# תרגיל 4

4 ביוני 2020

דויד פונרובסקי 208504050

## 1 תרגיל 1:

כיוון ראשון, נחשב את תוחלת השגיאה באופן ישיר: $\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(A\left(S\right)\right)\right]=\int_{[0,1]}L_{\mathcal{D}}\left(A\left(S\right)\right)\cdot\mathcal{P}\left(L_{\mathcal{D}}\left(A\left(S\right)\right)\right)$ כאשר $\mathcal{P}\left(L_{\mathcal{D}}\left(A\left(S\right)\right)\right)$ היא צפיפות ההסתברות של פונקציית המחיר. נקבל כי:

$$\begin{aligned}
&\int_{[0,1]}L_{\mathcal{D}}\left(A\left(S\right)\right)\cdot\mathcal{P}\left(L_{\mathcal{D}}\left(A\left(S\right)\right)\right)=\\
&\int_{0}^{\varepsilon}L_{\mathcal{D}}\left(A\left(S\right)\right)\cdot\mathcal{P}\left(L_{\mathcal{D}}\left(A\left(S\right)\right)\right)+\int_{\varepsilon}^{1}L_{\mathcal{D}}\left(A\left(S\right)\right)\cdot\mathcal{P}\left(L_{\mathcal{D}}\left(A\left(S\right)\right)\right)\\
&\leq\varepsilon\cdot\sup_{L_{\mathcal{D}}(A(S))\in[0,\varepsilon]}\left\{\mathcal{P}\left(L_{\mathcal{D}}\left(A\left(S\right)\right)\right)\right\}+\left(1-\varepsilon\right)\cdot\sup_{L_{\mathcal{D}}(A(S))\in[\varepsilon,1]}\left\{\mathcal{P}\left(L_{\mathcal{D}}\left(A\left(S\right)\right)\right)\right\}\\
&\leq\varepsilon\cdot M+\left(1-\varepsilon\right)\cdot\delta
\end{aligned}$$

ומכאן שנוכל ליבחור $m\left(\varepsilon,\delta\right)$ כך שהתוחלת תיהיה קטנה כרצוננו.
כיוון שני נניח כי התוחלת שואפת ל $0$ כאשר $m$ שואף לאינסוף ונקבל מאישיון מרקוב כי:

$$\begin{aligned}
\mathbb{P}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(A\left(S\right)\right)\leq\varepsilon\right]=1-\mathbb{P}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(A\left(S\right)\right)\geq\varepsilon\right]\geq\\
\geq1-\frac{\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(A\left(S\right)\right)\right]}{\varepsilon}
\end{aligned}$$

ומאחר $\lim_{m\to\infty}\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(A\left(S\right)\right)\right]=0$ אז ניתן ליבחור $m\left(\varepsilon,\delta\right)$ מספיק גדול כך ש - $\mathbb{E}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(A\left(S\right)\right)\right]\leq\varepsilon\cdot\delta$ ולקבל כי

$$\mathbb{P}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}\left(A\left(S\right)\right)\leq\varepsilon\right]\geq1-\delta$$

## 2 תרגיל 2:

נסמן ב $r_M$ את הרדיוס המקסימלי של נקודה בעלת ערך $1$ וב $r_m$ את רדיוס המנימלי של נקודה בעלת ערך 0. ונגדיר את $\varepsilon=\frac{r_m-r_M}{2}$. ניתבונן במכונה המחזירה את

$$A_{\varepsilon}\left(S\right)\mapsto h_{r_M+\varepsilon}$$

מכאן שקיים $x$ כך ש- $\mathcal{L}_D\left(h_{r_M+\varepsilon}\left(x\right)\right)\geq\varepsilon$ אם כל הנקודות בדגימה $S$ נפלו ברדיוס הקטן למ $x-\varepsilon$. כלומר .

$$\begin{aligned}
\mathcal{P}\left(\mathcal{L}_D\left(h_{r_M+\varepsilon}\left(x\right)\right)\geq\varepsilon\right)=\mathcal{P}^m\left(x'\in[0,x-\varepsilon]\right)=\\
\left(1-\mathcal{P}\left(x'\in[x-\varepsilon,x]\right)\right)^m\leq e^{-m\varepsilon}
\end{aligned}$$

