טל גאורגי

שאלה 1

נניח בשלילה כי- VCdim(H) הוא סופי, ונראה שהיא למידה PAC בהנחת ה- Realizable, ונראה שלכל A וכל
אלגוריתם אשר ילמד את ההתפלגות:
ההיפותזה הנכונה תהיה בכל ח, נניח את הכדור הקטן ביותר אשר מכסה את
כל הדוגמאות שלנו:

(הכדור הגדול יותר הוא B_0, המקורי
המקורי הנכון)

נניח כי S היא דוגמא שמובטח הגיע ממנה שונה מהדוגמאות מהכדור [illegible] הנכון, [illegible] כמו
כן נגדיר שהאלגוריתם יכול ללמוד רק [illegible] בתוך אותו מרחב הדוגמאות [illegible]
למידות [illegible] חיוביות.
[illegible] אז נסמן את הכדור שהתקבל בתור B_R, אז שגיאת האלגוריתם
היא: $ep(h_R)=P(B_0\setminus B_R)$
כעת, נניח שהאלגוריתם הוא מסוגל ללמוד PAC את ההתפלגות. יהיו (ϵ,δ).
אם $P(B_0)\le\epsilon$ אז ברור כי מתקיים $ep(h_R)=P(B_0\setminus B_R)\le\epsilon$ וזה מספיק.

אחרת, $\epsilon<P(B_0)$. כעת נגדיר את T להיות טבעת סביב מעגל שכל אחד מהם
ל-ϵ, ונחלק ל-$T_1,\dots,T_d$ מקבילות 2, ככה שכל אחד מהם [illegible] בהסתברות אחת. ניתן לעשות
זאת בצורה הזאת שהסתברות הטבעת היא רציפה, בהנחה לגבי ההתפלגות.
כעת נניח כי כל דוגמא שנקבל תהיה בתוך $B_0 \setminus D_R$ תהיה בתוך הטבעת, כלומר
ובנוסף שהסתברות של הכדור $B_0\setminus B_R$ קטנה מהסתברות של T, כלומר:
$ep(h_R)=P(B_0\setminus B_R)\le P(T)=\epsilon$
כי הגדרנו את T בצורה כזאת,
שהסתברות שלה היא ϵ.

לכן נקבל כי:

$$P(e_D(h_n) > \epsilon) \underset{?}{\le} (1-\epsilon)^n \le e^{-\epsilon n} = \delta$$

ההסתברות של n הדוגמאות הן מחוץ לריבוע

אי שוויון מוכר: $(1-x)^n \le e^{-xn}$

נבחר את n כך ש:

$$e^{-\epsilon n} = \delta \Rightarrow -n\epsilon = \ln(\delta) \Rightarrow n = \frac{-\ln(\delta)}{\epsilon} = \frac{1}{\epsilon}\ln\left(\frac{1}{\delta}\right)$$

ולכן הראנו כי עבור מדגם בגודל $n(\epsilon,\delta) = \frac{1}{\epsilon}\ln\left(\frac{1}{\delta}\right)$ נקבל למידות PAC על פי ההגדרה!

באופן כללי עבור כל התפלגות D שהיא realizable עבור $\mathbb{R}^d$ עם

$H_{box} \ni B_1, B_2$ כך ש נקבל את המלבן המינימלי עם $n(\epsilon,\delta) \ge n$ נקבל $P(e_D(h_n) > \epsilon) < \delta$

בשביל זה שאין בעיה בחירת δ של הגודל המלבן מגדלים, שווה $n(\epsilon,\delta)$

בכל זאת תלוי ב- δ.

□

שאלה 2

נניח L נתון.

⊕ אם H למידה, PAC עבור C-H למידה PAC בגרסה:

$$E[ep(A(S))] \overset{\text{תוחלת שלמה}}{=} E[ep(A(S)) \mid ep(A(S)) \le \epsilon] \underbrace{P(ep(A(S)) \le \epsilon)}_{\le 1} + \underbrace{E[ep(A(S)) \mid ep(A(S)) > \epsilon]}_{\le 1} \cdot \underbrace{P(ep(A(S)) > \epsilon)}_{\le \delta} \le \epsilon + \delta$$

נוסחה שלמה

$\le \epsilon$ כנתון מה שתנאי מבטיח

≤ 1 מהגדרת ep

$\le \delta$ מהגדרת PAC עבור m גדול מספיק

לכן ניתן לבחור את $\delta(\epsilon)$ בגרסה של יגל, ומתן מספר דוגמאות מספיק לקבל הצלחה $\ge \delta$, לקבל את הדרוש, ואפשר גם לבחור את ϵ כך שלא יהיה קבוע ולא, שתלוי לפי הגדרת PAC.

