[illegible] גרשוני

308350503

עקרונות למידה - תרגיל בית 4

1) נוכיח זאת ע"י הוכחת הגרירה בעזרת הערת השאלה, נתחיל בכך שנניח כי (b) מתקיים ונוכיח את (a):

מאי-שוויון מרקוב מתקיים $\underset{S\sim D^m}{\mathbb{P}}\left(L_D(A(S))>\varepsilon\right)\le \mathbb{E}\left(L_D(A(S))\right)/\varepsilon$.

מכיוון והנחנו את (b) אז מתקיים:

$$\underset{S\sim D^m}{\mathbb{P}}\left(L_D(A(S))>\varepsilon\right)=0$$

זאת מכיוון שהסתברות מוגדרת בין 0 ל-1 בלבד.

נרצה להפוך את אי-השוויון בתוך ההסתברות ונקבל:

$$\underset{S\sim D^m}{\mathbb{P}}\left(L_D(A(S))\le\varepsilon\right)=1$$

כעת, אנו יודעים כי $0<\delta<1$ ולכן $1-\delta\le 1$ ונקבל:

$$\underset{S\sim D^m}{\mathbb{P}}\left(L_D(A(S))\le\varepsilon\right)\ge 1-\delta$$

כנדרש.

נניח כעת את (a) ונוכיח את (b):

מהנחת (a) אנו יודעים כי לכל $\varepsilon,\delta\in(0,1)$ קיים $m(\varepsilon,\delta)$

כך שלכל $m(\varepsilon,\delta)\le m$ מתקיים $\underset{S\sim D^m}{\mathbb{P}}\left(L_D(A(S))\le\varepsilon\right)\ge 1-\delta$

ומכך שיכולנו לבחור כל ε,δ בקטע $(0,1)$ נוכל לבחור את $\delta=\varepsilon=\frac{1}{m}$ ולכן מתקיים:

$$\underset{S\sim D^m}{\mathbb{E}}\left(L_D(A(S))\right)=\int_{-\infty}^{\infty}x\underset{S\sim D^m}{\mathbb{P}}\left(L_D(A(S))=x\right)dx$$

כעת, נוסיף לכך ש-$L_D(A(S))$ היא פונקציה רציפה המוגדרת בין 0 ל-1 ולכן נוכל לתחום את האינטגרל לקטע הזה:

$$=\int_{0}^{1}x\underset{S\sim D^m}{\mathbb{P}}\left(L_D(A(S))=x\right)dx$$

↓

נפצל את האינטגרל על ε הנקודה:

$$= \int_0^{\varepsilon} x \, \mathbb{P}_{S\sim D^m}\left(L_D(A(S)) = x\right) dx + \int_{\varepsilon}^{1} x \, \mathbb{P}_{S\sim D^m}\left(L_D(A(S)) = x\right) dx$$

$$\leq \int_0^{\varepsilon} x \, \mathbb{P}_{S\sim D^m}\left(L_D(A(S)) \leq \varepsilon\right) dx + \int_{\varepsilon}^{1} x \, \mathbb{P}_{S\sim D^m}\left(L_D(A(S)) > \varepsilon\right) dx$$

$$= \int_0^{\varepsilon} x \, \mathbb{P}_{S\sim D^m}\left(L_D(A(S)) \leq \varepsilon\right) dx + \int_{\varepsilon}^{1} x \left(1 - \mathbb{P}_{S\sim D^m}\left(L_D(A(S)) \leq \varepsilon\right)\right) dx$$

נציב כעת את הנתון (a) ונקבל:

$$\leq \int_0^{\varepsilon} x (1-\delta) \, dx + \int_{\varepsilon}^{1} x (1-(1-\delta)) \, dx$$

נחשב את האינטגרל ונזכור כי הסתברות חסומה מלמעלה ע"י 1:

$$\leq \varepsilon + \delta - \varepsilon\delta \leq \varepsilon + \delta$$

כעת נעבור על הגבול של התוחלת כאשר $m \to \infty$:

$$\lim_{m\to\infty} \mathbb{E}_{S\sim D^m}\left(L_D(A(S))\right) \leq \lim_{m\to\infty} (\varepsilon + \delta) = \lim_{m\to\infty} \left(\frac{1}{m} + \frac{1}{m}\right) = 0$$

הגדרת
ε, δ

נוסיף על-כך שמדובר בתוחלת של משתנה מקרי אי-שלילי ולכן תוחלתו חסומה מלמטה ע"י 0 ומכלל הסנדוויץ' נקבל את הדרוש:

$$\lim_{m\to\infty} \mathbb{E}_{S\sim D^m}\left(L_D(A(S))\right) = 0$$

הראנו איך כל הגדרה גוררת את השנייה ולכן שקולות.

2) ראשית, נשים לב כי קבוצת ההיפותזות הינה בגודל אינסופי שכן ניתן להגדיר אינסוף עיגולים עם רדיוס $r\in\mathbb{R}_+$.

בהינתן m דגימות אקראיות ניתן להגדיר עיגול חדש עם רדיוס r' שייתן לנו את ההתאמה המדויקת ביותר לדגימות החיוביות. כלומר, העיגול בעל השטח הקטן ביותר שמכיל את כל הדגימות החיוביות ואפס מן הדגימות השליליות:

$$r' = \inf\{t \mid \mathbb{P}(t\le\|x\|\le r<\varepsilon\}$$

אם לא קיימות דגימות חיוביות אזי $r'=\phi$.

כעת, עלינו להראות כי לכל $0<\varepsilon,\delta<1$ קיים $m=m_H(\varepsilon,\delta)$ כך שלכל התפלגות D מעל $\mathbb{R}^2$ ולכל עיגול ברדיוס r, מעל הסתברות של לפחות $(1-\delta)$ מעל סט אימון S בגודל m לפחות, פונקציית ההפסד $L_{D,r}(r')$ היא לכל היותר ε.

נשים לב כי לכל דוגמא חיובית מתוך r' ולכל דוגמא שלילית מחוץ ל-r, פונקציית ההפסד שווה לאפס ולכן נבחן אך רק את הדגימות החיוביות הנמצאות בטבעת הנוצרת בין r ל-r'

כלומר, $Area=\{x \mid r'\le\|x\|\le r\}$

ולכן ניתן לומר כי $\mathbb{P}(Area)\ge\varepsilon$.

ע"פ הגדרה זו, ההסתברות שדוגמא מתוך ההתפלגות D תפספס הינה $(1-\varepsilon)$ ולכן ההסתברות ש-m דוגמאות כאלה יתפספסו הינה $(1-\varepsilon)^m$ וזה נובע מההסתברות של חוק הכפל של מאורעות שונים.

