הטכניון – מכון טכנולוגי לישראל

הפקולטה למדעי המחשב

מבוא למערכות לומדות (236756)

סמסטר חורף תשפ"ד – 12 באפריל 2024

מרצה: ד"ר יונתן בלינקוב

מבחן מסכם מועד א' – פתרון חלקי

שימו לב: הפתרונות המופיעים כאן הם חלקיים בלבד ומובאים בשביל לעזור לכם בתהליך הלמידה.

ייתכנו כאן חוסרים / ליקויים / טעויות של ממש.

בהצלחה!

שאלה 1: Maximum Likelihood Estimation [12 נק']

נתונות m דגימות חד-ממדיות $S=\{x_1,x_2,\dots,x_m\}$ שנדגמו i.i.d. ממשתנה מקרי $X\sim\mathrm{Geo}(p)$ עבור $p\in(0,1]$ לא ידוע.

נזכיר שמשתנה גאומטרי מאפיין את מספר ה"ניסויי ברנולי" (הטלות מטבע) בסיכוי p עד להשגת "הצלחה" אחת ומאופיין ע"י פונקציית ההסתברות הבאה:

$$\Pr(X=x_i|p)=(1-p)^{x_i-1}p.$$

שערכו (estimate) את הפרמטר p באמצעות Maximum Likelihood Estimation.

משמע, עליכם לפתח ביטוי עבור $\hat{p}=\mathrm{argmax}_p\, L(S;p)=\mathrm{argmax}_p \Pr(x_1,x_2,\dots,x_m;p)$.

תשובה:

$$\hat{p}=\mathrm{argmax}_p\, L(S;p)\overset{\text{(monotonicity)}}{=}\mathrm{argmax}_p \ln L(S;p)=\mathrm{argmax}_p \ln \Pr(x_1,x_2,\dots,x_m;p)$$

$$\overset{\text{(iid)}}{=}\mathrm{argmax}_p \ln\prod_{i=1}^{m}(1-p)^{x_i-1}p=\mathrm{argmax}_p\sum_{i=1}^{m}\ln\left((1-p)^{x_i-1}p\right)$$

$$=\mathrm{argmax}_p\sum_{i=1}^{m}(x_i-1)\ln\big((1-p)\big)+\ln(p)$$

Differentiate w.r.t. p: $0=\sum_{i=1}^{m}\left(\frac{(1-x_i)}{(1-p)}+\frac{1}{p}\right)\Leftrightarrow\frac{m}{p}=\frac{\sum_{i=1}^{m}(x_i-1)}{(1-p)}\Leftrightarrow\frac{m}{p}+\frac{m}{1-p}=\frac{\sum_{i=1}^{m}x_i}{(1-p)}$

$$\Leftrightarrow\frac{m(1-p)+mp}{p(1-p)}=\frac{m}{p(1-p)}=\frac{\sum_{i=1}^{m}x_i}{(1-p)}\Leftrightarrow\hat{p}=\frac{m}{\sum_{i=1}^{m}x_i}=\frac{1}{\bar{X}}$$

(One could also explain that this is a maximal point since the function is concave.)

שאלה 2: VC-dimension of Decision Trees [24 נק']

א. [4 נק'] להלן ההגדרה של "ניתוץ". השמטנו מההגדרה את הַכַּמָּתים.

השלימו את שלושת הכמתים החסרים. בכל מקום כתבו בבירור האם חסר בהגדרה $\forall$ או $\exists$.

$$\mathcal{H} \text{ shatters } C \iff \underbrace{\forall}_{\text{השלימו}} y_1, \dots, y_{|C|} \in \mathcal{Y}: \underbrace{\exists}_{\text{השלימו}} h \in \mathcal{H}: \underbrace{\forall}_{\text{השלימו}} x_i \in C: h(x_i) = y_i$$

מעתה והלאה בשאלה זו מרחב הדוגמאות הוא $\mathcal{X} = \mathbb{R}$ ומרחב התיוגים הוא $\mathcal{Y} = \{-1, +1\}$.

