



מבוא למערכות לומדות (236756)

סמסטר אביב תשפ"ג – 29 בדצמבר 2023

מרצה: ד"ר ניר רחנפלד

מבחן מסכם מועד ב'

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.
- קריאות:
- 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 15 עמודים ממוספרים סה"כ, כולל עמוד שער זה שמספרו 1 ושלושה עמודי טיוטה בסוף הגליון.
- נא לכתוב רק את המבוקש ולצרף הסברים קצרים עפ"י ההנחיות.
- בתום המבחן יש להגיש את שאלון זה בלבד.

בהצלחה!

שאלה 1: Nearest neighbors [20 נק']

היו יותר מדי תשובות אפשריות לשאלה,

ולכן לדעתנו היא אינה דוגמה טובה ללמוד ממנה ולא צירפנו אותה לכאן.

שאלה 2: VC-dimension, AdaBoost, פונקציות מיפוי [30 נק']נגדיר את מחלקת ההיפותזות \mathcal{H} של מקטעים (intervals) חד-ממדיים.משמע, $\mathcal{H} = \{h_{a,b} \mid a, b \in \mathbb{R} \text{ s.t. } b > a\}$ מעל $\mathcal{X} = \mathbb{R}$, כשהיפותזה בודדת מוגדרת בתור $h_{a,b}(x) = \mathbb{I}[a \leq x \leq b]$ פונקציית האינדיקטור

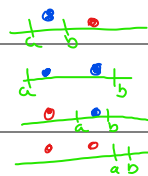
$$\text{VCdim}(\mathcal{H}) =$$

2א. [7 נק'] מהו ה-VC-dimension של \mathcal{H} ?

הוכיחו את תשובתכם.

הוכחה:

$$\text{VCdim}(\mathcal{H}) \geq 2$$

כלומר, הציגו סימול של כה' של
תוצאה של ניון לטיוט משמאל \Rightarrow נחלקנו יחד
המשמאל

$\text{VCdim}(\mathcal{H}) < 3$: "על סמך 3 נק' קיים תיוט שלם כן לא ניתן לטיוט נכון לכל הנתונים"

יהיו x_1, x_2, x_3 נק', בה"כ $x_1 < x_2 < x_3$ ~~נבחר~~ ^{נבחר} $y_1 = y_3 = 1, y_2 = -1$

(בס' כ' (ה"מחיר היסודי) $h_{a,b}(x_i) = y_i$ כקש $\forall x_i: h_{a,b}(x_i) = y_i$

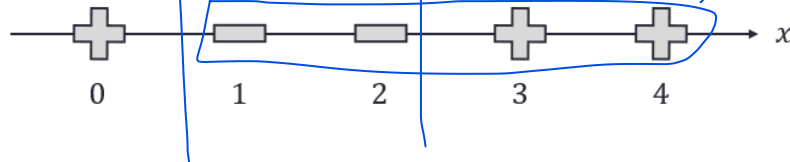
כלומר עבור $i=1$: $a < x_1$ ועבור $i=3$: $x_3 < b$ אך $x_1 < x_2 < x_3$

ואתקבל כי $h_{a,b}(y_i) = 1 \neq -1$ הסתירה לנתון ולכן

על סמך 3 נק' קיים תיוט שלם כן לא ניתן לטיוט נכון

$$\text{VCdim}(\mathcal{H}) < 3$$

עמ' 4

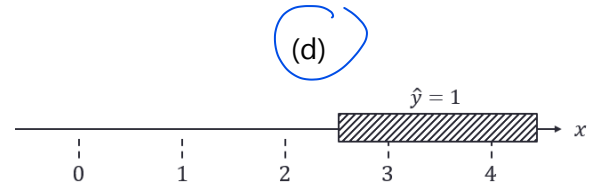
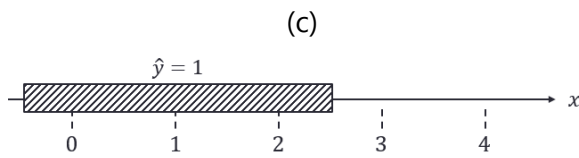
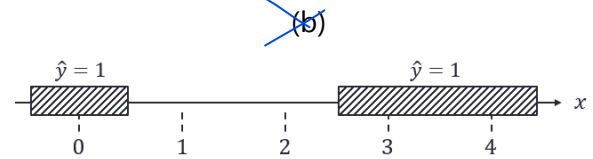
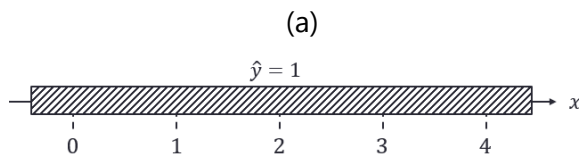


