1.1.1

נגדיר ←

$$V_{(d+1)\times 1} = \begin{bmatrix} 1 \\ w \\ 1 \\ b \end{bmatrix}, \quad Q = \begin{bmatrix} I_d & 0 \\ 0 & 0 \end{bmatrix}, \quad A_{m\times(d+1)} = \begin{bmatrix} -A_1- \\ -A_2- \\ \vdots \\ -A_m- \end{bmatrix}$$

$$d_{m\times 1} = \begin{bmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}, \quad a = 0_{1\times(d+1)}$$

$\forall i\in[m],\ A_i = -y_i[x_i^T\ \ 1]$

נשים לב כי ←

(I) $\frac{1}{2}V^TQV = \frac{1}{2}w^TI_dw = \frac{1}{2}w^Tw$

(II) $\forall i\in[m],\ y_i(\langle w,x_i\rangle+b)\geq 1 \iff -y_i(\langle w,x_i\rangle+b)\leq -1 \iff -y_i[x_i^T,\ 1]\cdot V\leq -1$
$\iff AV\leq d$

ולכן קיבלנו כי הביטויים שמחפשים למזער שקול משני המקרים, ושם
גם התנאים — לכן (1) שקול ל-(2) מש״ל

1.2.1

(a) נסמן ← $S_m=\{x_1,x_2,\dots,x_m\}$.

$$P_D[S_m\cap X'=\emptyset] = P_D[x_1\notin X'\wedge x_2\notin X'\dots x_m\notin X']$$
$$\overset{(I)}{=} \prod_{i=1}^{m} P_D[x_i\notin X'] = \prod_{i=1}^{m} 1-P_D[x_i\in X'] \overset{(II)}{<} \prod_{i=1}^{m}(1-\epsilon) = (1-\epsilon)^m \overset{(III)}{\leq} e^{-m\epsilon}$$

(I) S_m מורכב מ-m דגימות ב.ת.ל.

(II) X' מקיים כי $P_D[x\in X']>\epsilon$.

(III) ידוע ברנולי.

(b)

נגדיר את $Z_{\hat{f}_s}$ כהגדרתו בסעיף הקודם.

נשים לב כי $L_{D,f}(\hat{f}_s) > \varepsilon \iff P_D[x \in Z_{\hat{f}_s}] > \varepsilon$.

בנוסף, $S_m \cap Z_{\hat{f}_s} = \emptyset$ מכיוון הגדרת $Z_{\hat{f}_s}$ ומכיוון ש-$\hat{f}_s$ עקבי על S_m

ולכן מסעיף (a) יוצא ⇐ $P_D[S_m \cap Z_{\hat{f}_s} = \emptyset] \le e^{-m\varepsilon}$

כעת נשים לב כי $\{L_{D,f}(\hat{f}_s) > \varepsilon\} \subseteq \{S \cap Z_{\hat{f}} = \emptyset\}$

ולכן $$P_D\left[P_D[x \in Z_{\hat{f}_s}] > \varepsilon\right] = P_D[L_{D,f}(\hat{f}_s) > \varepsilon] \le P_D[S \cap Z_{\hat{f}_s} = \emptyset] \le e^{-m\varepsilon}$$

סה״כ ⇐ $P_D[L_{D,f}(\hat{f}_s) > \varepsilon] \le e^{-m\varepsilon}$.

(c)

נסמן פלט כ-$\hat{f}_s$ את האלגוריתם A שמחזיר מהמדגם S_m

משמע, $P_D[L_{D,f}(A(S_m)) < \varepsilon] = P_D[L_{D,f}(\hat{f}_s) < \varepsilon]$

נשים לב כי

$$P_D[L_{D,f}(\hat{f}_s) < \varepsilon] = 1 - P_D[L_{D,f}(\hat{f}_s) \ge \varepsilon] = 1 - \left(P_D[L_{D,f}(\hat{f}_s) > \varepsilon] + P_D[L_{D,f}(\hat{f}_s) = \varepsilon]\right)$$

$$\ge 1 - P_D[L_{D,f}(\hat{f}_s) < \varepsilon] \ge 1 - e^{-m\varepsilon}$$

סה״כ ⇐ $P_D[L_{D,f}(\hat{f}_s) < \varepsilon] \ge 1 - \delta$, $\delta = e^{-m\varepsilon}$.

