שלום לבודקים!
יום לפני ההגשה, המסמך הושחת ולא ניתן לפתוח אותו. ניסיתי לשחזר את הקובץ מגירסאות קודמות אך גם הן נמחקו, ולבסוף הורדתי תוכנה שמשחזרת קבצים (כפי שתראו בהמשך -בהצלחה שנוייה במחלוקת) וצילמתי את המסך כי לא ניתן לשחזר ללא מנוי😊

אז לעיתים הסימנים קצת התהפכו אבל אני חושבת שניתן להבין את הלוגיקה בהוכחות.

מחילה על תרגיל הפרנקנשטיין והצלחה בבדיקת התרגיל.

יום טוב! בר.

תרגיל IML 4

Bar Aloni

[redacted]

PAC Learnability

1. מכיח ש המשפטים הבאים שקולים:

(1) $\exists m(\epsilon,\delta)\ \forall\, \delta,\epsilon > 0\ \forall m \geq m(\epsilon,\delta)$ s.t: $\mathbb{P}_{S\sim D^m}\left[L_D(h_S) \leq \epsilon\right] \geq 1-\delta$

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

$(1)\to(2)$: נבחר $\epsilon, \delta > 0$, אז קיים $m(\epsilon,\delta) \in \mathbb{N}$ כך שלכל $m \geq m(\epsilon,\delta)$:

$\mathbb{P}_{S\sim D^m}\left[L_D(h_S) \leq \epsilon\right] \geq 1-\delta$ פונקציית ה-loss מקבלת ערכים בקבוצה $\{0,1\}$, אז פונקציית השגיאה המוכללת מקבלת ערכים בתחום $[0,1]$ (ובפרט $0 \leq X_S(h_S) := L_D$ מ"מ א"ש) ולכן תוחלתה גם היא בתחום הנ"ל, ובפרט: $0 \leq \mathbb{E}_{S\sim D^m}\left[L_D(h_S)\right] < \infty$.

מתקיים:

$$0 \leq \mathbb{E}_{S\sim D^m}\left[L_D(h_S)\right] = \int_{-\infty}^{\infty} x\cdot f_X(x)dx =$$

$$= \int_0^1 x\cdot f_X(x)dx = \int_0^\epsilon x\cdot f_X(x)dx + \int_\epsilon^1 x\cdot f_X(x)dx \leq$$

כאשר

$$\leq \int_0^\epsilon \epsilon\cdot f_X(x)dx + \int_\epsilon^1 f_X(x)dx = \epsilon F_X(\epsilon) + \int_0^1 f_X(x)dx - \int_0^\epsilon f_X(x)dx =$$

$$= \epsilon F_X(\epsilon) + 1 - F_X(\epsilon) = 1 - F_X(\epsilon)(1-\epsilon) = 1 - \mathbb{P}(X \leq \epsilon)(1-\epsilon) \leq$$

$$\leq 1 - (1-\delta)(1-\epsilon) = \epsilon + \delta - \delta\cdot\epsilon$$

א"ש האחרון מתאפשר כי אנחנו עומדי בתנאי א"ש מרקוב.

כלומר קיבלנו שישם $m(\epsilon,\delta) \in \mathbb{N}$ כך שלכל $m \geq m(\epsilon,\delta)$:
$0 \leq \mathbb{E}_{S\sim D^m}\left[L_D(h_S)\right] \leq \epsilon + \delta - \delta\cdot\epsilon$ ומכיוון שהדבר מתקיים לכל ϵ,δ קטנים כרצוננו (בפרט כאלו עבורם $\epsilon + \delta - \delta\cdot\epsilon$ א"ש) ובצירוף כלל הסנדוויץ', נקבל ש: $\lim_{m\to\infty} \mathbb{E}_{S\sim D^m}\left[L_D(h_S)\right] = 0$.

כנדרש.

$(2)\to(1)$: נבחין ש: $L_D =: X_S(h_S) \geq 0$ מכיון שפונקצית הloss א"ש, אז התוחלת שלה (פונקציית השגיאה המוכללת) היא א"ש.
כמו כן, מקיום הגבול ומכך שהוא שווה ל-0 מתקיים עבור $\epsilon, \delta > 0$, שנגדיר,
ו- $\hat{\epsilon} = \epsilon\cdot\delta > 0$ שנגדיר שקיים $m \in \mathbb{N}(\epsilon,\delta)$ כך שלכל $m \geq m(\epsilon,\delta)$:

$\mathbb{E}_{S\sim D^m}[L_D(h_S)] = \left|\mathbb{E}_{S\sim D^m}[L_D(h_S)] - 0\right| < \hat{\epsilon} = \epsilon \cdot \delta$ (*)

כאשר השוויון השמאלי ביותר נבע מכך שתוחלת של ממים א"ש גם היא א"ש ולכן הערך המוחלט מיותר. בפרט נשים לב שהראים ש $\mathbb{E}_{S\sim D^m}[X_S] < \infty$

אזי X הוא מ"מ א"ש בעל תוחלת סופית ולכן אנחנו יכולים לחסום אותו בעזרת א"ש מרקוב:

מרקוב:

מתקיים: $\mathbb{P}_{S\sim D^m}[X_S \le \epsilon] = 1 - \mathbb{P}_{S\sim D^m}[X_S > \epsilon] \le \frac{\mathbb{E}_{S\sim D^m}[X_S]}{\epsilon}$, ולכן:

$\mathbb{P}_{S\sim D^m}[X_S \le \epsilon] \ge 1 - \frac{\mathbb{E}_{S\sim D^m}[X_S]}{\epsilon}$.

מ-$(*)$ אנחנו מקבלים שלכל $\epsilon, \delta > 0$, קיים $m \in \mathbb{N}(\epsilon,\delta)$ כך שלכל $m \ge m(\epsilon,\delta)$: $\mathbb{E}_{S\sim D^m}[X_S] < \epsilon \cdot \delta$, כלומר:

$\mathbb{P}_{S\sim D^m}[X_S \le \epsilon] \ge 1 - \frac{\mathbb{E}_{S\sim D^m}[X_S]}{\epsilon} \ge 1 - \frac{\epsilon \cdot \delta}{\epsilon} = 1 - \delta$ כנדרש.

2. יהיו $\epsilon, \delta > 0$ התפלגות D על $\mathbb{R}^2$, ופונקציית תיוג $h_{r^*} \in \mathcal{H} = \{1_{[\|x\|_2 \le r]} : r \in \mathbb{R}_+\}$

נראה ש-$\mathcal{H}$ למידת PAC:

עבור $m \ge m(\epsilon,\delta)$ יהא המדגם $S = \{(x_i,y_i)\}_{i=1}^{m}$ שהדגימות בו נדגמו iid מ-D

נציע את המסווג $h_r \in \mathcal{H}$ עבור $r := \max_{y_i = 0} \|x_i\|_2$, ונראה שהוא אכן למדן-PAC למחלקה.

