תרגיל 4 - IML

שאלה 1:

נראה שקילות:

$b \Longleftarrow a$: נתון $\lim_{m\to\infty}\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]=0$. מתקיים

$$\begin{aligned}
&\lim_{m\to\infty}\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]=0 \iff \forall\varepsilon_0>0\ \exists m_0\ \forall m>m_0\ \left|\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]\right|<\varepsilon_0\\
\iff &\lim_{m\to\infty}\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]=0 \iff \forall\varepsilon_0>0\ \exists m_0\ \forall m>m_0\ \ \mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]<\varepsilon_0
\end{aligned}$$

יהיו $\delta,\varepsilon>0$. נבחר $\varepsilon_0=\varepsilon\delta$. אז קיים m_0 כך שלכל $m>m_0$ מתקיים

$$\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]<\varepsilon\delta$$

ממרקוב ($L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)$ הינו משתנה מקרי אי שלילי כי $L_{\mathcal{D}}$ מוגדר להיות תוחלת של פונקציית loss שהיא אי שלילית) מתקיים

$$\begin{aligned}
&\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)>\varepsilon\right)\leq\frac{\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]}{\varepsilon}<\frac{\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]}{\varepsilon}=\frac{\varepsilon\delta}{\varepsilon}=\delta\\
\iff &\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\geq\varepsilon\right)\leq 1-\delta
\end{aligned}$$

נסמן $m\left(\varepsilon,\delta\right)=m_0$ ונקבל את הדרוש.

$a \Longleftarrow b$: נתון שלכל $\varepsilon,\delta>0$ קיים $m\left(\varepsilon,\delta\right)$ כך שלכל $m\geq m\left(\varepsilon,\delta\right)$ מתקיים:

$$\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\leq\varepsilon\right)\geq 1-\delta$$

יהי $\varepsilon>0$. אז לכל $m\geq m\left(\frac{\varepsilon}{2},\frac{\varepsilon}{2}\right)$ מתקיים

$$(*)\,\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)>\frac{\varepsilon}{2}\right)<\frac{\varepsilon}{2}$$

כעת:

$$\begin{aligned}
\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]&=\int_{x=0}^{x=1}x\cdot\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)=x\right)\\
&=\int_{x=0}^{x=\frac{\varepsilon}{2}}x\cdot\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)=x\right)+\int_{x=\frac{\varepsilon}{2}}^{x=1}x\cdot\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)=x\right)\\
&\leq\int_{x=0}^{x=\frac{\varepsilon}{2}}\frac{\varepsilon}{2}\cdot\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)=x\right)+\int_{x=\frac{\varepsilon}{2}}^{x=1}1\cdot\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)=x\right)\\
&=\frac{\varepsilon}{2}\int_{x=0}^{x=\frac{\varepsilon}{2}}\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)=x\right)+\int_{x=\frac{\varepsilon}{2}}^{x=1}\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)=x\right)\\
&=\frac{\varepsilon}{2}\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\leq\frac{\varepsilon}{2}\right)+\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)>\frac{\varepsilon}{2}\right)\\
&<\frac{\varepsilon}{2}\cdot 1+\frac{\varepsilon}{2}=\varepsilon
\end{aligned}$$

כאשר המעבר האחרון נובע מ־$(*)$ ומכך שהסתברות חסומה ע"י 1. סה"כ הראינו שבהנתן $\varepsilon>0$ קיים $m_0=m\left(\frac{\varepsilon}{2},\frac{\varepsilon}{2}\right)$ כך שלכל $m\geq m_0$ מתקיים $\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]<\varepsilon$ וכיוון ש־ $L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)$ אי שלילי זה שקול לכך ש־ $\left|\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]\right|<\varepsilon$, וכל זה שקול ל־ $\lim_{m\to\infty}\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(\mathcal{A}\left(S\right)\right)\right]=0$ כדרוש.

שאלה 2:

יהיו $\mathcal{Y}=\{0,1\}$, $\mathcal{X}=\mathbb{R}^2$, $\mathcal{H}=\{h_r \mid r\in\mathbb{R}_+\}$ ו־$h_r(x)=1_{[||x||_2\leq r]}$.
צ"ל ש־$\mathcal{H}$ היא למידה PAC ושמתקיים:

$$m_{\mathcal{H}}(\varepsilon,\delta)\leq\frac{\log(\frac{1}{\delta})}{\varepsilon}$$

נגדיר את $\mathcal{A}$ להיות אלגוריתם ה־ERM הבא: בהנתן training set S מחזיר את h_S ־ המעגל עם הרדיוס הקטן ביותר שמכיל את כל הדגימות x_i ב־S עם $y_i=1$. נסמן את רדיוסו ב־ r_S.
נניח ריאלייזביליות. יהי h^* מעגל עם שגיאת הכללה 0. נסמן את הרדיוס של h^* ב־r^*.
יהיו $\varepsilon,\delta>0$. יהי $r\leq r^*$ עבורו $\mathcal{D}_{\mathcal{X}}\left(\{x\mid r\leq||x||\leq r^*\}\right)=\varepsilon$. נגדיר $E=\left\{x\in\mathbb{R}^2\mid r\leq||x||\leq r^*\right\}$.
כעת, ההסתברות ש־$L_{\mathcal{D}}\left(h_S\right)\geq\varepsilon$ חסומה מלמעלה ע"י ההסתברות שאף דגימה ב־S לא נמצאת ב־E. כלומר כל הנקודות ב־S נדגמו ברדיוס שקטן מ־r, כלומר שגיאה גדולה יותר מ־ε (לא יתכן שידגמו נקודות ברדיוס שגדול מ־r^* בגלל ש־h^* אופטימלית).
הסתברות לנקודה ב־E היא ε, ולכן ההסתברות לא ליפול שם היא $1-\varepsilon$. סה"כ, כיוון שיש m דגימות ב־S מתקיים

$$\mathcal{D}_{\mathcal{X}}\left(L_{\mathcal{D}}\left(h_S\right)\geq\varepsilon\right)\leq(1-\varepsilon)^m$$

מהתכונה הנתונה לנו בתרגיל: $1-\varepsilon\leq e^{-\varepsilon}$ נקבל $(1-\varepsilon)^m\leq e^{-\varepsilon m}$.
אנחנו רוצים שההסתברות תהיה קטנה מ־δ, נמצא איזה m נצטרך לפי החסם הזה:

$$\delta\geq e^{-\varepsilon m}\iff\delta e^{\varepsilon m}\geq1\iff e^{\varepsilon m}\geq\frac{1}{\delta}\iff\log_e\left(e^{\varepsilon m}\right)\geq\log_e\left(\frac{1}{\delta}\right)$$

$$\iff\varepsilon m\geq\log_e\left(\frac{1}{\delta}\right)\iff m\geq\frac{\log_e\left(\frac{1}{\delta}\right)}{\varepsilon}$$

לכן מהמינימליות של $m_{\mathcal{H}}$ נקבל ש־$m_{\mathcal{H}}\left(\varepsilon,\delta\right)\leq\frac{\log_e\left(\frac{1}{\delta}\right)}{\varepsilon}$. סה"כ הראינו שלכל $\varepsilon,\delta>0$ ולכל $m\geq\frac{\log_e\left(\frac{1}{\delta}\right)}{\varepsilon}\geq m_{\mathcal{H}}\left(\varepsilon,\delta\right)$ מתקיים $\mathcal{D}_{\mathcal{X}}\left(L_{\mathcal{D}}\left(h_S\right)\geq\varepsilon\right)\leq\delta$ ולכן $\mathcal{H}$ היא למידה PAC.