נתון סט אימון עם חמש דוגמאות ב- \mathbb{R} :

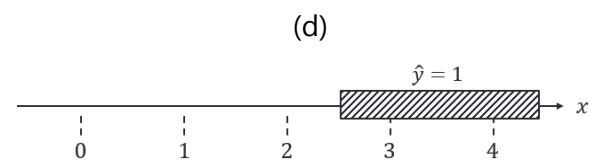
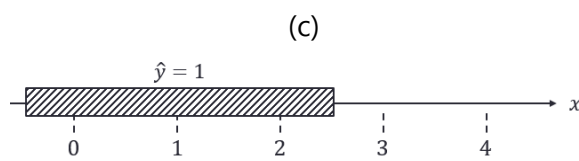
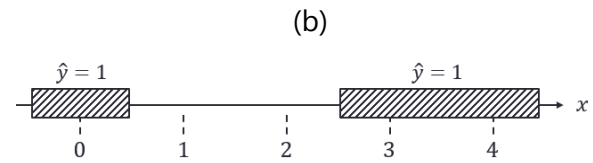
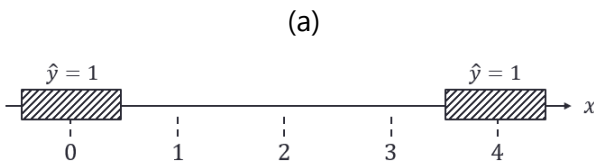
מריצים אלגוריתם AdaBoost על הדאטה הנתון, עם המחלקה \mathcal{H} שהגדרנו בתור מחלקת בסיס (מסווגים חלשים).
בכל איטרציה לומדים מסווג חלש עם ERM על ההתפלגות הנוכחית. המסווג החזק הוא ה-ensemble הממושקל שמתקבל.

בשני הסעיפים הבאים מופיעים תרשימים של כללי החלטה על הישר \mathbb{R} . הכללים חוזים $\hat{y} = 1$ רק במקטעים המקווקוים.
בכל סעיף, הקיפו בבירור את האות היחידה שמתאימה לתשובה הנכונה.

ב. [7 נק'] מבין הבאים – מה המסווג החזק (הממושקל) שמחזיר AdaBoost אחרי האיטרציה הראשונה? אין צורך בהסבר.

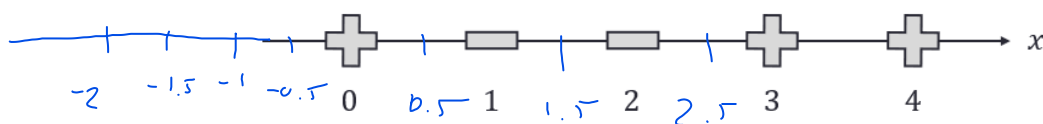


ג. [7 נק'] מבין הבאים – מה המסווג החזק (הממושקל) שמחזיר AdaBoost אחרי שתי איטרציות? הסבירו בקצרה.



הסבר קצר: נבחרו וזר האטו שמכיוונוכו סמוך
פלאו כיוון יש יתרה חשיבות לפלאו
פלאו + האטו

הסעיף הבא לא קשור ל-AdaBoost או למחלקה \mathcal{H} שהוגדרה, אלא רק לדאטה הנתון (מוצג שוב לנוחיתכם):



$$(-1.5)^2 \quad (-0.5)^2 \quad (0.5)^2 \quad (1.5)^2 \quad (2.5)^2$$

$$= + + + + +$$

ד. [9 נק'] אילו מבין פונקציות המיפוי הבאות הופכות את הדאטה הנתון לפריד ליניארית (לאו דווקא הומוגנית)? סמנו את כל התשובות המתאימות בבירור. סימון לא ברור יוביל לפסילת התשובה.

