הטכניון – מכון טכנולוגי לישראל

הפקולטה למדעי המחשב

מבוא למערכות לומדות (236756)

סמסטר חורף תשפ"ג – 15 בפברואר 2023

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.

הנחיות הבחינה:

- **משך הבחינה:** שלוש שעות.
- **חומר עזר:** המבחן בחומר סגור (ללא ספרים, מחברות, דפי נוסחאות).
- מחשבון: מותר.
- כלי כתיבה: עט בלבד.
- יש לכתוב את התשובות **על גבי שאלון זה**.
- מותר לענות בעברית או באנגלית.
- הוכחות והפרכות צריכות להיות פורמליות.
- קריאוּת:
  - תשובה בכתב יד לא קריא – **לא תיבדק**.
  - בשאלות רב-ברירה – הקיפו את התשובות בבירור. סימונים לא ברורים יביאו לפסילת התשובה.
  - לא יתקבלו ערעורים בנושא.
- במבחן 17 עמודים ממוספרים סה"כ, כולל עמוד שער זה שמספרו 1 ושלושה עמודי טיוטה בסוף הגיליון.
- נא לכתוב רק את המבוקש ולצרף הסברים קצרים עפ"י ההנחיות.

- **בתום המבחן יש להגיש את שאלון זה בלבד.**

בהצלחה!

חלק א' – שאלות פתוחות [94 נק']

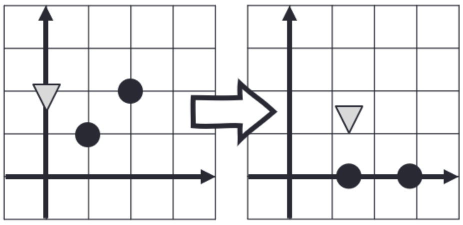
שאלה 1: רגישות של מסווגים לסיבובים [24 נק']

נתון סֵט אימון דו-ממדי עם תיוגים בינאריים, משמע לכל $i=1,\ldots,m$ מתקיים $\mathbf{x}_i\in\mathbb{R}^2, y_i\in\{-1,1\}$.

לומדים שני מסווגים:

- בשלב הראשון: לומדים מסווג על סט האימון המקורי ומחשבים עליו את דיוק האימון.
- בשלב השני:
  - מסובבים את כל הדאטה ב-45° סביב ראשית הצירים (ראו דוגמה).
  - מאמנים מסווג חדש על סט האימון המעודכן, ומחשבים עליו את דיוק האימון המעודכן.

סיבוב לדוגמה



תזכורת: הסיבוב של הדאטה יכול להתבצע ע"י מיפוי $\mathbf{x}_i\mapsto\mathbf{Qx}_i$, כאשר $\mathbf{Q}\in\mathbb{R}^{2\times2}$ הינה מטריצת סיבוב ממשית, שהיא מטריצה אורתונורמלית המקיימת $\mathbf{Q}^{-1}=\mathbf{Q}^\top$.

עבור כל אלגוריתם למידה, סמנו האם דיוק האימון של המסווג החדש על סט האימון המעודכן זהה בהכרח לזה של המסווג המקורי על סט האימון המקורי.

הסבירו בקצרה את תשובותיכם (2-4 משפטים בכל סעיף).

הניחו שאין צעדים אקראיים או שגיאות נומריות בריצת האלגוריתמים (בעיות קמורות מתכנסות לפתרון האנליטי במדויק).

א. kNN עם $k=3$ (דוגמה לא נחשבת שכנה של עצמה). דיוק האימון זהה בהכרח? **כן** / לא

הסבר: סיבוב של מערכת הצירים לא משפיע על מרחקים אוקלידים בין נקודות.