ולכן אם נדגום S_m בגודל $m = \frac{1}{\varepsilon}\ln\left(\frac{1}{\delta}\right)$ נקבל דיוק בהסתברות δ, ε.

האלגוריתם שלנו A, יפעל כך: בהינתן $S_m=\{x_i\}_1^m$ דגימות, $A(S_m)=h_{\hat r_s}$ s.t. $\hat r_s=\max\left(\{\|x_i\|^2\}_{i=1}^m\right)$. כלומר, יחזיר $\hat r_s$ שיהיה $\|x_i\|^2$ המקסימלי מתוך הדגימות.

נסמן ב- $X_{\hat r_s}$ את מרחב הדגימות $\{x\in\mathbb{R}^2 : \|x\|^2\le\hat r_s\}$

כמו שעשינו במקרה הריאליזבילי, נסמן את הפתרון האמיתי ב- $\underline{r}$
באופן דומה נגדיר את X_r.

נגדיר גם $Z_s=X_r\setminus X_{\hat r_s}$. שם זה $\underline{\text{כל}}$ ערכי r ל-S_m

נשים לב כי $h_{\hat r_s}$ הוא סימון שקול ל- $\hat f_s$ מהשאלה הקודמת, וכו'.

Z_s הוא כל $x\in\mathbb{R}^2$ כך ש- $x\in h_r \wedge x\notin h_{\hat r_s}$ לכן מהגדרה של השאלה הקודמת משמע $\hat f_s(x)\neq f(x)$ ← לכן Z_s שקול ל- $Z_{\hat f_s}$ מהגדרה של השאלה הקודמת.

לכן מתוך הוכחת השאלה הקודמת נקבל כי

$$P_D[L_{D,f}(h_{\hat r_s})<\epsilon]=P_D[L_{D,f}(\hat f_s)<\epsilon]=P_D[L_{D,f}(A(S_m))<\epsilon]>1-\delta$$

כאשר $\delta=e^{-m\epsilon}$ ⇐ $m=\frac{1}{\epsilon}\ln\left(\frac{1}{\delta}\right)\overset{\text{סימון}}{=:}\hat m$

משמע, לכל זוג (ϵ,δ) לכל $m>\hat m$ מתקיים כי $e^{-m\epsilon}<e^{-\hat m\epsilon}$

לכן $$P_D[L_{D,f}(A(S_m))<\epsilon]>1-e^{-m\epsilon}>1-e^{-\hat m\epsilon}=1-\delta$$

לכן לכל $m>\hat m$ אי-השוויון הרצוי מתקיים - משמע $m_H(\epsilon,\delta)\le\hat m=\frac{1}{\epsilon}\ln\left(\frac{1}{\delta}\right)$

$\boxtimes$

יהי S מדגם בגודל $m \ge m_H^{UC}(\epsilon/2, \delta)$.

מהנתון כך ש-H מקיימת את uniform convergence property אזי

$$(*) \overset{\text{סימון}}{:=} D^m\left(\{S \in \{X\times Y\}^m \mid S \text{ is } \tfrac{\epsilon}{2}\text{-representative}\}\right) \ge 1-\delta$$

נגדיר כעת כ-

(I) $h_D = \arg\min_{h\in H} L_D(h)$

(II) $h_S = \arg\min_{h\in H} L_S(h)$

מהנתון S כלשהו,
בנוסף לב כי מתקיים ש-h_S מוחזר מאלגוריתם את ה-ERM אזי $L_S(h_S) \le L_S(h_D)$.