מחלקת ההיפותזות $\mathcal{H}$ עליה נשאל הינה עצי החלטה מעומק $L \geq 1$.

ב. [10 נק'] עבור עומק כללי $L \geq 1$, הוכיחו שהחסם התחתון הוא $\text{VCdim}(\mathcal{H}) \geq 2^L$.

הוכחה: נבחר 2^L נקודות שונות על הישר, למשל $C = \{1,2, \dots, 2^L\}$. יהי $y_1, \dots, y_{|C|}$ תיוג שרירותי של הנקודות.

עץ החלטה בעומק L יכול בקלות לחלק את הישר הממשי לעד 2^L מקטעים משלימים (כל מקטע שייך לעלה).

(אפשר לפצל את השורש בנקודה $2^{L-1} + 0.5$ וכן הלאה.)

נקבע את הסיווג בעלה של x_i לפי y_i וההיפותזה (=העץ) מסכימה עם כל התיוגים. הראינו ש-$\mathcal{H}$ מנתצת את C.

(שימו לב שייתכנו צמתים שיוצאים מהם שני עלים עם אותו סיווג. זה לא סותר את ההגדרות בקורס.

אם רוצים להימנע מכך בכל זאת, אפשר לאחד אותם ולהחזיר את הסיווג בצומת האב, רקורסיבית)

מקרא טעויות נפוצות:

(A) הגדרת קבוצה C כללית מבלי להסביר מדוע ניתן לבחור כזו. בהוכחת חסם תחתון יש לבנות קבוצה C מפורשת.

(B) התייחסות חלקית/חסרה לתיוגים. למשל: הגדרה חלקית של העץ באמצעות x_i מבלי להתייחס ל-y_i.

(C) התייחסות למחלקת ההיפותזות $\mathcal{H}$ כקבוצה סופית.

(D) חסר בנימוק לכך שהעץ המתקבל הוא בעומק L.

ג. [10 נק'] עבור עומק כללי $L \geq 1$, הוכיחו שהחסם העליון הוא $\mathrm{VCdim}(\mathcal{H}) \leq 2^L$.

הוכחה: יהי אוסף שרירותי של $2^L + 1$ נקודות <u>שונות</u> על הישר. בה"כ נניח $x_1 \leq x_2 \leq \cdots \leq x_{2^L+1}$.

<u>נבחר</u> תיוג שמשתנה לסירוגין, משמע $y_1 = +1, y_2 = -1, y_3 = +1$ וכן הלאה.

נניח בשלילה שקיימת היפותזה במחלקה (עץ בעומק של לכל היותר L), שמסווגת נכונה את הנקודות.

בהיפותזה יש לכל היותר 2^L עלים. לפי עיקרון שובך ה-🕊, קיים עלה שבו לפחות שתי דוגמאות.

לפי אופן פעולת עצי החלטה, בהכרח מתקיים ששתיים מהדוגמאות בעץ הן עוקבות בסידור, משמע

x_i, x_{i+1}. אבל בתיוג שבחרנו מתקיים $y_i \neq y_{i+1}$ ולא ייתכן ששתי הדוגמאות מסווגות נכונה באותו עלה

שמחזיר סיווג יחיד! סתירה.

<u>מקרא טעויות נפוצות:</u>

<u>(A)</u> חסרה הגדרה מפורשת של תיוג על הנקודות.

<u>(B)</u> טיעון קומבינטורי במקום הוכחת המבוקש. פתרונות אלה כמעט תמיד כללו טענות שגויות – למשל שיש 2^{2^L} עצים אפשריים, בעוד שברור ש-$\mathcal{H}$ אינסופית. פתרונות שפירשו איך לצמצם את $\mathcal{H}$ לקבוצה סופית על ידי התייחסות לכל קבוצת מסווגים שמסווגת את דוגמאות C באופן זהה כמחלקת שקילות, כמעט תמיד כללו טעויות בספירה של מס' העצים או התיוגים האפשריים.