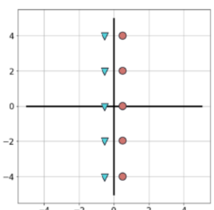
ב. Hard-SVM ליניארי לא הומוגני בהנחה שהדאטה המקורי פריד. דיוק האימון זהה בהכרח? **כן** / לא

הסבר: פרידות ליניארית נשמרת תחת סיבוב של הצירים. ניתן למפות כל (נורמל של) מסווג למסווג חדש $\mathbf{w}\mapsto\mathbf{Qw}$.

ג. AdaBoost with decision stumps, maximum of $T=10$ iterations. דיוק האימון זהה בהכרח? כן / **לא**

הסבר: מחלקת ההיפותזות של decision stumps מכילה מסווגים שמיושרים לפי הצירים. הסיבוב יכול להפוך דאטה שפריד בקלות לדאטה שקשה מאוד להפריד. לא מובטח שהאלגוריתם יצליח לסווג את כל הנקודות במספר צעדים נתון מוגבל (10).

אפשר גם לתת כדוגמה דאטה דומה לזה שהיה בתרגיל הבית:



ד. Homogeneous logistic regression with L2 regularization (ללא הנחה שהדאטה המקורי פריד):

$$\operatorname{argmin}_{\mathbf{w}\in\mathbb{R}^2}\sum_i \ln(1+\exp(-y_i\mathbf{w}^\top\mathbf{x}_i))+\lambda\|\mathbf{w}\|_2^2,\quad \text{for some } \infty>\lambda>0$$

דיוק האימון זהה בהכרח? **כן** / לא

הסבר: ניתן למפות באופן חח"ע כל (נורמל של) מסווג למסווג חדש $\mathbf{w}\mapsto\mathbf{Qw}$ בלי לשנות את ערך המטרה, כי מתקיים $\|\mathbf{w}\|_2^2=\mathbf{w}^\top\mathbf{w}=\mathbf{w}^\top\mathbf{Q}^\top\mathbf{Qw}=\|\mathbf{Qw}\|_2^2$ וגם $\ln(1+\exp(-y_i\mathbf{w}^\top\mathbf{x}_i))=\ln(1+\exp(-y_i(\mathbf{Qw})^\top\mathbf{Qx}_i))$.

שאלה 2: VC-dimension [20 נק']

א. [6 נק'] הוכיחו שלכל מחלקת היפותזות <u>סופית</u> $\mathcal{H}$ מתקיים בהכרח $\mathrm{VCdim}(\mathcal{H}) \le \log_2(|\mathcal{H}|)$.

<u>רמז</u>: ניתן להוכיח זאת בשלילה ובקצרה.

הוכחה (לרשותכם דפי טיוטה בסוף הגיליון):

נניח בשלילה $\mathrm{VCdim}(\mathcal{H}) > \log_2(|\mathcal{H}|)$.

משמע, קיימת קבוצה $\mathcal{C}$ של נק' שונות שאפשר לנתץ עם $\mathcal{H}$ ומתקיים $|\mathcal{C}| > \log_2(|\mathcal{H}|)$.

מספר ההשמות (labelings) האפשריות של תיוגים של הנק' בקבוצה הוא $2^{|\mathcal{C}|}$.

לכל השמה יש לפחות היפותזה אחת ב-$\mathcal{H}$ שמסכימה איתה. ומקבלים בסתירה

$$|\mathcal{H}| \ge 2^{|\mathcal{C}|} \Longrightarrow \log_2(|\mathcal{H}|) \ge |\mathcal{C}| \underbrace{> \log_2(|\mathcal{H}|)}_{\text{assumption}}$$

ב. [4 נק'] נגדיר מרחב דוגמאות סופי $\mathcal{X} = \{x_1, x_2\}$ (במרחב יש שתי דוגמאות במרחב כלשהו, לא נבטא במפורש את x_1, x_2)

ומחלקת היפותזות סופית $\mathcal{H} = \{h_1, h_2\}$, כאשר h_1, h_2 מוגדרות באופן הבא:

	h_1	h_2
x_1	$+1$	-1
x_2	-1	-1

(המשמעות של הטבלה היא, למשל, שמתקיים $h_1(x_1) = +1$.)