⊛ אם H למידה PAC בגרסה, נניח שהיא למידה PAC. מאי-שוויון מרקוב נקבל כי:

$$P(ep(A(S)) \ge \epsilon) \le \frac{E[ep(A(S))]}{\epsilon}$$

ולכן מהגדרת PAC בגרסה נקבל

נבחר את $\delta \cdot \epsilon \ge \eta$ ונקבל:

$$P(ep(A(S)) \ge \epsilon) \le \frac{\delta \cdot \cancel{\epsilon}}{\cancel{\epsilon}} \le \delta$$

ולכן H למידה PAC (רגיל).

הערה: בכל הגרסאות מדובר על התפלגות P שהיא realizable עבור H וקיימת שלמות שונה i.i.d ממנה דוגמה

שאלה 3

נניח שה-Vc-dim הוא 4. נניח שקיימת קבוצה בגודל 4 אשר מנותצת ע"י H. קיימת לה תת קבוצה בגודל 3 שמנותצת ע"י H, וכן, בפרט נוכל להסתכל על 3 נקודות באופן הבא:

אם x=0 [illegible] [illegible]. [illegible] [illegible] [illegible] [illegible]. [illegible] [illegible] [illegible] [illegible] בקודקודים שלו,

ניתן לתת [illegible] [illegible] [illegible] [illegible] [illegible]. [illegible] 2 קודקודים [illegible] ניתן [illegible] [illegible] [illegible] [illegible], [illegible] [illegible] [illegible] [illegible] [illegible] [illegible], [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] קודקודים [illegible] [illegible] [illegible].

[illegible] [illegible] [illegible] [illegible] [illegible] [illegible]. [illegible] [illegible] [illegible] [illegible] [illegible]: [illegible]... [illegible]. [illegible] [illegible] [illegible] [illegible] [illegible] 2 קודקודים [illegible] [illegible] ניתן לשים [illegible] [illegible], [illegible] לא ניתן [illegible] [illegible] [illegible] 4 [illegible]!

לכן הוכחנו שה-Vc-dim של H שווה ל-3. □

שאלה 4

נניח שקיים מודל של המחלקה הוא ~~X~~ ∞, ובשביל זה נראה שלכל מספר קיימת השמה (למא,...,רא) של קבוצות שמספקת את ידי המחלקה. נראה זאת עבור לכל השמה של משתנים סופית. את השמה כזו: (רל,מא ,..., (אל,רא)).

~~X~~ יהי עץ משולש של השמות. לכל קודקוד מסוים ~~X~~ יש n ~~X~~ x_i ו-n ~~X~~ שיורי z_i באופן הבא:

שיורי

(*) אם הקודקוד בענף של 1 ל-(x_i), אז נסמן $z_i = x_i$.

(*) אם הקודקוד בענף של 0 ל-(x_i), אז נסמן ~~X~~ $z_i = 1 - x_i$

כך לכל השמה ~~X~~ עם לפחות שיש סופית (מחלקה n-ית) שיורים את העץ, ולכן לכל ענף קיים סופית למחלקה שמקיים אותו, כלומר ההשמה מתקיימת!

לכן מימד ה-עץ הוא ∞.

□

שאלת בונוס

(9) נתונה לנו ההתפלגות:

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

וכעת נרצה למצוא את h אשר מקיימת $h = \arg\min_{h\in H} e_p(h)$.

נשים לב שאם $\hat{x}\in[0,0.2]\cup[0.4,0.6]\cup[0.8,1]$ מתקיים: $x=\hat{x}$

$$P(Y=1|\hat{x}) = 0.8 > P(Y=0|\hat{x}) = 0.2$$

ועבור שאר [illegible] עבור $\hat{x}$:

$$P(Y=0|X=\hat{x}) = 0.9 > P(Y=1|X=\hat{x}) = 0.1$$

מכיוון ש-X מתפלג יוני-פורמית ב-[0,1] ולכן כיוון שאנחנו בוחרים לפי שיטת ההסתברות, נבחר ל-1 מתקיים כאשר כן הסיכוי יותר גבוה להיות 1, ובאופן דומה נבחר ל-0 כאשר בו הסיכוי יותר גבוה להיות 0.