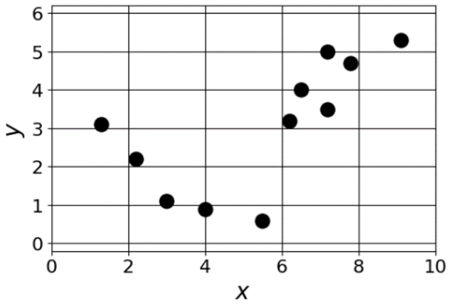
<u>(C)</u> פתרון מהצורה: "מעיקרון שובך היונים קיימות שתי דוגמאות שיסווגו על ידי אותו עלה ונגדיר תיוג הפוך לשתיהן". פתרון זה מבצע הנחה על הבניה של העץ עוד <u>לפני</u> בחירת התיוג של הדוגמאות. מעבר לכך שזהו סדר שגוי של ההוכחה, מדובר בטעות – אמנם <u>קיים</u> עץ שלא מצליח לסווג את שתי הנקודות עם התיוג הנ"ל, אך אין זה מוכיח כי <u>לא קיים עץ</u> אחר ב-$\mathcal{H}$ שכן מצליח לסווג אותן.

להמחשה, דוגמה נגדית: תהי קבוצה $C = \{x_i\}_{i=1}^{2^L+1}$ כלשהי (בה"כ ממוינות בסדר עולה), ויהיו x_i, x_{i+1} שתי דוגמאות עוקבות כנ"ל (מגיעות לאותו עלה בעץ <u>ספציפי</u>). נגדיר להן תיוג הפוך, למשל $y_i = 1, y_{i+1} = -1$. ברור שאפשר להגדיר את התיוג של יתר הנקודות כך: $y_{\leq i} = 1$, $y_{\geq i+1} = -1$ (כל הנקודות עם אינדקס קטן או שווה ל-i מתויגות כ-1 והשאר 1-). התיוג הנ"ל ניתן לסיווג באופן מושלם על ידי עץ החלטה בעומק 1.

<u>(D)</u> הנחת אלגוריתם בניה ספציפי לעץ. לדוגמה: מיון של נקודות C ואז חלוקה לפי החציון בדומה לבניה בסעיף ב'. אם ברצוננו להוכיח שלא קיים עץ ב-$\mathcal{H}$ שמסווג נכון את כל הנקודות, לא ניתן להניח הנחות על אופן הבניה שלו. הוכחה כזו תקפה רק על עצים שנבנו ע"י האלגוריתם שנבחר, ולא על עץ כללי ב-$\mathcal{H}$.

<u>(E)</u> התייחסות למקרה שקיימים $i \neq j$ כך ש-$x_i = x_j$. במקרה כזה הגודל של הקבוצה C היה קטן מ-$2^L + 1$. ובנוסף לא כך הגדרנו ניתוץ.

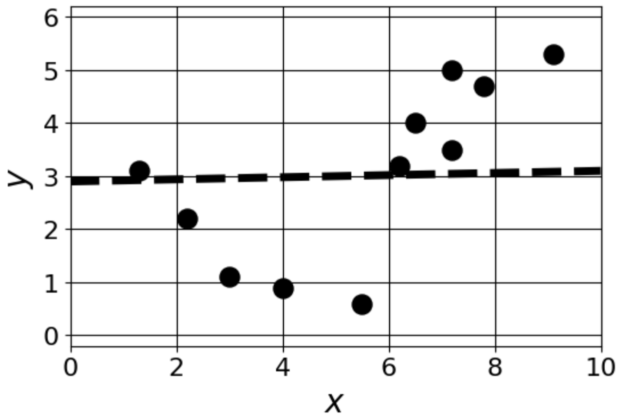
שאלה 3: Linear Regression [32 נק']



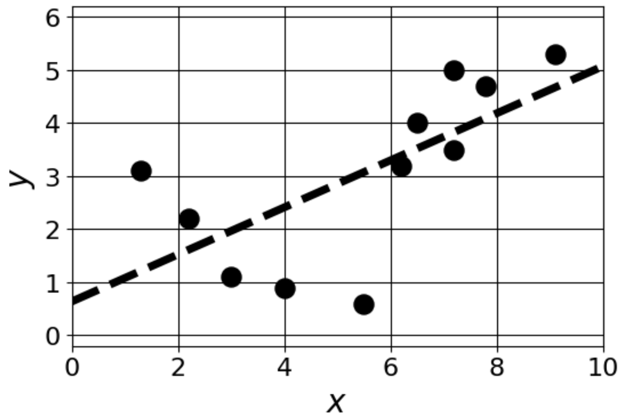
בתרשים שלפניכם מופיע מדגם אימון חד-ממדי $\mathcal{X}=\mathbb{R},\mathcal{Y}=\mathbb{R}$.