<u>ענו</u>: מתקיים $\mathrm{VCdim}(\mathcal{H}) = \boxed{1}$. <u>הוכיחו</u> את תשובתכם.

הוכחה:

את הקבוצה $\mathcal{C} = \{x_1\}$ אפשר לנתץ עם $\mathcal{H}$. לכן $\mathrm{VCdim}(\mathcal{H}) \ge 1$.

ולפי סעיף א' מתקיים $\mathrm{VCdim}(\mathcal{H}) \le \log_2(|\mathcal{H}|) = \log_2 2 = 1$.

רמז: גם בסעיפים הבאים ניתן להגדיר מרחב דוגמאות סופי $\mathcal{X} = \{x_1, \dots, x_n\}$ ואוסף סופי $\mathcal{H}$ של היפותזות $h: \mathcal{X} \to \{-1, +1\}$.

על ההיפותזות להיות **שונות** (distinct) במובן שאין שתי היפותזות ב-$\mathcal{H}$ שמחזירות פלט זהה על כָּל הדוגמאות ב-$\mathcal{X}$.

בשני הסעיפים אין צורך להוכיח את ה-VC-dimension של המחלקות שתציעו.

ג. [5 נק'] הראו מחלקת היפותזות $\mathcal{H}$ עבורה $|\mathcal{H}| = 4$ וגם $\mathrm{VCdim}(\mathcal{H}) = 1$.

תשובה (לרשותכם דפי טיוטה בסוף הגיליון):

	h_1	h_2	h_3	h_4
x_1	$+1$	$+1$	$+1$	-1
x_2	-1	-1	$+1$	-1
x_3	-1	$+1$	-1	-1

ד. [5 נק'] הראו מחלקות היפותזות $\mathcal{H}_1, \mathcal{H}_2$ עבורן $\mathrm{VCdim}(\mathcal{H}_1) = \mathrm{VCdim}(\mathcal{H}_2) = 1$ וגם $\mathrm{VCdim}(\mathcal{H}_1 \cup \mathcal{H}_2) = 3$.

תשובה (לרשותכם דפי טיוטה בסוף הגיליון):

	h_1	h_2	h_3	h_4
x_1	-1	-1	-1	$+1$
x_2	$+1$	$+1$	-1	$+1$
x_3	$+1$	-1	$+1$	$+1$

	h_1	h_2	h_3	h_4
x_1	$+1$	$+1$	$+1$	-1
x_2	-1	-1	$+1$	-1
x_3	-1	$+1$	-1	-1

שאלה 3: רגרסיה ליניארית [25 נק']

בשאלה זו מרחב הדוגמאות הוא $\mathcal{X} = \mathbb{R}^d$ ומרחב התיוגים הוא $\mathcal{Y} = \mathbb{R}$.

נתון מדגם $S = \{(\mathbf{x}_\mathrm{i}, y_i)\}_{i=1}^m$.

רגרסיה ליניארית המשלבת רגולריזציה מסוג L1 ומסוג L2 נקראת ElasticNet:

$$\mathbf{w}^\star \triangleq \operatorname{argmin}_{\mathbf{w}\in\mathbb{R}^d}\left(\underbrace{\sum_{i=1}^m(\mathbf{w}^\top\mathbf{x}_i - y_i)^2}_{\mathcal{L}(\mathbf{w};S)} + \underbrace{\lambda\alpha\|\mathbf{w}\|_2^2 + \lambda(1-\alpha)\|\mathbf{w}\|_1}_{R_{\lambda,\alpha}(\mathbf{w})}\right),$$

עבור $\lambda \geq 0, \alpha \in [0,1]$ כלשהם.

