מבוא ללמידה חישובית תרגיל בית 3

(1 נגדיר $m=d=\lfloor \frac{1}{\gamma^2} \rfloor$ ו- $x_i=e_i$ לכל $i=1,\dots,m$

(e_i - וקטור i בבסיס הסטנדרטי) של $\mathbb{R}^m$.

נניח באינדוקציה שאלגוריתם Perceptron יטעה

בכל איטרציה, כלומר $m=d=\lfloor \frac{1}{\gamma^2} \rfloor$ טעויות בסה"כ.

בסיס - טריוויאלי, באיטרציה הראשונה תמיד נקבל את

w כי $y_1=\pm 1$, $w_1\cdot x_1=0$

הנחה - נניח שעד האיטרציה ה-k היו רק טעויות.

כלומר, $w_k=\sum_{i=1}^{k-1} x_i$ לכן הכניסה ה-j ב-w_k היא 1 אם $j<k$

ו-0 אחרת. נקבל:

$$w_k\cdot x_k=\sum_{j=1}^{m} w_k(j)\cdot x_k(j)=\sum_{j=1}^{k-1} 1\cdot x_k(j)+\sum_{j=k}^{m} 0\cdot x_k(j)=\sum_{j=1}^{k-1} x_k(j)=0$$

מכיוון ש- $y_i=\pm 1$ $\forall 1\le i\le m$: נקבל כאן גם טעות

כנדרש.

נשאר רק למצוא את w^*. נבחר כך: $w^*=\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\cdot\frac{1}{\sqrt{\frac{1}{\lfloor\gamma^2\rfloor}}}$

נקבל ש- $\|w^*\|=1$ ו-

$$\frac{x_i y_i w^*}{\|w^*\|}=\frac{y_i x_i\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\cdot\frac{1}{\sqrt{\frac{1}{\lfloor\gamma^2\rfloor}}}}{1}=y_i^2\cdot\frac{1}{\sqrt{\frac{1}{\lfloor\gamma^2\rfloor}}}=\frac{1}{\sqrt{\frac{1}{\lfloor\gamma^2\rfloor}}}\ge\frac{1}{\sqrt{\frac{1}{\gamma^2}}}=\frac{1}{\frac{1}{\gamma}}=\gamma$$

כנדרש.

2) ראשית נבחין שיש $\frac{d(d-1)}{2}$ היפותזות ב-H.
מאחר לכל x_0 נתון גורר שרוב ההיפותזות יתנו 0.
(כי $d \ge 6$, $|n_0| = \frac{(d-1)(d-2)}{2}$, $|n_1| = d-1$ אז $|n_1| < |n_0|$)
אם נבדוק רק את אותה גרסה של H קטן
יותר לכן בטוח מספר הטעויות הסופי רק יקטן, כלומר
מספיק לחסום את מספר הטעויות במקרה שטעינו.
אם טעינו אז $|H_1| = d-1$.
כעת יהי $x_1 \ne x_0$. מאחר "שקיבלנו" את אחד הערכים, בה"כ
i יש כעת רק היפותזה אחת שנותנת $h(x_1)=1$ לעומת
d-2 היפותזות הפוכות. כלומר $|n_1|=1$, $|n_0|=d-2$.
לכן הטענה - אם נבדוק רק קטנות את גרסה עם
חסם של המחות לא נבדל יותר על מספר הטעויות, לכן
נניח שטעינו ולמרות שחזינו $\tilde{y}_1=0$ בעצם $y_1=1$,
אז נשארנו רק עם היפותזה אחת - $H_2=\{h_{x_0,x_1}\}$
ולכן מספיק ס"מ, כנדרש.

---

3) תהי סדרה $C=\{x_i\}_{i=1}^{m}$, $\forall i,j: x_i \ne x_j$. בה"כ $x_i < x_j \Leftrightarrow i<j$
נשים לב שבחירת הנקודות a,b שקולה לבחירה
של אחד מ-0 עד m איברים רצופים מ-C שיושבו של ידי
הקטע $[a,b]$. כלומר שקול לבחירת שני אינדקסים
i,j כך ש- $a<x_i<x_j<b$ כך ש- $h_{a,b}(x_k)=\begin{cases}1, & i\le k\le j\\ 0, & else\end{cases}$
בנוסף יש עוד אפשרות $x_m<a$ או $x_0>a$. לכן
סה"כ יש m+1 אפשרויות לבחירת כל אינדקס.
קיבלנו סה"כ $\frac{m(m+1)}{2}$ אפשרויות ל-$h_{a,b}$, כלומר
$$\Pi_H(m) = \frac{m(m+1)}{2}$$

$$P\left[e_p(ERM(S)) - \min_{h\in H} e_p(h) \geq \varepsilon\right] \leq 4\left(\frac{2en}{d}\right)^d \exp\left(-\frac{n\varepsilon^2}{32}\right)$$ (4

