# פתרון שיעורי בית 2 – רוני קוסיצקי 205817893

## שאלה 1

ניעזר ברמז.
**כיוון ראשון** – נוכיח שם $H$ היא PAC למידה אז $H$ היא PAC למידה בתוחלת(מקווה שזה התרגום הנכון למונח).

מכיוון שהיא PAC למידה אזי קיים אלגוריתם כלשהוא $A$ כך שמתקיים לכל $\epsilon, \delta > 0$ כך

$$\exists N(\epsilon, \delta)\ s.t\ \forall n \geq N(\epsilon, \delta), \forall P: \mathbb{P}\left[e_P\left(A(S_n)\right) \leq \epsilon\right] \geq 1 - \delta$$

יהיה $a \in (0,1)$ כלשהוא. מהגדרת הPAC קיים אלגוריתם כלשהוא A ופונקציה N ל $\epsilon = \frac{a}{2}, \delta = \frac{a}{2}$ כך ש

$$\forall n \geq N\left(\frac{a}{2}, \frac{a}{2}\right) = N^*\left(\frac{a}{2}\right), \forall P: \mathbb{P}\left[e_P\left(A(S_n)\right) > \frac{a}{2}\right] \leq \frac{a}{2}$$

מנוסחת ההסתברות השלמה נקבל ש

$$\mathbb{E}\left[e_P\left(A(S)\right)\right] = \mathbb{E}\left[e_P\left(A(S)\right) \middle| e_P\left(A(S)\right) \leq \frac{a}{2}\right] \cdot \mathbb{P}\left[e_P\left(A(S)\right) \leq \frac{a}{2}\right] + \mathbb{E}\left[e_P\left(A(S)\right) \middle| e_P\left(A(S)\right) > \frac{a}{2}\right] \cdot \mathbb{P}\left[e_P\left(A(S)\right) > \frac{a}{2}\right]$$

נמצא חסם עליו לכל ביטוי. ראשית מתקיים ש $\mathbb{E}\left[e_P\left(A(S)\right) \middle| e_P\left(A(S)\right) \leq \frac{a}{2}\right] \leq \frac{a}{2}$ כי $e_P\left(A(S)\right) \leq \frac{a}{2}$ אזי גם הגבול של התוחלת יהיה קטן מ$\frac{a}{2}$. כמון כן מתקיים $\mathbb{P}\left[e_P\left(A(S)\right) \leq \frac{a}{2}\right] \leq 1$, וזה מתקיים כי כל הסברות באשר היא חסומה ע"י 1.

כמו כן, מתקיים $e_p\left(A(S)\right) = P(A(S)(x) \neq Y)$ ולכן $e_P\left(A(S)\right) \leq 1$. בסה"כ $\mathbb{P}\left[e_P\left(A(S)\right) > \frac{a}{2}\right] \leq \frac{a}{2}$(לפי מה שהראנו קודם).

בסך הכל, נקבל ש:

$$\begin{aligned}\forall n \geq N^*\left(\frac{a}{2}\right), \forall P: \mathbb{E}\left[e_P\left(A(S)\right)\right] &= \mathbb{E}\left[e_P\left(A(S)\right) \middle| e_P\left(A(S)\right) \leq \frac{a}{2}\right] \cdot \mathbb{P}\left[e_P\left(A(S)\right) \leq \frac{a}{2}\right] + \mathbb{E}\left[e_P\left(A(S)\right) \middle| e_P\left(A(S)\right) > \frac{a}{2}\right] \\ &\cdot \mathbb{P}\left[e_P\left(A(S)\right) > \frac{a}{2}\right] \leq \frac{a}{2} \cdot 1 + 1 \cdot \frac{a}{2} = a\end{aligned}$$

לכן, מצאנו פונקציה N המקיימת $N(a) = N\left(\frac{a}{2}\right)$ לכל $a$ בין 0 ל1 ולכל הסתברות $P$ בהינתן סט אימון $S$ המקיים $|S| \geq N'(a)$:

$$\mathbb{E}\left[e_P\left(A(S)\right)\right] \leq a$$

לכן, בסה"כ קיבלנו שהיא למידה *PAC* בתוחלת לפי הגדרה.

