מבוא ללמידה חישובית – תרגיל 2

1

נניח H היא למידה PAC בתוחלת עבור אלגוריתם A כלשהו כלומר קיימת $N(a): (0,1) \to \mathbb{N}$ כך שמתקיים לכל $a \in (0,1)$ ולכל התפלגות P ול-S קבוצת דוגמאות מתוך P כך ש $|S| \geq N(a)$ מתקיים:

$$\mathbb{E}[e_p(A(S))] \leq a$$

נוכיח H למידה PAC , יהי $\epsilon, \delta \in (0,1)$ מאי שיוון מרקוב מתקיים

$$P\left[e_p\left(A(S)\right) > \epsilon\right] \leq P\left[e_p\left(A(S)\right) \geq \epsilon\right] \leq \frac{\mathbb{E}[e_p(A(S))]}{\epsilon} \leq \frac{a}{\epsilon}$$

מכיוון שזה נכון לכל a נבחר $a < \epsilon\delta$ ונקבל

$$P\left[e_p\left(A(S)\right) > \epsilon\right] < \delta$$

לכן ניתן להגדיר $N'(\epsilon_0, \delta_0) = N(a_0)$ כך שa_0 מקיים $a_0 < \epsilon_0\delta_0$ ונקבל את שH למידה PAC

נניח H למידה PAC כלומר קיימת $N(\epsilon, \delta): (0,1) \times (0,1) \to N$ כך שלכל $\epsilon, \delta \in (0,1)$ ולכל התפלגות P ול-S קבוצת דוגמאות מתוך P כך ש $|S| \geq N(\epsilon, \delta)$ מתקיים:

$$P\left[e_p\left(A(S)\right) > \epsilon\right] < \delta$$

יהי $a \in (0,1)$ מכיוון שH למידה PAC לכל $\epsilon, \delta \in (0,1)$ מתקיים עבור $\epsilon = \delta = \frac{a}{2}$

$$\mathbb{E}\left[e_p\left(A(S)\right)\right] \underbrace{=}_{1} \mathbb{E}\left[e_p\left(A(S)\right) \middle| e_p\left(A(S)\right) > \epsilon\right] P\left[e_p\left(A(S)\right) > \epsilon\right] + \mathbb{E}\left[e_p\left(A(S)\right) \middle| e_p\left(A(S)\right) \leq \epsilon\right] P\left[e_p\left(A(S)\right) \leq \epsilon\right] \underbrace{\leq}_{2} \epsilon + \delta = a$$

כאשר 1 נובע ממשפט התוחלת השלמה ו2 נובע מ:

$\mathbb{E}\left[e_p\left(A(S)\right) \middle| e_p\left(A(S)\right) > \epsilon\right] \leq 1$ כי ממוצע על e_p עם פונקציית הפסד 0-1 הוא לכל היותר 1

$P\left[e_p\left(A(S)\right) > \epsilon\right] < \delta$ כי H למידה *PAC*

$\mathbb{E}\left[e_p\left(A(S)\right) \middle| e_p\left(A(S)\right) \leq \epsilon\right] \geq \epsilon$ ממוצע של איברים שהם לכל היותר אפסילון הוא לכל היותר אפסילון

$P\left[e_p\left(A(S)\right) \leq \epsilon\right] \leq 1$ הסתברות קטנה שווה ל1

לכן ניתן להגדיר $N'(a) = N\left(\frac{a}{2}, \frac{a}{2}\right)$ ומתקיים

$$\mathbb{E}[e_p(A(S))] \leq a$$

כנדרש.