ראשית נבחין ש-S נדגם iid מ-$\mathbb{R}^2$, וש-h_r מסווג לפי תיוג h_{r^*} **על S**. כלומר $r \le r^*$: אם תייגנו את x ב-1, אז הוא בהכרח תוייג ב-1 לפי h_{r^*}. וגם השלילה מתקיימת-מתקיימת: אם x תוייג ב-0 לפי h_{r^*} אז הוא בהכרח תוייג ב-0 במסווג שלנו. כלומר: השגיאה שלנו יכולה להתרחש רק על x המקיים $\|x\|_2 \in (r, r^*)$.

נחלק למקרים:

- $D(\{x : \|x\|_2 \in (0, r^*)\}) < \epsilon$: אז $D(\{x : \|x\|_2 \in (r, r^*)\}) < \epsilon$ כלומר בהסתברות 1 מתקיים: $L_{D,h_{r^*}}(h_r) < \epsilon$.
- $D(\{x : \|x\|_2 \in (0, r^*)\}) \ge \epsilon$: ניתן לקבוע R המקיים:

$D(\{x : \|x\|_2 \in (R, r^*)\}) = \epsilon$. נבחין שאם יש x המקיים את התנאי הזה, אזי $R \le r$ (מבחירת $r := \max_{y_i=0} \|x_i\|_2$) ולכן $D(\{x : \|x\|_2 \in (r, r^*)\}) \le \epsilon$, כלומר $L_{D,h_{r^*}} \le \epsilon$.

כעת, נשים לב שההסתברות לקבל מדגם **ללא** נקודות כאלו היא $(1-\epsilon)^m$ וכי מתקיים: $(1-\epsilon)^m < 1-\delta$ אמ"ם $m \ge \frac{\ln(1\backslash\delta)}{\epsilon}$

כלומר קיבלם שהחלקה למידת PAC עבור $m(\epsilon,\delta) \ge \frac{\ln(1\backslash\delta)}{\epsilon}$, כנדרש.

VC dimension

3. נטען שמימד VC של $\mathcal{H}_{con}$ הוא d:

- נציג את הקבוצה $C = \{e_i\}_{i=1}^{d}$. נזהה את הוקטור $x \in \{0,1\}^d$ עם ההיפותזה $h \in \mathcal{H}_{con}$ כך שאם ההיפותזה לוקחת את x_i, אז $x_i = 0$ ואם היא לוקחת את $\overline{x_i}$ אז $x_i = 1$ ב-x. נבחין שמתקיים: $x \cdot I_d = x$ כלומר על כל הדוגמאות בקבוצה זו הפונקציה h המתוארת ע"י x (כל רצף אפסים ואחדות שנבחר) נותנת לנו את x ומכיוון שאנחנו יכולים ליצור כל רצף אפסים ואחדות שנרצה (מהגדרת h), אנו מקבלים עבור C כל תיוג שנבחר.

- תהא $C=\{x_i\}_{i=1}^{d+1}\subset\{0,1\}^d$. מעקרון שובך היונים קיימים לפחות 2 וקטורים שונים x_k, x_l עבורם הקואורדינטה ה-j מקבלת 1. (כלומר ישנה תלות בין הוקטורים). ולכן לכל היפותזה עבורה האחד משתערך ב-T, גם האחר, ולא נכל לקבל סט תוויות בהן $y_k = 1, y_l = 0$, אזי C לא מנופצת ע"י המחלקה.

Agnostic PAC

4. יהא S מדגם $\frac{\epsilon}{2}$-מייצג עבור $\mathcal{H}, D, \ell$.
נסמן $\hat{h}\in argmin_{h\in\mathcal{H}} L_D(h)$, $h_S\in ERM_{\mathcal{H}}$.
אז לכל $h\in\mathcal{H}$: $|L_S(h)-L_D(h)|<\frac{\epsilon}{2}$, ובפרט עבור h_S ועבור $\hat{h}$ מתקיים:
$L_D(h_S) < L_S(h_S)+\frac{\epsilon}{2}$, $L_S(\hat{h}) < L_D(\hat{h})+\frac{\epsilon}{2}$
כמו כן מכיוון ש$h_S\in ERM_{\mathcal{H}}$: $L_S(h_S)\le L_S(h)$ לכל $h\in\mathcal{H}$ ובפרט עבור $\hat{h}\in argmin_{h\in\mathcal{H}} L_D(h)$.
נאגד הכל ונקבל:

$$L_D(h_S) < L_S(h_S)+\frac{\epsilon}{2}\le L_S(\hat{h})+\frac{\epsilon}{2}\le\left(L_D(\hat{h})+\frac{\epsilon}{2}\right)+\frac{\epsilon}{2}=L_S(\hat{h})+\epsilon=$$
$$=\min_{h\in\mathcal{H}} L_D(h)$$

כנדרש.

5. בחרתי לא להגיש שאלה זו.

Theoretical Claim

6. בחרתי לא להגיש שאלה זו.

7. יהיו $\mathcal{H}_1,\mathcal{H}_2$ מחלקות היפותזה כך ש: $\mathcal{H}_2\subset\mathcal{H}_1$. לכל קבוצה $C\subset\mathcal{X}$ המנופצת ע"י $\mathcal{H}_1$, ומכך ש: $\mathcal{H}_2\subset\mathcal{H}_1$, מתקיים: $|\mathcal{H}_2^C|\le|\mathcal{H}_1^C|=2^{|C|}$. (C מנופצת ע"י $\mathcal{H}_2$). אז יש לם הכלה של קבוצות מנפצות ולכן הסופרימומים של גדלי הקבוצות המנפצות מקיימים את היחס הבא:

$$VC(\mathcal{H}_2)=\sup_{m\in\mathbb{N}}\{\exists C\ s.t\ |C|=m\ and\ C\ is\ shattered\ by\ \mathcal{H}_2\}\ge$$
$$\sup_{m\in\mathbb{N}}\{\exists C\ s.t\ |C|=m\ and\ C\ is\ shattered\ by\ \mathcal{H}_2\}=VC(\mathcal{H}_1)$$

כנדרש.