**כיוון שני** – מכיוון שH היא למידת PAC בתוחלת אנו יודעים שקיים אלגוריתם $A$ ופנוקציה $N(A)$ כך שלכל $a \in (0,1)$ והתפלגות P וסט אימון S כך שמתקיים $|S| \geq N'(a)$:

$$\mathbb{E}\left[e_P\left(A(S)\right)\right] \leq a$$

יהיה $\epsilon, \delta$ חיוביים כלשהם. נחבר את $a = \min\{\epsilon\delta, 1\}$.

ולכן נקבל ש $\mathbb{E}\left[e_P\left(A(S)\right)\right] \le \min\{\epsilon\delta, 1\}$ בהנתן התנאים שלמעלה. לפי מרקוב מתקיים:

$$\mathbb{P}\left[e_P\left(A(S_n)\right) > \epsilon\right] \le \frac{\mathbb{E}\left[e_P\left(A(S_n)\right)\right]}{\epsilon} \le \frac{\min\{\epsilon\delta, 1\}}{\epsilon}$$

אם $\delta\epsilon > 1$ נקבל שהביטוי שווה בדיוק ל1 ולכן בסה"כ:

$$\mathbb{P}\left[e_P\left(A(S_n)\right) > \epsilon\right] \le \frac{1}{\epsilon} = \frac{1}{\delta\epsilon} \cdot \delta \le 1 \cdot \delta = \delta$$

ואילו במקרה השני נקבל:

$$\mathbb{P}\left[e_P\left(A(S_n)\right) > \epsilon\right] \le \frac{\delta\epsilon}{\epsilon} = \delta$$

כלומר בשני המקרים אנו קטנים מ$\delta$. $N(a)$ היא פונקציה של $N(\epsilon, \delta)$ ולכן קיבלנו שהיא $PAC$ למידה על פי הגדרה.

## שאלה 2

נוכיח שממד ה-*VC* של $H$ הוא $2K$. על מנת להוכיח זאת, נראה ש-$H$ מנתצת קבוצה כלשהי של $2k$ נקודות, ואינה מנתצת אף קבוצה של נקודות נתונים בגודל $2K+1$.

יהיו $2k$ נקודות על הקו. ניקח כל רצף של דגימות חיוביות (סט של דגימות חיוביות ללא כל דגימה שלילית ביניהן) כמרווח. בשיטה זו, המרווחים המקסימליים שאנו יכולים לקבל הוא $k$ מרווחים. זה קורה כאשר 2 הנקודות הקרובות ביותר מכל צד של מדגם חיובי הן שליליות (כל המרווחים מוגדרים על ידי מדגם חיובי יחיד). כל שאר המקרים נותנים פחות מ-$k$ מרווחים, כך שהבחירה של כל תווית תתאים.

בהנחה שכל הנקודות שונות (אחרת, לא נוכל לבחור מדגם באותה נקודה להיות חיובית ועוד שלילית בו-זמנית, כך שלא ניתן לנפץ את המדגם).

נוכל לקחת את הדוגמה הבאה: $\{x1 < x2 < \cdots < x2k+1\}$. אם כל הדגימות שיש להן מדדים אי-זוגיים הן חיוביות, והשאר שליליות, אז נקבל תיוג לא חוקי - מכיוון שמספר המרווחים גדול מ-$k$. לפיכך, לא ניתן לנפץ כל דגימה שגדולה מ-$2k$.

## שאלה 3

נראה שעבור מלבנים ממימד $d$מימד הVC יהיה $2d$.