כיוון שאנחנו שואפים לעשות את זה רק מכיוון שההתפלגות כבר ידועה לנו מראש. אם לא היה לנו את ההתפלגות, הטעות [illegible] המינימלית (וזה מאפשר לנו לקבל את ה[illegible])

ולכן נבחר את המסווג האופטימלי $h\in H$ להיות:

$$h(x) = \begin{cases} 1 & \text{if } x\in[0,0.2]\cup[0.4,0.6]\cup[0.8,1] \\ 0 & \text{else} \end{cases}$$

□

ⓑ השיטה הראשונה יותר טובה כי k-NN מתחשבת בכל k-NN אתר, יש סיכוי
יותר שיהיה הדימיון הכי גבוה בין הקלטים שלו. השיטה השנייה
מגלה רגישות לכל k-NN אתר, כלומר מניין שבכל פיקסל יותר יכולה
לשנות, כי ממילא יותר להתאים את הצורה שלו.
בנוסף לכך, שיטה השיטה הזכות המקורית לא נכון, מה
שמתייחס עם חלק התכונות, הקלטים המשתנים, כלומר שבכל שינוי עם זה
הדמיון, השיטה השנייה מחשבת ומרגישה השיטה.

ⓒ השיטה השנייה יותר טובה כי יותר זול, זה אם אין השיטה השנייה
הראשונה היא פיקסלית. השיטה השנייה יותר לא תוכל את כלל הפיקסל (k=3)
ולכן יש מחיר רב יותר, כלומר "לפעמים" זה הדבר הכי נכון שיהיה k
ולכן כתבו זה שהכל מתכוון מודל לא יותר מלפרט לכל overfitting, כלומר
כאשר יהיה שלמדנו העוצמה את יותר קלטים הכוללים בהם מופע. וגם
יותר הדוגמא, אך לפעמים את התוצאה הטובה בין קלטים.
לטובת

ⓓ יותר יותר שמתי על k- pixels השיטה הראשונה
לכן מינימלי, כלומר יותר k שבין הפיקסלים המתאימים על
את הטווח המלא. לא k גבוה אין לכן יותר מתאים
שמתאים ל-C, שלכן יותר אין לה שמתאים את השיטה השנייה, כלומר לכן
זה יהיה מצב של overfitting כמו בסעיף C.

(e) [illegible] $k=3$ [illegible] ERM_1 [illegible] H_3

[illegible] Holdout [illegible].

[illegible] validation [illegible] training data,

[illegible] H_4 [illegible] Holdout [illegible]

[illegible]:

$$P_{S_2}\left(e_p(ERM(S_2)) - \min_{h\in H_3} e_p(h) > \epsilon\right) \le 4\,\Pi_{H_3}(2|S_2|)\cdot e^{-\frac{|S_2|\epsilon^2}{32}}$$

[illegible] VC [illegible] H_3 [illegible] 6 [illegible] 3 [illegible]:

$$\Pi_{H_3}(2|S_2|) \le \sum_{i=1}^{6}\binom{2|S_2|}{i} \le \left(\frac{2e|S_2|}{6}\right)^6 \quad (\text{Sauer's Lemma})$$

[illegible]:

$$P_{S_2}\left(e_p(h_H) - \min_{h\in H_3}(h) > \epsilon\right) \le 4\left(\frac{e|S_2|}{3}\right)^6\cdot e^{-\frac{\epsilon^2|S_2|}{32}} = \delta$$

[illegible] ϵ:

$$4\left(\frac{e|S_2|}{3}\right)^6 e^{-\frac{\epsilon^2|S_2|}{32}} = \delta$$

$$e^{-\frac{\epsilon^2|S_2|}{32}} = \frac{\delta}{4\left(\frac{e|S_2|}{3}\right)^6}$$

$$\frac{\epsilon^2|S_2|}{32} = \ln\left(\frac{4}{\delta}\left(\frac{e|S_2|}{3}\right)^6\right) = \ln\left(\frac{4}{\delta}\right) + 6\ln\left(\frac{e|S_2|}{3}\right)$$