אם-כך, נבחר m שיספק את אי-השוויון $(1-\varepsilon)^m\le\delta$ ועל ההסתברות $(1-\delta)$ מעל m דגימות אקראיות משקל השגיאה בתוך הטבעת $Area$ חסום ע"י ε ועם בעזרת אי-השוויון שנלמד, $1-x\le e^{-x}$ נקבל כי מתקיים:

$$m_H(\varepsilon,\delta)\le\frac{\log\left(\frac{1}{\delta}\right)}{\varepsilon}$$

וכיוון והנ"ל מתקיים נסיק כי H הינה PAC למידה, כנדרש.

3) ראשית, נשים לב כי מחלקת ההיפותזות המוגדרת בשאלה
הינה סופית, ולכן למידה PAC.
בהמשך לכך נראה כי גודלה של מחלקת ההיפותזות חסום
מלמעלה ע"י (3^d) שכן עבור כל משתנה מתוך d משתנים
יכול לקבל ערך אמת, שקר או כלל לא יהיה חלק מהביטוי
הבוליאני ולכן $|\mathcal{H}^d_{con}| \le 3^d$.
אי-לכך, נוכל לומר כי ע"י הוכחה לעיל $VCdim(\mathcal{H}_{con}) \le d\log 3$.
נראה כי מחלקת ההיפותזות מנפצת את קבוצת וקטורי היחידה
עבור $\{e_i \mid i \le d\}$ - ניבחן תיוג כל התצאות $(y_1,\dots,y_d)\in\{0,1\}^d$
ואת הקבוצת האינדקסים של הביטויים שלהם ערך 0 כלומר,

$$Zero=\{i \mid y_i=0\}$$

ניבחן בהיפותזה $h=\bigwedge_{i\in Zero}\bar{x}_i$ ונקבל כי לכל $i\in[d]$
מתקיים $h(e_i)=y_i$ ולכן ניתן להסיק כי $VCdim(\mathcal{H}_{con})\ge d$.
כעת, נותר להראות כי $VCdim(\mathcal{H}_{con})\le d$ ובכך נסיק כי
$VCdim(\mathcal{H}_{con})=d$, נעשה זאת ע"י הנחה בשלילה כי קיים
סט בגודל $d+1$: $A=\{a_1,\dots,a_{d+1}\}\subset X$ המנופץ ע"י $\mathcal{H}_{con}$.
אז קיים בפרט סט שנוכל לבחור היפותזות $h_1,\dots,h_{d+1}$ מתוך $\mathcal{H}$
כך שלכל $i,j\in[d+1]$:

$$h_i(a_j)=\begin{cases}0, & i=j\\ 1, & i\ne j\end{cases}$$

כלומר, הביטוי הבוליאני המתאים ל-h_i מכיל ליטרל המקבל
ערך שקר על a_i וערך אמת על a_j עבור כל $j\ne i$.
לכן, מעיקרון שובך היונים, מתוך $d+1$ ליטרלים קיימים לפחות
שניים המתאימים לאותו משתנה - נסמנם t_1,t_2.
כעת, אם $t_1=t_2$ אזי t_1 מקבל ערך אמת עבור a_1 שכן
t_2 מקבל ערך אמת על a_1 אך זאת בסתירה להנחה.
לעומת זאת, אם $t_1\ne t_2$ אז לא נוכל לקבל ערך אמת בשניהם
עבור a_3 ושוב הגענו לסתירה. $\Longrightarrow VCdim(\mathcal{H}_{con})=d$.

א) בהינתן כי ל-$\mathcal{H}$ ישנה את תכונת ה-Uniform Convergence

אז הפונקציה $m_{\mathcal{H}}^{UC}:(0,1)^2\to\mathbb{N}$, לכל $\varepsilon,\delta\in(0,1)$

אם קבוצת S בגודל מוגבל m כך שמתקיים:

$$m^{UC}(\varepsilon/2,\delta)\le m$$

אזי בהסתברות של לפחות $(1-\delta)$ נקבל ש-S הינה $\varepsilon/2$-מייצגת ולכן לכל השערה $h\in\mathcal{H}$ מתקיים:

$$L_D(h)\le L_S(h)+\varepsilon/2$$

ובפרט,

$$L_D(h)\le \min_{h'\in\mathcal{H}} L_D(h')+\varepsilon/2$$

ולכן גם,

$$L_D(h)\le \min_{h'\in\mathcal{H}} L_D(h')+\varepsilon$$

וקיבלנו כי $\mathcal{H}$ הינה למידה Agnostic-PAC כאשר:

$$m_{\mathcal{H}}(\varepsilon,\delta)=m^{UC}(\varepsilon/2,\delta)$$

כנדרש.

7) נניח כי $VCdim(H_1)=d$, אז-כך, קיימת קבוצה בגודל d
נקודות המנותצת ע"י H_1 כלומר, לכל תוצאה אפשרית y_i
עבור $i\in[d]$ קיימת היפותזה $h\in H_1$ המתאימה לה.
כעת, בהינתן כי H_2 מכילה את כל ההיפותזות מתוך H_1
אזי H_2 מנתצת את אותה קבוצה גם.
מכאן נסיק כי $VCdim(H_1) = d \le VCdim(H_2)$

8) (a) במילים, $\tau_H(m)$ היא המספר המקסימלי של היפותזות
שונות לקבוצה C אשר גודלה m על $\{\pm1\}$ הש"ט
בצמצום של H ל-C.

(b) בהינתן כי $VCdim(H)=\infty$ אזי נסיק כי גודלה המקסימלי
של קבוצה $C\subset X$ הניתנת לניתוץ ע"י H אינה חסומה
מלמעלה או במילים אחרות, H מנתצת כל גודל של
קבוצה. מכאן נוכל להסיק בהינתן קבוצה C בגודל m
כלשהו נקבל כי הצמצום של H ל-C כבר מכיל את
כל האפשרויות, כל ההיפותזות, עבור 2 תוצאות
אפשריות. בסוף-כך, מהגדרת $\tau_H(m)$ נקבל בגודל המקסימלי:

$$\tau_H(m) = 2^m$$

(c) בהמשך לסעיף הקודם, בהינתן כי $VCdim(H)=d$
ובהינתן קבוצה C בגודל $m\le d$ כלשהי נקבל כי גודל
המקסימלי להיפותזות עם 2 תוצאות אפשריות:

$$\tau_H(m) = 2^m$$

(ד) כעת נאמר כי $VCdim(\mathcal{H}) = d$ ונניח כי-$d<m$.

i. נבצע אינדוקציה על m גודל הדגימה כך שעבור בסיס האינדוקציה כאשר $m=1$ לכל מחלקת השערות $\mathcal{H}$ נקבל כי שני צדי המשוואה שלעיל יהיו שווים או 1 כאשר הסט בגודל רק 2 כאשר הסט ניתן לניפוץ.