Theoretical Claim

8. נענה:
a. $\tau_{\mathcal{H}}(m)$ לוקחת את המועמד המקסימלי מבין אלו שהוצעו ע"י כל תתי הקבוצות $C\subset\mathcal{X}$ מגודל m. המועמד שכל קבוצה מציעה הוא מספר היפותזות שלא

מזדהות עליה. לכל המשמעות של $\tau_{\mathcal{H}}(m)$ היא מספר ההיפותזות המקסימלי שלא מזדהות על תת קבוצה כלשהיא $C\subset\mathcal{X}$ מגודל m, על פני תתי קבוצות מגודל m.
b. נניח $VC(\mathcal{H})=\infty$. יהא $m\in\mathbb{N}$ אזי קיימת $C\subset\mathcal{X}$ כך ש- $|\mathcal{H}_C|=2^{|C|=m}$,

בנוסף נבחין כי $\mathcal{H} \subset \{\pm 1\}^m$ ולכן: $\le \mathcal{H}\ 2^m$. ולכן יש קבוצה שמגיעה לחסם העליון אז המקסימום הוא החסם העליון:

$$\max_{m \in \mathbb{N}} \{|\mathcal{H}_C| : C \subset \mathcal{X},\ |C| = m\} = 2^m = \tau_{\mathcal{H}}(m)$$

c. נניח $VC(\mathcal{H}) < \infty = d$. אזי לכל $m \in \mathbb{N}$, $d \ge m$:
כל $C \subset \mathcal{X}$ מגודל m מנותצת ע"י $\mathcal{H}$, כלומר קיימת C מגודל m כך ש- $|\mathcal{H}_C| = 2^{|C| = m}$. בנוסף נבחין כי $\mathcal{H} \subset \{\pm 1\}^m$ ולכן: $\le \mathcal{H}\ 2^m$. אזי מתקיים:

$$\max_{m \in \mathbb{N}} \{|\mathcal{H}_C| : C \subset \mathcal{X},\ |C| = m\} = 2^m = \tau_{\mathcal{H}}(m)$$

כי קיימת תת-קבוצה בקבוצה שממנה נקח מקסימום, שמגיעה לחסם העליון ולכן המקסימום הוא החסם העליון.

d. נוכיח:

i. נוכיח שעבור $C \subset \mathcal{X}$ סופית מתקיים:
$|\mathcal{H}_C| \le |\{B \subset C : \mathcal{H} \text{ shutters } B\}|$, באינדוקציה על $|C| = m$:

עבור m=1: נסמן $C = \{x\} \subset \mathcal{X}$. אז נבחין ש: $|\mathcal{H}_C| \le 2$ מכיוון שעל נקודה אחת ניתן לקבל עד 2 סיווגים שונים.

- אם $|\mathcal{H}_C| = 1$: אז $\forall h \in \mathcal{H}: h(x) = y$ קבוע. תתי הקבוצות היחידות של C הן C עצמה והקבוצה הריקה. מאחר ו-C לא מנופצת ע"י $\mathcal{H}$ ומאחר והקבוצה הריקה תמיד מנופצת ע"י כל $\mathcal{H}$ (באופן ריק) נקבל ש: $|\{B \subset C : \mathcal{H} \text{ shutters } C\}| = 1$.
- אם $|\mathcal{H}_C| = 2$: אז $\forall h \in \mathcal{H}: h(x) \in \{\pm 1\}$. כלומר C מנותצת ע"י $\mathcal{H}$ ומאחר והקבוצה הריקה תמיד מנופצת ע"י כל $\mathcal{H}$ (באופן ריק) נקבל ש: $|\{B \subset C : \mathcal{H} \text{ shutters } C\}| = 2$.

כלומר השוויון אכן מתקיים כמצופה.

נניח שהטענה מתקיימת לכל $k < m$.

נוכיח את הטענה עבור m:
תהא $\mathcal{H}$, $C = \{c_1, \dots, c_m\}$, ונסמן:

$$C' = \{c_2, \dots, c_m\}$$

$$Y_0 = \{(y_2, \dots, y_m) : (0, y_2, \dots, y_m) \in \mathcal{H}_C \text{ or } (1, y_2, \dots, y_m) \in \mathcal{H}_C\}$$

כלומר Y_0 מחזיקה את צמצום היפותזות מ- $\mathcal{H}_C$ על C' - כלומר היפותזות מ- $\mathcal{H}_{C'}$. אזי מה"א מתקיים אי השיוון הבא:

$$|Y_0| = |\mathcal{H}_{C'}| \le |\{B \subset C' : \mathcal{H} \text{ shutters } B\}| =$$

$$= |\{B \subset C : c_1 \notin B \text{ and } \mathcal{H} \text{ shutters } B\}|$$

והמעבר האחרון נבע מכך ש: $B \subset C'$ מנופצת ע"י $\mathcal{H} \Leftrightarrow B \subset C$ מנופצת ע"י $\mathcal{H}$ עבור B: $c_1 \notin B$.

נסמן:

$$Y_1 = \{(y_2, \dots, y_m) : (0, y_2, \dots, y_m) \in \mathcal{H}_C \text{ and } (1, y_2, \dots, y_m) \in \mathcal{H}_C\}$$

$$\mathcal{H} \supset \mathcal{H}' = \{h \in \mathcal{H} : \exists h' \in \mathcal{H} \text{ s.t } h'(c) = (-h(c_1), h(c_2), \dots, h(c_m))\}$$

כלומר Y_1 מחזיקה את כל ההיפותזות שמזדהות על C' ולא על C (קיימות 2 היפוזות שמחזירות ערכים שונים על c_1). נבחין שזו בדיוק ההגדרה של $\mathcal{H}'_C$. אזי מהנחת האינדוקציה:

$$|Y_1| = |\mathcal{H}'_C| \le |\{B \subset C' : \mathcal{H}' \text{ shutters } B\}| =$$

$$= |\{B \subset C' : \mathcal{H}' \text{ shutters } B \cup \{c_1\}\}| =$$

$$= |\{B \subset C : c_1 \in B \text{ and } \mathcal{H}' \text{ shutters } B\}| \le\ \le |\{B \subset C : c_1 \in B \text{ and } \mathcal{H} \text{ shutters } B\}|$$

והשוויון האחרון נבע מכך ש: $B \subset C'$ מנופצת ע"י $\mathcal{H}' \Leftrightarrow B \subset C$ מנופצת ע"י $\mathcal{H}'$ עבור B: $c_1 \in B$.
א"ש האחרון מתקיים מכיוון ש: $\mathcal{H}' \subset \mathcal{H}$.