מכאן שאם נסמן $\mathcal{P}\left(\mathcal{L}_D\left(h_{r_M+\varepsilon}\left(x\right)\right)\geq\varepsilon\right)=\delta$ נקבל כי לכל $m\geq m\left(\varepsilon,\delta\right)=\frac{\log\left(\frac{1}{\delta}\right)}{\varepsilon}$ ההסתברות ש $\mathcal{L}_D\left(h_{r_M+\varepsilon}\left(x\right)\right)\leq\varepsilon$ גדולה מ $1-\delta$ ■

# 3 תרגיל 3:

לפי הטענה מתירגול, מאחר ו $\mathcal{H}_{con}$ סופית, אז $VC\dim\left(\mathcal{H}_{con}\right) \leq \log\left(|\mathcal{H}_{con}|\right) = d$ נראה כי קיים סט $S \in \{0,1\}^{d+1}$ בגודל $d$ ה"מתפזר" ($shatters$) על פני כל הפונק' מ $\{0,1\}^d \to \{0,1\}$ :$h$. נסמן ב $S_i$ את הדגימה ה$i$ית וב $S_i^j$ את הקרודינטה ה$j$ית שלה. ניתבונן ב-

$$S = \left\{ (S_i, 0) \,|S_i^j = \delta_i^j \;\; \forall i \in [d] \right\}$$

.

ניתבונן ב $q$ פסוקית אפשרית. ונסמן ב $A = \{i|\bar{x}_i \in q\}$ (קבוצת האינדקסים של ליטרל שלילה) ובאופן זהה את $B = \{i|x_i \in q\}$. נראה ש $q\left(S_i\right) = 0$ לכל $i$.

$$q\left(S_i\right) = \bigwedge_{j\in B} x_j \left(S_i^j\right) \bigwedge_{j\in A} \bar{x}_j \left(S_i^j\right) = \bigwedge_{j\in B} x_j \left(\delta_i^j\right) \bigwedge_{j\in A} \bar{x}_j \left(\delta_i^j\right) = \begin{cases} 1 \bigwedge_{j\in B} x_j \left(\delta_i^j\right) \bigwedge_{j\in A, j\neq i} \bar{x}_j \left(\delta_i^j\right) & i \in A \\ 0 \bigwedge_{j\in B, j\neq i} x_j \left(\delta_i^j\right) \bigwedge_{j\in A} \bar{x}_j \left(\delta_i^j\right) & i \in B \end{cases}$$
$$= \begin{cases} 1 \bigwedge_{j\in B} 0 \bigwedge_{j\in A, j\neq i} 1 & i \in A \\ 0 \bigwedge_{j\in B, j\neq i} 0 \bigwedge_{j\in A} 1 & i \in B \end{cases} = 0$$

מאחר ו $q$ יכולה להיות כל פסוקית, קיבלנו כי האילוצים עדיין מאפשרים מיפוי אל כל הפונקציות האפשריות. ולכן $VC\dim\left(\mathcal{H}_{con}\right) \geq d$ מכאן ש $VC\dim\left(\mathcal{H}_{con}\right) = d$ ■

# 4 תרגיל 4:

מאחר ו $\mathcal{H}$ היא מתכנסת באופן אחיד ($uniform\, convergence$) אז קיים $m_{\mathcal{H}}^{UC}\left(\varepsilon, \delta\right)$ כך שלכל $m \geq m_{\mathcal{H}}^{UC}\left(\varepsilon, \delta\right)$ ולכל $\mathcal{D}$ התפלגות מעל $x$. מתקיים כי $\mathcal{P}\left(\{S \in (x,y)^m\}\; is\, \varepsilon\, representative\right) \geq 1-\delta$. ולכן

$$\mathcal{P}\left(\mathcal{L}_{\mathcal{D}}\left(h\right) \leq \varepsilon\right) = \mathcal{P}\left(\left\{\mathcal{L}_S\left(h\right) \leq \mathcal{L}_{\mathcal{D}}\left(h^*\right) + \varepsilon\right\} \bigcap \left\{S\;is\;\frac{\varepsilon}{2}\;representative\right\}\right) +$$
$$+\mathcal{P}\left(\left\{\mathcal{L}_S\left(h\right) \leq \varepsilon\right\} \bigcap \left\{\overline{S\;is\;\frac{\varepsilon}{2}\;representative}\right\}\right) \geq$$
$$\geq \mathcal{P}\left(\left\{\mathcal{L}_S\left(h\right) \leq \mathcal{L}_{\mathcal{D}}\left(h^*\right) + \varepsilon\right\} \bigcap \left\{S\;is\;\frac{\varepsilon}{2}\;representative\right\}\right)$$