א. [5 נק'] בטור הימני יש שלוש השָמות של λ, α, שהופכות את הבעיה לעיל לבעיות רגרסיה מוּכָּרות. בטור השמאלי, יש חמש שיטות גנרטיביות. במהלך הקורס ראינו שקילות בין שתי בעיות מימין לשתי שיטות משמאל (תחת הנחות מסוימות על הדאטה). את השקילות השלישית עליכם להסיק לבד.

מלאו את המשבצות הריקות של שלוש הבעיות בטור הימני במספרים של שלוש השיטות השקולות להן בטור השמאלי.

a. הבעיה לעיל כאשר $\lambda = 0$	i	i. שערוך MLE
		ii. שערוך Naïve Bayes
b. הבעיה לעיל כאשר $\lambda > 0, \alpha = 0$	v	iii. שערוך MAP עם Gaussian prior
		iv. שערוך MAP עם Binomial prior
c. הבעיה לעיל כאשר $\lambda > 0, \alpha = 1$	iii	v. שערוך MAP עם Laplace prior

פירוט: כאשר $\lambda = 0$ אין שום Prior ולכן מקבלים MLE. נאיב בייס זו הנחה על הדאטה בניגוד לרגולריזציה ש"מניחה" הנחות על הפתרונות. אין הגיון ב-(iv) כי זו התפלגות דיסקרטית.

הערות בדיקה: כל סימון שגוי הוריד 2 נקודות מתוך ה-5.

מעתה נניח $\boldsymbol{\lambda > 0, \alpha \in (0,1)}$ בלבד.

ב. [12 נק'] בסעיף זה נניח שיש $d = 2$ פיצ'רים והם **זהים** זה לזה, כלומר $\forall i \in [m]\colon\ \mathrm{x}_i[1] = \mathrm{x}_i[2]$ (וכאמור, $\lambda > 0, \alpha \in (0,1)$).

(a) [6 נק'] לכל $\mathbf{w} = \begin{bmatrix} \mathrm{w}[1] \\ \mathrm{w}[2] \end{bmatrix} \in \mathbb{R}^2$ נגדיר את $\widetilde{\mathbf{w}} = \begin{bmatrix} \mathrm{w}[2] \\ \mathrm{w}[1] \end{bmatrix}$.

<u>הוכיחו</u>: תחת ההנחות, לכל $\mathbf{w} \in \mathbb{R}^2$ מתקיים $\mathcal{L}(\mathbf{w}; S) + R_{\lambda,\alpha}(\mathbf{w}) = \mathcal{L}(\widetilde{\mathbf{w}}; S) + R_{\lambda,\alpha}(\widetilde{\mathbf{w}})$.

הוכחה:

נורמות L_1, L_2 אדישות לסדר של המקדמים ולכן מתקיים $R_{\lambda,\alpha}(\mathbf{w}) = R_{\lambda,\alpha}(\widetilde{\mathbf{w}})$.

מכיוון שהפיצ'רים זהים, מתקיים:

$$\mathcal{L}(\mathbf{w}; S) = \sum_{i=1}^{m} (\mathrm{x}_i[1]\mathrm{w}[1] + \mathrm{x}_i[2]\mathrm{w}[2] - y_i)^2 = \sum_{i=1}^{m} \left((\mathrm{w}[1] + \mathrm{w}[2])\,\mathrm{x}_i[1] - y_i\right)^2$$

ורואים שגם כאן אין השפעה לסדר של המקדמים.

<u>טענת עזר</u>: פונקציית הרגולריזציה $R_{\lambda,\alpha}$ היא **קמורה במובן החזק** (strictly convex). משמע, מתקיים באי-שיוויון חזק:

$$\forall \mathbf{w}_1 \neq \mathbf{w}_2 \in \mathbb{R}^d, \forall \beta \in (0,1)\colon\ \ \beta R_{\lambda,\alpha}(\mathbf{w}_1) + (1-\beta) R_{\lambda,\alpha}(\mathbf{w}_2) > R_{\lambda,\alpha}(\beta \mathbf{w}_1 + (1-\beta)\mathbf{w}_2).$$

(b) [6 נק'] <u>הוכיחו</u>: תחת ההנחות (2 פיצ'רים זהים, $\lambda > 0, \alpha \in (0,1)$), שני מקדמי הפתרון האופטימלי זהים.