כעת נניח כי הטענה נכונה לכל $k<m$ ונראה כי היא נכונה גם עבור כל m בצעד האינדוקציה:

נקבע מחלקת השערות $\mathcal{H}$ ונגדיר $C=\{c_1,\dots,c_m\}$ ונניח שנאמר לפי ברמה נחלק את C לשתי קבוצות כאשר הראשונה היא $\{c_1\}$ והשנייה $C'=\{c_2,\dots,c_m\}$ ובנוסף לכך נגדיר את הקבוצות הבאות:

$$A_1=\{(a_2,\dots,a_m) \mid (0,a_2,\dots,a_m)\in\mathcal{H}_C \vee (1,a_2,\dots,a_m)\in\mathcal{H}_C\}$$

$$A_2=\{(a_2,\dots,a_m) \mid (0,a_2,\dots,a_m)\in\mathcal{H}_C \wedge (1,a_2,\dots,a_m)\in\mathcal{H}_C\}$$

נשים-לב לפי בנייתם- האיחוד הזר כי מתקיים:

$$|\mathcal{H}_C| = |A_1| + |A_2|$$

והשני הוא כי ע"פ הגדרת C' מתקיים $A_1=\mathcal{H}_{C'}$.

מהנחת האינדוקציה אנו מקבלים:

$$|A_1| = |\mathcal{H}_{C'}| \le |\{B\subseteq C' \mid \mathcal{H} \text{ shatters } B\}| =$$

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

נרצה לחסום גם את גודלה של A_2, נעשה זאת ע"י הגדרת $\mathcal{H}_2\subseteq\mathcal{H}$ בצורה הבאה:

$$\mathcal{H}_2=\{h\in\mathcal{H} \mid \exists h_2\in\mathcal{H} \text{ st } (1-h_2(c_1),h_2(c_2),\dots,h_2(c_m)) =$$

$$= (h(c_1),h(c_2),\dots,h(c_m))\}$$

נשים לב כי אם H_2 מנתצת תת-קבוצה B של C'
אז היא מנתצת גם את תת-הקבוצה $B\cup\{c_1\}$
ובמשלים לכך, אם H_2 מנתצת את $B\cup\{c_1\}$ אז
היא מנתצת גם את B.

לכן, $A_2=H_{2C}$ ובהגדרת האינדוקציה, עבור H_2, C',
נקבל:

$$
\begin{aligned}
|A_2| = |H_{2C}| &\leq |\{B\subseteq C' \mid H \text{ shatters } B\}| = \\
&= |\{B\subseteq C' \mid H_2 \text{ shatters } B\cup\{c_1\}\}| = \\
&= |\{B\subseteq C \mid c_1\in B \wedge H_2 \text{ shatters } B\}| \leq \\
&\leq |\{B\subseteq C \mid c_1\in B \wedge H \text{ shatters } B\}|
\end{aligned}
$$

לסיום, נשים לב כי $|H_C|=|A_1|+|A_2|$ ולכן נותר
לחבר את $|A_2|, |A_1|$ ולקבל:

$$
\begin{aligned}
|H_C| &\leq |\{B\subseteq C \mid c_1\notin B \wedge H \text{ shatters } B\}| + \\
&+ |\{B\subseteq C \mid c_1\in B \wedge H \text{ shatters } B\}| = \\
&= |\{B\subseteq C \mid H \text{ shatters } B\}|
\end{aligned}
$$

והוכחנו את הטענה.

ii. במילים, מספר ההיפותזות שבצמצום של H לתת-קבוצה
C עבור H חסום מלמעלה ע"י מספר תת-הקבוצות
של C ש-H מנתצת כולן.

iii. נוכיח כי מתקיים:

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

נשים לב כי באגף השמאלי יש לנו גודל של קבוצה מתוך קבוצה שגודלה המקסימלי הוא d שכן הראינו בסעיפים קודמים שבהינתן VCdim אנו לא יכולים את הניתוץ מתוך מרחב בגודל מגודל m.
לעומת זאת, באגף הימני הביטוי מסמל את הסכום של כל תתי-הקבוצות האפשריות מגודל $k\le d$ שכן הם נלקחו ממרחב בגודל בגודל m.
לכן, מכיוון שבאגף השמאלי מהווה בגבול של האגף הימני וכיון שמצב אלו באין בגודל נקבל כי האגף הימני יכול להיות או גדול ממש ממש או שווה לו וזה קורה רק אם C מנותצת ע"י H ולכן כל תתי-הקבוצות מנותצות גם הן ע"י H.

iv. נשתמש בהגדרת $\tau_H(m)$ וכן בהוכחת סעיפים קודמים והמשוואה שקיבלנו בסעיף זה ונקבל:

$$\tau_H(m) = \max\{|H_C| \mid C\subseteq X, |C| = m\} \le$$

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

$$\le \left(\frac{em}{d}\right)^{d}$$

כנדרש.

Scanned with CamScanner

(e) בהמשך לסעיף (d) iii , אי-השוויון נשמר גם עבור

$m=d$ שכן ראינו כי במצב אחד $\tau_H(m) = 2^m$

ובמצב שני $\left(\frac{em}{d}\right)^d = e^m$ ולכן:

$$\tau_H(m) = 2^m \le e^m$$

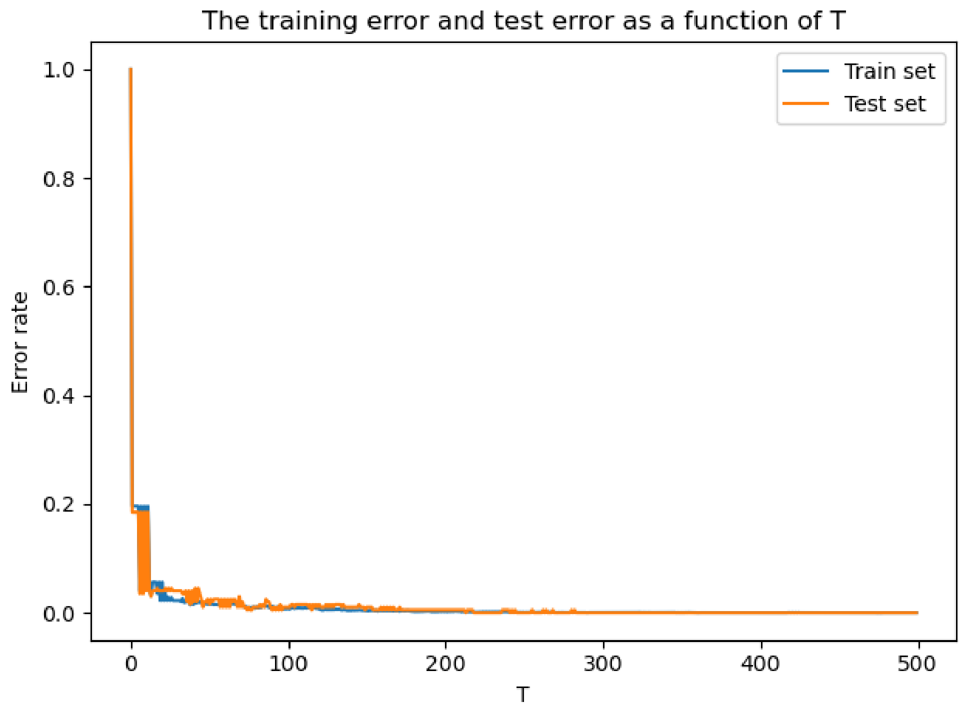
ונשים-לב כי חסם זה הדוק לכל $(ea)^m$ עבור $a=\frac{m}{d}>1$.