א. [2 נק'] לומדים רגרסיה לינארית לא הומוגנית (עם Least squares רגיל) על המדגם הנתון. הקיפו את האות של התרשים שמתאים לקו הרגרסיה שנלמד.

~~(A)~~

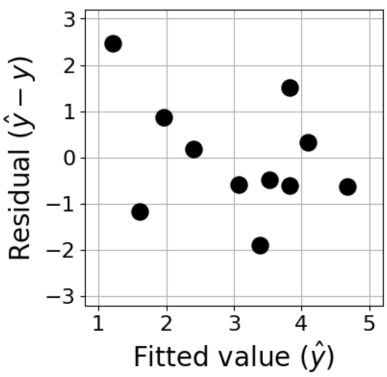


<u>(B) – רואים שה-residuals קטנים יותר</u>

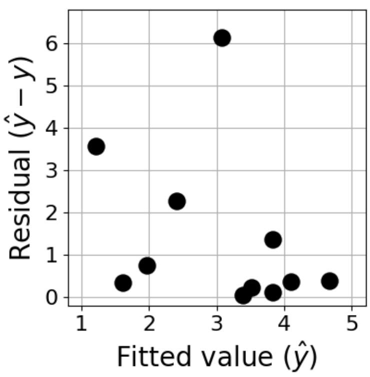


ב. [4 נק'] להלן תרשימי Residual analysis מהסוג שהוצג בהרצאה (שימו לב לצירי ה-y השונים).
הקיפו את האות של התרשים <u>היחיד</u> שמתאים לנתונים ולקו הרגרסיה מהסעיף הקודם.

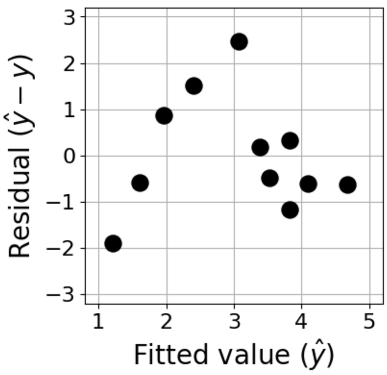
~~(A)~~



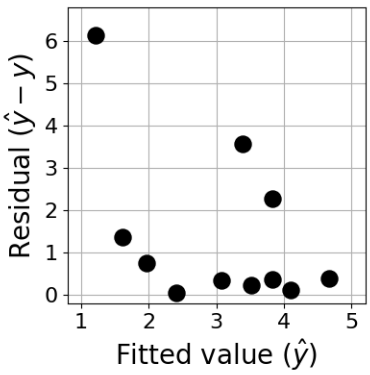
~~(B)~~



<u>(C)</u>



~~(D)~~

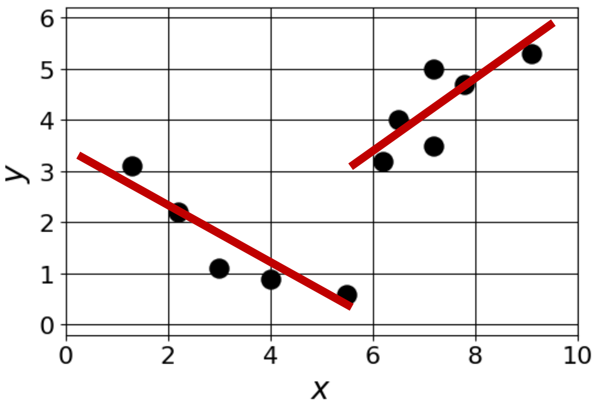


נגדיר מחלקת Thresholded linear regression המאפשרת התאמה של שני קווי רגרסיה נפרדים:

$$\mathcal{H}_1=\left\{h_{w_1,b_1,w_2,b_2,\tau}\,\middle|\,w_1,b_1,w_2,b_2,\tau\in\mathbb{R}\right\},\quad \text{where}\quad h_{w_1,b_1,w_2,b_2,\tau}(x)=\begin{cases}w_1x+b_1, & x<\tau\\ w_2x+b_2, & x\geq\tau\end{cases}$$