ראשית, נראה שקיימת קבוצה הניתנת לניתוץ בגודל $2d$. נבחר נקודות הנמצאות על וקטורי היחידה מהצירים, ואת הנקודות הנגדיות אליהן. לדוגמה, עבור מימד 3 נקבל $(0,1,0)$, $(-1,0,0)$, $(1,0,0)$..
אם אנחנו במרחב של $d$ מימדים אז נקבל בסה"כ $2d$ נקודות(בכיוון של הציר ובכיוון הנגדי שלו). קל להשתכנע שניתן לנתץ את כל הנקודות הללו.
בהנתן קובייה עם אורכי צלעות של 1 לפחות נוכל לקבל שכל הנקודות יקבלו תיוג חיובי. בהנתן שנרצה להשמיט את אחד הנקודות, נקצר את הצלע באותו כיוון כך שתהיה קצרה יותר מ1, ולכן תקבל תיוג חיובי. נוכל לשחק עם התיוגים כך שחלק מהצלעות יהיו קצרות וחלק בגודל של לפחות 1 וכך נוכל לשחק עם התיוגים של הנקודות בין חיובי לשלילי. אם נרצה שכל הנקודות יהיו מתוייגות כ0 אזי היפותזה שבה כל הצלעות קצרות מ1 תתן תיוגים שליליים לכולם.
נוכל להכליל את "וקטורי היחידה" האלו לכל הנקודות במרחב שכן כל נקודה במרחב בנוייה מ$2d$ וקטורי היחידה הללו ולכן נקבל שכל קבוצה בגודל $2d$ ניתנת לניתוץ.

נראה שאם קיימות $2d+1$ אזי הקבוצה לא ניתנת לניתוץ.
נסתכל בשנית על הדוגמה מלמעלה. תהי נקודה נוספת הנמצאת איפשהוא בין 0 ל1 באחד הצירים(זה לא משנה אם זה בפנים או יותר גדול מ1, ההוכחה תהיה זהה. הכתיבה קלה יותר משיקולי סימטריה). נניח בלי הגבלת הכלליות שהוא בכיוון החיובי. קל לראות שלא כל התיוגים אפשריים. נקרא לנקודה הקרובה $a$ ולנקודה הרחוקה $b$.
במידה ונרצה לתייג את $b$ יחד עם נקודה נוספת על אחד הצירים הנוספים כחיוביים ואת נקודה $a$ זה לא יהיה אפשרי כי המלבנים שלנו מקבילים לצירים ולכן יהיו חייבים לכלול את 2 הנקודות. לכן, בסה"כ הקבוצה לא ניתנת לניתוץ.
בדומה למקודם, בגלל שכל נקודה במרחב מורכבת מוקטורי היחידה אזי ניתן להכליל את ההוכחה הזו לכל המרחב שלנו ולראות שכל $2d+1$ בלתי ניתנות לניתוץ.

## שאלה 4

סעיף א-

יהיה $x \in X^n$ ונגדיר $A = h_1, B = h_2$. לכן מתקיים ש

$$\Pi_{A\cup B} = |\{h(x): h \in A \cup B\}| = |\{h(x): h \in A\}| \cup |\{h(x): h \in B\}|$$
$$\leq |\{h(x): h \in A\}| + |\{h(x): h \in B\}| = \Pi_{h_1} + \Pi_{h_2}$$

סעיף ב-

כעת מגדירים לנו את פעולת הכפל. נסתכל על 2 הקבוצות $\Pi_{h_1}, \Pi_{h_2}$.יהי $x$. $h_1 \otimes h_2(x) = 1$ אם ורק אם שתי ההיפותזות ייתנו לנו 1(כי אחרת נקבל כפל ב0 ולכן המכפלה תהיה 0).

נוכל בעצם לרשום את זה בצורה הבאה $h_1 \otimes h_2(x) = h_1\big(h_2(x)\big) = 1$(הסדר איננו משנה בכפל ולכן נניח בלי הגבלת הכלליות שאנו מפעילים את היפותזה 2 קודם).

לכן בסה"כ נקבל ש

$$\{h(x): h \in h_1 \otimes h_2(x)\} = \big\{q\big(p(x)\big): p \in h_1, q \in h_2\big\} = \bigcup_{p\in h_1} \{q\big(p(x)\big): q \in h_2\}$$

ומתקיים לפי *union bound*:

$$|\{h(x): h \in h_1 \otimes h_2(x)\}| \leq \sum_{p\in h_1} \left|\left\{q\big(p(x)\big): q \in h_2\right\}\right|$$