שאלה 3:

תהי $\mathcal{H}=\{h_1,...,h_N\}$ מחלקת היפותזות סופית. מימד ה־VC שלה בהכרח סופי (לא יתכן שקיימת תת קבוצה של $\mathcal{H}$ בגודל יותר מ־$|\mathcal{H}|$ שמנותצת). נסמן $VC-dim\left(\mathcal{H}\right)=x$. אז בהכרח $2^x\leq|\mathcal{H}|$ (מהגדרת ניתוץ). **האם אפשר להניח ש־ $\mathcal{Y}=\{0,1\}$?)** ולכן $x\leq\log_2|\mathcal{H}|$ וכיוון ש־x מספר שלם מתקיים $x\leq\lfloor\log_2|\mathcal{H}|\rfloor$

שאלה 4:

יהיו $\mathcal{X}=\{0,1\}^n$ ו־$\mathcal{Y}=\{0,1\}$. לכל $I\subseteq[n]$ נגדיר את הparity function:

$$h_I(x)=\left(\sum_{i\in I}x_i\right)\mod 2$$

$\mathcal{H}_{\text{parity}}=\{h_I\mid I\subseteq[n]\}$. נטען כי $VC-dim\left(\mathcal{H}_{\text{parity}}\right)=n$. מתרגיל 3 אנו יודעים שמתקיים

$$VC-dim\left(\mathcal{H}_{\text{parity}}\right)\leq\lfloor\log_2|\mathcal{H}_{\text{parity}}|\rfloor=\log_2 2^n=n$$

כעת נראה שקיימת תת קבוצה $C\subset\mathcal{X}$ בגודל n אשר מנותצת ע"י $\mathcal{H}$. נגדיר

$$C=\left\{(x_1,...,x_n)\in\{0,1\}^n\mid\forall i\in[n],\forall j\neq i,x_i=1,x_j=0\right\}=\{e_1,...,e_n\}$$

מספר הפונקציות $h:C\to\mathcal{Y}$ הוא 2^n. לכל $I\subseteq[n]$ הפונקצייה $h_I\in\mathcal{H}$ מקיימת

$$h_I\left(x\right)=\left(\sum_{i\in I}x_i\right)\mod 2$$

נטען שהצמצום של h_I ל־C מקיים

$$h_I\left(x\right)=\begin{cases}1 & x\in C_I:=\{e_i\mid i\in I\}\\0 & x\in C\setminus C_I\end{cases}$$

וכך נראה ש־$\mathcal{H}$ מנתצת את C. עבור $I \subseteq [n]$, לכל $e_j \in C$ מתקיים

$$h_I\left(e_j\right) = \left(\sum_{i\in I} x_i\right) \mod 2 = \begin{cases} 1 & j \in I \\ 0 & \text{else} \end{cases}$$

ולכן בהנתן $I \subseteq [n]$, הפונקציה ששולחת את הוקטורים $C_I := \{e_i \mid i \in I\}$ ל־1 ואת שאר הוקטורים ב־C ל־0 היא בדיוק h_I (מצומצמת ל־C). סה"כ הראינו שכל הפונקציות $h : C \to \mathcal{Y}$ קיימות ב־$\mathcal{H}$ (אחרי צמצום ל־C) ולכן C מנותצת ע"י $\mathcal{H}$.

שאלה 5:

בהנתן מספר k, תהי $\left([a_i, b_i]\right)_{i=1}^{k}$ קבוצה של k אינטרוולים על $\mathbb{R}$ ונגדיר את האיחוד שלהם $A = \bigcup_{i=1}^{k} [a_i, b_i]$ (נניח שהם זרים זה לזה).
מחלקת ההיפותזות $\mathcal{H}_{k-intervals}$ כוללת את הפונקציות

$$h_A(x) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases}$$

נטען כי $\text{VC-dim}\left(\mathcal{H}_{k-intervals}\right) = 2k$. נקח

$$C = \{x_1, ..., x_{2k}\}$$

עבור $h_A \in \mathcal{H}_{k-intervals}$ אשר מקיימת

$$h_A\left(x_1\right) = 1, h_A\left(x_2\right) = 0, h_A\left(x_3\right) = 1, ..., h_A\left(x_{2k-1}\right) = 1, h_A\left(x_{2k}\right) = 0$$

נשים לב שיש k איברים אותם h_A שולחת ל־1 ובין כל שני איברים כאלה יש איבר שנשלח ל־0. כלומר צריך k אינטרוולים ב־A. לכל h_A אחרת צריך פחות אינטרוולים.
נראה כי $\text{VC-dim}\left(\mathcal{H}_{k-intervals}\right) < 2k+1$. נניח בשלילה ש־ $\text{VC-dim}\left(\mathcal{H}_{k-intervals}\right) \geq 2k+1$. אז קיימת קבוצה $C = \{x_1 < ... < x_{2k+1}\} \subseteq \mathbb{R}$ שמנותצת ע"י $\mathcal{H}_{k-intervals}$. בפרט קיימת הפונקציה $h_A \in \mathcal{H}$ כך ש־ $A = \bigcup_{i=1}^{k} [a_i, b_i]$, אשר מקיימת

$$h_A\left(x_1\right) = 1, h_A\left(x_2\right) = 0, h_A\left(x_3\right) = 1, ..., h_A\left(x_{2k}\right) = 0, h_A\left(x_{2k+1}\right) = 1$$

נשים לב שיש $k+1$ איברים ש־h_A שולחת ל־1. בין כל שני איברים ש־h_A שולחת ל־1, יש איבר שהיא שולחת ל־0. כלומר יש סה"כ $k+1$ אינטרוולים **זרים** ב־A, בסתירה לכך שיש k בלבד כאלו.

כעת נניח ש־A היא כל איחוד סופי של אינטרוולים, כלומר k לא מוגבל. אזי $\text{VC-dim}\left(\mathcal{H}_{intervals}\right) = \infty$. נוכיח: מתקיים

$$\mathcal{H}_{intervals} = \bigcup_{k\in\mathbb{N}} \mathcal{H}_{k-intervals}$$

נניח בשלילה ש־ $\text{VC-dim}\left(\mathcal{H}_{intervals}\right) = 2k-1$ עבור $2 \leq k \in \mathbb{N}$ כלשהו.. אבל קודם הוכחנו כי $\text{VC-dim}\left(\mathcal{H}_{k-intervals}\right) = 2k$ לכן קיימת $C \subset \mathbb{R}$ שמנותצת ע"י $\mathcal{H}_{k-intervals}$ וכיוון שמתקיים ש־ $\mathcal{H}_{k-intervals} \subset \mathcal{H}_{intervals}$, בפרט C מנותצת ע"י $\mathcal{H}_{intervals}$ - כלומר $\text{VC-dim}\left(\mathcal{H}_{intervals}\right) \geq 2k$ בסתירה להנחה.

