הטכניון – מכון טכנולוגי לישראל

הפקולטה למדעי המחשב

מבוא למערכות לומדות (236756)

סמסטר אביב תשפ"ג – 29 בדצמבר 2023

מרצה: ד"ר ניר רוזנפלד

מבחן מסכם מועד ב' – 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: 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 $h_{a,b}(x) = \underbrace{\mathbb{I}[a \le x \le b]}_{\text{פונקציית האינדיקטור}}$.

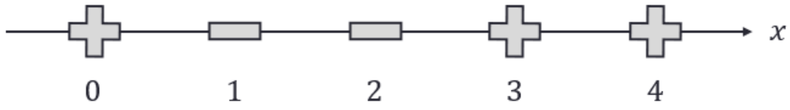
א. [7 נק'] מהו ה-VC-dimension של $\mathcal{H}$? $\text{VCdim}(\mathcal{H}) = \boxed{2}$.

הוכיחו את תשובתכם.

הוכחה:

כמו בהרצאה.

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

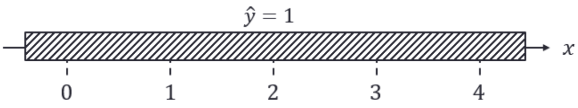


מריצים אלגוריתם AdaBoost על הדאטה הנתון, עם המחלקה $\mathcal{H}$ שהגדרנו בתור מחלקת בסיס (מסווגים חלשים).
בכל איטרציה לומדים מסווג חלש עם ERM על ההתפלגות הנוכחית. המסווג <u>החזק</u> הוא ה-ensemble הממושקל שמתקבל.

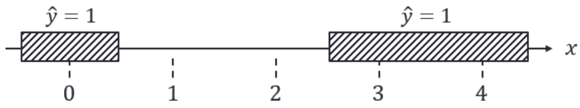
בשני הסעיפים הבאים מופיעים תרשימים של כללי החלטה על הישר $\mathbb{R}$. הכללים חוזים $\hat{y}=1$ רק במקטעים המקווקווים.
בכל סעיף, הקיפו <u>בבירור</u> את האות <u>היחידה</u> שמתאימה לתשובה הנכונה.

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

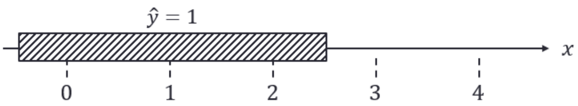
(a)



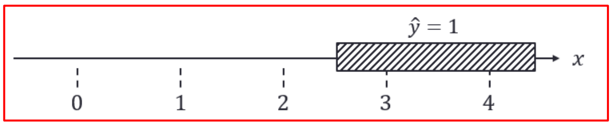
(b)



(c)

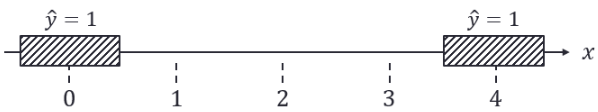


(d)

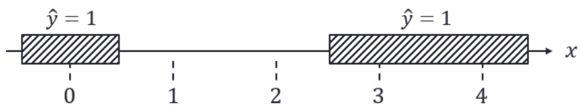


ג. [7 נק'] מבין הבאים – מה המסווג <u>החזק</u> (הממושקל) שמחזיר AdaBoost אחרי שתי איטרציות? <u>הסבירו</u> בקצרה.

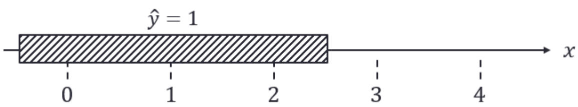
(a)



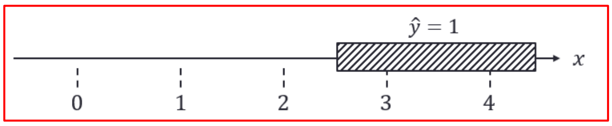
(b)



(c)



(d)



הסבר קצר:

מסווג שממשקל 2 היפותזות בינאריות תמיד פועל כמו היפותזה <u>יחידה</u> (זו עם המשקל הגבוה יותר).

ראינו תופעה כזו ב-demo בתרגול 10.

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

לרשותכם דפי טיוטה בסוף הגיליון.

i. $\mathbb{R} \ni \phi(x) = x - 1.5$

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

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

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

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

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

שאלה 3: רגרסיה ורגולריזציה [20 נק']