בנוסף, מהנתון ש-S הינו $\frac{\epsilon}{2}$-representative אזי $L_D(h_S) - L_S(h_S) < \frac{\epsilon}{2} \Leftarrow |L_S(h_S) - L_D(h_S)| < \frac{\epsilon}{2}$

וגם $L_S(h_D) - L_D(h_D) < \frac{\epsilon}{2} \Longleftarrow |L_S(h_D) - L_S(h_D)| < \frac{\epsilon}{2}$.

לכן נקבל $\Leftarrow$

$$L_D(h_S) < \frac{\epsilon}{2} + L_S(h_S) \le \frac{\epsilon}{2} + L_S(h_D) < \frac{\epsilon}{2} + \frac{\epsilon}{2} + L_D(h_D) = L_D(h_D) + \epsilon$$

לכן עבור S כלשהו אזי מתקיים $L_D(h_S) < L_D(h_D) + \epsilon$.

בנוסף כי $\min_{h\in H} L_D(h) = L_D(h_D)$ מהגדרת h_D.

לכן

$$1-\delta \le (*) < P_D[L_D(h_S) < L_D(h_D) + \epsilon] = P_D[L_D(h_S) < \min_{h\in H} L_D(h) + \epsilon]$$

ולכן קיבלנו כי H למידה APAC עם $m_H(\epsilon, \delta)$

ומתקיים $m_H(\epsilon,\delta) \le m_H^{UC}(\frac{\epsilon}{2}, \delta)$ ∎

1.3.1

VCD של H_{parity} הוא n.

<u>הוכחה:</u>

נשים לב כי $h_I(x) = \langle w_I, x \rangle \bmod 2$, $\forall I \subseteq [n]$

כאשר w_I הוא וקטור בינארי כך ש- $w_I^i = 1 \iff i \in I$.

לכן $|H_{parity}| = |\{w : |w| = n\}| = 2^n$, ולכן ידוע כי $VCD(H_{parity}) \le \log_2(2^n) = \underline{\underline{n}}$.

כעת נראה כי קיים סט דוגמאות S כך ש- $|S| = n$ וגם H_{parity} מנתצת את S.

$S = \{(00\ldots1), (00\ldots10), (00\ldots100), \ldots, (10\ldots00)\}$ (ז"א סט וקטורים כך שלכל $s_i \in S$ מתקיים $s_i = 0^{n-i}\,1\,0^{i-1}$.

נשים לב כי לכל $I \subseteq [n]$, הפונקציה h_I ממפה את s_j ל-1 אמ"מ $j \in I$.

לכן, לכל תיוג של S קיים $I \subseteq [n]$ כך ש- h_I מקיים תיוג זה שכן כיוון $j \in I$ אמ"מ התיוג של s_j הוא 1.

לכן H_{parity} מנתצת את S, ומכיוון ו- $|S| = n$ אזי $VCD(H_{parity}) \ge n$

ומכך הכל קיבלנו $VCD(H_{parity}) = \underline{\underline{n}}$ ∎

1.3.2

נסמן ⇐ $VCD(H_1) = d_1$, $VCD(H_2) = d_2$.

מהגדרת VCD, ידוע כי קיים סט S כך ש- $|S| = d_1$ ועבור H_1 מנתצת — כלומר לכל תיוג של S קיימת $h \in H_1$ כך ש- $h(S)$ מקיימת תיוג זה — אבל מכיוון ו- $H_1 \subseteq H_2$ אזי גם h נכללת בנוסף מקיימת $h \in H_2$ ולכן ידוע כי לכל תיוג של S קיימת $h_2 \in H_2$ כך ש- $h_2(S)$ מקיימת תיוג זה, ולכן H_2 מנתצת את S, ולכן $VCD(H_2) = d_2 \ge d_1 = VCD(H_1)$.

כנדרש ∎.