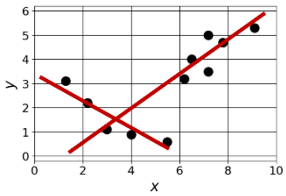
כמו במקרה הלינארי, נגדיר היפותזה אופטימלית ב-$\mathcal{H}_1$ בתור היפותזה שמביאה למינימום את ה-MSE על הדאטה.

משמע מושגת שגיאה: $\min_{w_1,b_1,w_2,b_2,\tau\in\mathbb{R}}\frac{1}{m}\sum_{i=1}^{m}\left(h_{w_1,b_1,w_2,b_2,\tau}(x_i)-y_i\right)^2$.



ג. [2 נק'] ציירו על גבי התרשים שמשמאל היפותזה אופטימלית ב-$\mathcal{H}_1$
(זאת שאלת הבנה, לא נדרשת בציור רמת דיוק של מחשב).

הערה: היפותזה כאן צריכה להיות פונקציה מ-$\mathbb{R}$ ל-$\mathbb{R}$ ולכן לא נכון לצייר משהו כמו:



ד. [8 נק'] ברשותכם אלגוריתם קופסה שחורה LS המקבל אוסף שרירותי של דוגמאות חד-ממדיות והתיוגים שלהן ופותר את בעיית ה-Least squares הלינארית הלא-הומוגנית הרגילה שלמדנו בקורס.
עבור דאטה חד-ממדי כללי (ולא רק למדגם הנתון בסעיפים הקודמים), הציעו אלגוריתם למידה המוצא פרמטרים w_1,b_1,w_2,b_2,τ שמייצגים היפותזה אופטימלית ב-$\mathcal{H}_1$. אין צורך להוכיח את נכונות האלגוריתם.

תשובה:

נמיין את הנקודות.

נבחר מספר בין כל שתי נקודות עוקבות (וגם לפני הראשונה ואחרי האחרונה) וננסה אותו בתור τ.

לכל ערך τ כזה נלמד שני קווי רגרסיה עם LS, אחד לנקודות מימין ואחד לנקודות משמאל.

בסוף נחזיר את הצירוף של הפרמטרים שהחזיר את ה-MSE הנמוך ביותר (ממושקל לפי מס' הדוגמאות בכל תת קבוצה).

מקרא טעויות נפוצות:

a) שימוש ב-cross validation לקביעת τ. שימו לב ש-τ הוא פרמטר נלמד ולא היפרפרמטר (כמו w_1,b_1,w_2,b_2), צריך לקבוע אותו באופן שימזער את השגיאה על ה-train set. ראו איך הגדרנו היפותזה "אופטימלית" לעיל. השאלה כאן לא עוסקת כלל בהכללה.

b) קביעת τ לפי המינימום של סכום ה-MSE של שתי תתי-הבעיות (במשקל שווה). שימו לב שבאופן כללי מתקיים $\frac{1}{|S_1\cup S_2|}\sum_{i\in S_1\cup S_2}(h(x_i)-y_i)^2\neq\frac{1}{|S_1|}\sum_{i\in S_1}(h(x_i)-y_i)^2+\frac{1}{|S_2|}\sum_{i\in S_2}(h(x_i)-y_i)^2$.
היה צריך להתייחס מפורשות לעניין זה בפתרון ולהביא למינימום את $\frac{1}{m}\sum_{i=1}^{m}\left(h_{w_1,b_1,w_2,b_2,\tau}(x_i)-y_i\right)^2$.

c) לא לציין במפורש אילו ערכים של τ האלגוריתם בודק.

