מועד א' פתרון חישובית תרגיל בית 2

1) נשקיע פונקציה $A(S)$ עם $S=\{(x_1,y_1),\dots,(x_m,y_m)\}$

באופן הבא:

- את ק"ס ו ק ש- $y_i=1$ נחזיר את h_{x_i}
- אחרת נחזיר h^-

הוכחה:

נ"ש שלגבי אין"א פונקציה שמסווגת את $e_S(h)$ את $A(S)$.

נניח שאין כזו ש- $y_i=1$. כלומר $A(S)=h^-$ ולכן $e_S(h^-)=\sum_{j=1}^{m}\Delta_{0z}(0,y_j)=0$ משום שהנאמנה נותנת הסברים כשגרים.

נ"ש שכן ק"ס ו שעבורו $y_i=1$. כלומר $A(S)=h_{x_i}$ ולכן $e_S(h_{x_i})=\sum_{j=1}^{m}\Delta_{0z}(h_{x_i}(x_j),y_j)=\left[\sum_{\substack{j=1\\ j\neq i}}^{m}\Delta_{0z}(0,y_j)\right]+\Delta_{0z}(1,1)=$

$$=\sum_{\substack{j=1\\ j\neq i}}^{m}\Delta_{0z}(0,y_j)$$

אותו כשגרם שיקרו y_j שאינו 1 כך שגם עבור x_i מתקיים $\Delta_{0z}(0,y_i)=\Delta_{0z}(0,0)=0$ שכן סה"כ ערכי $e_S(h_{x_i})$ כשגרים, שכן $e_S(A(S))=0$ כפי שרצינו להוכיח.

2) כלומר $1-$ שעה λ-קטן $\mathbb{E}[e_p(A(S))]$ (*) ונכנה λ-H
PAC ניתנת. (*) מתקיים $|S| > N(\lambda)$
יהי $\delta/3$. נקרא $N = N(\lambda \cdot \delta/3)$. יהי S כך ש$\overset{\text{גודל}}{}$
$\lambda \cdot |S| > N$ ונקבל:

$$\delta = \frac{\delta}{3} \geq \frac{\mathbb{E}[e_p(A(S))]}{\lambda} > \mathbb{P}[e_p(A(S)) > \lambda]$$

כנדרש.

הוא מקרה הפשוט שבו אלג' מוציא תמיד אותו מ.ס. כך
שלכל $\delta > 0$ קיים $\lambda > 0$. (כלומר N בלתי תלוי ב-δ)
כלומר $2-$ מתקיים עבור $\delta/3$ הנתון מבקרת S. כך
ש- $(\delta/3)N < |S|$ מתקיים $\mathbb{P}[e_p(A(S)) > \lambda] < \delta$.
~~[redacted]~~. נבדיל $N: (0,1) \to \mathbb{N}$ כך- $N(\lambda) = N(\frac{\lambda}{2}, \frac{\delta}{2})$
נקבל: לכל מ.ס S כך ש- $|S| > N(\lambda)$:

$$\mathbb{E}[e_p(A(S))] = \mathbb{E}[e_p(A(S)) | e_p(A(S)) < \frac{\lambda}{2}] \cdot \mathbb{P}[e_p(A(S)) < \frac{\lambda}{2}] +$$
$$\mathbb{E}[e_p(A(S)) | e_p(A(S)) \geq \frac{\lambda}{2}] \cdot \mathbb{P}[e_p(A(S)) \geq \frac{\lambda}{2}] +$$

נשארנו עם אלגוריתם הלמ"ה ונסיים
ב אחד מהמקרים:

1) $\mathbb{E}[e_p(A(S)) | e_p(A(S)) < \frac{\lambda}{2}] \leq 1$
(כלומר - כל קשר שנוצר עם d_p, תא"ז הוצעת
על $d_p(A(S))$ וכל נ"ן 0 ל-1 מוקצית d_p
2) $\mathbb{P}[e_p(A(S)) > \frac{\lambda}{2}] < \frac{\delta}{2}$
(כלומר - לכל הגדרה של PAC-Learnable מתר
$\frac{\delta}{2} = \delta/3$
3) $\mathbb{E}[e_p(A(S)) | e_p(A(S)) \geq \frac{\lambda}{2}] \leq 1$
(כלומר - גודלתן $d_p(A(S))$ קבוע ואת הוצעת על N
על $d_p(A(S))$ וכל שלל יתכן $\frac{\lambda}{2}$ הוצעת על N
שנבה שלוות שקולה יותר מהמכ"ם ושאול אותה N
4) $\mathbb{P}[e_p(A(S)) \geq \frac{\lambda}{2}] \leq 1$
(כלומר - הסתברות, תא"ז קשה את אחד 1-ל