.2

נוכיח כי ה*VC-dimension* של משפחת היפותזות *H* המוגדרת כל ידי איחוד של *k* קטעים על הישר הממשי הוא *2k*

תחילה נראה כי *VC-dimension(H)* $\geq 2k$:

נסתכל על קבוצה *X* המכילה *2k* נקודות על הישר הממשי נמספר אותם על פי המיקום של על הישר משמאל ימין כאשר את הכי שמאלי נסמן בx_1 נראה שניתן לנפץ את *X* על ידי *H* נסמן את המרחק המינימלי בין שתי

נקודות בm נסתכל על הקטעים $[x_i, x_{i+1}]$ עבור i אי זוגי בתחום $[1,2k-1]$ נשים לב h המגודרת על ידי הטעים הללו מחזירה 1 עבור כל $x_i \in X$ אם נרצה לתייג נקודה בעלת אינדקס t זוגי ב0 פשוט ניקח את הקטע $\left[x_{t-1}, x_t - \frac{m}{2}\right]$ עבור t אי זוגי הקטע המתאים יהיה $\left[x_t + \frac{m}{2}, x_t\right]$ כמובן אם רוצים לתייג 2 המגדירות אותו קטע ב0 הקטע המתאים יהיה $\left[x_t + \frac{m}{2}, x_{t+1} - \frac{m}{2}\right]$ לכן ניתן לנתץ את X כלומר

$$VC-dimension(H) \geq 2k$$

נראה שלא ניתן לנתץ X בגודל *2k+1*:

נמספר את הנקודות משמאל לימין על פי המיקום שלהם על הישר הממשי כאשר הראשונה מסומנת x_1 נראה שאין K קטעים כך שכל הנקודות עם האינדקס הזוגי מקבלות לייבל 0 ואינקס אי-זוגי מקבל לייבל 1 נשים לב שיש *k+1* נקודות עם אינדקס אי זוגי לכן חייב להיות קטע שמכיל שתי נקודות אבל אם קיימם קטע שמכיל שתי נקודות אז הוא מכיל גם נקודה בעלת אינדקס זוגי ולכן הלייבל שלה שונה מ0 לכן H לא מנתצת X בגודל *2k+1*.

3.

נראה ש*vc-dimensions* של המחלקה הנתונה הוא *2d*

נסתכל על הנקודות $x_1, x_2, \dots, x_{2d} \in \mathbb{R}^d$ השונות המקיימות $x_{i_j} = \begin{cases} t_i \ if \ i \ mod \ d = j \\ 0 \ if \ i \ mod \ d \neq j \end{cases}$ כאשר t_i הוא מספר ממשי כלשהו שונה מ0 שמקיים שאם $i \leq d$ אז t_i שלילי אחרת t_i חיובי

נשים לב לכל קורדינטה j יש רק שני נקודות בהם הקורדינטה אינה 0.

יהיה $s_1, s_2, \dots, s_{2d} \in \{0,1\}$ לייבלים כלשהם נראה שניתן להגדיר $h_{a,b}$ כך ש$h_{a,b}(x_i) = s_i$ או במילים אחרות H מנתצת את $x_1, x_2, \dots, x_{2d}$

אם $s_i = 0$ נסמן $i \ mod \ d = c$ ונסתכל על הווקטור הנוסף שהקורדינטה במקום הc איננה אפס אם ערכה גדול מערך בx_i נבחר $\boldsymbol{a}$ שמקיים $0 > \boldsymbol{a}_c > x_{i_c}$ אם אחרת נבחר $\boldsymbol{b}$ שמקיים $0 < b_c < x_{i_c}$.

אם $s_i = 1$ נסמן $i \ mod \ d = c$ ונסתכל על הווקטור הנוסף שהקורדינטה במקום הc איננה אפס אם ערכה גדול מערך בx_i נבחר $\boldsymbol{a}$ שמקיים $0 > \boldsymbol{a}_c < x_{i_c}$ אם אחרת נבחר $\boldsymbol{b}$ שמקיים $0 < b_c > x_{i_c}$.

ונקבל $h_{a,b}$ כך ש$h_{a,b}(x_i) = s_i$ כלומר *vc-dimension(H)*$\geq 2d$

יהיו $x_1, x_2, \dots, x_{2d}, x_{2d+1} \in \mathbb{R}^d$ שונים נראה שלא ניתן לנתץ אותם:

נסתכל על ערכי המקסימום והמינימום עבור כל קודינטה יש d ערכי מקסימום וd ערכי מינימום כלומר קיים x_i כך שאינו מכיל אף מינימום או מקסימום נרצה למצוא h כך שעבור כל x פרט לx_i מחזיר אחד ועבור x_i מחזיר 0 מכיוון ששאר הווקטורים מכילים את הערכי המינימום והמקסימום של עבור כל קודינטה חייב להתקיים $a_k = \min_{j \in \{1,\dots,2d+1\}} x_{j_k}$, $b_k = \max_{j \in \{1,\dots,2d+1\}} x_{j_k}$ עבור כל $k \in \{1, \dots, d\}$ אבל מתקיים לכל k $a_k \leq x_{i_k} \leq b_k$ לכן גם $h(x_i) = 1$ ולכן לא ניתן למצוא h כזו כלומר H לא מנתצת את $x_1, \dots, x_{2d+1}$ לכן

Vc-dimension(H) = 2d

4.

(a)

$$\Pi_{H_1 \cup H_2}(n) = \max_{C \subseteq X, |C| = n} |(H_1 \cup H_2)_C|$$

כמו כן:

$$
\begin{aligned}
(H_1 \cup H_2)_c &= \{[h(x_1), h(x_2), \dots, h(x_n)] | h \in H_1 \cup H_2\} \\
&= \{[h(x_1), h(x_2), \dots, h(x_n)] | h \in H_1\} \cup \{[h(x_1), h(x_2), \dots, h(x_n)] | h \in H_2\} \\
&= H_{1_c} \cup H_{2_c}
\end{aligned}
$$

לכן:

$$
\begin{aligned}
&\max_{C\subseteq X, |C|=n} |(H_1 \cup H_2)_C| \\
&\quad = \max_{C\subseteq X, |C|=n} \left|H_{1_c} \cup H_{2_c}\right| \le \max_{C\subseteq X, |C|=n} \left|H_{1_c}\right| + \max_{C\subseteq X, |C|=n} \left|H_{2_c}\right| \\
&\quad - \max_{C\subseteq X, |C|=n} \left|H_{1_c} \cap H_{2_c}\right| \le \max_{C\subseteq X, |C|=n} \left|H_{1_c}\right| + \max_{C\subseteq X, |C|=n} \left|H_{2_c}\right| = \Pi_{H_1}(n) + \Pi_{H_2}(n)
\end{aligned}
$$

(b)

עבור *n* מתקיים:

$$
\begin{aligned}
|(H_1 \otimes H_2)_c| &= |\{[h_1(x_1)h_2(x_1), \dots, h_1(x_n)h_2(x_n)] | h_1 \in H_1, h_2 \in H_2\}| \\
&= \sum_{[h_2(x_1), \dots, h_2(x_n)] \in H_{2_c}} \sum_{[h_1(x_1), \dots, h_1(x_n)] \in H_{1_c}} 1 = |H_{1_c}||H_{2_c}|
\end{aligned}
$$

כמו כן

$$
\begin{aligned}
\Pi_{H_1 \otimes H_2}(n) &= \max_{C\subseteq X, |C|=n} |(H_1 \otimes H_2)_c| = \max_{C\subseteq X, |C|=n} (|H_{1_c}||H_{2_c}|) \le \max_{C\subseteq X, |C|=n} |H_1| \max_{C\subseteq X, |C|=n} |H_2| \\
&= \Pi_{H_1}(n)\Pi_{H_2}(n)
\end{aligned}
$$

חלק מעשי

(a).