[illegible] $|S_2| = 0.2\cdot 1500 = 300$ [illegible]:

$$\frac{300\epsilon^2}{32} = \ln\left(\frac{4}{\delta}\right) + 6\ln(100) + 6$$

$$\epsilon = \sqrt{\frac{8}{75}\ln\left(\frac{4}{\delta}\right) + \frac{16}{25}(1+\ln(100))}$$

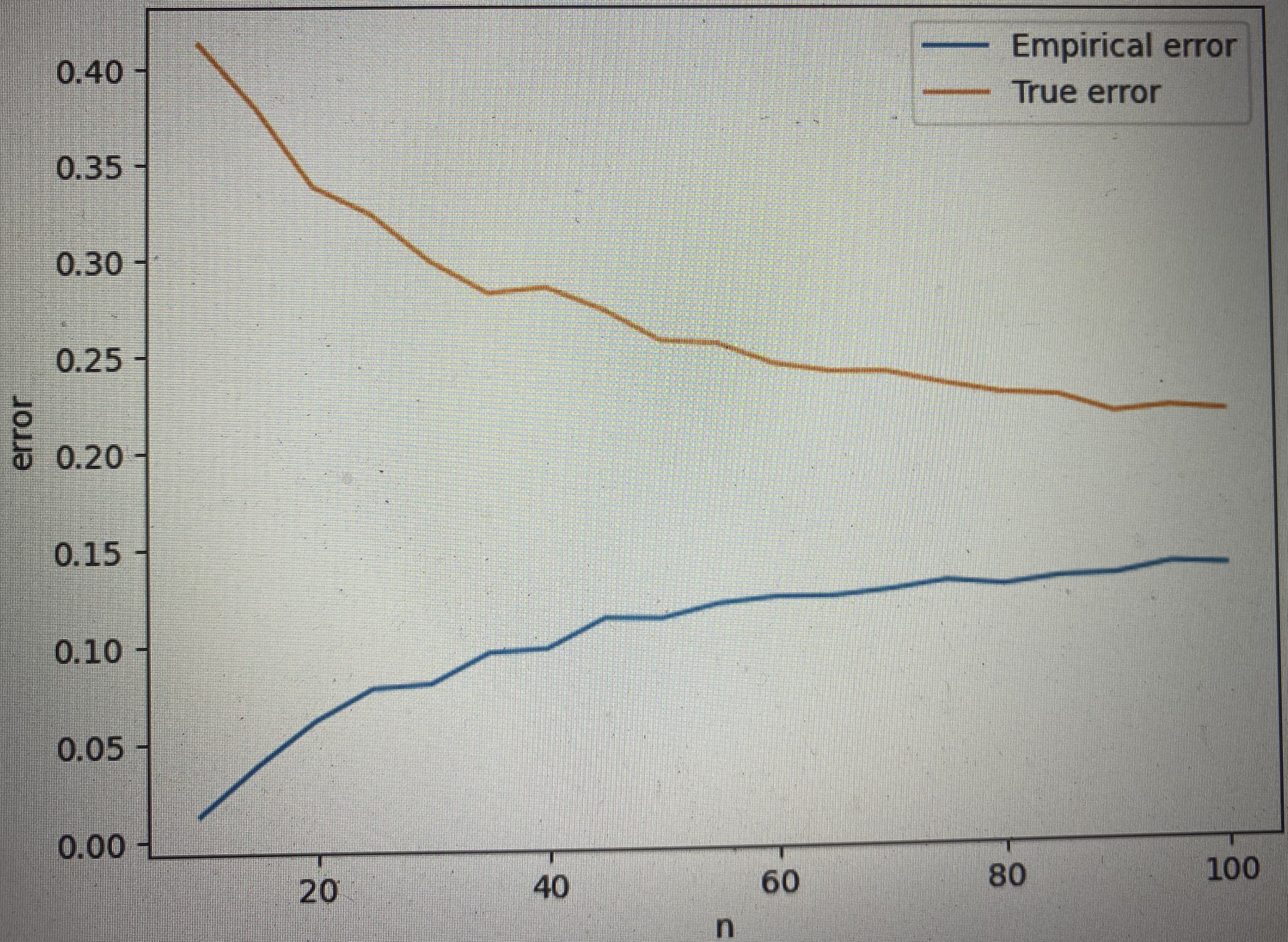
[illegible]:

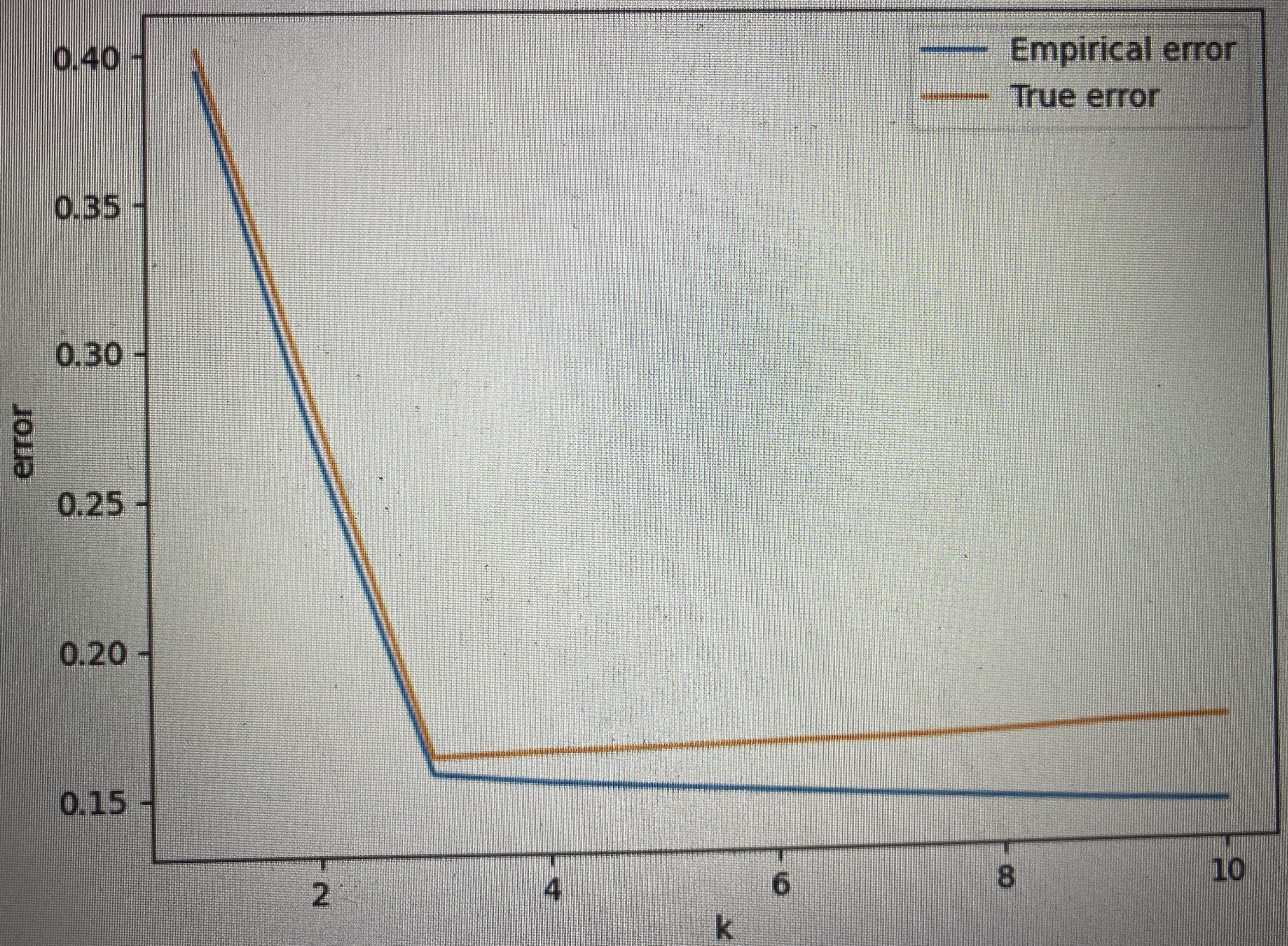
$$e_p(h_H) - \min_{h\in H_3} e_p(h) < \sqrt{\frac{8}{75}\ln\left(\frac{4}{\delta}\right) + \frac{16}{25}(1+\ln(100))}$$