שאלה 6:

שאלה 7:

נתון כי $\mathcal{H}$ היא בעלת תכונת UC. יהיו $\delta, \varepsilon > 0$. מתקיים, לכל $m \geq m_{\mathcal{H}}^{UC}\left(\frac{\varepsilon}{2}, \delta\right)$

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

נוכיח את הטענה הבאה: יהיו $\mathcal{H}, \mathcal{D}, \ell$. אם $S \in (\mathcal{X} \times \mathcal{Y})^m$ היא $\frac{\varepsilon}{2}$ representative אז בהנתן $h_S \in \underset{h\in\mathcal{H}}{\text{ERM}}(S)$ מתקיים

$$L_{\mathcal{D}}\left(h_S\right) \leq \min_{h\in\mathcal{H}} L_{\mathcal{D}}(h) + \varepsilon$$

הוכחה: מכך ש־S היא $\frac{\varepsilon}{2}$ representative מתקיים

$$\forall h\in\mathcal{H}\ \ \left|L_S\left(h\right)-L_{\mathcal{D}}\left(h\right)\right|<\frac{\varepsilon}{2}$$

נסמן $h_S\in\min_{h\in\mathcal{H}}L_S\left(h\right)$, $h^*\in\min_{h\in\mathcal{H}}L_{\mathcal{D}}\left(h\right)$. מתקיים

$$\left|L_S\left(h^*\right)-L_{\mathcal{D}}\left(h^*\right)\right|<\frac{\varepsilon}{2}\iff-\frac{\varepsilon}{2}<L_S\left(h^*\right)-L_{\mathcal{D}}\left(h^*\right)<\frac{\varepsilon}{2}\Longrightarrow(*)\,L_S\left(h^*\right)<L_{\mathcal{D}}\left(h^*\right)+\frac{\varepsilon}{2}$$

$$\left|L_S\left(h_S\right)-L_{\mathcal{D}}\left(h_S\right)\right|<\frac{\varepsilon}{2}\iff-\frac{\varepsilon}{2}<L_S\left(h_S\right)-L_{\mathcal{D}}\left(h_S\right)<\frac{\varepsilon}{2}\Longrightarrow L_{\mathcal{D}}\left(h_S\right)<L_S\left(h_S\right)+\frac{\varepsilon}{2}\le L_S\left(h^*\right)+\frac{\varepsilon}{2}\overset{(*)}{<}L_{\mathcal{D}}\left(h^*\right)+\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\min_{h\in\mathcal{H}}L$$

מהטענה שהוכחנו נקבל:

$$\left\{S\in\left(\mathcal{X}\times\mathcal{Y}\right)^m\mid S\text{ is }\frac{\varepsilon}{2}\text{ representative}\right\}\subseteq\left\{S\in\left(\mathcal{X}\times\mathcal{Y}\right)^m\mid L_{\mathcal{D}}\left(h_S\right)\le\min_{h\in\mathcal{H}}L_{\mathcal{D}}\left(h\right)+\varepsilon\right\}$$

ולכן לכל $m\ge m_{\mathcal{H}}^{UC}\left(\frac{\varepsilon}{2},\delta\right)$

$$\mathcal{D}^m\left(\left\{S\in\left(\mathcal{X}\times\mathcal{Y}\right)^m\mid L_{\mathcal{D}}\left(h_S\right)\le\min_{h\in\mathcal{H}}L_{\mathcal{D}}\left(h\right)+\varepsilon\right\}\right)\ge\mathcal{D}^m\left(\left\{S\in\left(\mathcal{X}\times\mathcal{Y}\right)^m\mid S\text{ is }\frac{\varepsilon}{2}\text{ representative}\right\}\right)\ge1-\delta$$

כלומר $\mathcal{H}$ היא Agnostic-PAC עם פונ' $m_{\mathcal{H}}$. כיוון ש־ $m_{\mathcal{H}}$ מחזירה את מספר הדגימות המינימלי, מתקיים $m_{\mathcal{H}}\left(\varepsilon,\delta\right)\le m_{\mathcal{H}}^{UC}\left(\frac{\varepsilon}{2},\delta\right)$.

שאלה 8:

שאלה 9:

נתון ש־ $\mathcal{H}$ היא למידה PAC עם סיבוכיות מדגם $m_{\mathcal{H}}$. נתונים $0<\varepsilon_1\le\varepsilon_2<1$ ו־$\delta\in(0,1)$. נסמן $m_{\mathcal{H}}\left(\varepsilon_1,\delta\right)=m_1, m_{\mathcal{H}}\left(\varepsilon_2,\delta\right)=m_2$. נסמן ב־$\mathcal{A}$ את האלג'. מהגדרת למידות PAC, לכל התפלגות $\mathcal{D}$ ופונ' תיוג f מתקיים שלכל $m\ge m_1$

$$\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D},f}\left(\mathcal{A}\left(S\right)\right)\le\varepsilon_1\right)\ge1-\delta$$

כיוון ש־ $\varepsilon_1\le\varepsilon_2$ מתקיים $\left\{S\mid L_{\mathcal{D},f}\left(\mathcal{A}\left(S\right)\right)\le\varepsilon_1\right\}\subseteq\left\{S\mid L_{\mathcal{D},f}\left(\mathcal{A}\left(S\right)\right)\le\varepsilon_2\right\}$ ולכן

$$\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D},f}\left(\mathcal{A}\left(S\right)\right)\le\varepsilon_2\right)\ge\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D},f}\left(\mathcal{A}\left(S\right)\right)\le\varepsilon_1\right)\ge1-\delta$$

תכונה זו גם נכונה לכל $m\ge m_2$. מהמינימליות של m_2 מתקיים $m_2\le m_1$.

נתונים $0<\delta_1\le\delta_2<1$ ו־$\varepsilon\in(0,1)$. נסמן $m_{\mathcal{H}}\left(\varepsilon,\delta_1\right)=m_1, m_{\mathcal{H}}\left(\varepsilon,\delta_2\right)=m_2$. נסמן ב־$\mathcal{A}$ את האלג'. מהגדרת למידות PAC, לכל התפלגות $\mathcal{D}$ ופונ' תיוג f מתקיים שלכל $m\ge m_1$

$$\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D},f}\left(\mathcal{A}\left(S\right)\right)>\varepsilon\right)\le\delta_1\le\delta_2$$

התכונה $\mathbb{P}_{S\sim\mathcal{D}^m}\left(L_{\mathcal{D},f}\left(\mathcal{A}\left(S\right)\right)>\varepsilon\right)\le\delta_2$ גם נכונה לכל $m\ge m_2$ וממינימליות m_2 מתקיים $m_2\le m_1$.