משמע, מתקיים: $\mathrm{w}^\star[1] = \mathrm{w}^\star[2]$.

הוכחה: נניח בשלילה שקיים פתרון אופטימלי $\mathbf{w}^\star$ שעבורו $\mathbf{w}^\star[1] \neq \mathbf{w}^\star[2]$.

ניצור פתרון חדש עם מקדמים מוחלפים $\widetilde{\mathbf{w}^\star}$. לפי הנחת השלילה מתקיים $\widetilde{\mathbf{w}^\star} \neq \mathbf{w}^\star$.

הראינו במהלך ההוכחה של סעיף (a) שמתקיים $R_{\lambda,\alpha}(\mathbf{w}^\star) = R_{\lambda,\alpha}(\widetilde{\mathbf{w}^\star})$.

ניצור פתרון ממוצע $\mathbf{u} = \frac{1}{2}(\mathbf{w}^\star + \widetilde{\mathbf{w}^\star})$.

לפי טענת העזר והאמור לעיל, מתקיים $R_{\lambda,\alpha}(\mathbf{w}^\star) = \frac{1}{2} R_{\lambda,\alpha}(\mathbf{w}^\star) + \frac{1}{2} R_{\lambda,\alpha}(\widetilde{\mathbf{w}^\star}) > R_{\lambda,\alpha}(\mathbf{u})$.

כמו בסעיף הקודם, $\mathcal{L}(\mathbf{u}; S) = \sum_{i=1}^{m} \left((\mathrm{u}[1] + \mathrm{u}[2])\,\mathrm{x}_i[1] - y_i\right)^2$.

מאופן הבנייה של $\widetilde{\mathbf{w}^\star}, \mathbf{u}$, מתקיים $\mathrm{u}[1] + \mathrm{u}[2] = \mathbf{w}^\star[1] + \mathbf{w}^\star[2]$ ולכן $\mathcal{L}(\mathbf{w}^\star; S) = \mathcal{L}(\mathbf{u}; S)$.

בסה"כ קיבלנו $\mathcal{L}(\mathbf{w}^\star; S) + R_{\lambda,\alpha}(\mathbf{w}^\star) > \mathcal{L}(\mathbf{u}; S) + R_{\lambda,\alpha}(\mathbf{u})$ בסתירה לאופטימליות של $\mathbf{w}^\star$.

להלן משפט. **הבינו אותו היטב.** בפרט, השתכנעו שהוא מכליל את שהוכחנו בסעיף הקודם.

משפט[1]: תחת הנחות (לא מגבילות במיוחד) על הנתונים, כל שני מקדמים של פתרון ה-Elastic Net מקיימים:

$$\forall a,b\in[d]:\ \left(|\mathbf{w}^{\star}[a]-\mathbf{w}^{\star}[b]|\le\frac{1}{\lambda\alpha}\sqrt{\left(1-\rho_{a,b}\right)\cdot c}\right),$$

כאשר $\rho_{a,b}\triangleq\frac{\mathrm{Cov}(\mathbf{x}[a],\mathbf{x}[b])}{\sigma_a\sigma_b}\in[-1,1]$ הוא מקדם המְתְאָם (קורלציה) של פִּירסון בין הפיצ'רים a,b (על פני כל הדוג') ו-$c>0$ הוא קבוע כלשהו.