(f) לסיכום, נבחין כי התנהגות של $\tau_H(m)$ משתנה בהתאם
לגודל של m כך שכאשר $m > VCdim(H)$ נקבל התנהגות
פולינומית, בעוד שכאשר $m \le VCdim(H)$ נקבל התנהגות
אקספוננציאלית.
במילים אחרות, קיימת תת-קבוצה $C \subset X$ הפולינומיאלית
ב-m גדלה ללא תלות ממספר הפונקציות, בעוד
שאז היא אקספוננציאלית ב-m כי גדלה עם מספר
הפונקציות H שהיא עשירה מספיק עבורן.
ניתן להגדיר במילים את $VCdim(H)$ כגודל הקבוצה
המינימלית אותה ניתן לנתץ.

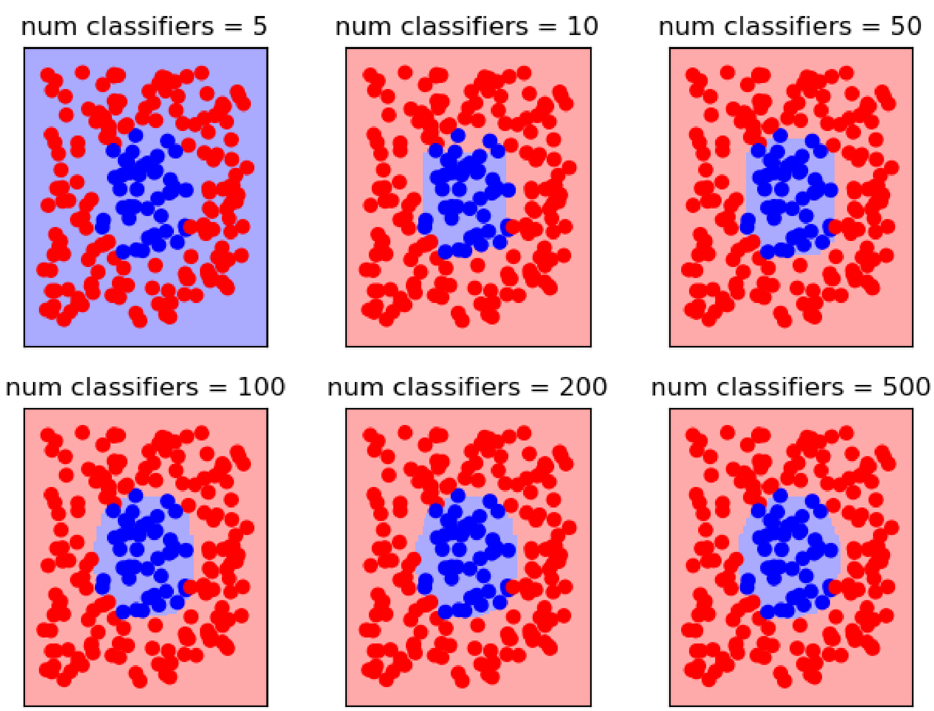
לירון גרשוני
308350503

מערכות לומדות – המשך תרגיל בית 4

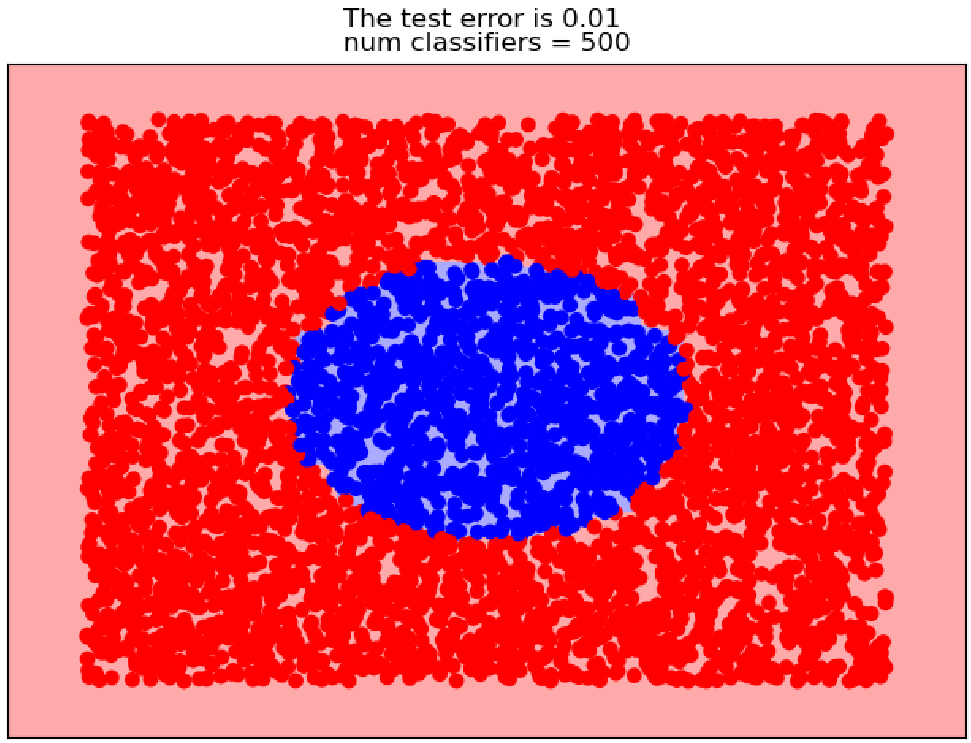
(10



(11

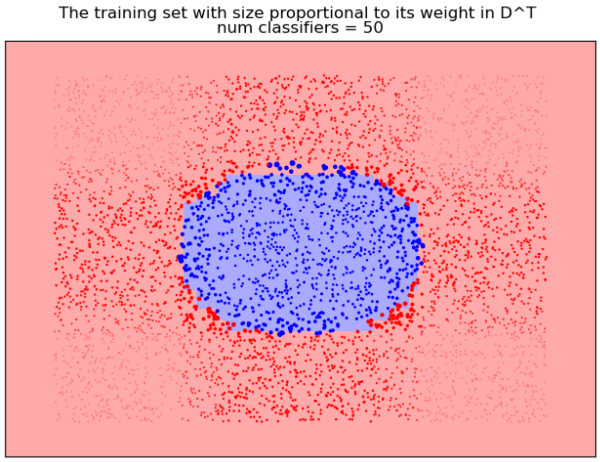


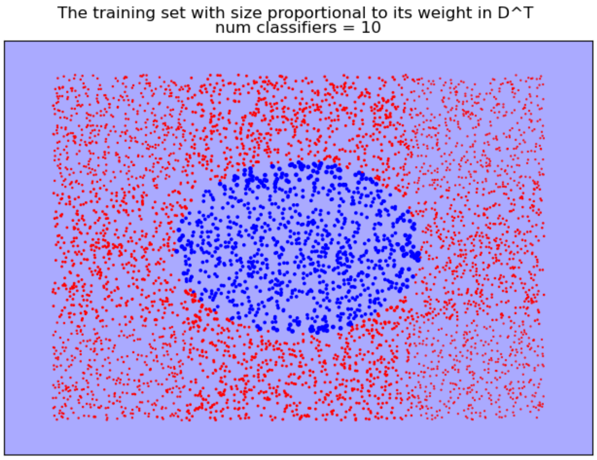
12)

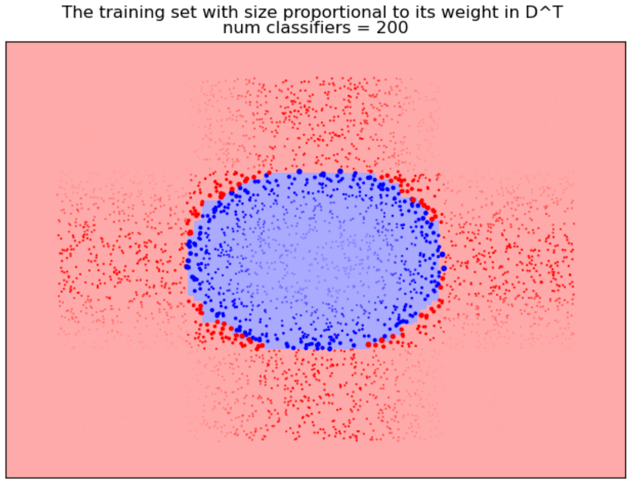


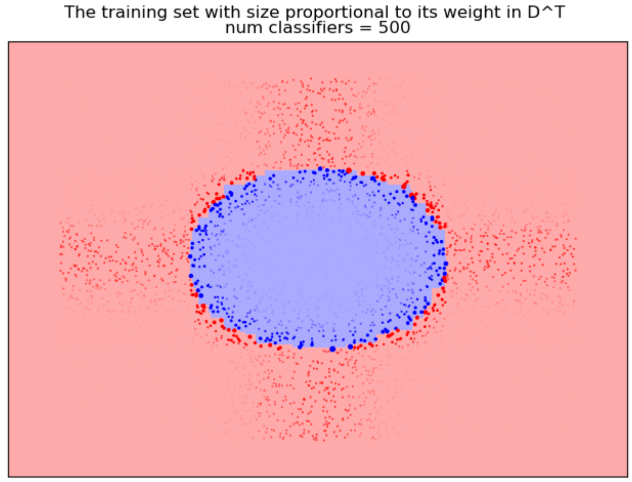
להלן אחת התוצאות שהתקבלו לי בשאלה זו בה התוצאות מראות כי מספר המסווגים המביאים למינימום את השגיאה הינו 500 וכן השגיאה המינימלית הינה 0.01. אני מציין זאת כיוון ובכל הרצה של הפונקציה הזו התקבלו תוצאות שונות כך שמספר המסווגים נע בין 50 ל-500 כלומר, הפער גדול ואינו קונסיסטנטי וכן השגיאה נעה בין 0.005 ל-0.01 כלומר, בטווח די קטן.

13)