לרשותכם דפי טיוטה בסוף הגיליון.

$$\mathbb{R}^2 \ni \phi(x) = \begin{bmatrix} x \\ 1.5 \end{bmatrix} \quad \text{.iv} \quad \times$$

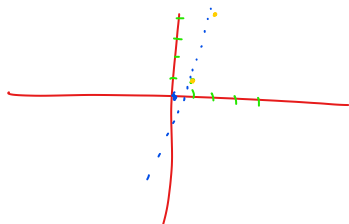
$$\mathbb{R} \ni \phi(x) = x - 1.5 \quad \text{i} \quad \times$$

$$\mathbb{R}^2 \ni \phi(x) = \begin{bmatrix} x^2 \\ 1.5 \end{bmatrix} \quad \text{.v} \quad \times$$

$$\mathbb{R} \ni \phi(x) = x^2 - 1.5 \quad \text{.ii} \quad \times$$

$$\mathbb{R}^2 \ni \phi(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad \text{.vi} \quad \circ$$

$$\mathbb{R} \ni \phi(x) = (x - 1.5)^2 \quad \text{.iii} \quad \circ$$



שאלה 3: גרסיה ורגולריזציה [20 נק']

עבור $\mathbf{X} \in \mathbb{R}^{m \times d}$, $\mathbf{y} \in \mathbb{R}^m$, פונקציה $R: \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$ וסקלר $\lambda > 0$, הגדרנו בעיות גרסיה לינארית הומוגנית עם רגולריזציה:

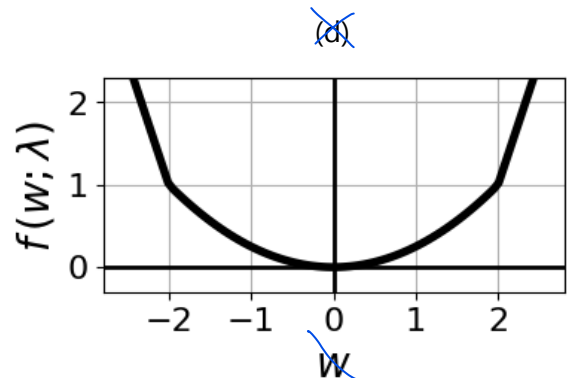
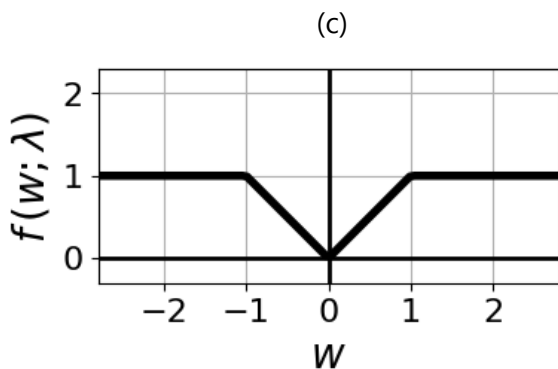
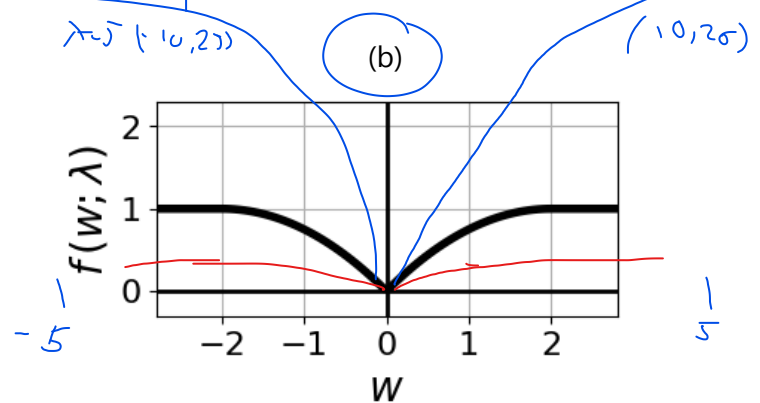
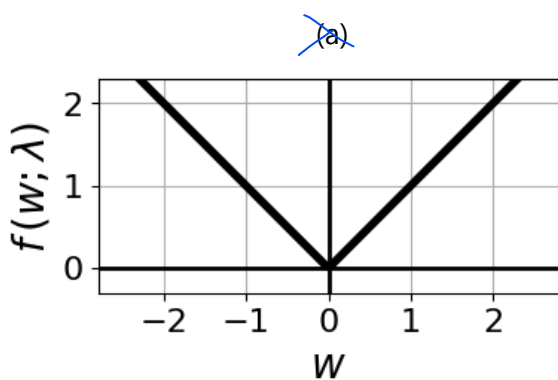
$$\min_{\mathbf{w} \in \mathbb{R}^d} \left(\frac{1}{2m} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + R(\mathbf{w}; \lambda) \right)$$

כאשר משתמשים בפונקציה $R_{LS}(\mathbf{w}; \lambda) \triangleq 0$ מקבלים בעיית Least squares רגילה.