מיצד שני ואנו מוצאים את $h$ לפי כלל $REM_{\mathcal{H}}\left(S\right)$ שמבטיח כי $h = \arg\min_{h\in\mathcal{H}}\left\{\mathcal{L}_S\left(h\right)\right\}$ אז $\mathcal{L}_S\left(h\right) \leq \mathcal{L}_S\left(h^*\right)$ כאשר

$$h^* = \arg\min_{h\in\mathcal{H}}\left\{\mathcal{L}_{\mathcal{D}}\left(h\right)\right\}$$

נזכיר שוב כי אנו מסתכלים על המצב בו $S$ הוא מייצג $\frac{\varepsilon}{2}$ פעם נוספת. ונקבל כי

$$\mathcal{L}_S\left(h\right) \leq \mathcal{L}_S\left(h^*\right) \leq \mathcal{L}_{\mathcal{D}}\left(h^*\right) + \frac{\varepsilon}{2}$$

ולכן אם $\mathcal{L}_S\left(h\right) \leq \frac{\varepsilon}{2}$ אז גם $\Leftarrow$ $\mathcal{L}_S\left(h\right) \leq \mathcal{L}_{\mathcal{D}}\left(h^*\right) + \varepsilon$. כלומר קיבלנו כי $\left\{\mathcal{L}_S\left(h\right) \leq \mathcal{L}_{\mathcal{D}}\left(h^*\right) + \varepsilon\right\} \subseteq \left\{S\;is\;\frac{\varepsilon}{2}\;representative\right\}$ ולכן :

$$\mathcal{P}\left(\left\{\mathcal{L}_S\left(h\right) \leq \mathcal{L}_{\mathcal{D}}\left(h^*\right) + \varepsilon\right\} \bigcap \left\{S\;is\;\frac{\varepsilon}{2}\;representative\right\}\right) = \mathcal{P}\left(\left\{S\;is\;\frac{\varepsilon}{2}\;representative\right\}\right) \geq 1-\delta$$

לכל $m$ הגדול מ $m_{\mathcal{H}}^{UC}\left(\frac{\varepsilon}{2}, \delta\right)$ נסיק לסיום $m\left(\varepsilon, \delta\right) \leq m_{\mathcal{H}}^{UC}\left(\frac{\varepsilon}{2}, \delta\right)$ ■

# 5 תרגיל 6:

ניתבונן ב $m_1 = m_{\mathcal{H}}\left(\varepsilon_1, \delta\right)$ ובאופן דומה נגדיר את $m_2$. לפי ההגדרה לכל $m \geq m_1$ מתקיים כי $\mathcal{P}\left(\mathcal{L}_{\mathcal{D}^m}\left(h\right) \leq \varepsilon_1\right) \geq 1-\delta$ לכל התפלגות $\mathcal{D}$. נניח כי $\varepsilon_2 \geq \varepsilon_1$ מכאן ש $\mathcal{P}\left(\mathcal{L}_{\mathcal{D}^m}\left(h\right) \leq \varepsilon_1\right) \leq \mathcal{P}\left(\mathcal{L}_{\mathcal{D}^m}\left(h\right) \leq \varepsilon_2\right)$ מיצד שני $\mathcal{P}\left(\mathcal{L}_{\mathcal{D}^m}\left(h\right) \leq \varepsilon_1\right) \geq 1-\delta$ לכל $m \geq m_1$ ולכן נקבל כי לכל $m \geq m_1$ מתקיים בנוסף כי $\mathcal{P}\left(\mathcal{L}_{\mathcal{D}^m}\left(h\right) \leq \varepsilon_2\right) \geq 1-\delta$ ולכן $m_1 \geq m_{\mathcal{H}(\varepsilon_2,\delta)} =$ $m_2$. (אם לא היה גדול מימנו אז היה ניתן לבחור אותו בתור $m_{\mathcal{H}(\varepsilon_2,\delta)}$ וזאת הייתה סתירה למינמליות) ■

# 6 תרגיל 8:

## 6.1 סעיף א:

$\tau_{\mathcal{H}}(m) = \max\{|\mathcal{H}_C| : C \subset \mathcal{X}, |C| = m\}$ יהיה למעשה המספר הממקסימלי של פונקציות מ $\mathcal{H}$ תחת סט אילוצים $C$ בגודל $m$.