סה"כ נקבל: $\lambda = 1 \cdot \frac{\lambda}{2} + \frac{\lambda}{2} \cdot 1 \geq \mathbb{E}[e_p(A(S))]$

כנדרש.

ושל"ו גדול והשלם, כלומר שלם יש בין שלם את 3 הנק' אותו מהן מחוץ למשולש הקו יצטרך שיעבור את הצלע השלישית שלו). אותו תמיד יהיו את שני הנקודות שבאמצע שלוש הקרובות ביותר. אותו 3 נקי על אותו של סה"כ (3 אשר שוכן 2 מהצלע. כאשר שאלנו עם הנתון - הצלע שלם את הנקודות שנותרו עם קו ישר אחד שמוכן את שני שלא יה עוד האלה תמיד אפשר שכיסינו את כל הנקודות $\hat{x}=\hat{y}=0$, כאשר השאלה על המשולש היא $(0,0)$ ולכן גם משולש שאנחנו שאלת שהנקודות על ה"ש). אלה"כ עוד $\hat{y}\geq y$ ו- $\hat{x}\geq x$. אלה"כ $\hat{x}=x$ (כאשר $\hat{y}=y$ ע"י שינינו השאלה תהיה א- $(\hat{x},\hat{y})$ כך שלם ל משולש שיכלו את שני הנקודות והצלע מוכרח נאמנה (x_0,y_0) ונקודה הכי תחתונה, (x_1,y_1).

המשלם יהיו 5 נקודות. אלה"כ יש נקודה הכי שמאלית נאמנה (x_0,y_0) ונקודה הכי תחתונה, (x_1,y_1).

4) רק 5 נקודות עם תווית 1-

3) 4 נקודות עם תווית 1-

2) נקודה אחת עם תווית 1- שיכלו את כל הנקודות עם תווית 1/0 כאלו שיכלו את שני $\{(-1,0),(0,1),(1,0),(0,-1)\}=S$. כאשר השאלה היא:

אי"מ ו- אם הוכח y. נראה קושאה על 4 נקודות

4) הוא"מ (ו/ו) 2. הוכח: $\{(x_1,y_1),(x_2,y_2)\}$, בהינתן אותה
- קל לראות שאחרת ע"י שקילות ת"א ניתן
לדרוש שמתקיים עבור אחת הנקודות ה-Label האפשרי:
• מ"ש: $f(x_1,y_1)=f(x_2,y_2)=0$ - נניח כי מתקיים בנקודה בצילום
(x_2,y_2)
• מ"ש: $f(x_1,y_1)=f(x_2,y_2)=1$ - נניח שיתקיים את אחד הנקודות
$\{(x_1,y_1),(x_2,y_2),(x_1,y_2),(x_2,y_1)\}$ ונסמן $\hat{x}$ אחת
שמתקיים עבורן: $f(x_1,y_1)=f(\hat{x},y_1)$. ברור כי $x_1<\hat{x}<x_2$, נסביר את הבחירה ב- $\frac{x_1+x_2}{2}$ ו"אמ"
ע"פ משפט ערך הביניים קיים $\hat{x}$ כזה שבין x_1 ל-x_2 ו"אמ"
אחת הנקודות $(\hat{x},y_1)$ ו-$(\hat{x},y_2)$ מתקיים $f(\hat{x},y_1)\neq f(x_1,y_1)$.
• מ"ש: $f(x_1,y_1)=f(\hat{x},y_1)=1$. בלי ל"ש $f(x_1,y_1)=1$
- נניח ש"י מתקיים ש-3 נקודות שכן נתון שלפחות
ע"פ משפט כנ"ל, נניח בה"כ שקיימות $\hat{x}\hat{y}$ ושהן: $\{(\hat{x},\hat{y}),(\hat{x},\hat{y}+2),(\hat{x},\hat{y}+1)\}$
ע"פ פונקציה - $f(\hat{x},\hat{y})=0$, $f(\hat{x},\hat{y}+2)=1$, $f(\hat{x},\hat{y}+1)=f$
מלמ' משהו כנ"ל - נקודה אחת = Label 1.

