מבוא ללמידה חישובית תרגיל בית 6

1)

א. נשים לב שניתן לפרק כל פונקציה ב- F
כשרשור (concatenation) של d פונקציות ב- H:

$f\in F \Rightarrow f=f_1\times f_2\times\ldots\times f_d$, $\quad \forall 1\le i\le d: f_i\in H$

הוכחנו בתרגול ש- $\Pi_F(n)\le\Pi_{F_1}(n)\cdot\Pi_{F_2}(n)$ כאשר

$F=\{(f_1\times f_2) \mid f_1\in F_1, f_2\in F_2\}$

לכן:

$$\Pi_F(n)\le\Pi_H(n)\cdot\ldots\cdot\Pi_H(n)=\Pi_H(n)^d$$

נניח $n\ge d+1$ לכן מהלמה של Sauer-Shelah:

$$\Pi_F(n)\le\Pi_H(n)^d\le\left(\left(\frac{en}{d+1}\right)^{d+1}\right)^d=\left(\frac{en}{d+1}\right)^{d(d+1)}$$

כנדרש

ב. נשים לב שהמחלקה C מורכבת מהרכבה של
$L-1$ שכבות של Hyperplanes במימד d ועוד שכבה אחת
במימד 1. כל שכבה כזו מסעיף א' היא
שרשור של פונקציות ב- H. כלומר:

$$C=\{f^{(0)}\circ f^{(1)}\circ f^{(2)}\circ\ldots\circ f^{(L-1)} \mid f^{(0)}\in H\wedge\forall i\ge 1: f^{(i)}=f_1^{(i)}\times\ldots\times f_d^{(i)}\wedge\forall 1\le j\le d: f_j^{(i)}\in H\}$$

הוכחנו בתרגול שלכל הרכבה מתקיים אי-שוויון
בנוגע ל- $\Pi_C(n)$: $\Pi_C(n)\le\Pi_{C_1}(n)\cdot\Pi_{C_2}(n)$ כאשר
$C=\{f_1\circ f_2 \mid f_1\in C_1\wedge f_2\in C_2\}$.

לכן:

$$\Pi_C(n)\le\Pi_H(n)\cdot\left[\prod_{i=1}^{L-1}\Pi_{F^{(i)}}(n)\right]\le\left(\frac{en}{d+1}\right)^{d+1}\left[\left(\frac{en}{d+1}\right)^{d(d+1)}\right]^{L-1}=$$

$$=\left(\frac{en}{d+1}\right)^{d+1+d(d+1)(L-1)}$$

1ה) נסמן $N=d+1+d(d+1)(L-1)$. לפי ~~משפט~~ מסעיף ג'

$$\Pi_C(n) \le \left(\frac{en}{d+1}\right)^N \le (en)^N$$

נסמן $n=VCdim(C)$. לפי:

$$2^n = \Pi_C(n) \le (en)^N \xRightarrow{\text{מסעיף ג'}} n \le 2N\log_2(eN)$$

כלומר $\boxed{VCdim(C) \le 2N\log_2(eN)}$ כנדרש

1ב) [ספירת על הפרמטרים במספר, הכולל מופיע אחרי טיוטה ושכחתי את סעיף ג' :)]

הפרמטרים ברשת הם המשקולות W וה-bias-ים b.
המשקולות $W^{(t)}$ בין השכבה ה-t לשכבה ה-$t+1$ היא
כן מטריצה שמגדירה פונקציה מ-$R^d \to R^d$ לכן
בגודל d^2. יש $L-1$ כאלו ועוד מטריצה שמייצגת
פונקציה מ-$R^d \to R$ בשכבה האחרונה, כלומר בגודל
d. בנוסף לכל נוירון יש bias (b) בגודל 1.
כלומר יש $d^2(L-1)+d$ פרמטרים ב-W-ים
ועוד $d(L-1)+1$ ב-b-ים. סה"כ:

$$N = d+1+d(d+1)(L-1)$$

ג) נעקוב אחר ריצת האלגוריתם:

• חישוב $j \leftarrow \underset{i\in\{1,2,3\}}{\arg\max}\, G(S,i)$:

$$G(S,1) = L(Y) - P(X_1=1)\cdot L(Y|X_1=1) - P(X_1=0)\cdot L(Y|X_1=0) =$$

נציב -

- $L(Y) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$
- $P(X_1=1) = \frac{3}{4} \quad P(X_1=0) = \frac{1}{4}$
- $L(Y|X_1=1) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0.918$
- $L(Y|X_1=0) = -0\log 0 - 1\log 1 = 0$