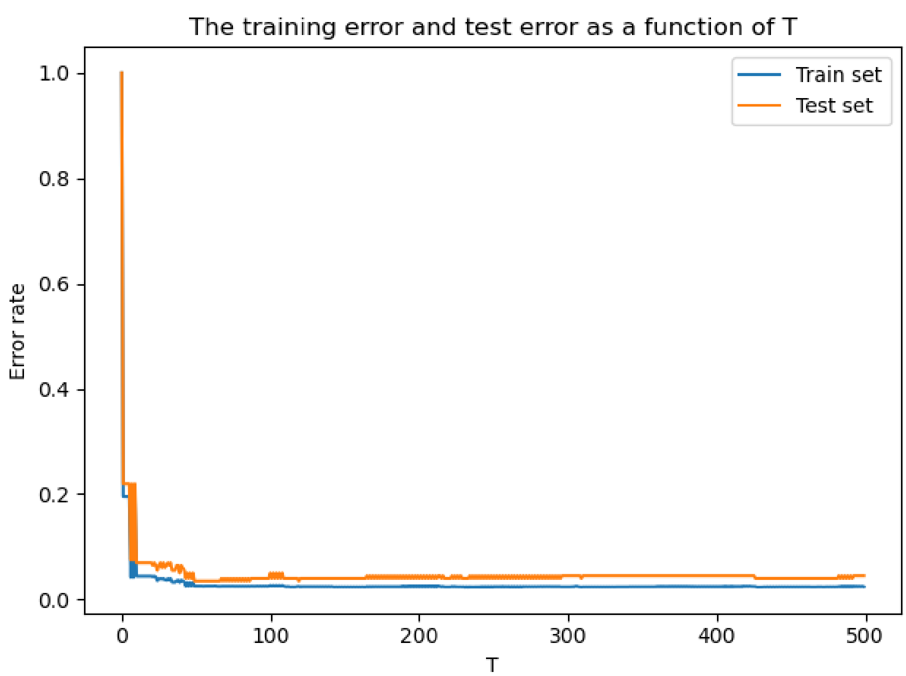


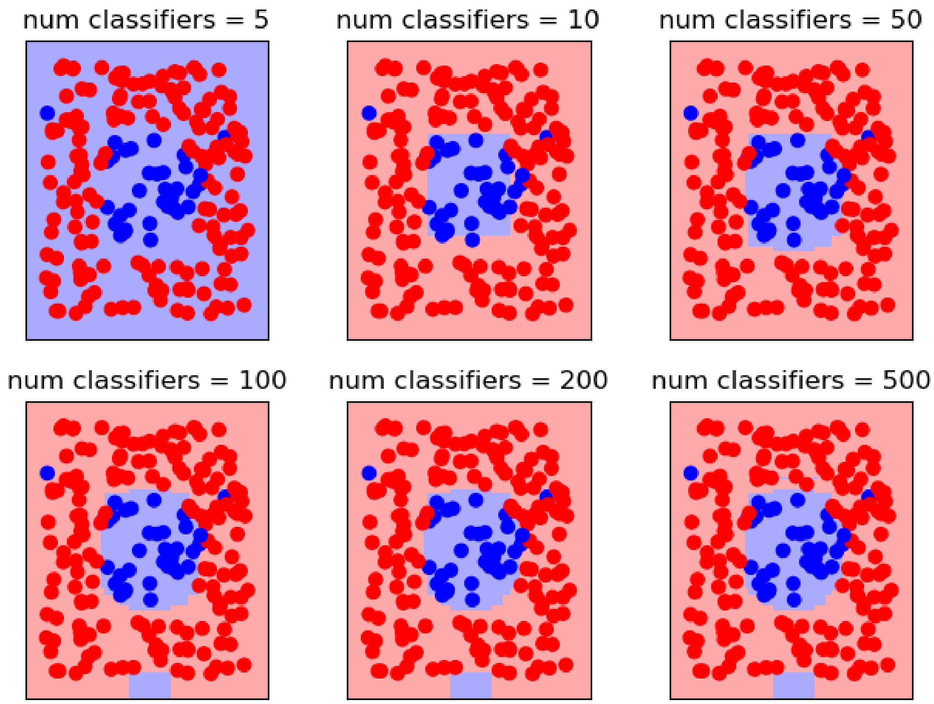


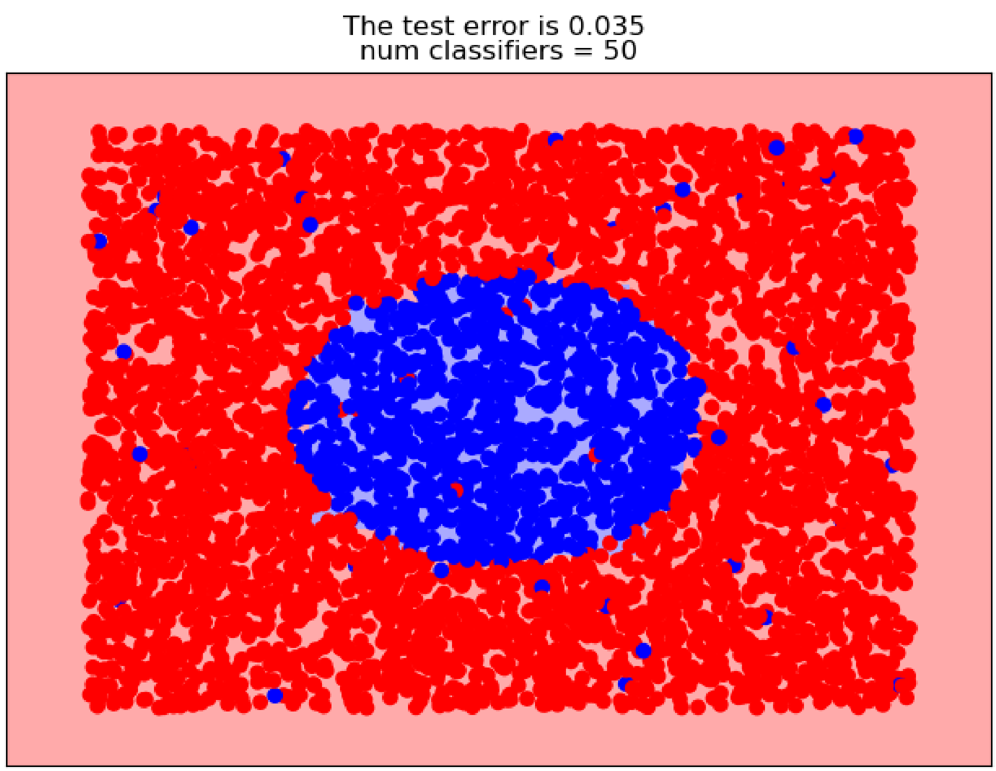


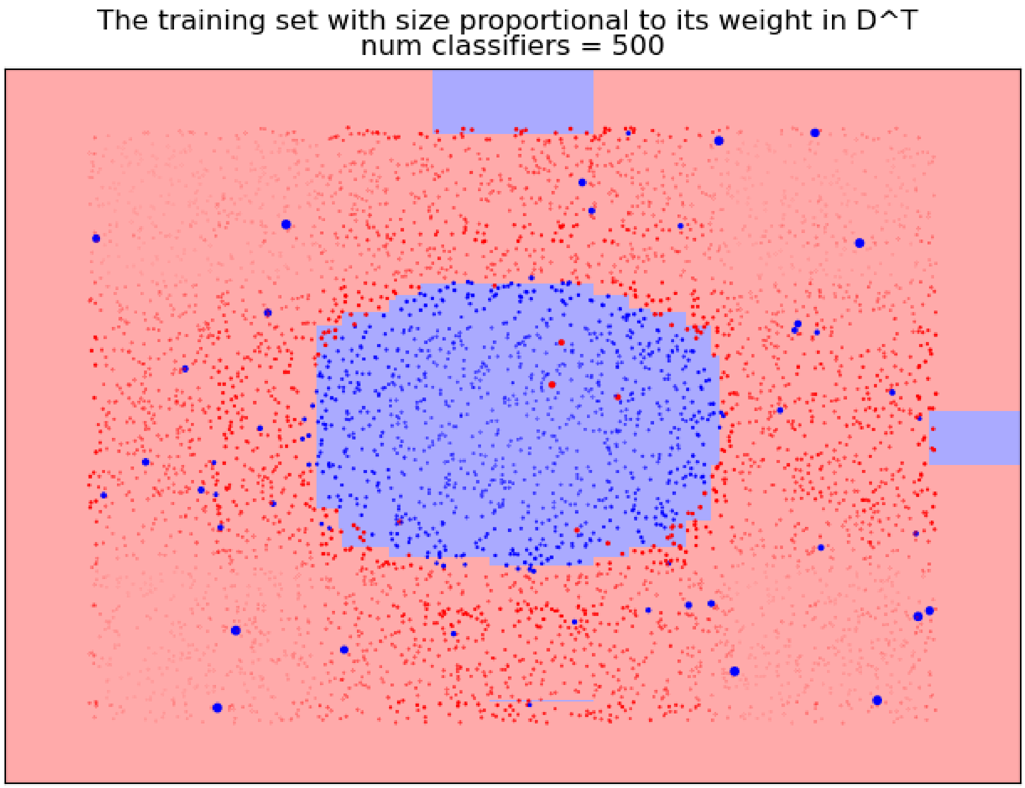
מעלה נוכל לראות גרפים המייצגים את נקודות האימון המייצגות את משקלן ע״פ גודלן על-פני גבולות ההחלטה בצבע כאשר אני משווה את אותו סוג הגרף עבור מספר שונה של מסווגים וכל זה תוך נרמול המשקלים כפי שייעצו לנו בתרגיל וזאת על-מנת שנוכל להבחין בהן.
נבחין אם כך בכמה נקודות שוני בין הגרפים- ראשית, כשמספר המסווגים קטן לא קיבלנו סיווג מהימן (כל הרקע בצבע כחול) וכן קשה להבחין בשוני בגודל של הנקודות כלומר, במשקלן.