## 6.2 סעיף ב:

אם $VC\dim(\mathcal{H}) = \infty$ אז למעשה כל סט אילוצים $C$ ממאפשר לפחות $2^{|C|}$ פונקציות (מידע לא מוריד אופציות) ולכן עבור בעיות כלסיפיקציה נקבל כי $\tau_{\mathcal{H}}(m) = \max\{|\mathcal{H}_C| : C \subset \mathcal{X}, |C| = m\} = 2^{|C|} = 2^m$.

## 6.3 סעיף ג:

נראה כי $\frac{\tau_{\mathcal{H}}(m)}{2^m}$ מונוטונית יורדת, ונסיק כי אם כמו בסעיף הקודם $2^m$, אחרת אם $\tau_{\mathcal{H}}(m<d) \neq 2^m$ אז זאת סתירה לכך ש $VC\dim(\mathcal{H}) = d$.
הוכחה,

$$\begin{aligned}
\frac{\tau_{\mathcal{H}}(m+1)}{2^{m+1}} &= \frac{1}{2^{m+1}} \max_C \left\{ |\mathcal{H}_C| : C \subset \mathcal{X}, |C| = m+1 \right\} = \\
&\frac{1}{2^{m+1}} \max_{C',x} \left\{ |\mathcal{H}_C| : \{x\} \cup C' \subset \mathcal{X}, |C'| = m, x \in \mathcal{X}, |x| = 1 \right\} = \\
&\leq \frac{1}{2^{m+1}} \max_{C'} \left\{ |\mathcal{H}_{C'}| : C' \subset \mathcal{X}, |C'| = m \right\} \times \max_x \left\{ |\mathcal{H}_x| : x \subset \mathcal{X}, |x| = 1 \right\} = \\
&\leq \frac{\tau_{\mathcal{H}}(m)}{2^{m+1}} \cdot \max_x \left\{ |\mathcal{H}_x| : x \subset \mathcal{X}, |x| = 1 \right\} \leq \frac{\tau_{\mathcal{H}}(m)}{2^{m+1}} \cdot 2 = \frac{\tau_{\mathcal{H}}(m)}{2^m}
\end{aligned}$$

מכאן שאם קיים $m < d$ כך שלכל סט $C$ בגדול $m$ מתקיים כי $|\mathcal{H}_C| < 2^m$. אז גם לכל $m' \geq m$ מתקיים כי $\frac{\tau_{\mathcal{H}}(m')}{2^{m'}} < 1$ ובפרט עבור $d$ בסתירה לכך ש $VC\dim(\mathcal{H}) = d$ ■

## 6.4 סעיף ד:

נראה תחילה כי $|\mathcal{H}_C| \leq |\{B \subset C : \mathcal{H}\, shutters\, B\}|$. נניח באינדוקציה עבור $|C'| < m$ וניתבונן ב $C$ בגודל $m$. נסמן ב $c \in C$ דגימה שרירותית נסמן ב $\xi$ את כל הפונקציות ב $\mathcal{H}_C$ כך ש $c$ לא מוסיפה אינפורמציה עלייהם (תחת האילוצים האחרונים ב $C$) כלומר

$$\xi = \{h \in \mathcal{H}_C : h(c, C/\{c\}) = h(\bar{c}, C/\{c\}) = 1\}$$

נסמן ב $\zeta$ את קבוצת כל הפונקציות ש $c$ מהווה אילוץ אמיתי עבורם כלומר

$$\zeta = \left\{ h \in \mathcal{H}_C : h(c, C/\{c\}) = \overline{h(\bar{c}, C/\{c\})} \right\}$$

. כמובן ש$|\mathcal{H}_C| = |\xi| + |\zeta|$.
נחסום את $\xi$ אם $h \in \xi$ אז כמובן שגם $h \in \mathcal{H}_{C/\{c\}}$ ולכן לפי ההנחה $|A| \leq |\{B \subset C/\{c\} : \mathcal{H}\, shutters\, B\}|$.
נחסום את $\zeta$. ניתבונן במחלקת ההיפותזות $\mathcal{H}'$ שמיצד אחד מקיימת $\forall h \in \mathcal{H}' \subset \mathcal{H}, \forall C' \in \mathcal{X} : |C'| \leq m$ מתקיים כי $h(\bar{c}, C') = 0$. כלומר כל הפונקציות ב$\mathcal{H}$ כך ש $h(c) = 1$ הוא תנאי סף עבורם. **לכן דגימה של** $c$ **אינה מוסיפה אינפורמציה שמשפיעה על החסם של** $\mathcal{H}'_{C' \cup \{c\}}$. ולכן חזרנו אל המקרה הראשון ונקבל כי

$$\begin{aligned}
|\zeta| = |&\left\{ h \in \mathcal{H}_C : h(c, C/\{c\}) = \overline{h(\bar{c}, C/\{c\})} \right\}| = \\
&|\mathcal{H}'_C| = |\mathcal{H}'_{C/\{c\}}| = |\{B \subset C/\{c\} : \mathcal{H}'\, shutters\, B\}| = |\left\{ B \subset C \bigwedge \{c\} \in B : \mathcal{H}\, shutters\, B \right\}|
\end{aligned}$$