א"ש האחרון מתקיים מכיוון ש: $\mathcal{H}' \subset \mathcal{H}$.
בנוסף נבחין ש: $|\mathcal{H}_c| = |Y_0| + |Y_1|$ מכיוון ש- Y_1 מחזיקה את כל ההיפותזות שיש להן זוג מקביל לתיוג ב c_1, ו- Y_0 מחזיקה את כל ההיפותזות שאין להן. לכן:
$|\mathcal{H}_c| = |Y_0| + |Y_1| \le |(|$

8. <u>**נמשיך:**</u>

d.

i. לכן:

$$\begin{aligned}|\mathcal{H}_C| &= |Y_0| + |Y_1| \\ &\le |\{B \subset C: c_1 \notin B \; and \; \mathcal{H} \; shatters \; B\}| \\ &+ |\{B \subset C: c_1 \in B \; and \; \mathcal{H} \; shatters \; B\}| \\ &= |\{B \subset C: (c_1 \notin B \; and \; c_1 \\ &\in B \;) \; and \; \mathcal{H} \; shatters \; B\}| \\ &= |\{B \subset C: \mathcal{H} \; shatters \; B\}|\end{aligned}$$

כנדרש.

ii. במילים: מספר התיוגים האפשריים על C הוא לכל היותר כמספר תתי הקבוצות של F שמנופצות ע"י מחלקת ההיפותזות.

iii. נרצה להראות ש: $|\{B \subset C: \mathcal{H} \; shatters \; B\}| \le \sum_{i=0}^{d} \binom{m}{i}$.
נבחין שמכך ש: $VCdim(\mathcal{H}) \le d < \infty$ אז כל קבוצה סופית $C \subset \mathcal{X}$ המקיימת: $|C| = m > d$ איננה מנופצת ע"י $\mathcal{H}$. לכן C מגודל מקסימלי d, וכך גם תתי הקבוצות שלה.
לכן:

$$\begin{aligned}&|\{B \subset C: \mathcal{H} \; shatters \; B\}| \; = \\ &= \left| \cup_{i=0}^{d} \{B \subset C: |B| = i \; and \; \mathcal{H} \; shatters \; B\} \right| \le \\ &\le \; \left| \cup_{i=0}^{d} \{B \subset C: |B| = i\} \right| = \Sigma_{i=0}^{d} |\{B \subset C: |B| = i\}| \; = \\ &= \; \sum_{i=0}^{d} \binom{m}{i}\end{aligned}$$

מכיון ש- $\binom{m}{i}$ מציין את מספר תתי הקבוצות של m (הגודל של C) מגודל i, כנדרש.

iv. תחת הנחות הסעיפים הקודמים מתקיים עבור $m > d = VCdim(\mathcal{H})$:

$$\begin{aligned}&\tau_{\mathcal{H}}(m) = \max\{|\mathcal{H}_C|: C \subset \mathcal{X}, |c| = m \} \le \\ &\le \max\{|\{B \subset C: \mathcal{H} \; shatters \; B\}|: C \subset \mathcal{X}, |c| = m \} \le \\ &\le \; \sum_{i=0}^{d} \binom{m}{i} \le \left(\frac{em}{d}\right)^{d}\end{aligned}$$

כאשר את כל המעברים הוכחנו בסעיפים הקודמים וא"ש האחרון נתון לנו בהנחות השאלה. (ויתרנו על המקסימום מאחר שאם כל איבר

בקבוצה קטן-שווה לחסם כלשהוא אז גם המקסימום של הקבוצה הנ"ל
קטן-שווה לו כאיבר של הקבוצה).
והוכחנו כנדרש.

e. עבור $d = VCdim(\mathcal{H})$ מתקיים $\tau_{\mathcal{H}}(d) = 2^d$ מכיוון שזהו הגודל המקסימלי
עבור $\mathcal{H}_C$ עם C מגודל d, ומפני שאכן קיימת קבוצה C המקיימת $|\mathcal{H}_C| = 2^d$
(מהגדרת מימד VC ומכך שהוא שווה ל-d).
נבחין שמכיוון ו- $2 \leq e$, שניהם א"ש ו-$d > 0$, אז מתקיים:

$$\tau_{\mathcal{H}}(d) = 2^d \leq \; e^d \left(\frac{ed}{d}\right)^d$$

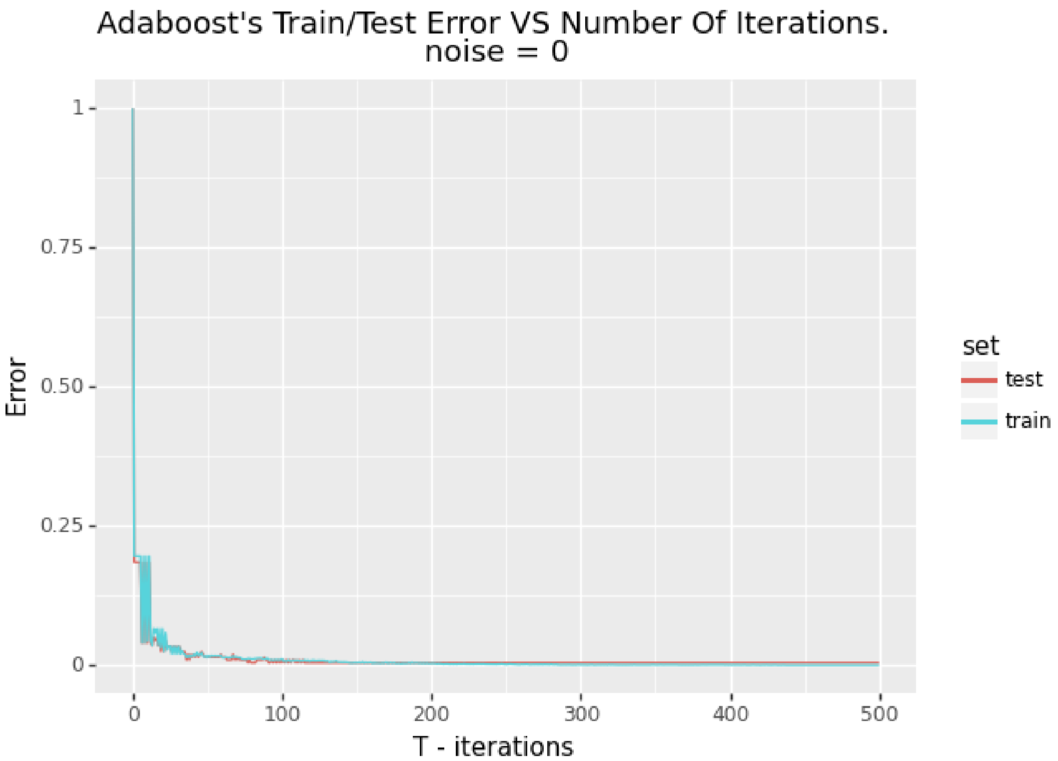
כלומר א"ש נכון גם עבור המקרה הזה, אך הוא לא חסם הדוק.