כעת, ניצור מחלקה דומה המכילה רק את הפונקציות הרציפות מ-$\mathcal{H}_1$. נעשה זאת ע"י הגדרת $\mathcal{H}_2$ עם פרמטריזציה שונה:

$$\mathcal{H}_2 = \left\{ h'_{u_1,c_1u_2,c_2} \,\middle|\, u_1, c_1 u_2, c_2 \in \mathbb{R} \right\}, \quad \text{where} \quad h'_{u_1,c_1u_2,c_2}(x) = u_1 \max\{0, x - c_1\} + u_2 x + c_2$$

ה. [8 נק'] הוכיחו (מתמטית) שמתקיים $\mathcal{H}_2 \subseteq \mathcal{H}_1$.

הוכחה: תהא היפותזה כלשהי ב-$h'_{u_1,c_1u_2,c_2} \in \mathcal{H}_2$. נדרוש:

$$h'_{u_1,c_1u_2,c_2}(x) = \begin{cases} u_2 x + c_2, & x < c_1 \\ u_1(x - c_1) + u_2 x + c_2, & x \geq c_1 \end{cases} = \begin{cases} w_1 x + b_1, & x < \tau \\ w_2 x + b_2, & x \geq \tau \end{cases} = h_{w_1,b_1,w_2,b_2,\tau}(x)$$

משמע, אם נבחר $\tau = c_1, w_1 = u_2, b_1 = c_2, w_2 = u_1 + u_2, b_1 = c_2 - c_1 u_1$

אזי ההיפותזה $h_{w_1,b_1,w_2,b_2,\tau} \in \mathcal{H}_1$ שקולה ל-$h'_{u_1,c_1u_2,c_2} \in \mathcal{H}_2$.

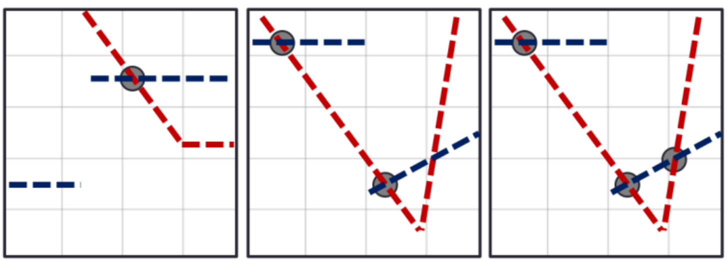
ו. [8 נק'] נתון דאטה חד-ממדי עם תיוגים רציפים $(\mathbf{x}_i \in \mathbb{R},\ y_i \in \mathbb{R})$ מחולק באופן אקראי לסט אימון ולסט מבחן.
פותרים את בעיית הרגרסיה על סט האימון ע"י מציאת היפותזות אופטימליות ב-$\mathcal{H}_1$ וב-$\mathcal{H}_2$ (ביחס ל-Training MSE).
כל שורה בטבלה מתארת תוצאות מתהליך הלמידה המתואר על דאטה נתון כלשהו.
חלק מהשורות מתארות תוצאות אפשריות וחלק מתארות תוצאות בלתי אפשריות.
לכל אחת מארבע השורות, סמנו האם היא אפשרית או לא.

Optimal hypothesis of $\mathcal{H}_1$		Optimal hypothesis of $\mathcal{H}_2$		
Training MSE	**Test MSE**	**Training MSE**	**Test MSE**	**האם אפשרית?**
17	55	25	33	**כן** / לא
17	35	15	37	כן / **לא**
17	35	42	51	**כן** / לא
0	11	0	9	**כן** / לא

מפתח: על כל סימון שגוי ירדו 3 נק' (עד למקסימום של 8 נק').

הסבר: השורה השנייה לא אפשרית כי לא ייתכן שהמחלקה השנייה תביא לשגיאת אימון נמוכה יותר (הוכחנו $\mathcal{H}_2 \subseteq \mathcal{H}_1$).
שאר השורות יכולות להתאים למקרים של overfitting (הראשונה) או underfitting (השלישית).