נקבל -

$$= 1 - \frac{3}{4}\cdot 0.918 \approx 0.311$$

במקרה של $i=2$ ו- $i=3$ יש שקילות מוחלטת לכן נחשב רק את $i=2$:

$$G(S,2) = L(Y) - P(X_2=1)\cdot L(Y|X_2=1) - P(X_2=0)\cdot L(Y|X_2=0) =$$

נציב -

- $L(Y) = 1$
- $P(X_2=1) = \frac{1}{2},\ P(X_2=0) = \frac{1}{2}$
- $L(Y|X_2=1) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$
- $L(Y|X_2=0) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$

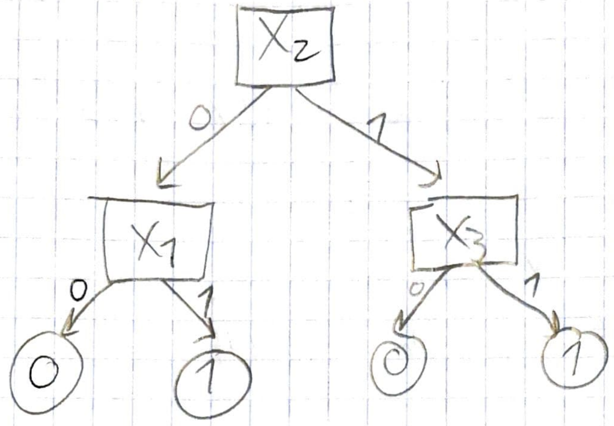
נקבל -

$$= 1 - \frac{1}{2} - \frac{1}{2} = 0$$

לכן $j=1$. $S_0=\{(0,0,1)\}$, $S_1=\{(1,1,1),(1,0,0),(1,1,0)\}$

עבור צאצא הענף של S_0 – כל הערכים, (כלומר הערך היחיד ב-S_0), שייכים ל-0 לכן נחזיר עלה עם ערך 0.
בתת העץ של S_1 ניתן לראות בקלות שאין הבדל בין בחירת $j=2$ ל-$j=3$ בשניהם כל בחירה של $X_j=0$ או $X_j=1$ מקבלים קבוצות עם שתי התוויות כלומר בכל מקרה נטעה בדוגמה אחת.
סה"כ נטעה בדוגמה אחת מתוך הארבע כנדרש.

ב. 2) עץ מתאים עם training error=0;



ג. 5) נכתוב את בעיית המינימיזציה-

$$\min \|w\|^2 \quad s.t.\ x_1 w = y_1, \dots, x_n w = y_n$$

ניתן לפתור עם לגרנז'יאן:

$$L(w;\lambda) = \|w\|^2 + \sum_{i=1}^{n} \lambda_i (x_i w - y_i)$$

נגזור ונשווה ל-0:
$$\frac{\partial L}{\partial w} = 2w + \sum_{i=1}^{n} \lambda_i x_i = 0 \Longrightarrow w = -\frac{1}{2}\sum_{i=1}^{n} \lambda_i x_i$$

לכן הוקטור מקדמים הוא $\alpha = -\frac{1}{2}\begin{pmatrix}\lambda_1 \\ \vdots \\ \lambda_n\end{pmatrix}$

ג. יהי וקטור $x \in \mathbb{R}^d$. נרצה לחשב את $\hat{x}^t w^*$:

$$\hat{x}^t w^* = \hat{x}^t (X^t \alpha) =$$
$$= (\hat{x}^t X^t)\alpha =$$
$$= [\hat{x}\cdot x_1^t, \dots, \hat{x}\cdot x_n^t]\begin{bmatrix}\alpha_1 \\ \vdots \\ \alpha_n\end{bmatrix}$$

$\hat{x} \in \mathbb{R}^d, X \in \mathbb{R}^{nd}$:(נזכיר)

כאשר כל מכפלה בגבולן היא dot product של שני וקטורים ב-$\mathbb{R}^d$. לכן אם $\hat{x} = x_j$ (השורה ה-j ב-X) אז:

$$x_j w^* = [x_j \cdot x_1, \dots, x_j \cdot x_n]\cdot\begin{bmatrix}\alpha_1 \\ \vdots \\ \alpha_n\end{bmatrix} = K_{s_j}\cdot\alpha$$

כנדרש.