f. כאשר $m > d = VCdim(\mathcal{H})$: הפונקציה עולה באופן פולינומילאי.
כאשר $m \leq d = VCdim(\mathcal{H})$: הפונקציה עולה באופן אקספוננציאלי.
אציע הגדרה חלופית למימד VC של מחלקת היפותזות:
$VCdim(\mathcal{H}) = \max_{m \in \mathbb{N}}\{\tau_{\mathcal{H}}(m) = 2^m\}$ מכיוון שעד מימד VC של המחלקה,
פונקציית הגדילה שלה עולה אקספוננציאלית ולאחריו היא גדלה פולינומיאלית.
ובמימד VC אנחנו מעוניינים בגודל הקבוצה המקסימלית שמקייימת תכונה זו,
כלומר בדיוק בנקודת השינוי של הפונקציה.

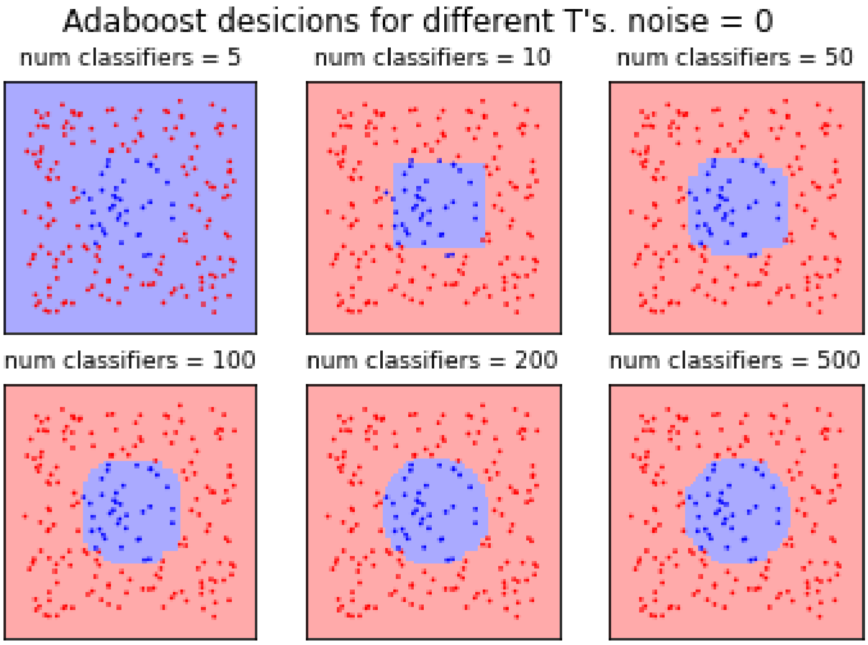
Adaboost

9. המימוש נמצא בקובץ adaboost.py.

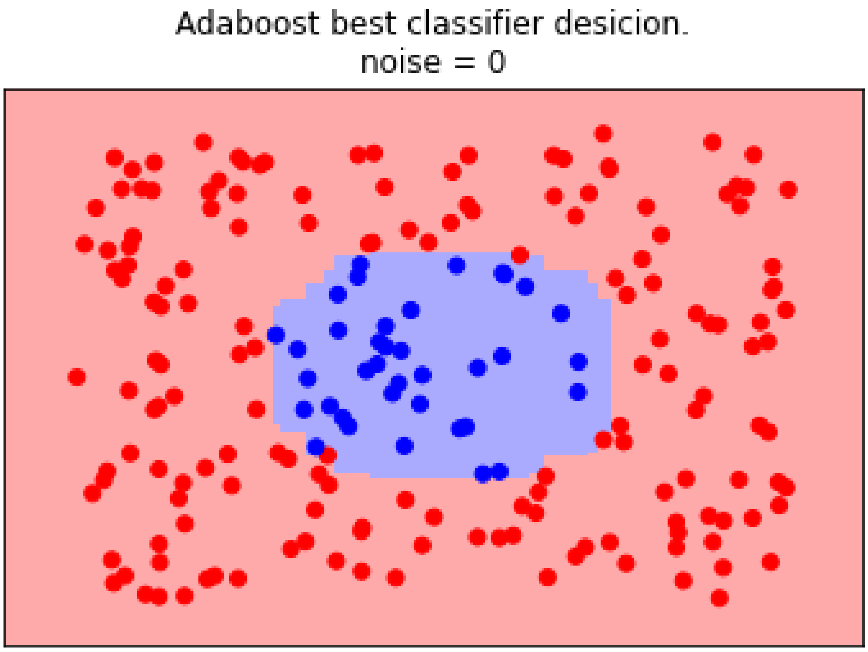
10.