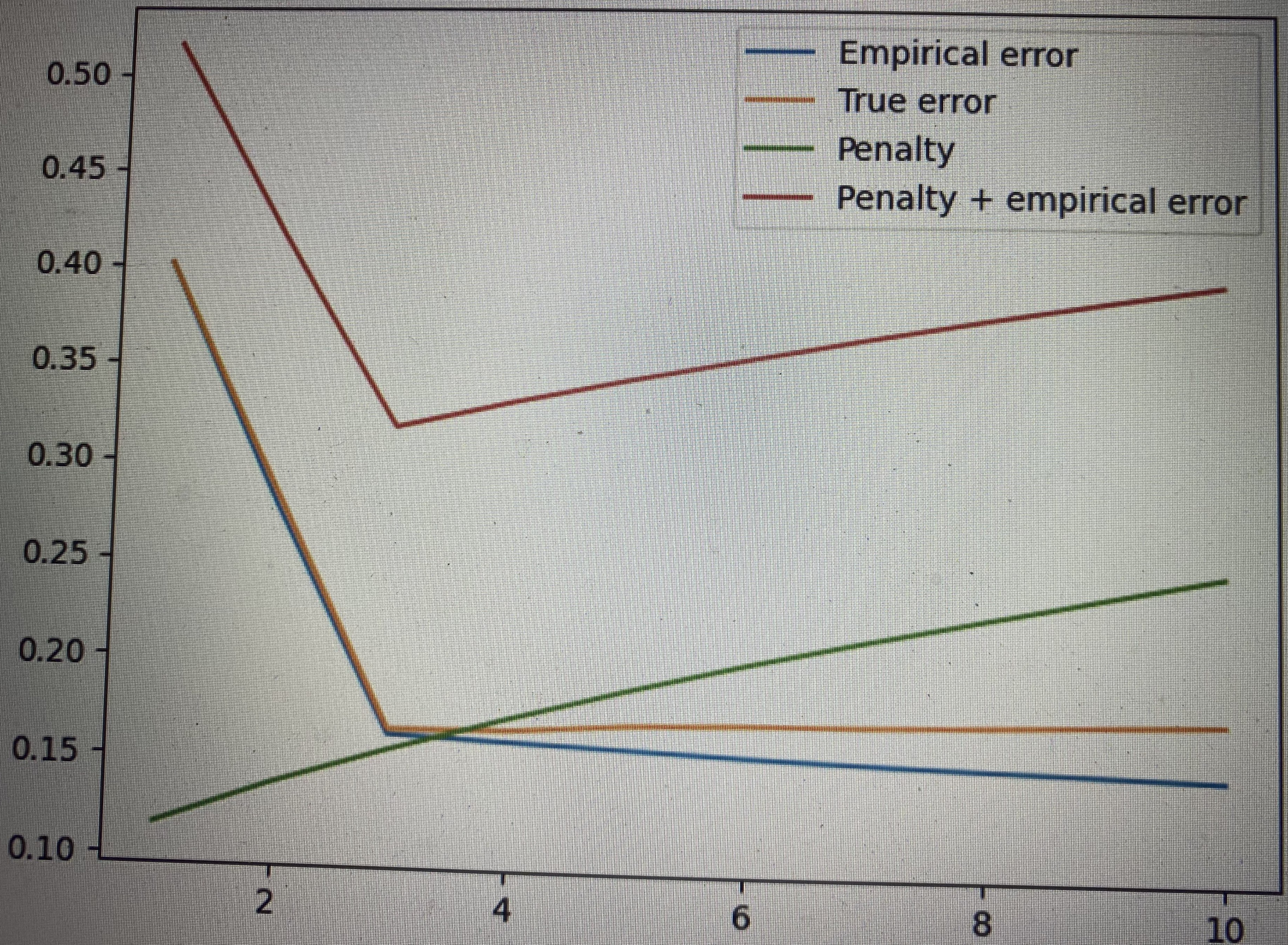
[illegible] 3 [illegible] ERM [illegible]:

[(0.0000329, 0.19016295), (0.4006[illegible], 0.5997[illegible]), (0.7004[illegible], 0.9988[illegible])]

([illegible]) □


Empirical error
True error
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
error
20
40
60
80
100
n


Empirical error
True error
0.40
0.35
0.30
0.25
0.20
0.15
error
2
4
6
8
10
k



Empirical error
True error
Penalty
Penalty + empirical error
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
2
4
6
8
10