שאלה 10:

נתון $\mathcal{H}_1\subseteq\mathcal{H}_2$. נניח בשלילה כי

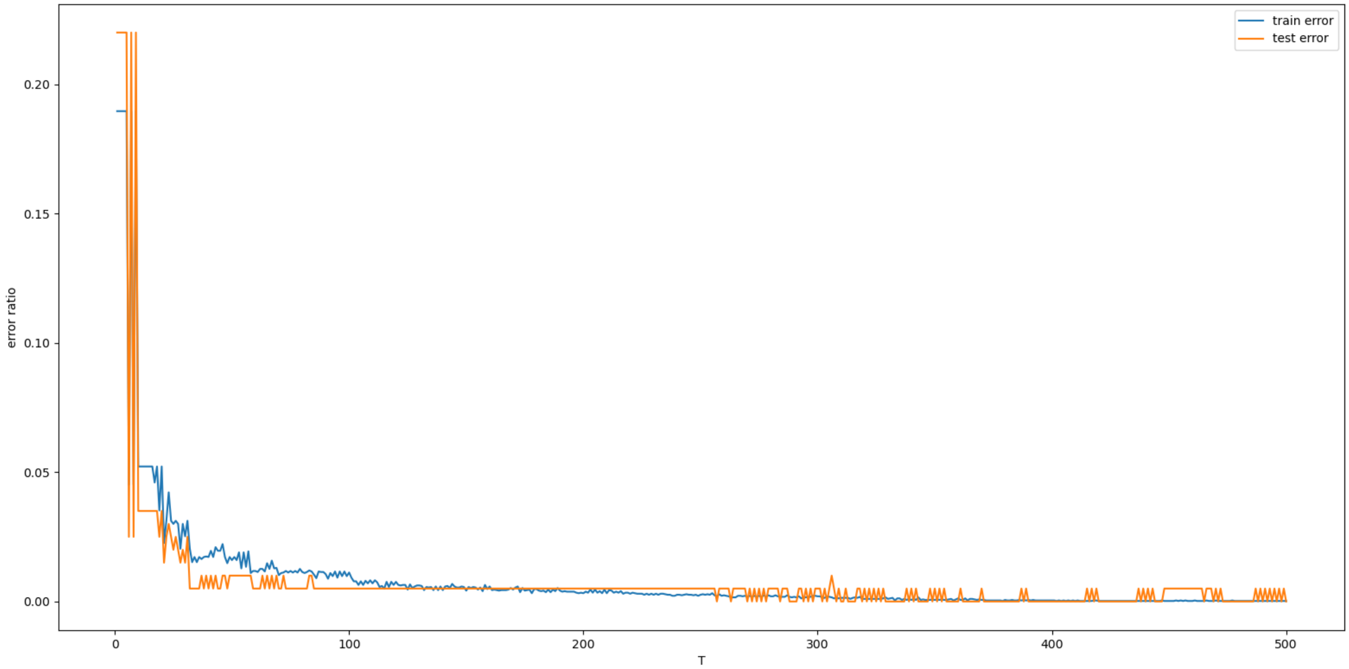
$$d=\text{VC-dim}\left(\mathcal{H}_2\right)<\text{VC-dim}\left(\mathcal{H}_1\right)=k$$

אז קיימת ב־$\mathcal{H}_1$ תת קבוצה בגודל k שמנותצת ע"י $\mathcal{H}_1$ וכיוון ש־ $d<k$, בהכרח קבוצה זו לא קיימת ב־$\mathcal{H}_2$. אך זו סתירה כי $\mathcal{H}_1\subseteq\mathcal{H}_2$.

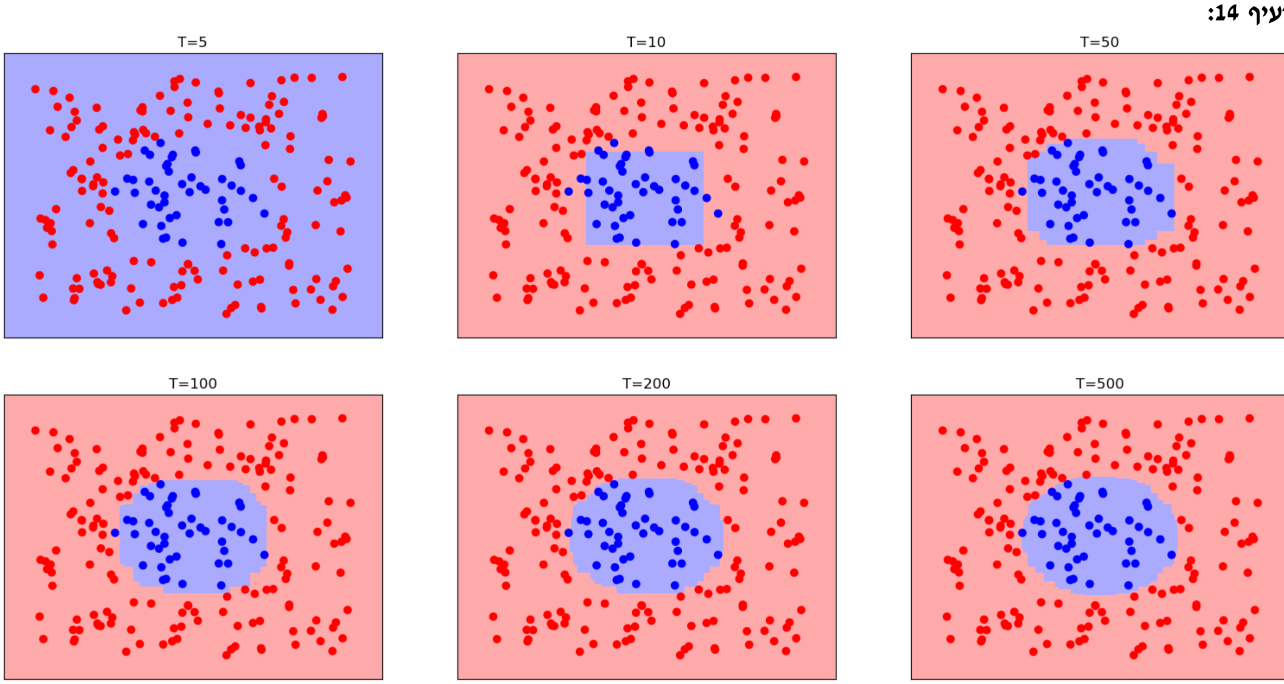
AdaBoost

סעיפים 13-16 עם רעש 0:

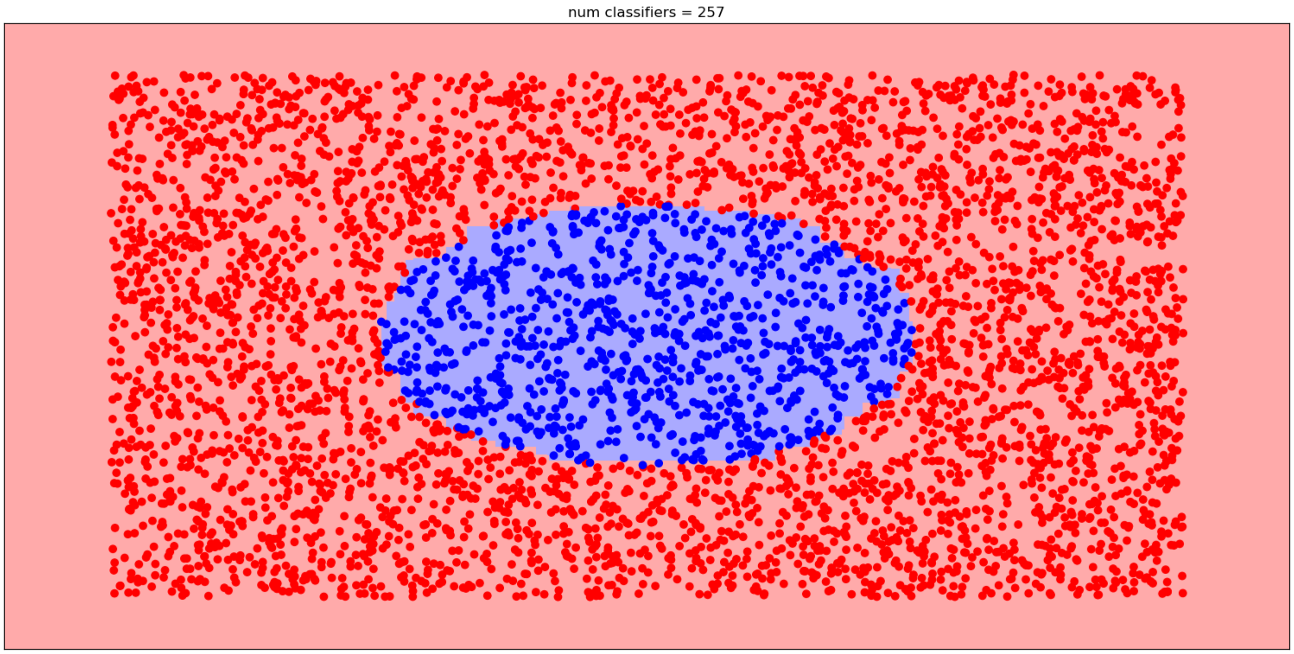
סעיף 13:



סעיף 14:

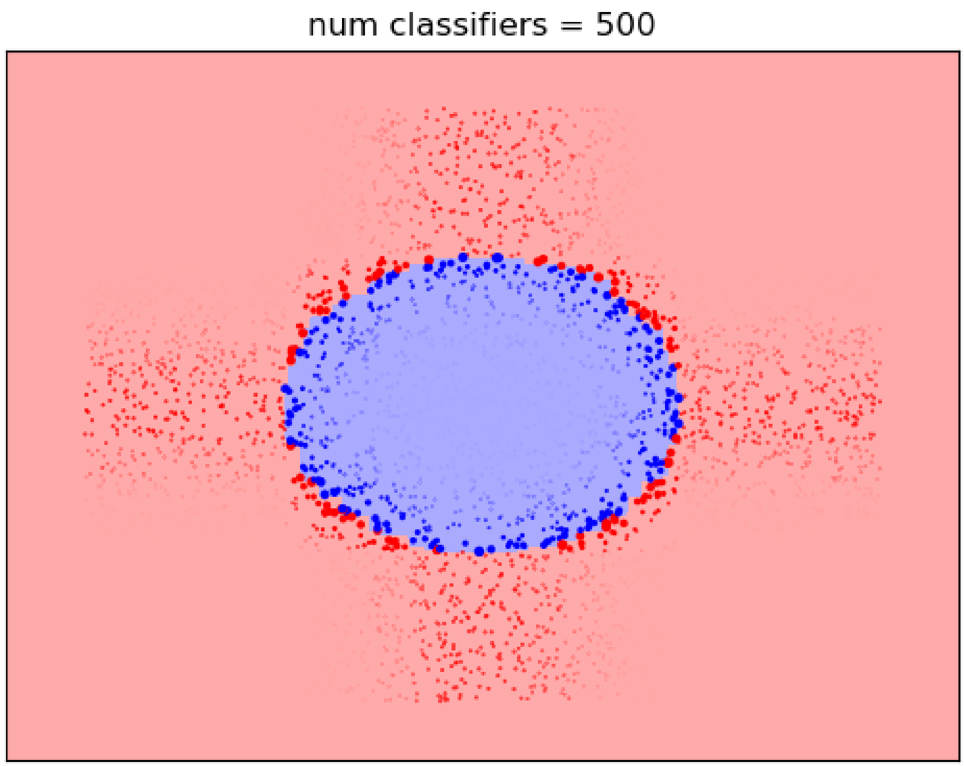


סעיף 15:



ה־T שמזער את השגיאה הוא 257 עם שגיאה 0.

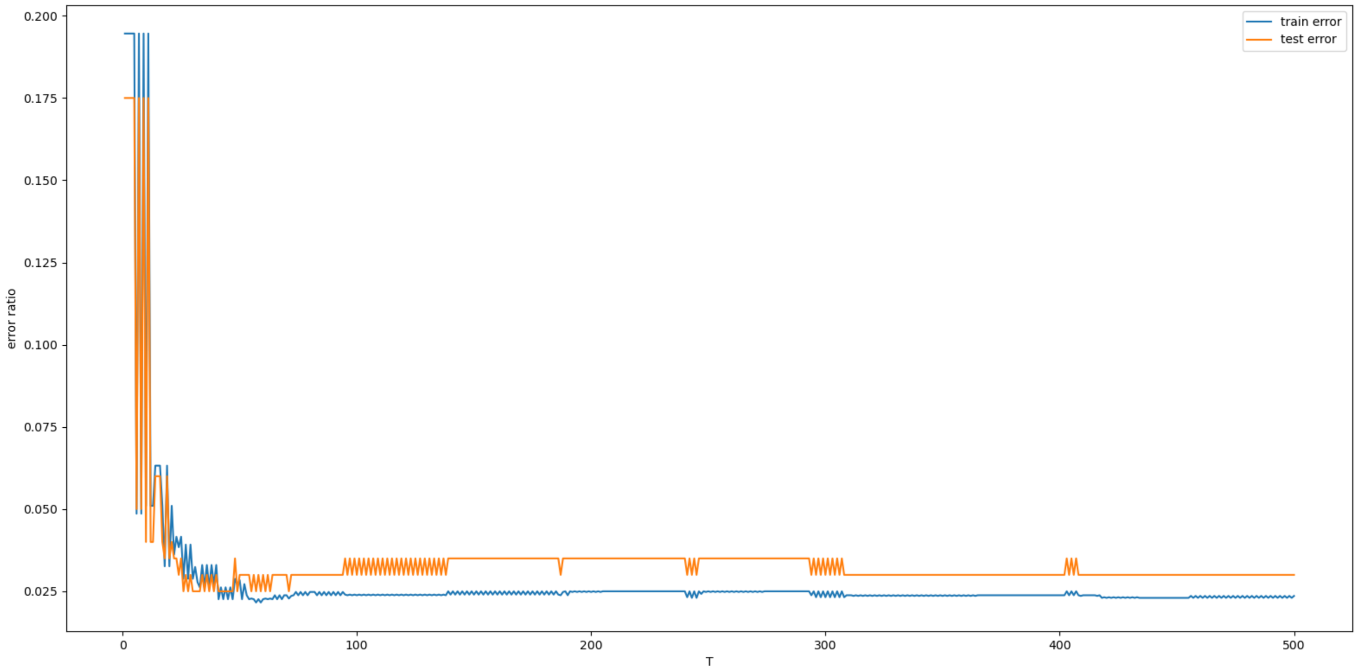
סעיף 16:



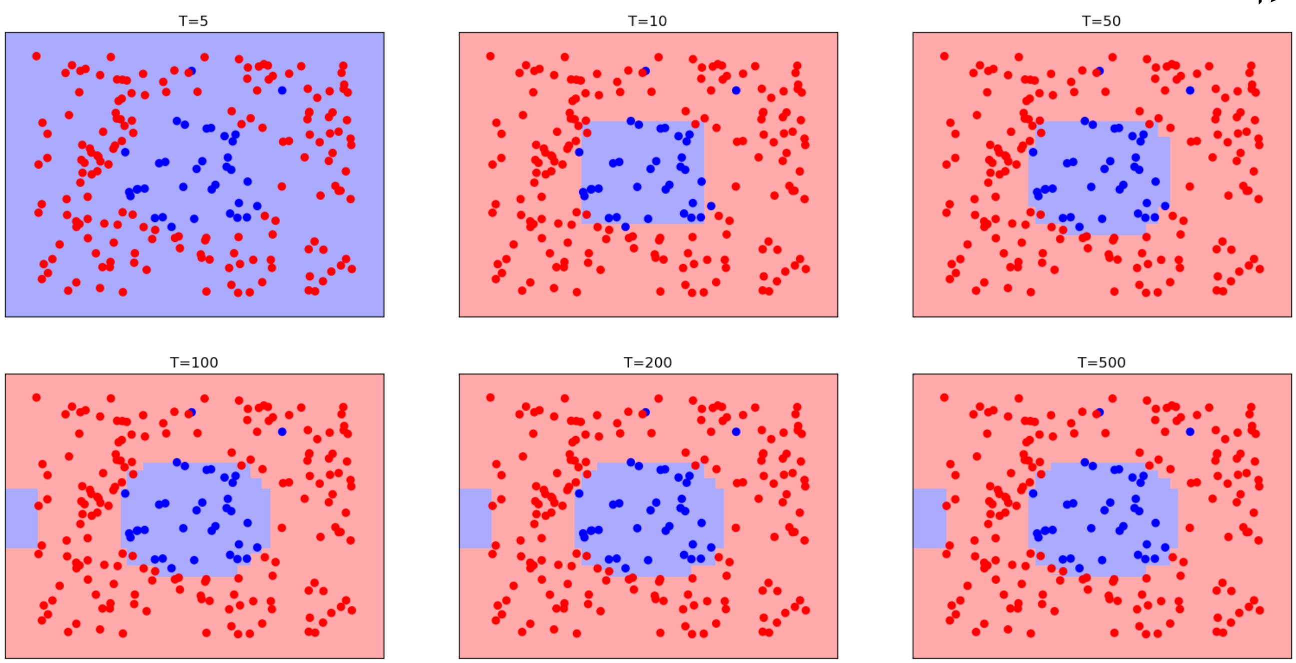
ניתן לראות שסביב העיגול הכחול המשקלים גבוהים יותר ־ זהו איזור בו המסווג לא בטוח בהחלטתו (איזור בו יש ריכוז גבוה של שתי המחלקות) וכנראה טעה הרב פעמים במהלך האיטרציות ולכן המשקלים גדולים.
לגבי 4 האיזורים האדומים עם משקלים גדולים יותר ־

סעיפים 13-16 עם רעש 0.01:

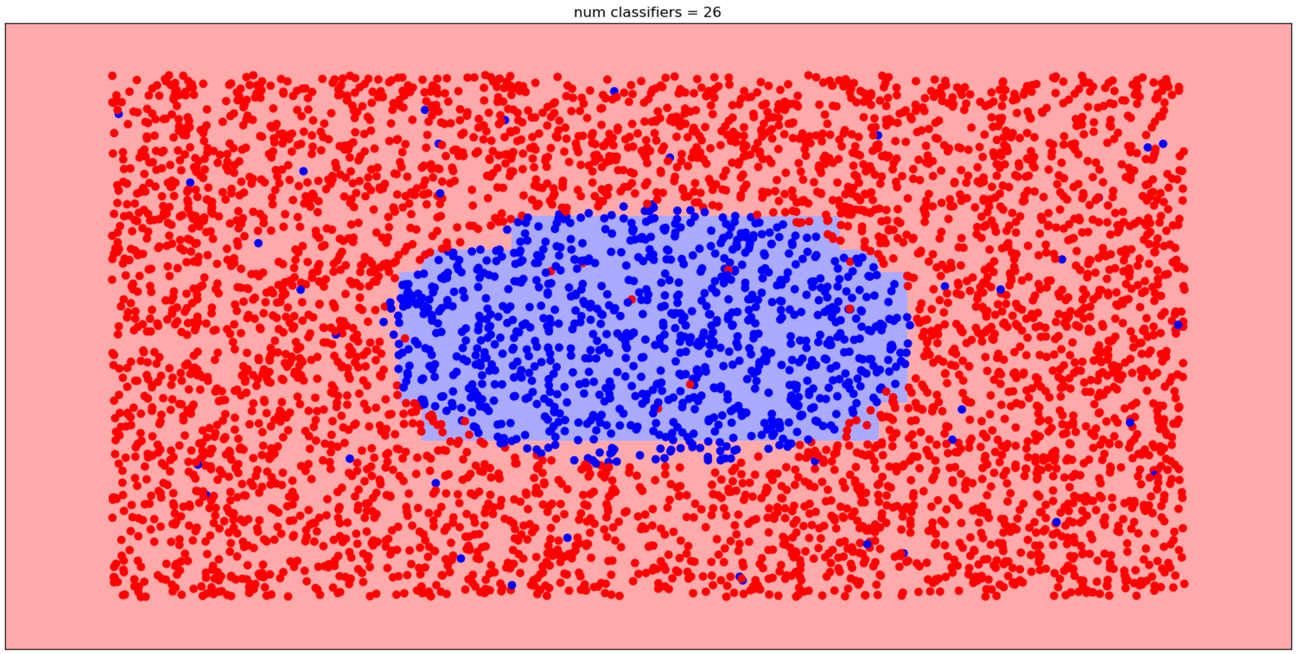
סעיף 13:



סעיף 14:

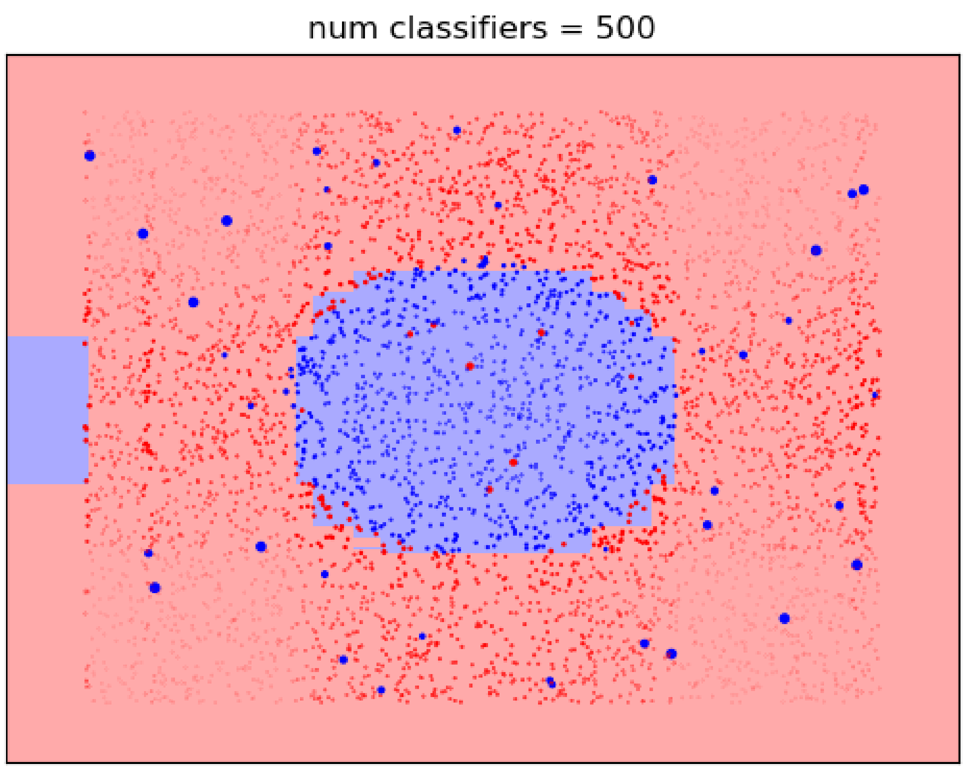


סעיף 15:



ה־T שמזער את השגיאה הוא 26 עם שגיאה 0.025.

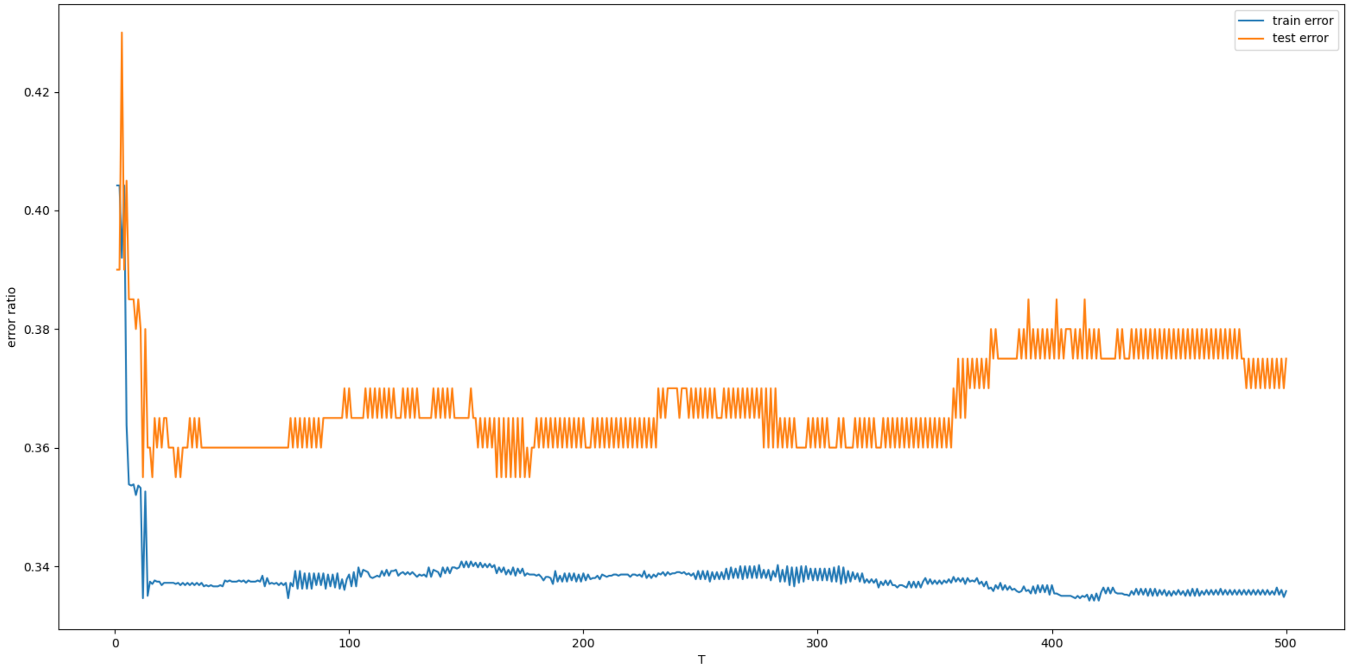
סעיף 16:



הנקודות הכחולות על הרקע האדום הם בעצם רעש ואנו רואים שהמסווג מתקשה לסווג אותן, ובגרף זה ניתן לראות בהתאם שהמשקלים שהוא נותן לנקודות האלה גדולים יחסית. זה גם נכון לנקודות האדומות שנמצאות בתוך העיגול הכחול.