שאיור ניתן מקרים

שקיל ל-x, נסמן
(x_2,y_2), (x_1,y_1) ח"ג ע"י $y_2=y_1$, $x_2>\hat{x}>x_1$
כך שיתקיים ניצב כזה $(\hat{x},\hat{y})$. אך נשים ש"י שקיים
שבו $(\hat{x},\hat{y}+2)$ הוא הנמצא הנמוך ביותר $\{(x_1,y_1),(x_2,y_2),(x_3,y_3)\}$
$g(\hat{x})>\hat{y}+2$ כאשר g פונקציה המוגדרת הישר (x_1,y_1)
(כאשר הישר עובר בין הנקודות הנ"ל). ומכיוון שקיים
$g(\hat{x})<\hat{y}+1$ אזי בין $(\hat{x},\hat{y}+1)$ הנמצא הנמוך ביותר
ל-g בנקודה.

3) יהי k מספר טבעי. הוכיחו כי $VCdim(H_k)=2k$

- יהי $S=\{x_1,\dots,x_{2k+1}\}$ קבוצה של $2k+1$ נקודות, כך ש- $x_1<x_2<\dots<x_{2k+1}$

נראה שלא ניתן לפרק את S ע"י k מלבנים. נגדיר $f(x_i)=\begin{cases}1, & i\in Nodd\\ 0, & i\in Neven\end{cases}$. נניח בשלילה שקיים $h\in H_k$ כך ש- $h(x_i)=f(x_i)$ עבור כל $1\le i\le 2k+1$.
אז קיימים $1\le j\le k$ כך ש- $h(x_{2j})=0$ ו- $h(x_{2j+1})=1$.
נסמן ע"י $I_j=(a_j,b_j)$ את המלבן שמכיל את x_{2j} ולא מכיל את x_{2j+1}. סה"כ $k+1$ מלבנים, בסתירה.

- נראה שקיים מצולע בעל $2k$ צלעות שניתן "לנפץ" ע"י k מלבנים. נשים לב שכל $N-2k$ צלעות של המצולע הוא קטן מספיק כך שכל הקודקודים שלו יהיו בתוך העיגול שמסומן בציור. נסמן את הקודקודים של העיגול ב- $x_1,\dots,x_{2k}$ בכיוון השעון. נגדיר $f:[0,1]\to\{0,1\}$ פונקציית Labeling כאשר $f(x_i)=h(x_i)$ עבור $h\in H_k$ כך ש- $f(x_i)\neq f(x_{i+1})$ לכל $i<2k$ שמסמן "מצולע", כלומר לכל קודקוד i כך ש- $f(x_i)\neq f(x_{i+1})$ או $i=2k$, כלומר הקודקוד האחרון.
נסמן את הקודקודים הללו, כלומר את "המצולע" ב- $\hat{X}=\{\hat{x}_1,\dots,\hat{x}_m\}$ כאשר $m\le 2k$. נגדיר את הקודקודים כך ש- $\hat{x}_j<\hat{x}_{j+1}$ לכל $j\le m$. נגדיר $h(x_i)=h(\hat{x}_j)$ עבור $x_i\in[\hat{x}_j,\hat{x}_{j+1})$. כלומר $h(x_j)=1$ ו- $h(x_{j+1})=0$ את הקודקוד x_i שמאלית ל- $\hat{x}_j$ מאחר ו- $x_i<\hat{x}_j<x_{i+1}$ בנוסף לכך שכל x שנמצא בין x_i לבין $\hat{x}_j$ מקבל את אותו ערך $h(x)=h(x_i)$ ולא $h(x)\neq f(x)$ כנדרש.
את המלבנים נגדיר כך שכל $I_j=[0,1]\setminus(\hat{x}_j,\hat{x}_{j+1})$ ונקבל את ההפרדה הנדרשת.