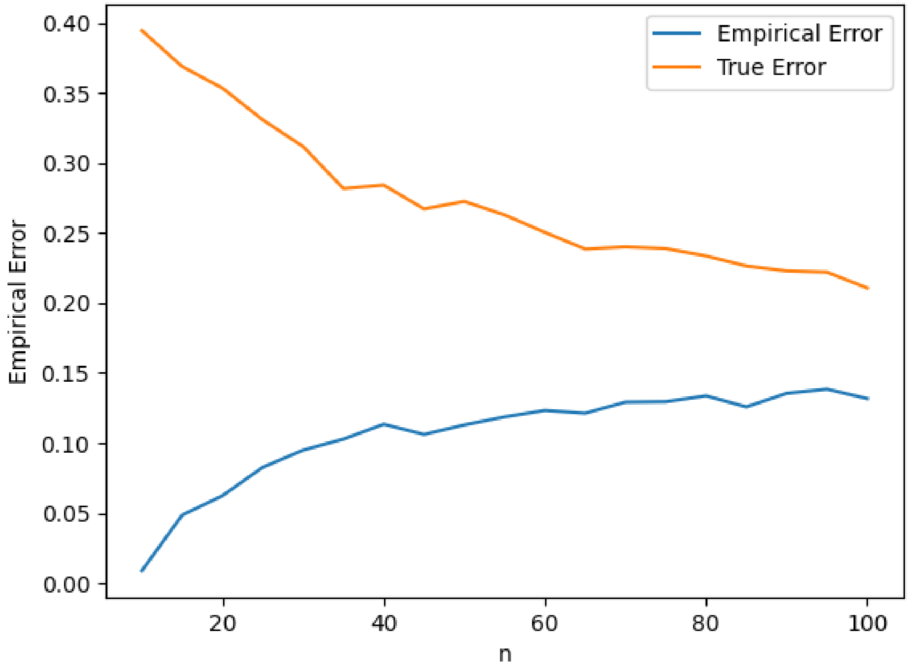
ראינו בהרצאה ש*h* עם השגיאה האמיתית הנמוכה ביותר כאשר פונקציית היא *zero-one loss* מתקבלת כאשר:

$$
h(x) = \begin{cases} 1 & \mathbb{P}(Y = 1 | X = x) > \mathbb{P}(Y = 0 | X = x) \\ 0 & otherwise \end{cases}
$$

לכן ה $h \in H_{10}$ בעל השגיאה המינימלית הוא:

$$
h(x) = \begin{cases} 1 & if\ x \in [0,0.2] \cup [0.4, 0,6] \cup [0.8, 1] \\ 0 & otherwise \end{cases}
$$

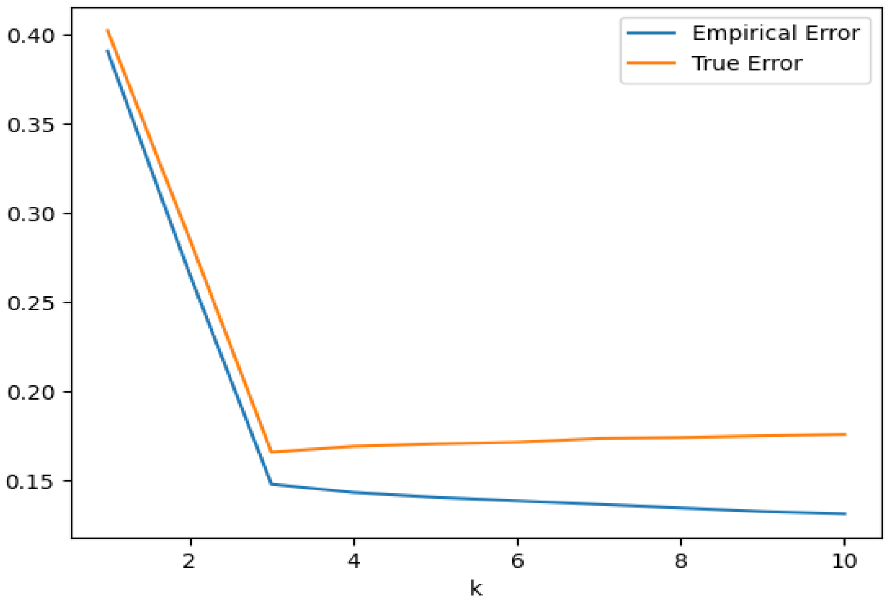
(b)



תחילה נשים לב כי כצפוי כלל שגודל הדגימה עולה השגיאה האימפרית מתקרבת לשגיאה האמיתית כי שראינו בשיעור.