3
ג

$$Z_t=\sum_{j=1}^{n} D_t[j]\cdot e^{-y_j w_t h_t(x_j)}$$

נציב $W_t=\frac{1}{2}\ln\frac{1-\varepsilon_t}{\varepsilon_t}$

נגזור את Z_t כפונקציה של W_t ונציב:

$$\frac{\partial Z_t}{\partial w_t}=\sum_{j=1}^{n} D_t[j]\cdot(-y_j)\, h_t(x_j)\cdot e^{-y_j w_t h_t(x_j)}$$

$$=-\sum_{j=1}^{n} D_t[j]\, y_j h_t(x_j)\cdot e^{-y_j\cdot h_t(x_j)\cdot\frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)}=$$

$$=-\sum_{j=1}^{n} D_t[j]\, y_j h_t(x_j)\left[\frac{(1-\varepsilon_t)}{\varepsilon_t}\right]^{\frac{-y_j h_t(x_j)}{2}}=$$

$$=-\underbrace{\sum_{\substack{j=1\\ h_t(x_j)=y_j}}^{n} D_t[j]}_{\varepsilon_t}\sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}+\underbrace{\sum_{\substack{j=1\\ (h_t(x_j)=-y_j)}}^{n} D_t[j]}_{1-\varepsilon_t}\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}=$$

$$=-\varepsilon_t\sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}+(1-\varepsilon_t)\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}=0$$

כנדרש.

3) ג. נגדיר את הפונקציה -

$$g(i):=\begin{cases}1, & h_t(x)\neq y_i\\ 0, & else\end{cases}$$

מתקיים: $\underset{x\sim D_{t+1}}{P}[h_t(x)\neq y]=\sum_{i=1}^{n}D_{t+1}[i]g(i)$

נזכיר את ההגדרה של D_{t+1}:

$$D_{t+1}[i]=\frac{D_t[i]\cdot e^{-y_iw_th_t(x_i)}}{\sum_{j=1}^{n}D_t[j]e^{-y_iw_th_t(x_j)}}\underset{\text{מסעיף א'}}{=}\frac{D_t[i]\cdot e^{-y_iw_th_t(x_i)}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}}$$

נציב חזרה ונקבל:

$$\underset{x\sim D_{t+1}}{P}[h_t(x)\neq y]=\sum_{i=1}^{n}\frac{D_t[i]\cdot e^{-y_iw_th_t(x_i)}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}}\cdot g(i)=$$

$$\overset{(*)}{=}\sum_{\substack{i=1\\g(i)=1}}^{n}\frac{D_t[i]e^{-y_i\cdot w_t\cdot(-y_i)}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}}=\sum_{\substack{i=1\\g(i)=1}}^{n}\frac{D_t[i]\,e^{w_t}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}}=$$

$$=\frac{e^{w_t}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}}\cdot\underbrace{\sum_{\substack{i=1\\g(i)=1}}^{n}D_t[i]}_{=\varepsilon_t}=\frac{e^{w_t}\cdot\varepsilon_t}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}}\underset{\text{מסעיף א'}}{=}\frac{\sqrt{\varepsilon_t(1-\varepsilon_t)}}{2\sqrt{\varepsilon_t(1-\varepsilon_t)}}=\frac{1}{2}$$

כנדרש.

(*) — נובע מכך שאם $g(i)=1$ אז $h_t(x)=-y_i$

ד. נובע (כמעט) ישירות מ-ג' - השגיאה של h_t עבור
ההתפלגות ה-t היא $\frac{1}{2}$. אבל גם אנחנו מניחים
שבוחרים ל-h_{t+1} מסווג עם שגיאה $\frac{1}{2}-\gamma$ עבור
$0<\gamma\leq\frac{1}{2}$. כלומר השגיאה של h_{t+1} צריכה להיות קטנה
ממש מחצי ולכן שונה $h_{t+1}\neq h_t$

ע) ניקח התפלגות על וכ"ש הנתון $(w_i := \alpha)$:

$$\mathbb{E}_{i\sim D}\left[y_i \sum_{j=1}^{k} w_j h_j(x_i)\right] \geq \mathbb{E}_{i\sim D}[\gamma]$$

כמובן ש- γ קבוע לכן $E[\gamma]=\gamma$.