.

$$\begin{aligned}
|\mathcal{H}_C| &\leq |\xi| + |\zeta| \leq |\left\{ B \subset C/\{c\} : \mathcal{H}\, shutters\, B \right\}| + |\left\{ B \subset C : \{c\} \in B \bigwedge \mathcal{H}\, shutters\, B \right\}| \\
&= |\left\{ B \subset C/\{c\} : \mathcal{H}\, shutters\, B \right\}| + \left\{ B \subset C \bigwedge c \in B \bigwedge \mathcal{H}\, shutters\, B \right\} = |\left\{ B \subset C : \mathcal{H}\, shutters\, B \right\}|
\end{aligned}$$

ומכאן כמובן נקבל את החסם הדרוש $\tau_{\mathcal{H}}(m) = \max\{|\mathcal{H}_C| : C \subset \mathcal{X}, |C| = m\} \leq \sum_0^d \binom{m}{i} \leq \left(e\frac{m}{d}\right)^d$ ■

## 6.5 סעיף ה:

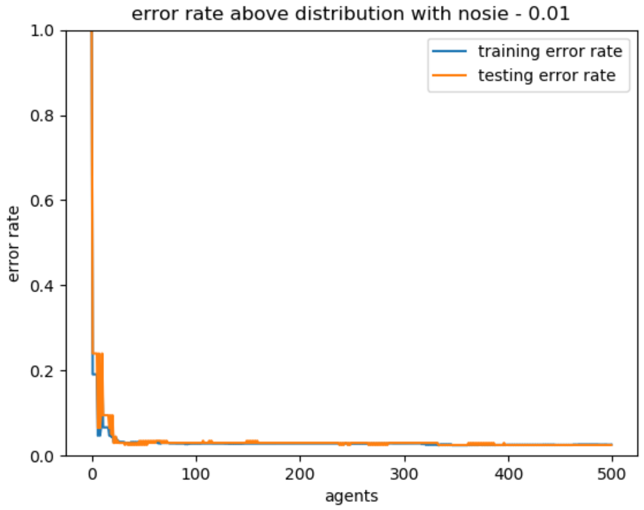
כמובן שמחזיק וכמובן שלא הדוק : $\tau_{\mathcal{H}}(d) = 2^d \leq e^d$.

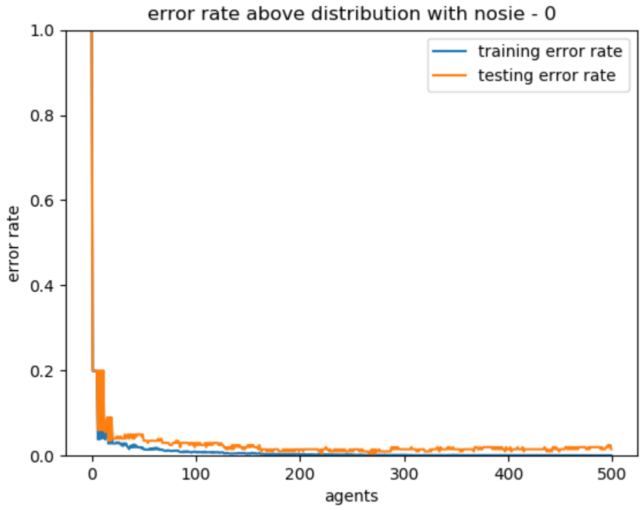
### 6.6 סעיף ו:

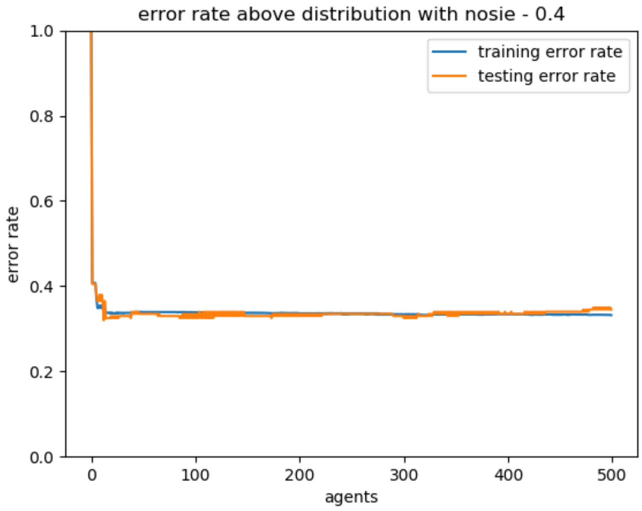
ניתן להגדיר את $VC\dim\left(\mathcal{H}\right)$ בתור נקודת "המפנה" בה $\tau_{\mathcal{H}}$ עוברת מלהיות אקספוננציאלי ב $m$ לפולניאמלי ב $m$.

# 7 תרגיל תיכנותי :

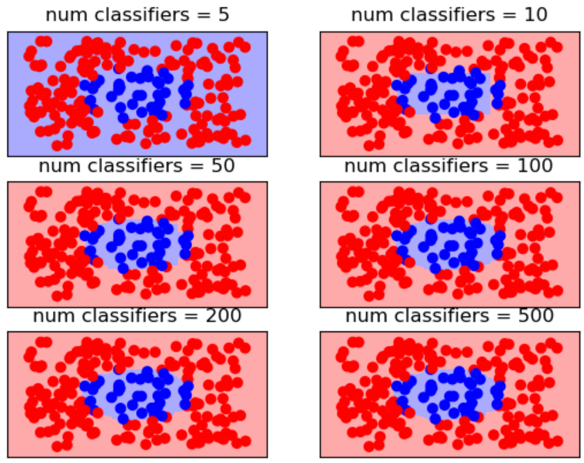
תוצאות מרשימות ביותר, כמובן כפי שהיינו מצפים עבור רעש גדול באופן משמעותי למימדי ההתפלגות מקבלים רף שמימנו כבר לא יורדים ( עבור רעש כזה פונקציית מדרגה אינה מהווה כבר $\gamma - weak\, learner$ ולמעשה אני בכלל לא בטוח שהנחת ה - $realizable$ עומדת ). בכל אופן גם במצב הזה, חסם השגיאה על האימון הדוק לחסם שגיאת המבחן שזה נחמד.

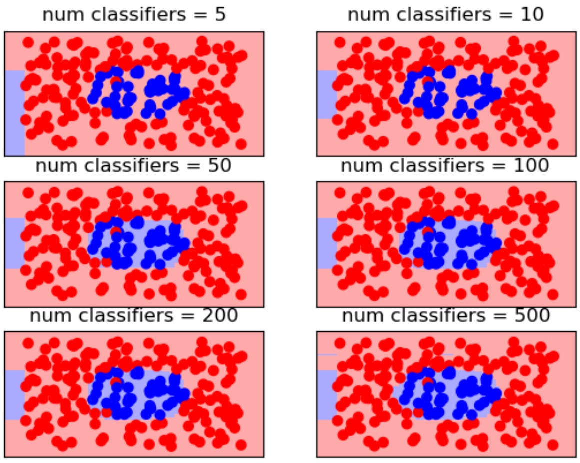


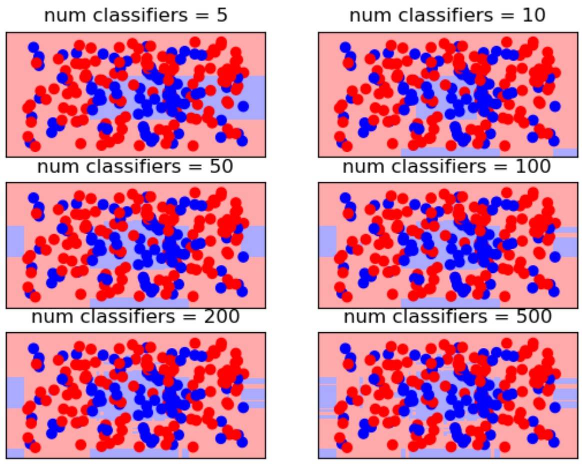



כאוף ניתן לראות, את הצלחת הכלפסיפציה.



כאן ניתן ליראות המשקלים שניתנים על ידי המכונות לנקודות.