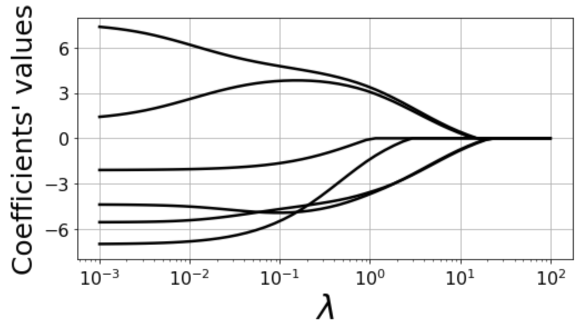
ג. [8 נק'] עבור dataset עם 6 פיצ'רים, אימנו ElasticNet עם ערכי $\alpha\in\{0.01,0.5,0.99\}$ וערכי $\lambda>0$ שונים. כל עקומה בגרפים מראה את ערכיו של מקדם יחיד (מתוך 6) בפתרון שנוצר מצירוף של α (ברמת הגרף) ו-λ (בציר x).

נתון: ישנם שני זוגות של פיצ'רים עם קורלציה גבוהה (בכל זוג, קורלציה גבוהה בין שני הפיצ'רים של אותו זוג).

ליד כֹּל גרף סמנו את ערך α המתאים לו ביותר והסבירו בקצרה את בחירותיכם. היעזרו במשפט לעיל בהסברים.

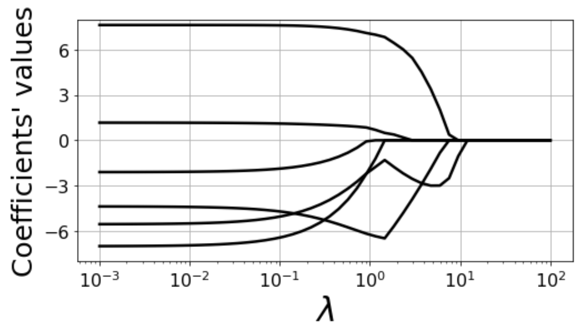
לאיזה ערך α מתאים הגרף? 0.01 / **0.5** / 0.99

הסבר: רואים שני זוגות של מקדמים שנעשים קרובים ככל ש-λ גדל (לכן α לא יכול להיות קטן מדי). ראינו בתרגול ש-L_2 רגיל **לא** מאפס מקדמים (כמו **שכן** קורה כאן), אלא רק גורם להם לדעוך לאפס.



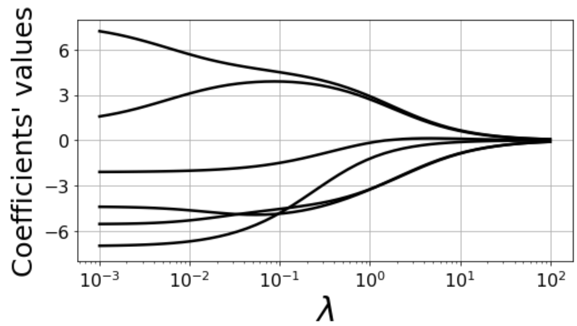
לאיזה ערך α מתאים הגרף? **0.01** / 0.5 / 0.99

הסבר: לא רואים בגרף תופעה שמקדמים נעשים קרובים ככל ש-λ גדל. נראה שלקורלציה כמעט אין השפעה. זה מסתדר עם ערך קטן של α. מעבר לכך, ראינו בתרגול ש-L_1 מאפס מקדמים.



לאיזה ערך α מתאים הגרף? 0.01 / 0.5 / **0.99**

הסבר: אותו הסבר כמו בתרשים הראשון.



[1] גרסה פשטנית של משפט מהמאמר [Zou and Hastie. 2005] "Regularization and variable selection via the elastic net"

שאלה 4: מסווגים ליניאריים [25 נק']

תזכורת: כלל החלטה $h\colon \mathbb{R}^d \to \{-1,+1\}$ נקרא ליניארי אם ורק אם קיימים $\mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}$ כך ש-$\forall \mathbf{x} \in \mathbb{R}^d\colon\ h(\mathbf{x}) = \operatorname{sgn}(\mathbf{w}^\top \mathbf{x} + b)$.