כלומר:

$$\sum_{i=1}^{n} D(i)\cdot y_i \sum_{j=1}^{k} w_j h_j(x_i) \geq \gamma \implies \sum_{j=1}^{k} w_j \sum_{i=1}^{n} D(i) y_i h_j(x_i) \geq \gamma$$

מכיוון ש- $\sum_{j=1}^{k} w_j = 1$ מתקיים שק"מ להיות ג' כלשהו

$\hat{j}$ עבורו $\gamma \leq \sum_{i=1}^{n} D(i) y_i h_{\hat{j}}(x_i)$. לכן:

$$\gamma \leq \sum_{i=1}^{n} D(i) y_i h_{\hat{j}}(x_i) =$$

$$= \sum_{\substack{i=1 \\ h_{\hat{j}}(x_i)=y_i}}^{n} D(i) y_i^2 + \sum_{\substack{i=1 \\ h_{\hat{j}}(x_i)=-y_i}}^{n} D(i)(-y_i^2) = \sum_{\substack{i=1 \\ h_{\hat{j}}(x_i)=y_i}}^{n} D(i) - \sum_{\substack{i=1 \\ h_{\hat{j}}(x_i)=-y_i}}^{n} D(i) =$$

$$= P_{i\sim D}[h_{\hat{j}}(x_i)=y_i] - P_{i\sim D}[h_{\hat{j}}(x_i)\neq y_i] =$$

$$= 1 - 2P_{i\sim D}[h_{\hat{j}}(x_i)\neq y_i]$$

נעביר אגפים ונקבל את הנדרש - $P_{i\sim D}[h_{\hat{j}}(x_i)\neq y_i] \leq \frac{1}{2} - \frac{\gamma}{2}$

ד) נבנה את H לפי הרמז - נגדיר $2d$

היפותזות על פי המלבן $[a_j,b_j]$ באופן הבא-

$$\forall 1\le j\le 2d.\ j\in N_{odd}\rightarrow h_j(x)=\begin{cases}1, & x\ge a_j\\ -1, & x<a_j\end{cases},\quad j\in N_{even}\rightarrow h_j(x)=\begin{cases}1, & x\le b_j\\ -1, & x>b_j\end{cases}$$

כלומר יש לנו בעצם d זוגות של היפותזות, כל זוג מגדיר

ביחד את ההיפותזה ש- $h(x)=1 \Leftrightarrow x\in[a_j,b_j]$ עבור כל d

הגבולות $[a_j,b_j]$. נשאר עוד $2d-1$ היפותזות קבועות-

$\forall j>2d\ h_j(x)=-1$. ונשתמש ב- $K=4d-1$ לפי הרמז.

כעת נגדיר $\forall j.\ \alpha_j=\frac{1}{4d-1}$ ו- $\gamma=\frac{1}{4d-1}$. כמובן ש- $\sum_{j=1}^{K}\alpha_j=1$

ו- $\gamma>0$ כנדרש.

כעת נראה נכונות אם"ם התנאי מתקיים עבור

עבור כל $(\hat{x},\hat{y})$ במדגם שמקיים $\hat{y}=1$ חיובי, כלומר

$\hat{x}$ בתוך המלבן ולכן, $\forall 1\le j\le 2d: h_j(x)=1$ ו- $\forall j>2d: h_j(x)=-1$ לכן:

$$\hat{y}\sum_{j=1}^{K}\alpha_j h_j(\hat{x})=\sum_{j=1}^{2d}\alpha_j+\sum_{j=2d+1}^{K}-\alpha_j=\frac{2d}{4d-1}-\frac{2d-1}{4d-1}=\frac{1}{4d-1}=\gamma\ge\gamma$$

כנדרש.

ולכל דוגמא כזאת $\hat{y}=-1$, כלומר $\hat{x}$ לא במלבן, נסמן בה"כ

ב-k^* את האינדקס שעבורו-

$$\forall 1\le j\le k^*: h_j(\hat{x})=1$$

$$\forall k^*<j\le 2d: h_j(\hat{x})=-1$$

ונשים לב $\forall 2d<j\le K: h_j(\hat{x})=-1$

ונקבל:

$$\hat{y}\sum_{j=1}^{K}\alpha_j h_j(\hat{x})=-\sum_{j=1}^{k^*}\alpha_j-\sum_{j=k^*+1}^{2d}-\alpha_j-\sum_{j=2d+1}^{4d-1}-\alpha_j=$$

$$=\frac{-k^*}{4d-1}+\frac{2d-k^*}{4d-1}+\frac{4d-1-2d}{4d-1}=$$

$$=\frac{1}{4d-1}\left[4d-1-2d+2d-k^*-k^*\right]=\frac{1}{4d-1}\left[4d-1-2k^*\right]\ge\frac{1}{4d-1}=\gamma$$

והמעבר הלפני אחרון נובע מכך ש- $k^*<2d$ לכן

$4d-1-2k^*\ge 1$. קיבלנו שבשני המקרים - $\hat{x}$ בתוך המלבן או $\hat{x}$

מחוצה לו, מתקיים התנאי הנדרש.