השורה הרביעית אפשרית במקרים רבים כאשר אין יחידות בהיפותזה האופטימלית. זה אפשרי ללא תלות במספר הנקודות בסט האימון (אחת, שתיים, שלוש או יותר).
למשל עבור שלושת ה-training sets הבאים, בכל אחת מהמחלקות ישנן אינסוף היפותזות אופטימליות (שמגיעות לשגיאת אימון אפס; נסמן היפותזה אופטימלית כלשהי מ-$\mathcal{H}_1$ בכחול כהה ומ-$\mathcal{H}_2$ באדום) עם שגיאת מבחן שונה:



כמו כן שימו לב שהשאלה לא עוסקת כלל באלגוריתם למציאת היפותזה אופטימלית, אלא רק במחלקות עצמן.

שאלה 4: נושאים שונים בפרספטרון וב-SVM [32 נק']

לפניכם אלגוריתם הפרספטרון (ההומוגני) המקבל כקלט מדגם אימון $S=\{(\mathbf{x}_i,y_i)\}_{i=1}^m$, כאשר $\mathcal{X}=\mathbb{R}^d$ וְ-$\mathcal{Y}=\{-1,+1\}$.

```
1       w = 0_d
2       errorFound = True
3
4       for epoch=1 to EPOCHS_NUM:
5         errorFound = False
6
7         for i=1 to m:
8           ŷ_i = sign(w^T x_i)
9
10          if y_i != ŷ_i:
11            w = w + y_i x_i
12            errorFound = True
13
14        if not errorFound:
15          break
```

ראינו בתרגול שפרספטרון שקול ל-SGD על $\sum_i \max\{0,-y_i\,\mathbf{w}^\top\mathbf{x}_i\}$. כעת נראה תכונה דומה ל-Margin Perceptron.

נגדיר $\mathcal{L}(\mathbf{w})=\sum_i \ell(\mathbf{w};\mathbf{x}_i,y_i)$ כאשר $\ell(\mathbf{w};\mathbf{x}_i,y_i)=\max\{0,1-y_i\,\mathbf{w}^\top\mathbf{x}_i\}$. נזכיר שהפונקציה $\ell(\mathbf{w};\mathbf{x}_i,y_i)$ קמורה ב-$\mathbf{w}$.

א. [5 נק'] הציעו subgradient של ℓ לפי $\mathbf{w}$, משמע, $\nabla_{\mathbf{w}}\ell(\mathbf{w};\mathbf{x}_i,y_i)$ (כתבו רק תשובה סופית).

תשובה: זו פשוט הנגזרת של ה-hinge loss,

$$\nabla_{\mathbf{w}}\ell(\mathbf{w};\mathbf{x}_i,y_i)=\begin{cases}-y_i\,\mathbf{x}_i, & y_i\,\mathbf{w}^\top\mathbf{x}_i<1\\ \mathbf{0}_d, & y_i\,\mathbf{w}^\top\mathbf{x}_i\ge 1\end{cases}$$

ב. [9 נק'] הציעו שינוי נקודתי לאלגוריתם הפרספטרון (ציינו את מספרי השורות שמתעדכנות ואת השינויים הנדרשים) כך שהאלגוריתם המעודכן יהיה שקול ל-SGD על $\mathcal{L}(\mathbf{w})$ שהגדרנו. יש להסביר בקצרה כיצד השקילות מתקיימת.

תשובה: נשנה את התנאי שבשורה 10 ל-

```
if y_i w^T x_i < 1:
```

לאחר השינוי, מתקיים בכל צעד באלגוריתם

$$\mathbf{w}=\underbrace{\begin{cases}\mathbf{w}+y_i\,\mathbf{x}_i, & y_i\,\mathbf{w}^\top\mathbf{x}_i<1\\ \mathbf{w}, & y_i\,\mathbf{w}^\top\mathbf{x}_i\ge 1\end{cases}}_{\text{Margin Perceptron Algorithm}}=\mathbf{w}-\begin{cases}-y_i\mathbf{x}_i, & y_i\,\mathbf{w}^\top\mathbf{x}_i<1\\ \mathbf{0}_d, & y_i\,\mathbf{w}^\top\mathbf{x}_i\ge 1\end{cases}=\underbrace{\mathbf{w}-\overbrace{\eta}^{=1}\nabla_{\mathbf{w}}\ell(\mathbf{w};\mathbf{x}_i,y_i)}_{\text{SGD}}$$

ללא תלות בסעיפים הקודמים, נחזור לעסוק באלגוריתם הפרספטרון הרגיל כפי שמופיע בתחילת השאלה.