5) נרצה ל: $\varepsilon = \sqrt{\frac{1}{2n}\ln\left(\frac{2}{\delta \cdot w(h)}\right)}$

א. הוכיחו שמתקיים $\forall h \in E: P[e_p(h) - e_s(h) > \varepsilon] \le 2e^{-2n\varepsilon^2}$

נציב ונקבל:

$$P[e_p(h) - e_s(h) > \varepsilon] \le 2e^{-2n \cdot \frac{1}{2n}\ln\left(\frac{2}{\delta \cdot w(h)}\right)} =$$

$$= 2e^{-\ln\left(\frac{2}{\delta \cdot w(h)}\right)} = \frac{2}{\frac{2}{\delta \cdot w(h)}} = \delta \cdot w(h)$$

נקבע אחוד H כלשי:

$$P\left[\bigcup_{i \in \mathbb{N}} e_p(h_i) - e_s(h_i) > \varepsilon\right] \le \sum_{i \in \mathbb{N}} P[e_p(h_i) - e_s(h_i) > \varepsilon] \le$$

(Union Bound ←)

$$\le \sum_{i \in \mathbb{N}} \delta \cdot w(h_i) = \delta \cdot \sum_{i \in \mathbb{N}} w(h_i) \le \delta$$

כיוון ש- $\sum_{h \in H} w(h) \le 1$ הוא טור חיובי מתכנס (ראינו בתרגול)

שאלה: עד איזה ערך נוכל לבחור δ ו-δ יהיה עדיין δ-נכון?

נרצה $\delta = \delta$ ונשאל באמת מהו ערך δ המינימלי.

ב. עבור h כלשהו, מתקיים בוודאות δ-נכון:

$$e_p(h) \le e_s(h) + T_h \quad \text{כאשר } T_h = \sqrt{\frac{1}{2n}\ln\left(\frac{2}{\delta \cdot w(h)}\right)}$$

$$|e_p(h) - e_s(h)| \le T_h$$

נשתמש בסעיף הקודם ונקבל:

$$e_s(h) \le e_p(h) + T_h \quad (*) \qquad e_p(h) \le e_s(h) + T_h \quad (**)$$

נסמן ב- h_{srm} את הכלל שמחזיר את h המינימלי:

$$e_s(h_{srm}) + T_{h_{srm}} \le e_s(h) + T_h \quad (***)$$

נציב את $(*)$ ו- $(**)$ ב- $(***)$:

$$e_s(h_{srm}) + e_s(h) + T_{h_{srm}} \le e_p(h) + e_s(h) + 2T_h$$

נסיק את החסם עבור h_{srm}:

$$e_p(h_{srm}) \le e_s(h_{srm}) + T_{h_{srm}} \Rightarrow e_p(h_{srm}) + e_s(h) \le e_p(h) + e_s(h) + 2T_h$$

ולכן $e_p(h_{srm}) - e_p(h) \le 2T_h$ כאשר $h = h_{min}$ (כמו שראינו)

$$E[\Delta_b(Y,h(x))] = \sum_{x,y} P[X=x,Y=y]\cdot\Delta_b(y,h(x)) =$$

$$= \sum_x P[X=x,Y=0]\cdot\Delta_b(0,h(x)) + P[X=x,Y=1]\cdot\Delta_b(1,h(x)) =$$

$$= \sum_x P[X=x]\Big(\underbrace{P[Y=0|X=x]\cdot\Delta_b(0,h(x)) + P[Y=1|X=x]\cdot\Delta_b(1,h(x))}_{\circledast}\Big)$$

נבדוק מתי הביטוי $\circledast$ ממזער עבור x נתון, $\hat{x}$.

נפריד למקרים את הביטוי $\circledast$:

$$\circledast = \begin{cases} P[Y=0|X=\hat{x}], & h(\hat{x})=1 \\ P[Y=1|X=\hat{x}]\cdot\frac{1}{2}, & h(\hat{x})=0 \end{cases} = \begin{cases} 1-P[Y=h(\hat{x})|X=\hat{x}], & h(\hat{x})=1 \\ \frac{1}{2}\left(1-P[Y=h(\hat{x})|X=\hat{x}]\right) & h(\hat{x})=0 \end{cases}$$

מכאן נובע כי האופטימלי h:

$$h(\hat{x}) = \arg\max_{y\in\{0,1\}}\Big[P[Y=y|X=\hat{x}]\cdot b(y)\Big]$$

כאשר $b(y)=\begin{cases}\frac{1}{2}, & y=1\\ 1, & y=0\end{cases}$