כאשר משתמשים בפונקציה $R_{\ell_1}(\mathbf{w}; \lambda) \triangleq \lambda \|\mathbf{w}\|_1$ מקבלים את בעיית ה-LASSO.

$$f(w; \lambda) \triangleq \begin{cases} \frac{1}{2} \lambda |w| - \frac{w^2}{4}, & |w| \leq 2\lambda \\ \lambda^2, & |w| > 2\lambda \end{cases} \quad \text{עבור } R_{CP}(\mathbf{w}; \lambda) \triangleq \sum_{j=1}^d f(w_j; \lambda)$$

א. [3 נק'] עבור $\lambda = 1$, הקיפו בבירור את האות המתאימה לתרשים שמתאר את $f(w; \lambda)$.



ב. [3 נק'] מה ניתן לומר על הקמירות של הפונק' $f(w; \lambda)$ כאשר $\lambda > 0$? הקיפו את התשובה בבירור.

i. קמורה

ii. קעורה

iii. תלוי בערך של λ

iv. לא קמורה ולא קעורה

מעתה, נסמן ב- $\hat{\mathbf{w}}_{LS}$ את פיתרון ה-Least squares, ב- $\hat{\mathbf{w}}_{\ell_1}$ את פיתרון ה-LASSO וב- $\hat{\mathbf{w}}_{CP}$ את פיתרון הרגרסיה תחת רגולריזציה של הפונק' R_{CP} שהגדרנו.

הנחה: מעתה נניח שהעמודות של \mathbf{X} אורתוגונליות כך שמתקיים $\mathbf{X}^T \mathbf{X} = m \mathbf{I}_{d \times d}$.

נתונה טענה 1: תחת ההנחה, מתקיים

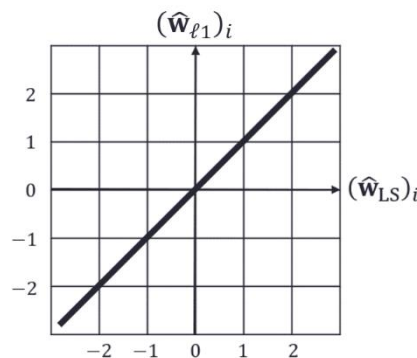
$$(\hat{\mathbf{w}}_{\ell_1})_i = \begin{cases} \text{sign}((\hat{\mathbf{w}}_{LS})_i) \cdot (|(\hat{\mathbf{w}}_{LS})_i| - \lambda), & |(\hat{\mathbf{w}}_{LS})_i| > \lambda \\ 0, & |(\hat{\mathbf{w}}_{LS})_i| \leq \lambda \end{cases}$$

נתונה טענה 2: תחת ההנחה, מתקיים

$$(\hat{\mathbf{w}}_{CP})_i = \begin{cases} (\hat{\mathbf{w}}_{LS})_i, & |(\hat{\mathbf{w}}_{LS})_i| \geq 2\lambda \\ 2 \cdot \text{sign}((\hat{\mathbf{w}}_{LS})_i) \cdot (|(\hat{\mathbf{w}}_{LS})_i| - \lambda), & \lambda < |(\hat{\mathbf{w}}_{LS})_i| < 2\lambda \\ 0, & |(\hat{\mathbf{w}}_{LS})_i| \leq \lambda \end{cases}$$

ג. [4 נק'] עבור כניסה i שרירותית וערך $\lambda = 1$, השתמשו בטענות וציירו באופן ברור על גבי התרשימים הבאים את העקומות של $(\hat{\mathbf{w}}_{CP})_i$ ו- $(\hat{\mathbf{w}}_{\ell_1})_i$ כפונקציה של $(\hat{\mathbf{w}}_{LS})_i$ בכל התחום $[-3, 3]$.