משפט ההתכנסות של הפרספטרון:

יהי דאטה פריד לינארית הומוגנית $S=\{(\mathbf{x}_i,y_i)\}_{i=1}^{m}$. נסמן את הנורמה המרבית של דוגמה ע"י $R\triangleq \max_i\|\mathbf{x}_i\|_2$.

יהי $\mathbf{w}_{\text{SVM}}$ פיתרון ה-Hard SVM על הדאטה, משמע: $\mathbf{w}_{\text{SVM}}\triangleq \operatorname{argmin}_{\mathbf{w}\in\mathbb{R}^d}\|\mathbf{w}\|_2 \;\; s.t.\;\forall i: y_i\mathbf{w}^\top\mathbf{x}_i\geq 1,$.

אזי, אלגוריתם הפרספטרון יעשה לכל היותר $(R\|\mathbf{w}_{\text{SVM}}\|_2)^2$ טעויות חיזוי בזמן האימון שלו.

ג. [9 נק'] נתון דאטה $S=\{(\mathbf{x}_i,y_i)\}_{i=1}^{100}$ פריד לינארית הומוגנית ש"הבטחת" המשפט עבורו הינה $(R\|\mathbf{w}_{\text{SVM}}\|_2)^2=10$.
יוצרים דאטה חדש ע"י הכפלת <u>כָּל</u> הדוגמאות ב-2, משמע $S'=\{(2\mathbf{x}_i,y_i)\}_{i=1}^{100}$.
מהי הבטחת המשפט לדאטה החדש? משמע, כמה טעויות חיזוי (לכל היותר) יעשה אלגוריתם הפרספטרון על הדאטה החדש? כתבו ערך <u>מספרי</u> מפורש והסבירו <u>בקצרה</u> את תשובתכם.

תשובה:

ברור שמתקיים $R'=2R$. אבל מתקיים גם $\|\mathbf{w}'_{\text{SVM}}\|_2=\frac{1}{2}\|\mathbf{w}_{\text{SVM}}\|_2$ ולכן ההבטחה המעודכנת נותרת 10.

(כדי להבין את השוויון השני - הסתכלו על הגדרת $\mathbf{w}_{\text{SVM}}$; רעיונות דומים הופיעו במבחני עבר רבים.)

ד. [9 נק'] נתון דאטה $S=\{(\mathbf{x}_i,y_i)\}_{i=1}^{100}$ פריד לינארית הומוגנית ש"הבטחת" המשפט עבורו הינה $(R\|\mathbf{w}_{\text{SVM}}\|_2)^2=10$.
נסיר את הדוגמה האחרונה ונקבל דאטה חדש $S'=\{(\mathbf{x}_i,y_i)\}_{i=1}^{99}$. נסמן את החסם המעודכן בעזרת $(R'\|\mathbf{w}'_{\text{SVM}}\|_2)^2$.
בטענה שלפניכם, כתבו את היחס $(\geq,\leq,=)$ המדויק <u>ביותר</u> כך שתהיה נכונה <u>בהכרח</u>: $10 \underbrace{\;\leq\;}_{\text{השלימו}} (R'\|\mathbf{w}'_{\text{SVM}}\|_2)^2$.

הסבירו <u>בקצרה</u>:

ברור שמתקיים $R'\leq R$. כמו כן, מהגדרת ה-SVM מתקיים $\|\mathbf{w}'_{\text{SVM}}\|_2\leq\|\mathbf{w}_{\text{SVM}}\|_2$.

(הסרת דוגמה שקולה להסרת אילוץ על ה-SVM, ולכן יכולים להתקבל רק פתרונות בנורמה נמוכה יותר.)