כעת, ככל ומספר המסווגים עולה אנו עדים לגבולות החלטה יותר מדויקים, נוכל לראות גם סוג של ״התכנסות״ של הנקודות אל צורת צלב וזאת אני משער נובע מאופן ההחלטות בכל איטרציה שכן דווקא הנקודות שנמצאות על הצירים מסווגות הרבה פעמים לא נכון ולכן משקלן עולה ובעקבות זה גם גודלן בגרף לעומת נקודות שנמצאות רחוק מן הצירים ודי בהחלטות הראשונות מסווגות נכון ולכן גודלן כה קטן שלא ניתן להבחין בהן כמעט.
בנוסף, נשים-לב לנקודה מעניינת והיא שדווקא הנקודות שנמצאות על הגבול בין ההחלטות כלומר, בין הצבע הכחול לאדום הן אלו שכשמספר המסווגים גבוה מקבלות את המשקל הגבוה ביותר ומיוצגות כנקודות הכי גדולות בגרפים וזה עושה שכל כיוון וככל שאנו מנסים להיות יותר ויותר מדויקים בהחלטה שלנו אנו ״משחקים״ הכי הרבה עם הנקודות הללו עד למציאת הגבול המתאים ולמעשה ״טועים״ בהחלטה לגביהן ומעלים את משקלן.