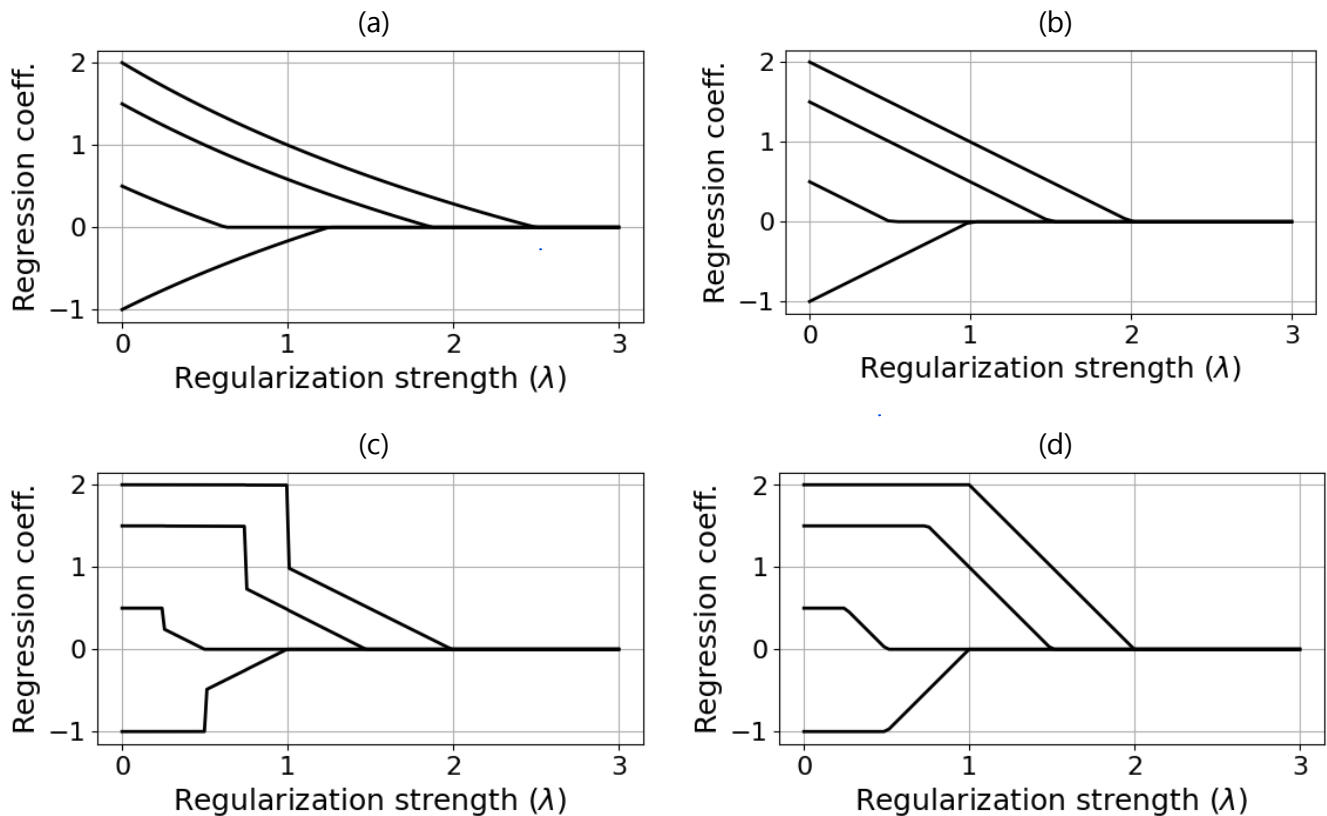
דוגמה: לו היה מתקיים $(\hat{\mathbf{w}}_{\ell_1})_i = (\hat{\mathbf{w}}_{LS})_i$, היה עליכם לצייר:



תשובות (ציירו על גבי התרשימים בכל התחום $[-3, 3]$):

$(\hat{\mathbf{w}}_{\ell_1})_i$ כפונקציה של $(\hat{\mathbf{w}}_{LS})_i$ כאשר $\lambda = 1$	$(\hat{\mathbf{w}}_{CP})_i$ כפונקציה של $(\hat{\mathbf{w}}_{LS})_i$ כאשר $\lambda = 1$

כעת, פותרים בעיית רגרסיה בארבעה ממדים ($\mathbf{X} \in \mathbb{R}^{m \times 4}$) המקיימת את הנחת האורתוגונליות, משמע $\mathbf{X}^T \mathbf{X} = m \mathbf{I}_{4 \times 4}$. התרשימים מתארים את ארבעת המקדמים (ציר אנכי) שמתקבלים עבור ערכי λ שונים (אופקי) תחת פונק' רגולריזציה שונות.



ד. [10 נק'] הקיפו את התשובות הנכונות והסבירו את בחירתכם.