נגדיר: $\operatorname{sgn}(z) = \begin{cases} -1, & z \le 0 \\ +1, & z > 0 \end{cases}$ (כך שלא מתקיים אפס לשום קלט)

תזכורת: $\sigma(z) = \frac{1}{1+\exp\{-z\}}$

ובפרט מתקיים:

$\sigma(0) = 0.5,\quad \sigma(1) \approx 0.73,\quad \sigma(2) \approx 0.88,\quad \sigma(3) \approx 0.95$

$\sigma(4) \approx 0.98,\quad \sigma(5) \approx 0.99,\quad \sigma(6) \approx 1$

א. [5 נק'] הוכיחו שכלל ההחלטה של logistic regression (עבור $\mathbf{w}_1 \in \mathbb{R}^d, b_1 \in \mathbb{R}$ כלליים) הוא ליניארי:

$$h_1(\mathbf{x}) = \begin{cases} +1, & \sigma(\mathbf{w}_1^\top \mathbf{x} + b_1) > 0.5 \\ -1, & \text{אחרת} \end{cases}.$$

הוכחה: מתקיים $h_1(\mathbf{x}) = 1 \Leftrightarrow \sigma(\mathbf{w}_1^\top \mathbf{x} + b_1) > 0.5 \Leftrightarrow \mathbf{w}_1^\top \mathbf{x} + b_1 > 0 \Leftrightarrow \operatorname{sgn}(\mathbf{w}_1^\top \mathbf{x} + b_1) = 1$ כנדרש.

טענת עזר: כלל החלטה ליניארי h צריך לקיים $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d\colon\ \underbrace{h(\mathbf{x}_1) = h(\mathbf{x}_2)}_{\text{if}} \Longrightarrow \underbrace{h\left(\tfrac{1}{2}\mathbf{x}_1 + \tfrac{1}{2}\mathbf{x}_2\right) = h(\mathbf{x}_1)}_{\text{then}}$.

ב. [10 נק'] נתונים $\mathbf{w}_3 \in \mathbb{R}^d, b_3 \in \mathbb{R}$ כלליים כלשהם.

הפריכו: כלל ההחלטה הבא ליניארי:

$$h_3(\mathbf{x}) = \begin{cases} +1, & \sigma(\mathbf{w}_3^\top \sigma(\mathbf{x}) + b_3) > 0.5 \\ -1, & \text{אחרת} \end{cases},$$

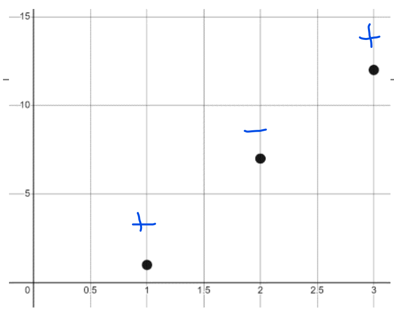
כאשר הפעולה $\sigma(\mathbf{x})$ מחזירה וקטור $[\sigma(x[1]), \sigma(x[2]), \dots, \sigma(x[d])]^\top$.

רמז: כדאי להשתמש בטענת העזר. ניתן לבחור $\mathbf{w}_3, b_3$ ספציפיים כדי להפריך.

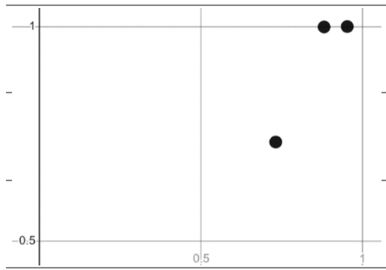
הפרכה: יש הרבה פתרונות אפשריים. הרעיון הוא להראות דאטה לא פריד ליניארית שניתן להפרדה ע"י h_3.

נבחר למשל $\mathbf{x}_1 = (1,1),\ \ \mathbf{x}_2 = (3,13),\ \ \mathbf{x}_3 = (2,7)$ ונתייג $y_1 = y_2 = +1$ וגם $y_3 = -1$.