מתקיים שאנו סוכמים על כל האיברים בקבוצה שהיפותזה $h_1$ נותנת לנו אבל היא לכל היותר תיהיה שווה ל$\Pi_{h_1}$ על פי הגדרה. כמו כן בסכום של כל אחד מהם אנחנו נסכום לכל היותר את כל האיברים שההיפותזות של $h_2$ ייתנו לנו, כלומר $\Pi_{h_2}$. לכן בסה"כ נקבל:

$$\Pi_{h_1\otimes h_2} \leq \Pi_{h_1} \cdot \Pi_{h_2}$$

**<u>חלק מעשי</u>**

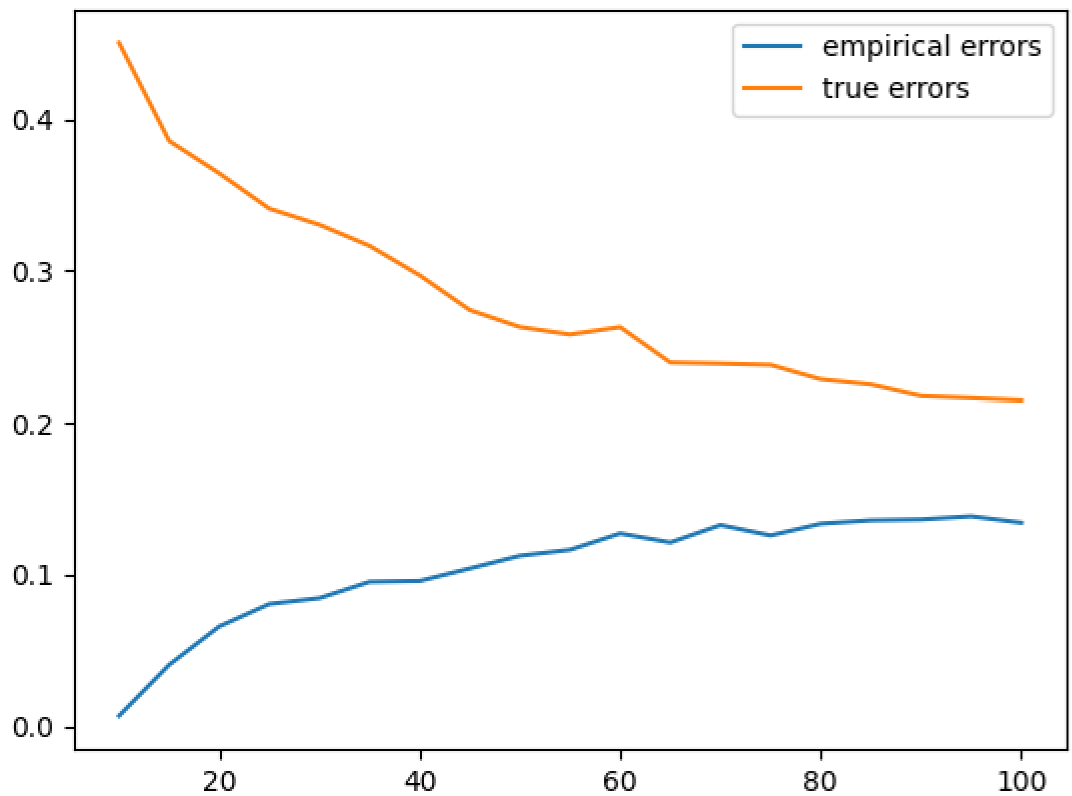
**א.** אנו יודעים את ההתפלגות(כאשר $x$ מתפלג אחיד) –

$$\boldsymbol{P[y=1|x]} = \begin{cases} \boldsymbol{0.8 \; if \; x \in [0,0.2] \cup [0.4,0.6] \cup [0.8,1]} \\ \boldsymbol{0.1 \; if \; x \in (0.2,0.4) \cup (0.6,0.8)} \end{cases}$$

נשתמש במסווג האופטימלי $e_p(h) = E\left[l_{0-1}\big(Y,h(x)\big)\right] = P[Y=y|X=x]$ ולכן נחליט שאם $x$ שייך לחלקי המקטעים $\boldsymbol{[0,0.2] \cup [0.4,0.6] \cup [0.8,1]}$ אזי המסווג ייתן לנו 1, אחרת 0. בסה"כ נקבל