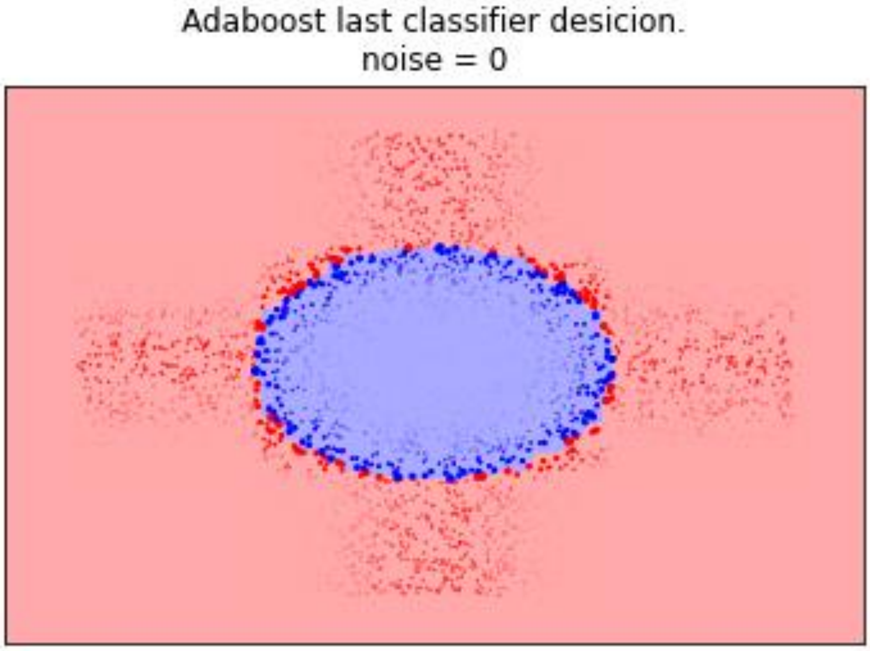
.11



12. השגיאה המינימלית היא: 0.005
ה-T המינימלי שמגיע לשגיאה המינימלית הוא: 77



.13



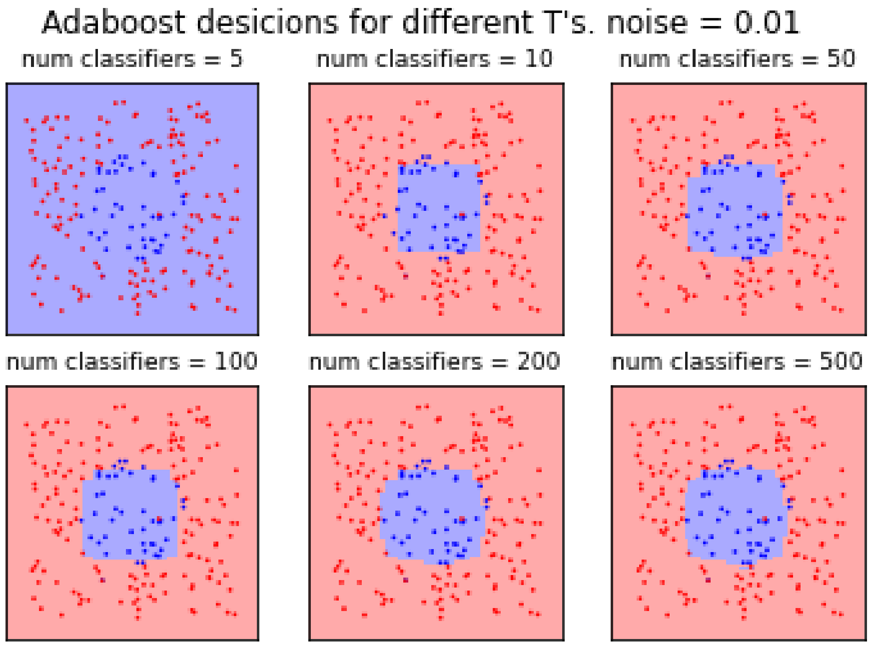
ניתן לראות שגודל הנקודות שעל ספ ההחלטה גדול מאשר הנקודות משני צדדיו. כמו כן ניתן לראות סימן + של נקודות המתויגות אדום הממושקלות יותר, ועיגול כחול שככל שמתקרבים לסף שלו משקלי הנקודות עולה. מהגדרת האלגוריתם, הוא מגדיל את ה"קנסות" על נקודות בהן הוא שגה בעבר, לכן מה שאנחנו רואים לפנינו תואם את האינטואיציה: נקודות הקרובות לסף (שפ העיגול) הן נקודות מסובכות לתיוג, נמצא את הגבול המדוייק ביניהן על ידי ניסוי וטעייה (כלומר הגברת המשקל). כמו כן האלגוריתם שלנו הוא גדם עץ ולכן הוא קובע תיוגים ע"י חלוקה לאורך הצירים. כך, עבור מפר חברי ועדה קטן הוא ינסה למסגר את העיגול שבמרכז ולכן ייצור חלוקה שמסמנת את האיזור המרכזי בכחול, ולכן יישגה על האיזורים שנמצאים בזרועות סימן ה:+.

14. להלן הגרפים:

- **<u>רעש = 0.01:</u>**

שאלה 10:

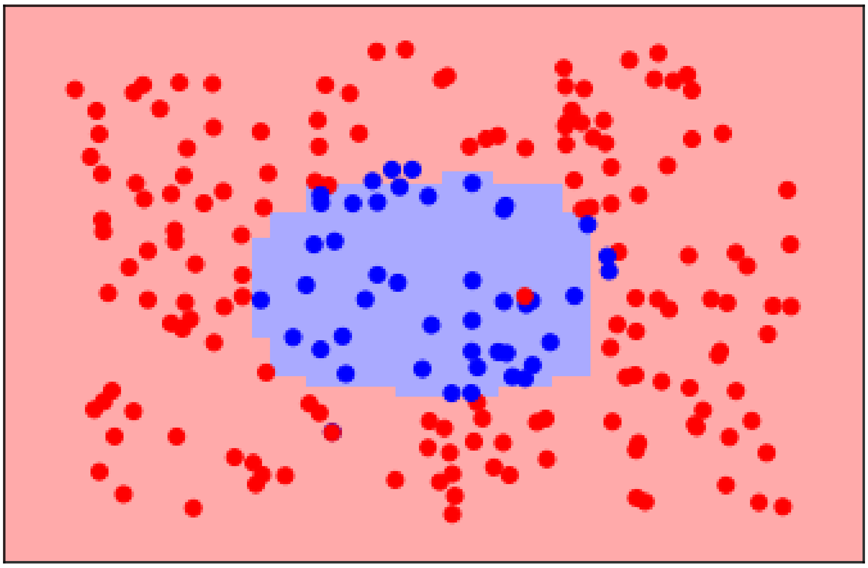
- שאלה 11:



- שאלה 12:

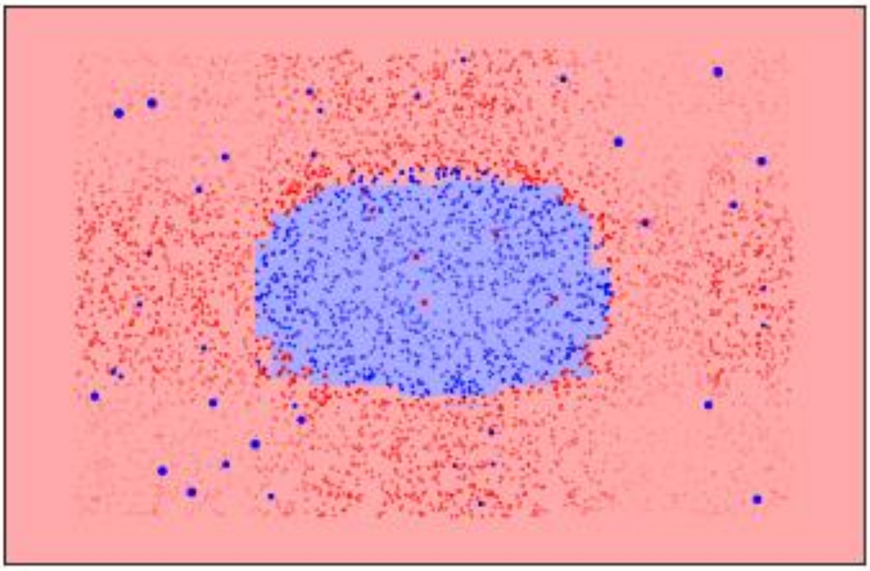
השגיאה המינימלית היא: 0.04
ה-T שמגיע לשגיאה המינימלית הוא: 225