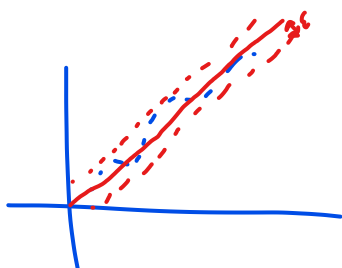
- (a) (b) (c) (d) התרשים שמתאים למקדמים של $\hat{\mathbf{w}}_{\ell_1}$ הוא:
- (a) (b) (c) (d) התרשים שמתאים למקדמים של $\hat{\mathbf{w}}_{\text{CP}}$ הוא:

הסבר:

(מקום נוסף בעמוד הבא)

המשך להסבר (במידת הצורך):

שאלה 4: Support Vector Regression [30 נק']

שאלה זו עוסקת בגרסה לינארית מ- \mathbb{R}^d ל- \mathbb{R} , אותה נפתור בשלבים, בדרך שדומה יותר ל-SVM מאשר ל-Least squares.נגדיר בעיית Hard-SVR עבור m דוגמאות $(\mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R})$ והיפר-פרמטר $\epsilon > 0$. קראו את הבעיה והבינו אותה היטב.

$$\operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}} \|\mathbf{w}\|_2^2$$

$$\text{s.t. } \mathbf{w}^T \mathbf{x}_i + b \leq y_i + \epsilon, \forall i \in [m]$$

$$\mathbf{w}^T \mathbf{x}_i + b \geq y_i - \epsilon, \forall i \in [m]$$

א. [6 נק'] כאשר $\epsilon \rightarrow 0$, עבור אילו סוגי דאטה קיים פיתרון לבעיית ה-Hard-SVR? הסבירו בקצרה.

תשובה והסבר קצר: $\operatorname{argmin} \|\mathbf{w}\|_2^2 \quad \text{s.t. } \mathbf{w}^T \mathbf{x}_i + b = y_i, \forall i \in [m]$

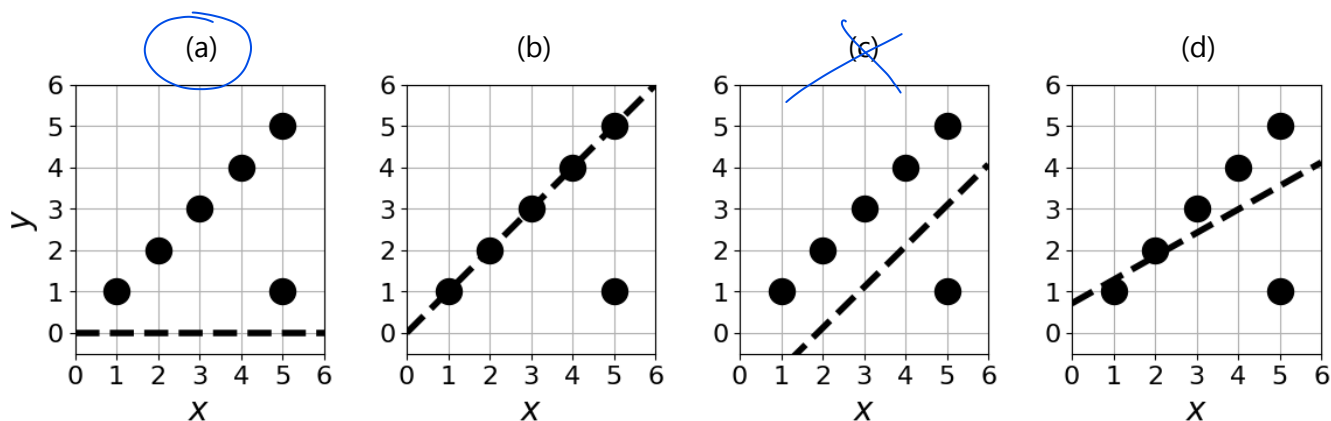
כלומר, אנחנו מחפשים למצוא את העמם כך שכל הנקודות נמצאות על הישר (הזן/הי).

כדי להבטיח שלכל דאטה יהיה פיתרון, נגדיר בעיית Soft-SVR עבור m דוגמאות $(\mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R})$ והיפר-פרמטר $\epsilon > 0$:

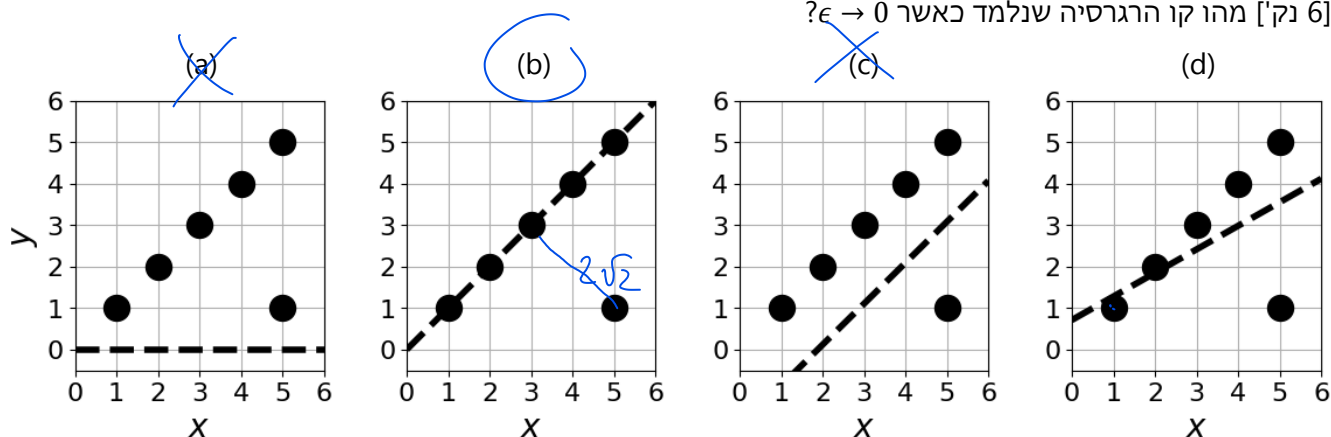
$$\begin{aligned} \underset{\substack{\mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R} \\ \forall i \in [m]: \xi_i, \xi_i^* \geq 0}}{\operatorname{argmin}} \quad & \|\mathbf{w}\|_2^2 + \sum_{i=1}^m (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & \mathbf{w}^T \mathbf{x}_i + b \leq y_i + \epsilon + \xi_i, \quad \forall i \in [m] \\ & \mathbf{w}^T \mathbf{x}_i + b \geq y_i - \epsilon - \xi_i^*, \quad \forall i \in [m] \end{aligned}$$

בסעיפים הבאים נתונים תרשימים של דאטה חד-ממדי $(x_i, y_i \in \mathbb{R})$ וקווי רגרסיה שונים. בכל סעיף כתוב ערך של ההיפר-פרמטר ϵ . הקיפו בבירור את האות שמתאימה לקו הרגרסיה שנלמד על ידי Soft-SVR.

ב. [6 נק'] מהו קו הרגרסיה שנלמד כאשר $\epsilon \rightarrow \infty$?



ג. [6 נק'] מהו קו הרגרסיה שנלמד כאשר $\epsilon \rightarrow 0$?



ד. [6 נק'] בשלב הזה, כשפתרנו בכיתה את בעיית ה-Soft-SVM, עברנו מבעיית אילוצים ללא אילוצים.

עשינו זאת ע"י הגדרה של $\ell_{\text{hinge}}(\mathbf{w}, b; \mathbf{x}_i, y_i) = \max\{0, 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)\}$ ופיתרון הבעיה הבאה:

$$\operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}} \left(\|\mathbf{w}\|_2^2 + \sum_{i=1}^m \ell_{\text{hinge}}(\mathbf{w}, b; \mathbf{x}_i, y_i) \right)$$

בדומה, הציעו פונקציית loss $\ell(\mathbf{w}, b; \mathbf{x}_i, y_i)$ רציפה וקמורה שמתאימה לפיתרון בעיות Soft-SVR ללא אילוצים. הסבירו בקצרה.

תשובה והסבר קצר:

$$\ell_{\text{loss}}(\mathbf{w}, b; \mathbf{x}_i, y_i) = \begin{cases} |y_i - (\mathbf{w}^T \mathbf{x}_i + b)| - 1, & |y_i - (\mathbf{w}^T \mathbf{x}_i + b)| > 1 \\ 0, & \text{otherwise} \end{cases}$$

ידוע שהבעיה הדואלית לבעיית ה-Soft-SVR שהגדרנו היא הבעיה הקעורה הבאה:

$$\operatorname{argmax}_{\substack{\sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0 \\ \forall i \in [m]: \alpha_i, \alpha_i^* \in [0, C]}} \left(\sum_{i=1}^m y_i (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^m (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i,j=1}^m (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \mathbf{x}_i^T \mathbf{x}_j \right)$$

ה. [6 נק'] האם הבעיה הדואלית לעיל מתאימה להפעלת טריק הקרנל, בדומה למה שעשינו ב-SVM?