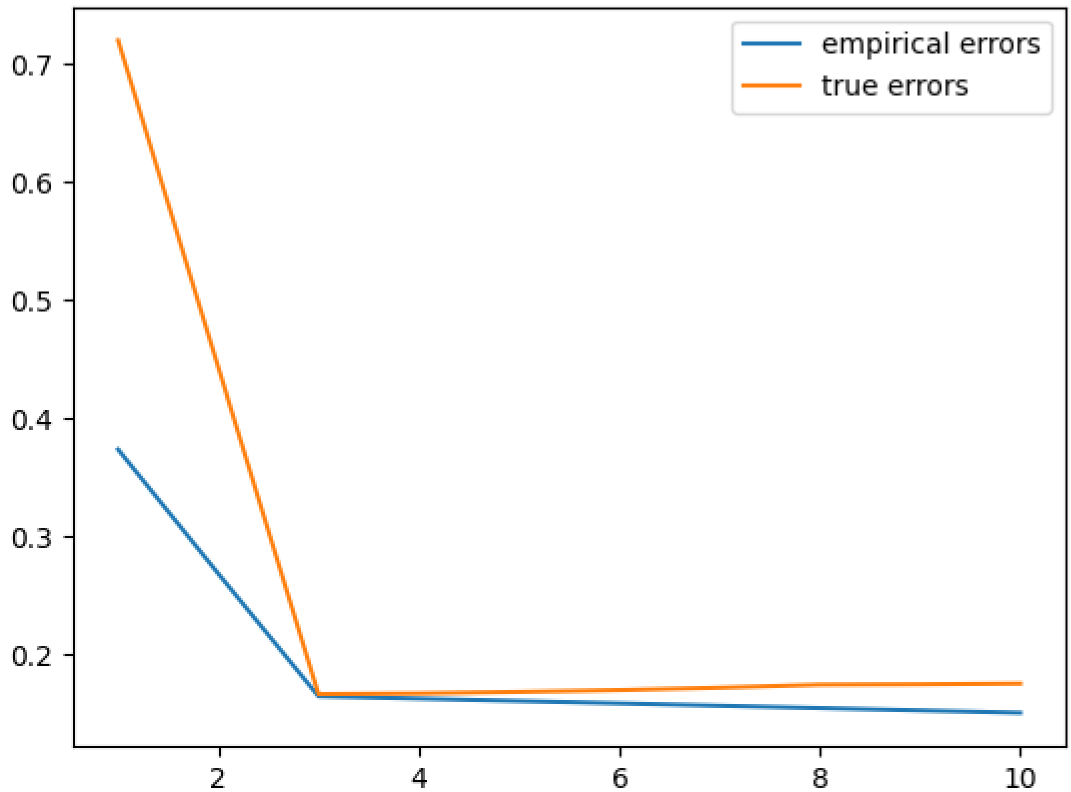
$$\begin{aligned} e_p(h) &= 1 \cdot P[Y=0,h(x)=1] + 1 \cdot P[Y=0,h(x)=0] \\ &= P[Y=0] \cdot P[y=0] + P[Y=1] \cdot P[y=1] \\ &= 0.2 \cdot 0.48 + 0.52 \cdot 0.1 = 0.148 \end{aligned}$$