Adaboost best classifier desicion.
noise = 0.01



- שאלה 13:

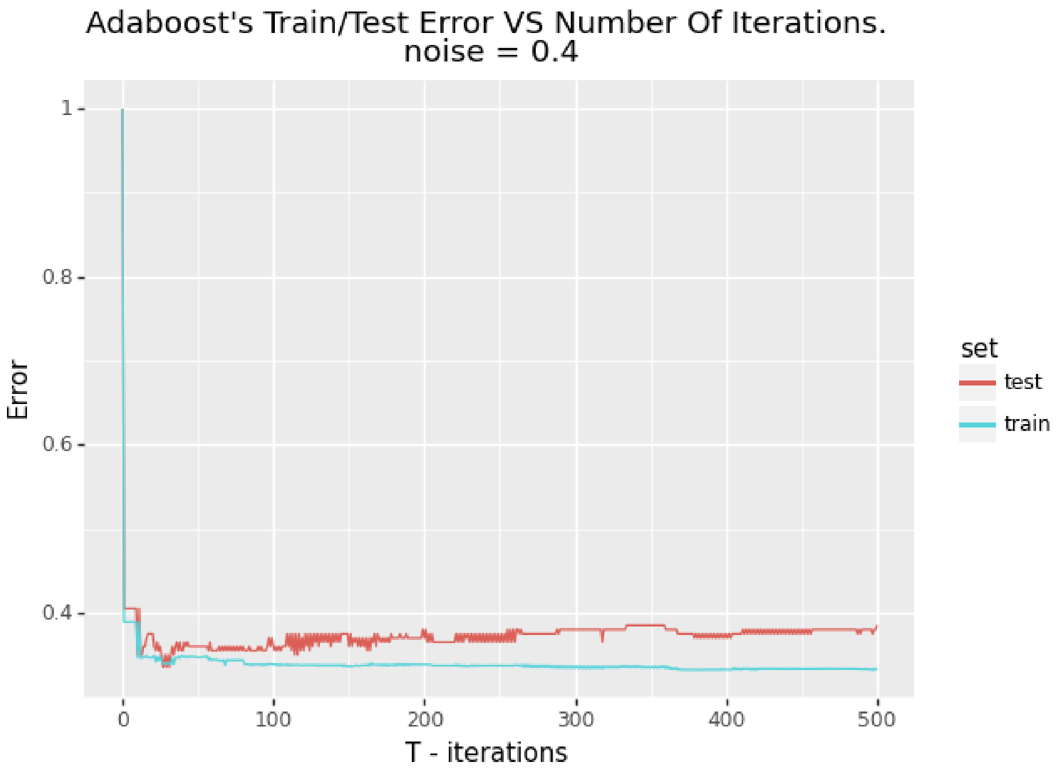
Adaboost last classifier desicion.
noise = 0.01



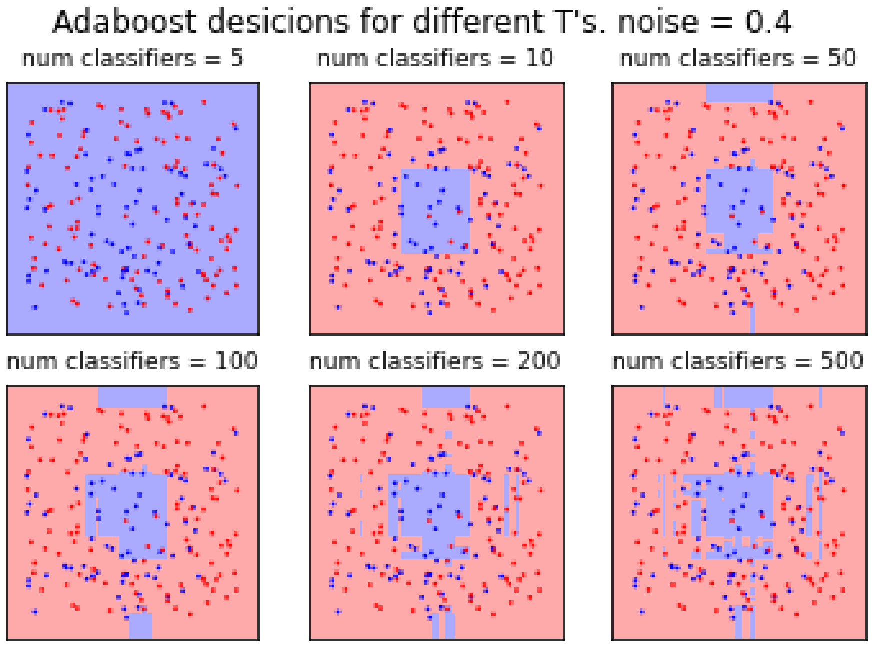
ניתן לראות כמו עבור רעש 0 שנקודות הגבול וזרועות ה:+ ממשוקלות יותר (מתהליך ניסוי וטעייה), אך להן נוספו הנקודות הרועשות שיצאו מתחומן ועברו לעומק תחום התיוג הנגדי. נקודות אלו נקנסות באופן תמידי (מצויות ברוב מוחלט של נקודות בעלות תיוג נגדי) עולה ויורד (מאחר והאלגוריתם לומד מטעויות כבדות שנעשו בעבר) ולכן לעיתים ייסמן את האיזור שלהן בצבע האמיתי שלהן, ויישגה על האיזור מסביב להן. לכן אנחנו רואים את סימן ה: + באופן קצת פחות מובחן.