ע"פ טענת העזר, הדאטה לא פריד ליניארית. ניתן גם לצייר זאת :



לאחר המיפוי:



$\sigma(\mathbf{x}_1) \approx (0.73, 0.73),\ \ \sigma(\mathbf{x}_2) \approx (0.95, 1),\ \ \sigma(\mathbf{x}_3) \approx (0.88, 1)$.

הדאטה החדש כן פריד, כפי שניתן לראות מהציור הבא.

(אפשר גם להציע מפריד ספציפי)

ג. [10 נק'] נתונים $u_2 \in \mathbb{R}, \mathbf{w}_2 \in \mathbb{R}^d$ כלליים כלשהם ומתקיים $u_2 > 1$.

הוכיחו/הפריכו: כלל ההחלטה הבא ליניארי: $h_2(\mathbf{x}) = \begin{cases} +1, & \sigma(u_2\sigma(\mathbf{w}_2^\top \mathbf{x}) - 1) > 0.5 \\ -1, & \text{אחרת} \end{cases}$.

תשובה:

הטענה נכונה, כי מתקיים כנדרש:

$$h_2(\mathbf{x}) = 1 \Leftrightarrow \sigma(u_2\sigma(\mathbf{w}_2^\top \mathbf{x}) - 1) > 0.5 \Leftrightarrow u_2\sigma(\mathbf{w}_2^\top \mathbf{x}) - 1 > 0$$

$$\Leftrightarrow \sigma(\mathbf{w}_2^\top \mathbf{x}) > {}^1/_{u_2}$$

ובגלל שהפונקציה $\sigma: \mathbb{R} \to (0,1)$ הינה חח"ע ומתקיים ${}^1/_{u_2} \in (0,1)$, אזי

$$h_2(\mathbf{x}) = 1 \Leftrightarrow \mathbf{w}_2^\top \mathbf{x} > \sigma^{-1}({}^1/_{u_2})$$

חלק ב' – שאלה אמריקאית [6 נק']

בשאלה הבאה סמנו את התשובות המתאימות (לפי ההוראות). בחלק זה אין צורך לכתוב הסברים.

א. בתרגילי הבית ראינו איך לבצע tuning על היפר-פרמטרים רבים בעזרת Grid search cross validation. נתון אלגוריתם למידה $\mathcal{A}$ עם אוסף $\mathcal{P}$ של היפר-פרמטרים שדורשים tuning. נניח שלכל היפר-פרמטר יש בדיוק v ערכים שנבדקים. נסמן את מספר ה-folds בתהליך ה-CV בתור k ואת מספר דוגמאות האימון בתור m.

סמנו את **כל** הטענות הנכונות בהכרח ביחס לריצה של תהליך ה-Grid search CV:

a. מספר הקריאות ל-$\mathcal{A}$ גדל בצורה ליניארית ב-$|\mathcal{P}|$.

b. מספר הקריאות ל-$\mathcal{A}$ גדל בצורה ליניארית ב-v.

c. <u>מספר הקריאות ל-$\mathcal{A}$ גדל בצורה ליניארית ב-k.</u>

d. סיבוכיות זמן הריצה של קריאה בודדת ל-$\mathcal{A}$ גדלה בצורה ליניארית ב-m.

$\frac{k-1}{k} m$

<u>הערות בדיקה:</u> מספר הקריאות ל-$\mathcal{A}$ הינו $k \cdot v^{|\mathcal{P}|}$. הסיבוכיות של קריאה בודדת ל-$\mathcal{A}$ תלוי באלגוריתם עצמו שאינו ידוע.

לתשובה d הוקצתה נקודה אחת מתוך 6.

בתשובות a-c סימון שגוי אחד הוריד 3 נקודות, שני סימונים שגויים ומעלה הורידו את כל ה-5.