**ב.**



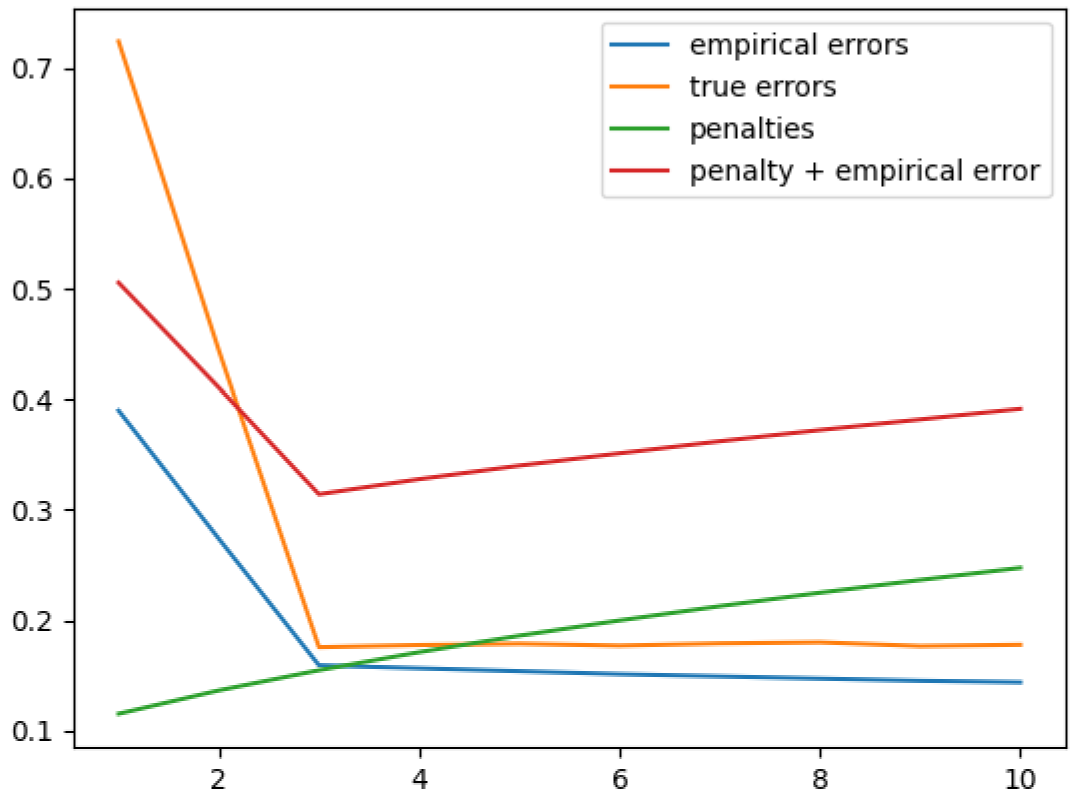

ככל ש-$n$ גדל, יש יחסית פחות דגימות המשתנות מהתפלגות $P$, ולכן המסווג שנקבל מ$ERM$ יהיה מדויק יותר והשגיאה האמיתית פוחתת. עם זאת, ככל ש-$n$ גדל קשה יותר למצוא מסווג שיצליח בכל הדגימות, לכן השגיאה האמפירית גדלה.

**ג.** הגרף המתקבל




השגיאה יורדת עם $k$. ניתן לראות שעבור $k \geq 3$ השגיאה האמיתית תהיה מינימלית. זה הגיוני כי אנו יודעים שההיפותזה הנכונה מכילה לפחות 3 אינטרוולים. מכיוון שהם חולקים את אותם קצוות אנחנו יכולים לפצל אותם ל3 אינטרוולים עבור כל $k > 3$. מצד שני, השגיאה האמפירית יורדת עבור $k > 3$ מכיוון שככל ש $k$ גדול יותר יש לנו יותר דרגות חופש ולכן אנו מצליחים לבצעי התאמה טובה יותר.

**ד.** נקבל ש $penalty = 2\sqrt{\frac{2k+\ln(20)}{1500}}$ (VC הוא $2k$ כפי שראינו בתחילת התרגיל). הגרף המתקבל




ה-$k$ הטוב ביותר לפי הגרף הוא $k = 3$. זוהי תוצאה טובה יותר בהשוואה לסעיף ג' מאחר שאנו יודעים שמספר המרווח המינימלי הוא 3.
כל ערכי $k$ שגדולים מ-3 גורמים לתוצאות "*over fitting*".

**ה.** שגיאה מינימלית – 0.13
ההיפותזות הטובות ביותר - *[(0.00023840229406962843, 0.20035292350454564), (0.40065092235390676, 0.5989577004040771), (0.7998975724496623, 0.9999078968344264)]*

קיבלנו שגיאה שהיא קרובה לשגיאה האמיתית וההיפותזות שלנו קרובות למציאות.