14) **<u>עבור רעש בגודל 0.01:</u>**

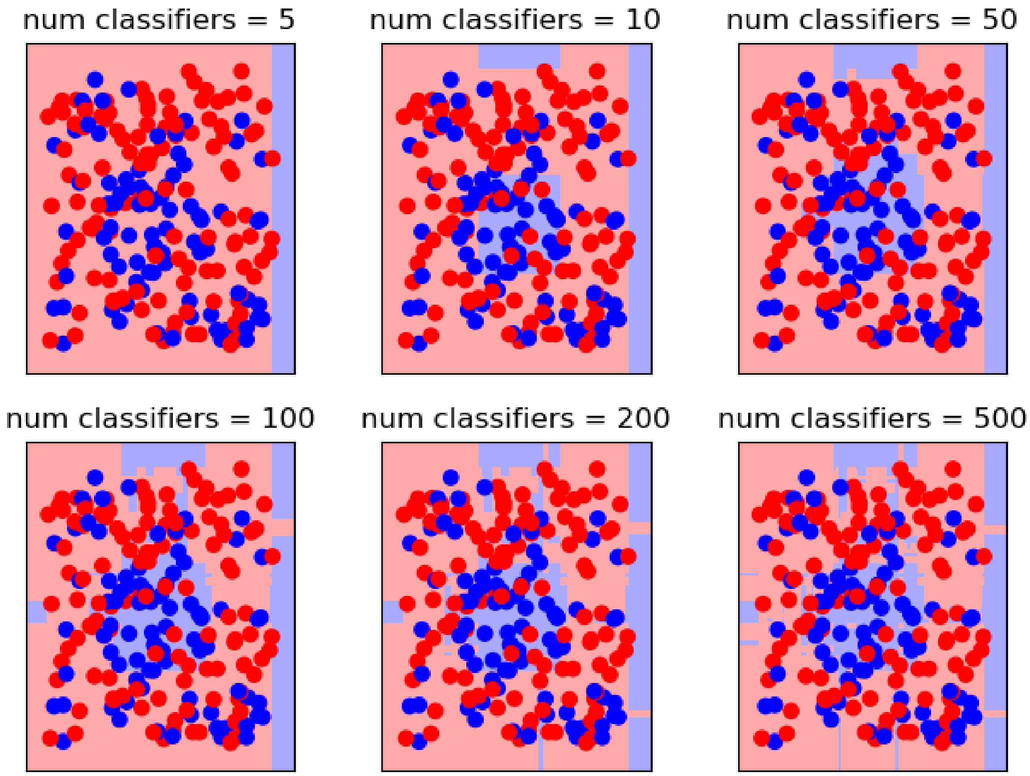


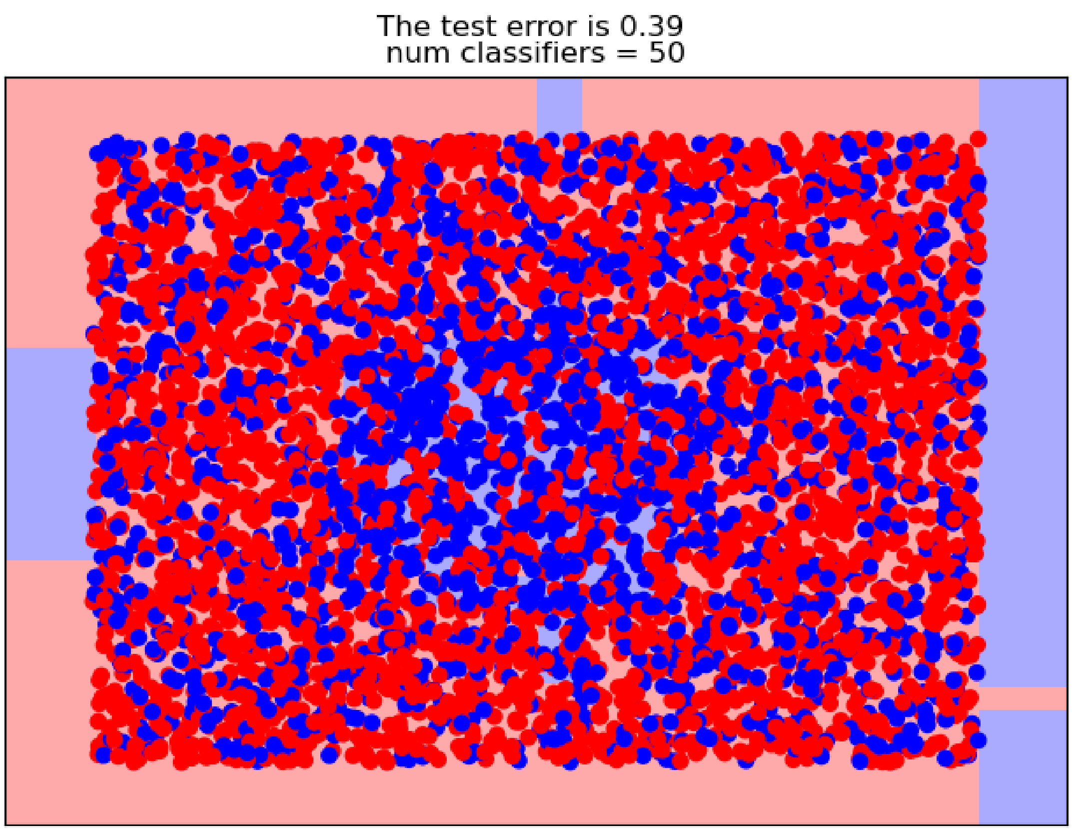

num classifiers = 5
num classifiers = 10
num classifiers = 50
num classifiers = 100
num classifiers = 200
num classifiers = 500