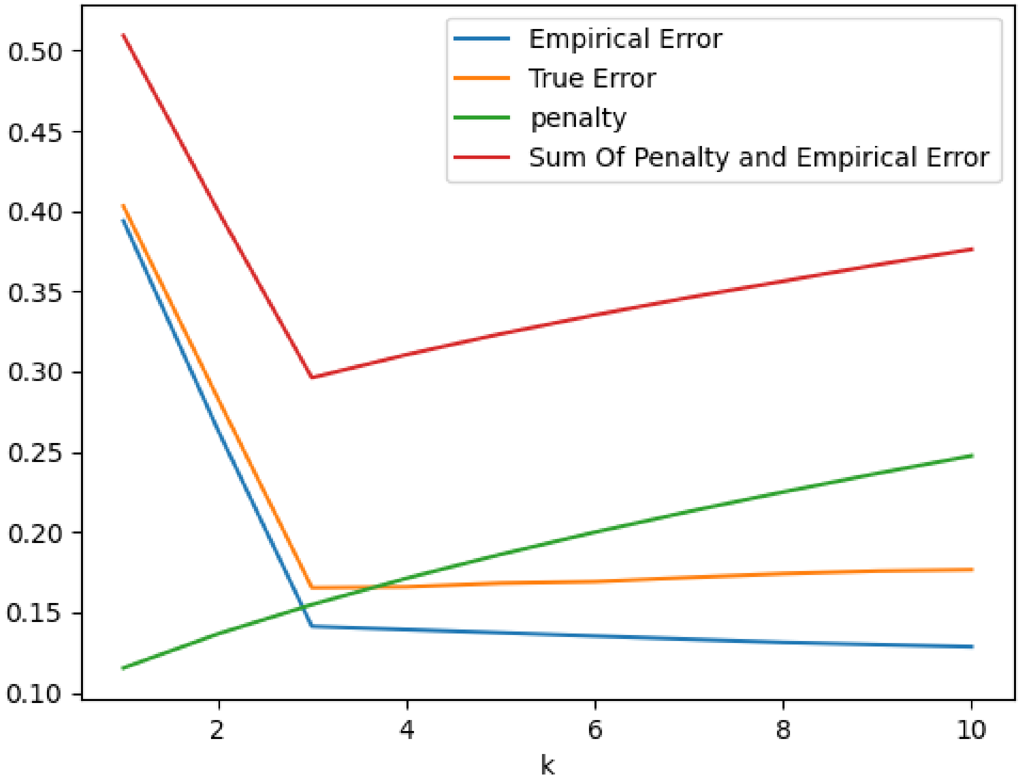
השגיאה האמפרית יורדת כי ככל שיש יותר דוגמאות ככה אנו מתקרבים להתפלגות האמיתית הנתונה כמו כן ככל שמספר הדוגמאות גדל השונות של הדוגמאות גדולה יותר לכן יש יותר טעויות למסווג כלומר שגיאה אמפרית גדולה יותר

(c)



השגיאה האמפירית יורדת כלל שכמות הקטעים עולה ופרט הא K הטוב ביותר עבור השגיאה האמפירית הוא 10 אבל ניתן לראות שהשגיאה האמיתית הנמוכה ביותר מתקבלת עבור $K = 3$ ניתן לראות שהיה *over fitting* כלומר נתנו מענה טוב לקבוצת המדגם אבל פיספסנו את התפלגות הטיבעית.

(d)



ניתן לראות שהשגיאה האמפירית עם הקנס האופטימלית היא כאשר k שווה שלוש כמו השגיאה האמיתית

(e)

לאחר הרצה של ה*cross validation* הk האופטימלי יצא 3 כפי שציפינו ננסה לתת חסם למרחק של היפותזה ממשפחה H_3 לשגחאה האמיתית.

נסמן את קבוצת ה*validation* בS_2 ונסתכל על היפותזות מH_3

נעבוד עם החסם הכללה מהרצאה

$$P_S\left[e_P(\mathrm{ERM}(S)) - \min_{h\in\mathcal{H}} e_P(h) > \epsilon\right] \leq 4\left(\frac{2en}{d}\right)^d e^{\frac{-n\epsilon^2}{32}}$$

כאשר *d=6* וגודל S_2 הוא 300 לכן ניתן לבודד את אפסילון

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

וזהו חסם עליון למרחק בין השגיאה האמפירית על S_2 לשגיאה האמיתית