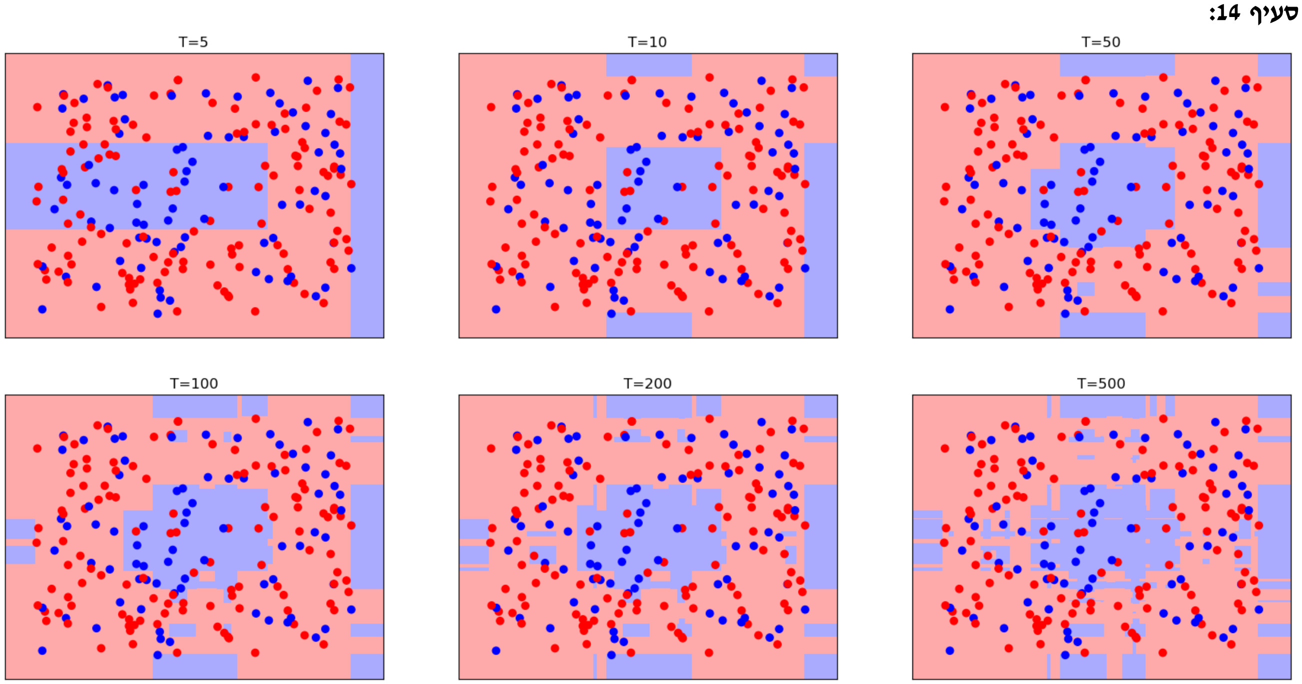
סעיפים 13-16 עם רעש 0.4:

סעיף 13:

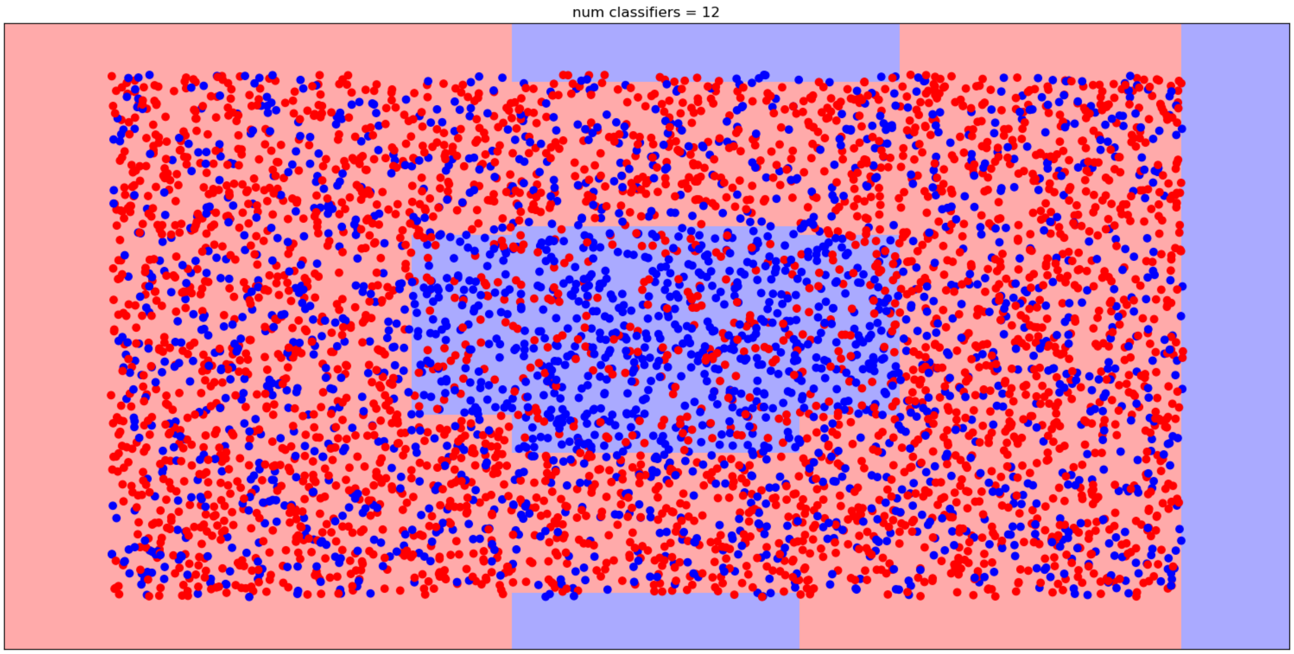


הדאטה רועש,

סעיף 14:

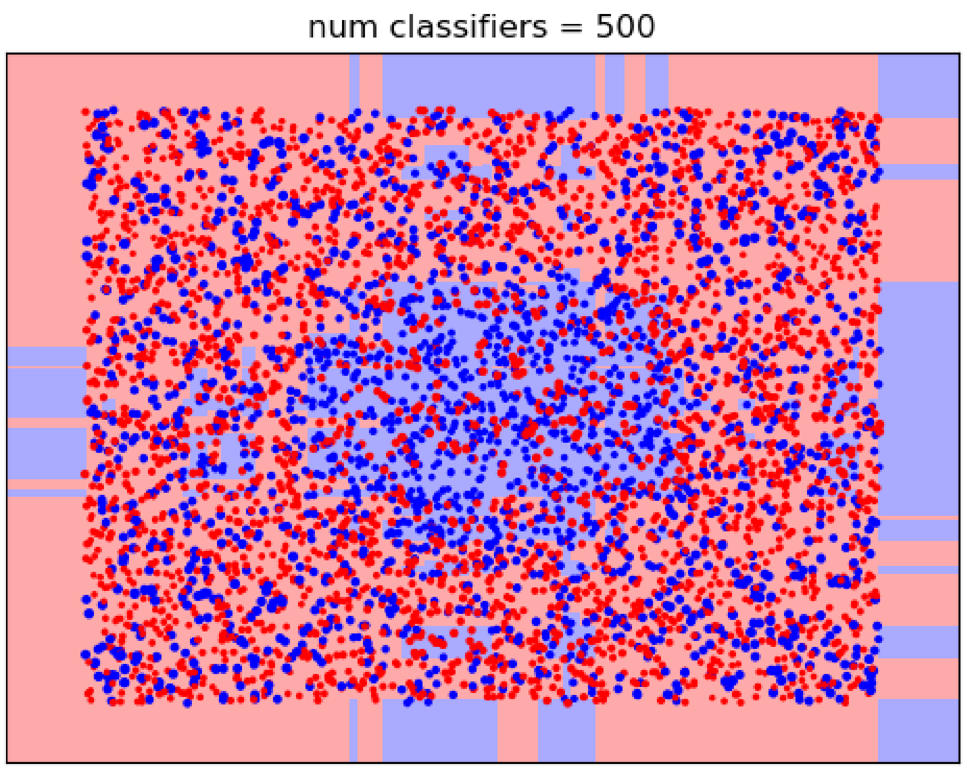


סעיף 15:



ה־T שמזער את השגיאה הינו 12 עם שגיאה 0.355.

סעיף 16:



ככל שהדאטה רועש יותר כך המסווג טועה יותר.