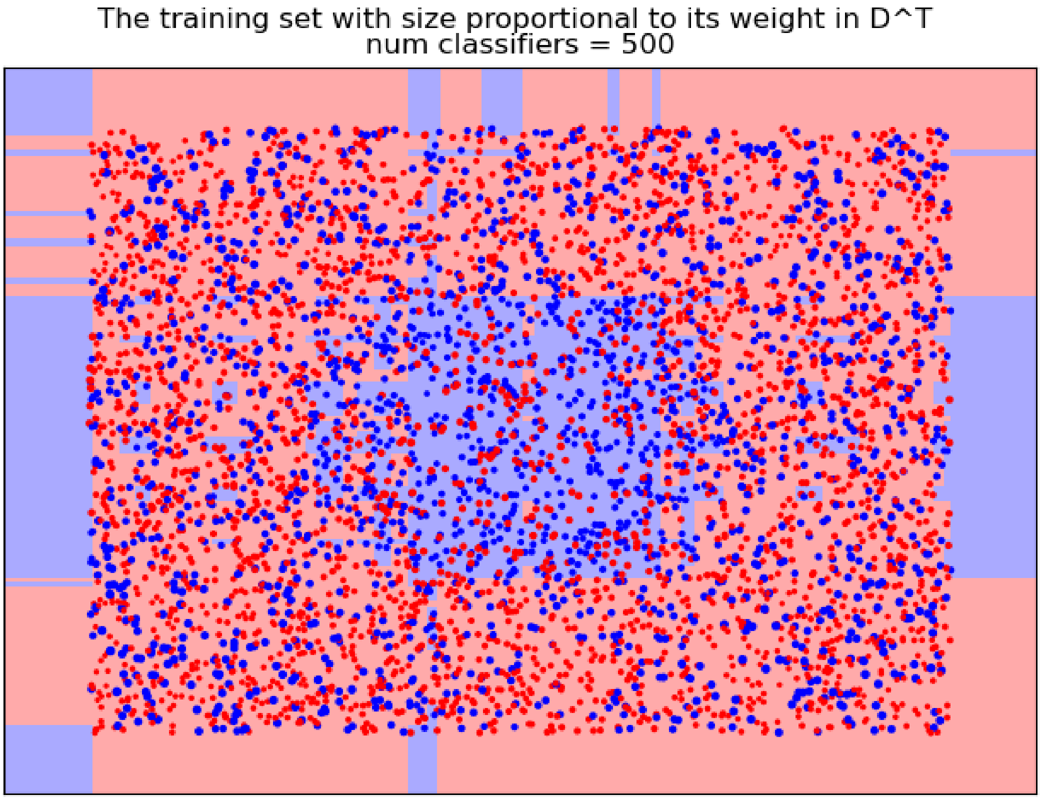

The test error is 0.035
num classifiers = 50



עבור רעש בגודל 0.4:


num classifiers = 5
num classifiers = 10
num classifiers = 50
num classifiers = 100
num classifiers = 200
num classifiers = 500


The test error is 0.39
num classifiers = 50



<u>**מענה על השאלות:**</u>

ראשית, בשאלה זו הרצנו את אותם תהליכים שביצענו לפני כן ללא רעש אך הפעם עם רעש. בנוסף, הרצנו פעם אחת רעש יחסית נמוך (0.01) ובפעם השנייה רעש יחסית גבוה (0.4) ולכן, מבלי לראות את התוצאות הסופיות אפילו, ציפיתי לקבל יחס שגיאה גבוה יותר עבור נקודות בדיקה, אי-התאמה בסיווג הנקודות וכן שגיאה הולכת וגדלה כאשר מספר המסווגים גבוה.

עבור ההבדלים בשאלה 10 נשים-לב לדבר מעניין- כאשר הרעש יחסית קטן אנחנו מקבלים באופן קונסיטיטוטיבי עלייה בשגיאה עבור כל מספר T כאשר מדובר בדגימות המבחן לעומת דגימות האימון בעוד שכאשר הרעש גבוה יחסית ניתן לראות שדווקא כאשר T קטן השגיאה של דגימות המבחן קטנה לעומת השגיאה של דגימות האימון אך בין T=200-300 המגמה מתהפכת.

באשר לנושא bias complexity tradeoff ניתן לראות שכפי שראינו בהרצאה, כאשר אנו מעלים את כמות הדגימות (T גבוה יותר) אנו מעשירים את מחלקת ההיפותזות ולכן מעלים את השונות ומורידים את ה-bias ולכן גם השגיאה שלנו עולה- ללא רעש שגיאה של כ-0.005 לעומת שגיאות של 0.035 וכן 0.39 כאשר אנו מוסיפים רעש למערכת. כל זה לעומת מחלקת היפותזות קטנה, כאשר T נמוך, בה השונות קטנה וה-bias גבוה.

עבור ההבדלים בשאלה 12 נשים-לב כי, לעומת הדגימות שביצענו ללא רעש, הפעם ההיפותזה הטובה ביותר בעלת יחס השגיאה הנמוך ביותר מתקבלת עבור מספר מסווגים נמוך שכן כאשר מספר המסווגים גבוה אנו נכנסים למצב שקראנו לו over fitting ולכן אינו מסביר טוב את הדגימות החדשות וזאת לעומת בחירה במספר מסווגים נמוך (50) אשר משאיר אותנו במצב של under fitting. בנוסף לכך, אנו ״משלמים״ על-כך שהכנסנו רעש למערכת בזה שקיבלנו כי יחס השגיאה גבוה בהרבה ועולה בצורה אקספוננציאלית (יחס שגיאה של 0.39 עבור רעש בגובה 0.4).

עבור ההבדלים בשאלה 13 ניתן לראות כי כאשר הרעש יחסית נמוך מתקבלות נקודות אשר סווגו לא נכון בכמות מועטת אך בהחלט ניתן לראות כי משקלן גבוה בעקבות כך וניתן למעשה להצביע על כל אחת כזו בקלות ביחס לאחרות סביבה שסווגו נכון ובעלות משקל נמוך. זאת לעומת רעש גבוה בו ישנם המון טעויות בסיווג הנקודות ולכן משקלן של רוב הנקודות גבוה ולמעשה לא מתקבלת החלטה חד-משמעית.