משפט שהוכחנו (מהרצאות במודל). לכן מספיק לדרוש שהביטוי בצד ימין קטן או שווה ל-δ. כלומר:

$$4\left(\frac{2en}{d}\right)^d \exp\left(-\frac{n\varepsilon^2}{32}\right) \leq \delta$$

$$\Downarrow$$

$$\ln 4\left(\left(\frac{2en}{d}\right)^d \cdot \frac{1}{\delta}\right) \leq \frac{n\varepsilon^2}{32} \implies n \geq \frac{32}{\varepsilon^2}\left(\ln\frac{4}{\delta}\left(\frac{2en}{d}\right)^d + \ln(n^d)\right) =$$

$$= \frac{32}{\varepsilon^2}\ln\frac{4}{\delta}\left(\frac{2e}{d}\right)^d + \frac{32d}{\varepsilon^2}\ln(n)$$

נסמן $x = n$, $a = \frac{32d}{\varepsilon^2}$ ו- $b = \frac{32}{\varepsilon^2}\ln\frac{4}{\delta}\left(\frac{2e}{d}\right)^d$.

קיבלנו את צד ימין של הנוסחה $x \geq a\ln(x) + b$.

לכן נדרוש את צד השמאלי של הנוסחה ונמשיך:

$$x \geq 4a\ln(2a) + 2b \implies n \geq \frac{128d}{\varepsilon^2}\ln\left(\frac{64d}{\varepsilon^2}\right) + \frac{64}{\varepsilon^2}\ln\frac{4}{\delta}\left(\frac{2e}{d}\right)^d$$

$$= \frac{128d}{\varepsilon^2}\ln\left(\frac{64d}{\varepsilon^2}\right) + \frac{64}{\varepsilon^2}\ln(4) + \frac{64}{\varepsilon^2}\ln\left(\frac{1}{\delta}\right) + \frac{64}{\varepsilon^2}d(\ln 2e) + \frac{64}{\varepsilon^2}d\ln\left(\frac{1}{d}\right) =$$

$$= \frac{64d}{\varepsilon^2}\ln\left(\frac{128d}{\varepsilon^2}\right)^2 + \underbrace{\frac{64}{\varepsilon^2}\ln\left(\frac{1}{\delta}\right) + \frac{64}{\varepsilon^2}\left(\ln(4) + \ln(e)\right)}_{(*)} + \frac{64d}{\varepsilon^2}\ln\left(\frac{1}{d}\right) =$$

$$\leq \frac{64d}{\varepsilon^2}\left[\ln\left(\frac{128}{\varepsilon^2}\right)^2 + \ln d^2 + \ln\frac{1}{d}\right] + \frac{256}{\varepsilon^2}\ln\left(\frac{1}{\delta}\right) =$$

(*) – δ קטן כידוע, מספיק בשביל המספר הזה.

$$= \frac{128}{\varepsilon^2}d\left[2\ln\left(\frac{128}{\varepsilon^2}\right) + \ln d\right] + \frac{256}{\varepsilon^2}\ln\left(\frac{1}{\delta}\right) =$$

$$= \frac{256}{\varepsilon^2}d\ln\left(\frac{128d}{\varepsilon^2}\right) + \frac{256}{\varepsilon^2}\ln\left(\frac{1}{\delta}\right)$$

כנדרש.

5) מההרצאה נקבל:

$$E[\Delta(Y,h(x;\theta))] = \sum_{x,y} p[X=x,Y=y]\cdot\Delta(y,h(x;\theta))$$

עבור x קבוע:

$$p[X=x,Y=0]\cdot\Delta(0,h(x;\theta)) + p[X=x,Y=1]\cdot\Delta(1,h(x;\theta)) =$$

$$= p[X=x]\left[P(Y=0|X=x)\cdot(-\ln(1-h(x;\theta))) + P(Y=1|X=x)\cdot(-\ln(h(x;\theta)))\right] =$$

$$= p[X=x]\cdot\left[\overset{p_0}{P(Y=0|X=x)}\ln\left(\frac{e^{\theta(x)}}{e^{\theta(x)}+1}\right) - \overset{p_1}{P(Y=1|X=x)}\cdot\ln\left(\frac{1}{1+e^{\theta(x)}}\right)\right]$$

כי סקלרי אכן המינ' הנ"ל מספיק להתעלם
מ-$P(X=x)$, נסמן $z=\theta(x)$, ונגזור לפי z, ונשווה ל-0:

$$f(z) = p_0\ln\left(\frac{e^z}{e^z+1}\right) - p_1\ln(1+e^z)$$

$$\Downarrow$$

$$f'(z) = p_0\frac{1}{e^z+1} - p_1\frac{e^z}{e^z+1}$$

$$\Downarrow$$

$$f'(z)=0 \Longrightarrow p_0\frac{1}{e^z+1} = p_1\frac{e^z}{e^z+1}$$

$$\Downarrow$$

$$p_0 = p_1 e^z$$

$$\Downarrow$$

$$e^z = \frac{p_0}{p_1}$$

$$\Downarrow$$

$$\theta(x) = z = \ln\left(\frac{p_0}{p_1}\right) = \ln\left[\frac{P(Y=0|X=x)}{P(Y=1|X=x)}\right]$$