC וזו סתירה לנתון

לכן PAC ← אגנוסטית PAC

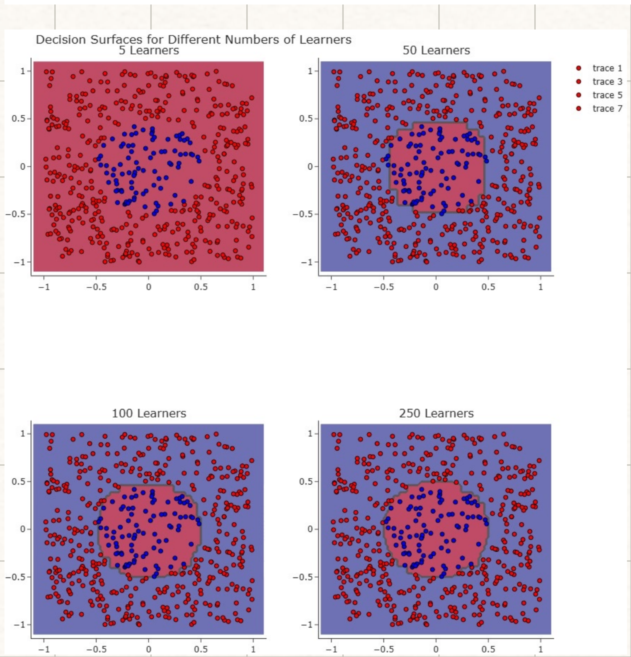
ולכן ה-H לא אגנוסטית PAC

practical part

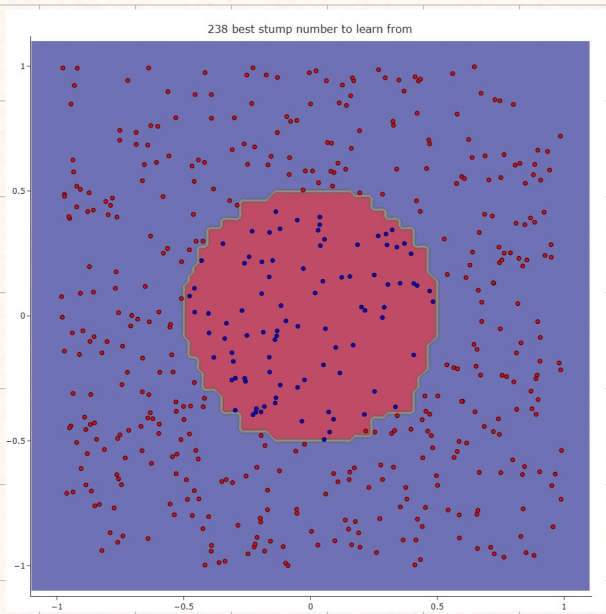


3.1

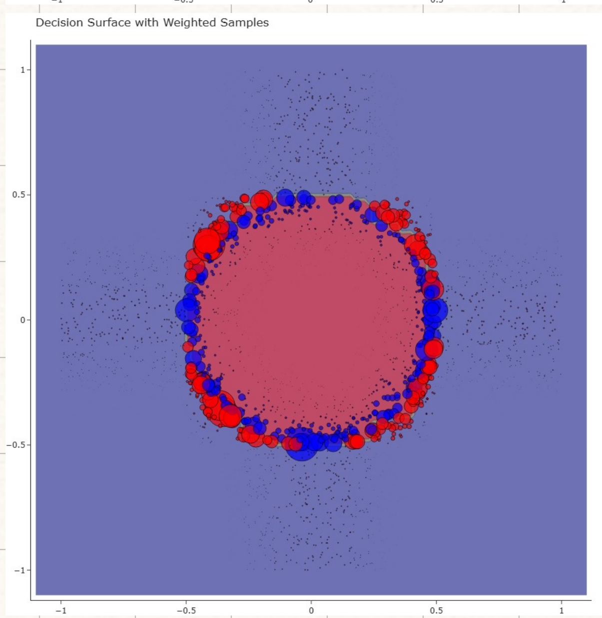
1. ניתן לראות מהתוצאות שקודם כל
הטעות יורדת ככל שמוסיפים מאמנים ביותר
לומדים שלא בדיוק מה שאמור מבחין
התכונות, ושינוי, שאין בה סבר פרמטרים,
כי זה יותר כמו הכללות.



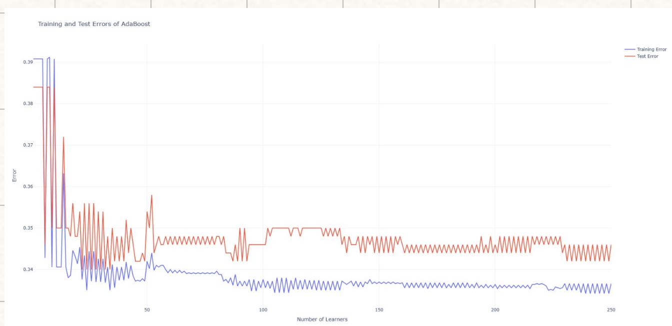
2. בכ־ 2 לומדים שלא רואה לחלוטין
שהמידע שלנו מוכר למציאות ולא מקרי
והמודל יוכל ויכול לכל שינוי שמוסיפים יותר
לומדים המידע עדיין יותר ברור וככה
כל לומד אינו יכולים להשאר בדרך מוכר
ליניארית וכך כך סגור כל התכונות האפשרויות
של לומדים, עד לשם המידע מסווג כבר בסוג
יותר



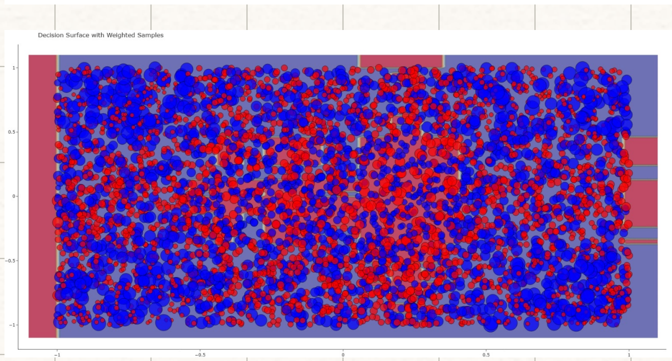
3 ניתן לראות שלאחר מספר הלומד מאתיים
נקבל את הצורה הכי נכונה היינו 238



4 כאשר אנו מסתכלים על המשקל הגדול ביותר
מתוקנים שהצורה המעגלית ככל הנראה מאוד.
זה מעיד לנו לראות שהדוגמאות הן על הגבול
ולמודל היה יותר קשה. הדוגמאות שלא בתוך
המעגל הן יותר קלות לסווג, הדוגמאות האלו
אשר נמצאות הן יותר קשות לסווג. ולבסוף
הדוגמאות אשר נמצאות מחוץ להמעגל של המעגל
הן הקשות לסווג, כמובן הגבול של
ה-y וה-x.



5 הוספת רעש לדאטה גורמת להפרעה
למודל של המודל, כפי שציפינו.
הצורה של הסוף לא בולטת כפי שראינו וההפרש
של ה-error, ככל שמתקדמים הבייס לא
משתנה של המודל, ולכן יהיה אותו דבר)
ולכן הסוף של הכי טוב הוא לא בסוף.



אין הפרדה ברורה בין שני הקלאסים,
לפי מה שנראה כי קשה להבדיל הדוגמאות ללמוד
וניתן לראות שאין לפי גבול הגדרה
והמודל לא מצליח לזהות את מה שאמרנו
למעלה.