<u>רעש = 0.4:</u>

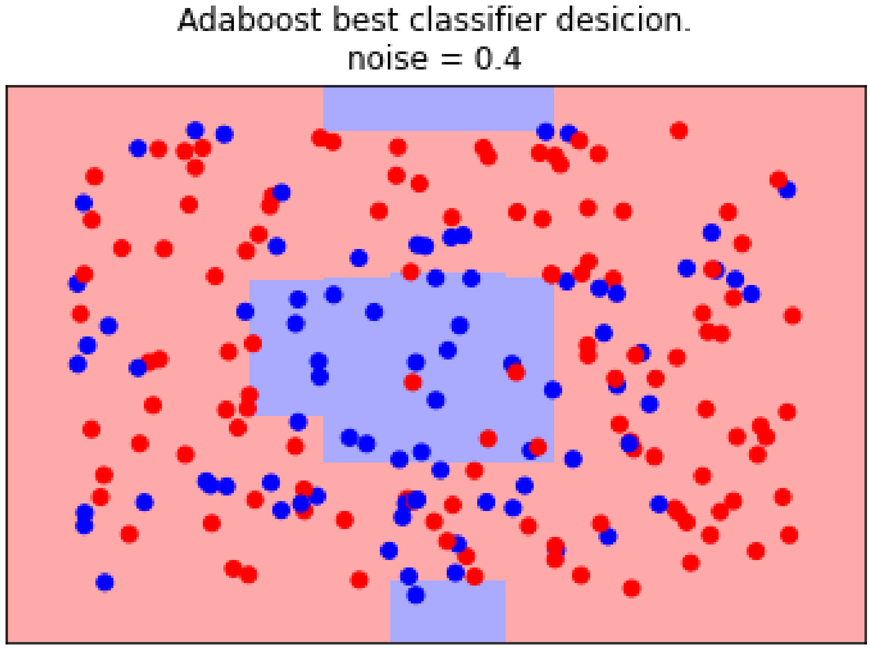
- שאלה 10:



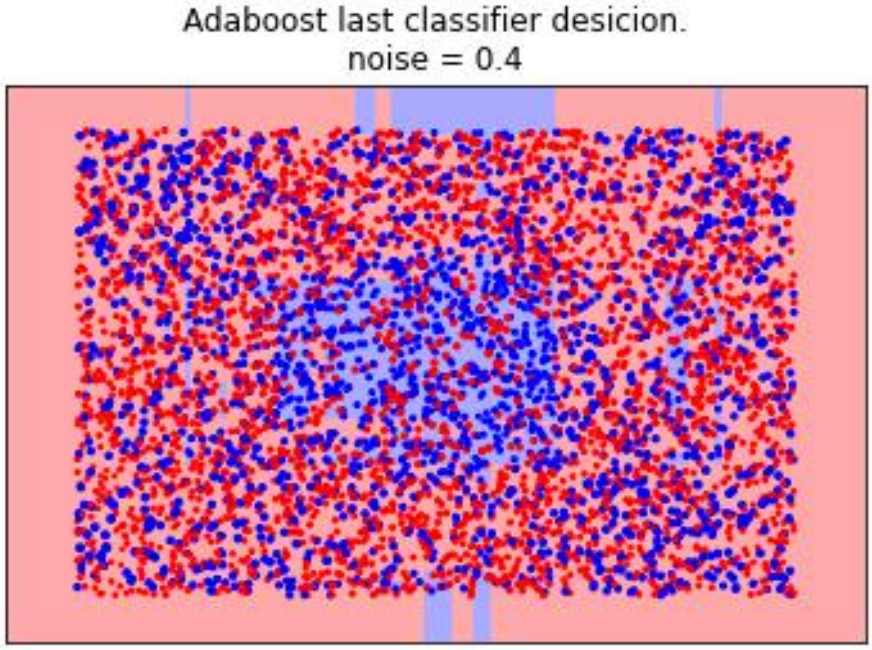
- שאלה 11:



- שאלה 12:
השגיאה המינימלית היא: 0.335
ה-T שמגיע לשגיאה המינימלית הוא: 27



- שאלה 13:



כאן גבר הרעש, ולכן אין צורה ממשית למדגם, והמשקולות מאוד מובחנות, כלומר יש הרבה טעויות. כמו כן מהסיווג הסופי ניתן להניח שיש רוב דחוס של דגימות כחולות במרכז.

- ניתן לראות שבאופן כללי עבור **מדגמים חסרי רעש (או עם רעש זניח)** השגיאה פוחתת ככל שמספר חברי הועדה עולה. כמו כן, ערכי השגיאה המתקבלים עבור מדגמים כאלה נמוכים משמעותית מאלו המתקבלים עבור מדגמים רועשים.
  הסיווג גם מתאים לתיוג האמיתי של הדגימות ויש פחות שגיאות ממושקלות.
  לעומת זאת ככל שמוסיפים **רעש**, ניתן לראות את שגיאת הבדיקה עולה על שגיאת האימון (ככל הנראה כתוצאה מלמידת יתר). נבחין שהסיווג לא תואם לציוג האיתי של הדגימות, ומצב זה הולך ומחריף ככל שהרעש עולה.

- הסבר של שאלה 10 במונחי **bias-variance tradeoff**:
  ניתן לראות שככל ש-**T** עולה, ה-**bias** פוחת והשונות עולה: אנחנו מגדילים את מחלקת ההיפותזות שלנו בכך שאנחנו ממצעים כללי החלטה שונים ולכן מייצרים למעשה כללי החלטה נוספים. אז ללמדן מסויים יש הרבה אפשרויות לבחירת היפותזה- מה שמקטין את ה**bias** ומגדיל את השונות כאמור. כמו כן, ככל שהרעש עולה יש סטיות רבות יותר, ונצטרך כושר הבעה גדול יותר כדי לייצר היפותזה שמסבירה את הנתונים. כלומר ששוב מחלקת ההיפותזות גדלה, ורואים זאת ע"י דעיכה של ה**bias** ועלייה של השונות.
  נבחין שבשלב זה בעת האימון, ננסה להתקרב להיפותזה שמייצגת את **מדגם האימון** ולכן נראה שככל **שמספר חברי הועדה גדל והרעש גדל**, יש אמנם התכנסות של השגיאות אך שגיאות הבדיקה עולות על שגיאות האימון, זאת אומרת שיש **overfitting**.

- ניתן לראות מהשוואת תוצאות שאלה 12 שהשגיאה המינימלית גדלה ככל שהרעש גובר. בויזואליזציה של השאלות ניתן לראות את ההסבר לתופעה הזו: הרעש מסית דגימות מאיזור התיוג האמיתי עבורן אל האיזור השני, מה שיוצר ערבובייה של דגימות מ-2 התיוגים באותו איזור. החלטת המסווג החלש שונה, והשגיאה עולה, כי קשה יותר להפריד בין הדגימות. כמו כן, אנחנו רואים שעבור רעש גדול (0.4) מספר חברי הועדה האופטימלי הוא נמוך: ניתן להסביר הבחנה זו בעזרת ה-**overfitting** שציינו מקודם. הרעש מסית את הלמידה ללמידה ספציפית של הרעש ולכן למידה אפקטיבית יותר תתבצע לפני שהגענו לשלב הזה.