## מבוא ללמידה חישובית תרגיל בית 3

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

(ה-$e_i$ - 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 כנדרש.

בסיס - טריוויאלי, באיטרציה הראשונה תמיד נעדכן את

$w$ כי $y_1=\pm1$ , $w_1\cdot x_1=0$

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

צעד: 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- $y_i=\pm1$ $\forall 1\le i\le m$: נקבל כי זו טעות

כנדרש.

נשאר רק להראות את $w^*$. נבחר כך:

$$w^*=\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\cdot\frac{1}{\sqrt{\frac{1}{\lfloor\gamma^2\rceil}}}$$

נקבל ש- $\|w^*\|=1$

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

כנדרש.

2) ראשית נבחין שיש $\frac{d(d-1)}{2}$ היפותזות ב-H.
מאחר שלכל $x_0$ נתון ברור שרוב ההיפותזות יתנו 0.
(כי $d\geq 6$, $|n_0|=\frac{(d-1)(d-2)}{2}$, $|n_1|=d-1$ אז $|n_0|>|n_1|$)
אם נבדוק נקבל את אותה גודל של H קטן
יותר לכן בטוח מספר הטעויות הסופי רק יקטן, כאשר
מספיק לחסום את מספר הטעויות במקרה שטעינו.
אם טעינו אז $|H_1|=d-1$.
כעת יהי $x_1\neq x_0$. מאחר שקיבלנו את אחד הערכים, בה"כ
i יש כעת רק היפותזה אחת שנותנת $h_{x_0x_1}=1$ לעומת
$d-2$ היפותזות אחרות. כלומר $|n_1|=1$, $|n_0|=d-2$.
לכן הסקנו - אם נבדוק כך קיבלנו את גודל עד
חסום לכל הגדלות לא גדל יותר של מספר הטעויות, לכן
נניח שטעינו ולמרות שחזינו $\hat{y}_1=0$ בעצם $y_1=1$,
אז נשארנו רק עם היפותזה אחת - $H_2=\{h_{x_0,x_1}\}$
ולכן מספר ט"מ, כנדרש.

---

3) תהי סדרה $C=\{x_i\}_{i=1}^{m}$, $\forall i,j: x_i\neq x_j$. בה"כ $x_i<x_j \iff i<j$
נשים לב שבחירת נקודות $a,b$ שקולה לבחירת
של בין 0 ל-m איברים רצופים מ-C שיסווגו על ידי
הקטע $[a,b]$. כלומר שקול לבחירת שני אינדקסים
$i,j$ כך ש- $a<x_i<x_j<b$ כך ש- $h_{a,b}(x_k)=\begin{cases}1, & i\leq k\leq j\\ 0, & else\end{cases}$
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

$$\Pi_H(m)=\frac{m(m+1)}{2}$$

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

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

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

$$\Downarrow$$

$$\ln 4\left(\left(\frac{2en}{d}\right)^d \cdot \frac{1}{\delta}\right) \le \frac{n\varepsilon^2}{32} \Longrightarrow n \ge \frac{32}{\varepsilon^2}\left(\ln\frac{4}{\delta}\left(\frac{2en}{d}\right)^d + \ln(n^d)\right) =$$

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

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 $x \ge a\ln(x)+b$.
לכן נציב את צד שמאל של הנוסחה ונקבל:

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

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

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

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

(*) - δ קטן כרצוננו, מספיק שנגדיל המספר גדול.

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

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

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:

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

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

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

כדי למצוא את המינ' הנ"ל מספיק להתעלם $P$

מ- $P(X=x)$, נסמן $Z=\theta(x)$, נגזור לפי $Z$ ונשווה ל-0:

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

$$\Downarrow$$

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

$$\Downarrow$$

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

$$\Downarrow$$

$$p_0 = p_1 e^z$$

$$\Downarrow$$

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

$$\Downarrow$$

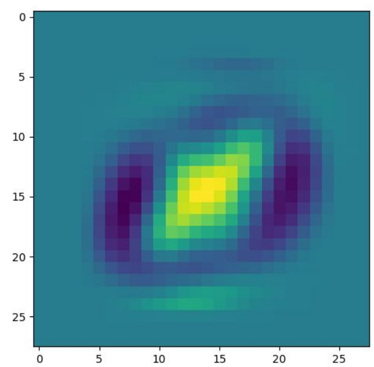
$$\theta(x) = z = \ln\Big(\frac{p_0}{p_1}\Big) = \ln\Big[\frac{P(Y=0|X=x)}{P(Y=1|X=x)}\Big]$$

Question 1:

| N | Mean Accuracy | 5% Percentile | 95% Percentile |
|---|---|---|---|
| 5 | 0.9180348004094167 | 0.8380501535312179 | 0.9626407369498464 |
| 10 | 0.9440327533265098 | 0.9067809621289662 | 0.971417604912999 |
| 50 | 0.9621033776867965 | 0.9528659160696008 | 0.9759467758444217 |
| 100 | 0.9646315250767654 | 0.9554503582395087 | 0.9739252814738997 |
| 500 | 0.9655987717502561 | 0.9626151484135107 | 0.9682702149437052 |
| 1000 | 0.9659570112589558 | 0.9631525076765609 | 0.968295803480041 |
| 5000 | 0.9671750255885364 | 0.9662231320368475 | 0.9677584442169908 |

Question 2:

Here is a visual representation of the weight vector:



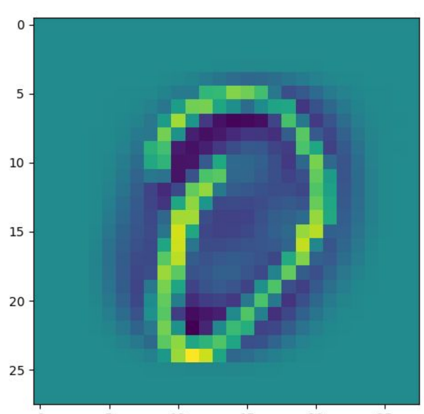

My intuition is that it prioritizes the center of the images over the margins.

Question 3:

The accuracy I received was 0.9672466734902764

Question 4:

I got a misclassification for the following image:




This makes sense - in this image the focus is on the margins and the center is “empty”, exactly the opposite of the Perceptron’s results.