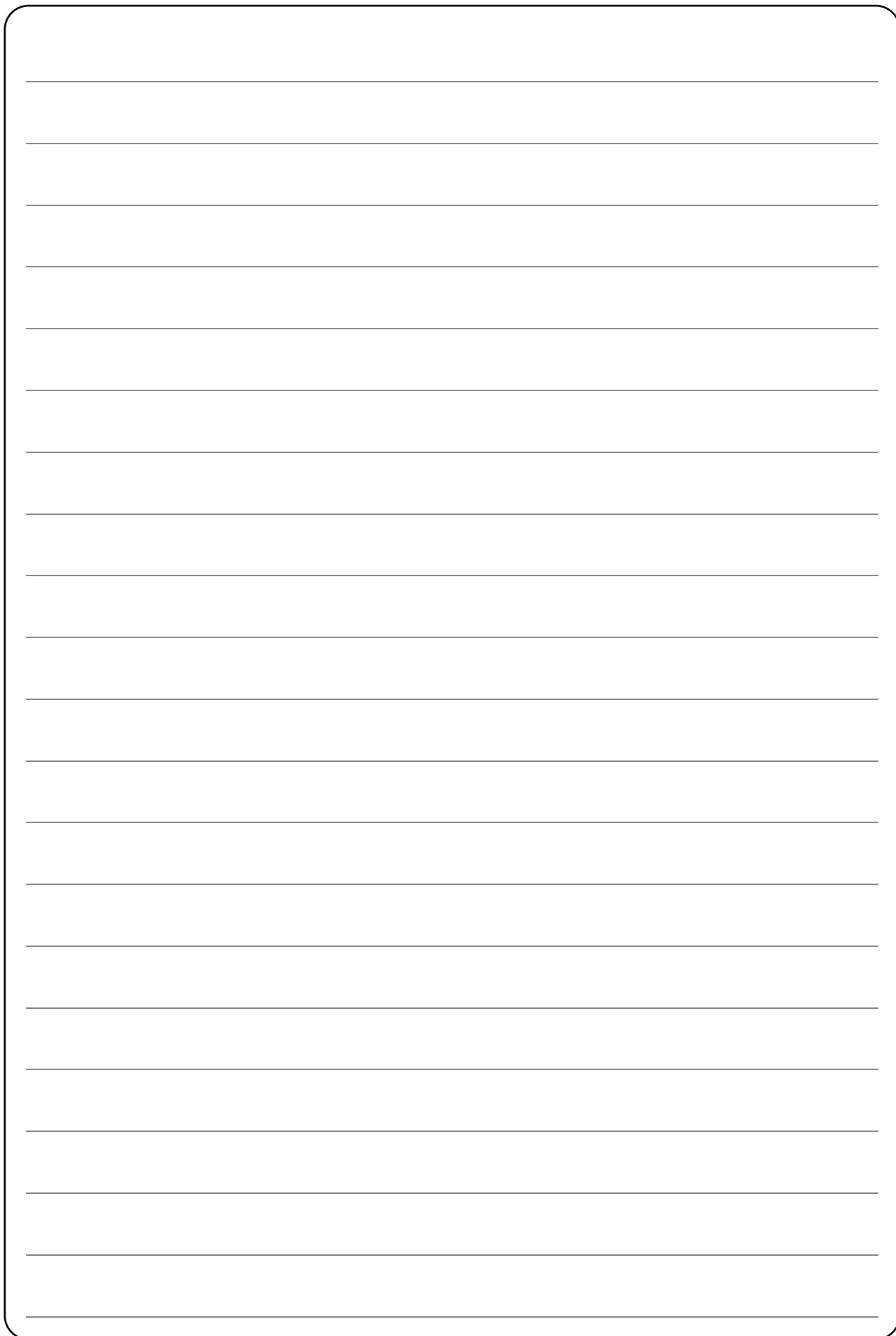
אם כן – הסבירו בקצרה באיזה אופן. אם לא – הסבירו בקצרה מדוע. הבהרה: השאלה אינה עוסקת בקמירות/קעירות.

תשובה והסבר קצר:

$$K_{\text{cr}}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$$

כן

מסגרת נוספת (יש לציין אם מדובר בטיוטה או בהמשך לתשובה אחרת):

A large rectangular box with rounded corners, containing 25 horizontal lines for writing. The lines are evenly spaced and extend across the width of the box. The box is intended for a student to provide a second answer or a clarification to the question above.

מסגרת נוספת (יש לציין אם מדובר בטיוטה או בהמשך לתשובה אחרת):

A large rectangular box with rounded corners, containing 25 horizontal lines for writing. The lines are evenly spaced and extend across the width of the box. The box is intended for a student to provide a second answer or clarification if needed.

מסגרת נוספת (יש לציין אם מדובר בטיוטה או בהמשך לתשובה אחרת):

A large rectangular box with rounded corners, containing 25 horizontal lines for writing. The lines are evenly spaced and extend across the width of the box. The box is intended for a student to provide a second answer or clarification to a question.