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

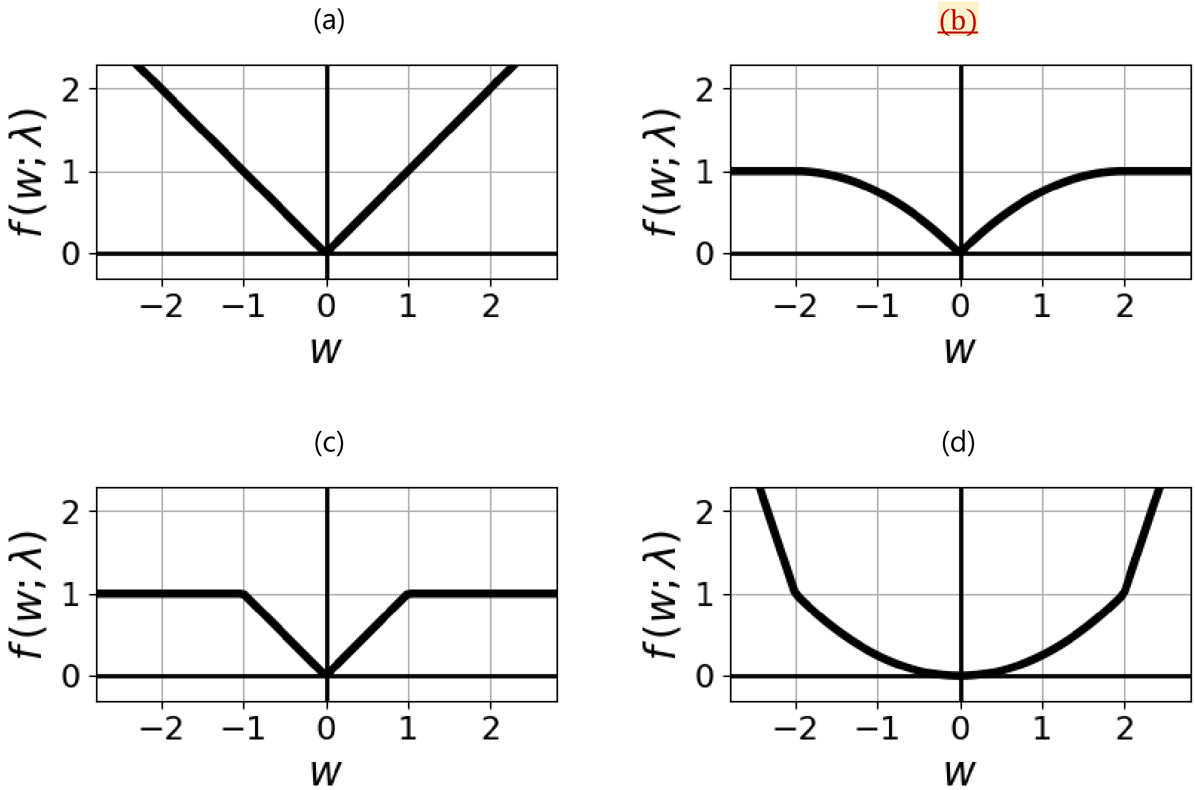
$$\min_{\mathbf{w}\in\mathbb{R}^d}\left(\frac{1}{2m}\|\mathbf{X}\mathbf{w}-\mathbf{y}\|_2^2+R(\mathbf{w};\lambda)\right).$$

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

כאשר משתמשים בפונקציה $R_{\ell 1}(\mathbf{w};\lambda)\triangleq\lambda\|\mathbf{w}\|_1$ מקבלים את בעיית ה-LASSO.

כעת נגדיר פונק' חדשה $R_{\mathrm{CP}}(\mathbf{w};\lambda)\triangleq\sum_{j=1}^{d}f(w_j;\lambda)$ עבור $f(w;\lambda)\triangleq\begin{cases}\lambda|w|-\frac{w^2}{4}, & |w|\leq 2\lambda\\ \lambda^2, & |w|>2\lambda\end{cases}$.

א. [3 נק'] עבור $\lambda=1$, הקיפו <u>בבירור</u> 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 <u>בבירור</u>.

i. קמורה
ii. קעורה
iii. תלוי בערך של λ
iv. *לא קמורה ולא קעורה*

מעתה, נסמן ב-$\hat{\mathbf{w}}_{\mathrm{LS}}$ את פיתרון ה-Least squares, ב-$\hat{\mathbf{w}}_{\ell 1}$ את פיתרון ה-LASSO וב-$\hat{\mathbf{w}}_{\mathrm{CP}}$ את פיתרון הרגרסיה תחת רגולריזציה של הפונק' R_{CP} שהגדרנו.

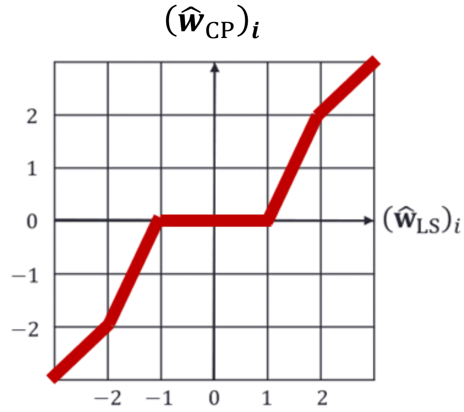
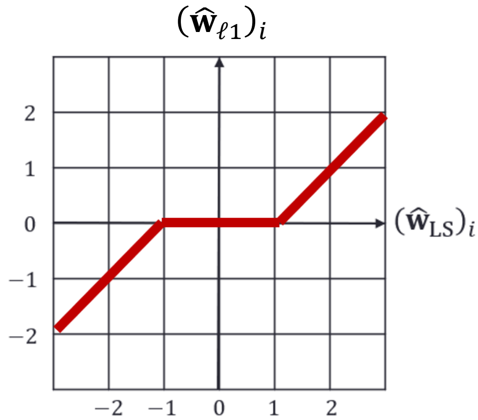
<u>הנחה:</u> מעתה נניח שהעמודות של $\mathbf{X}$ אורתוגונליות כך שמתקיים $\mathbf{X}^{\top}\mathbf{X} = m\mathbf{I}_{d\times d}$.

<u>נתונה טענה 1:</u> תחת ההנחה, מתקיים $(\hat{\mathbf{w}}_{\ell 1})_i = \begin{cases} \mathrm{sign}((\hat{\mathbf{w}}_{\mathrm{LS}})_i)\cdot(|(\hat{\mathbf{w}}_{\mathrm{LS}})_i| - \lambda), & |(\hat{\mathbf{w}}_{\mathrm{LS}})_i| > \lambda \\ 0, & |(\hat{\mathbf{w}}_{\mathrm{LS}})_i| \le \lambda \end{cases}$.

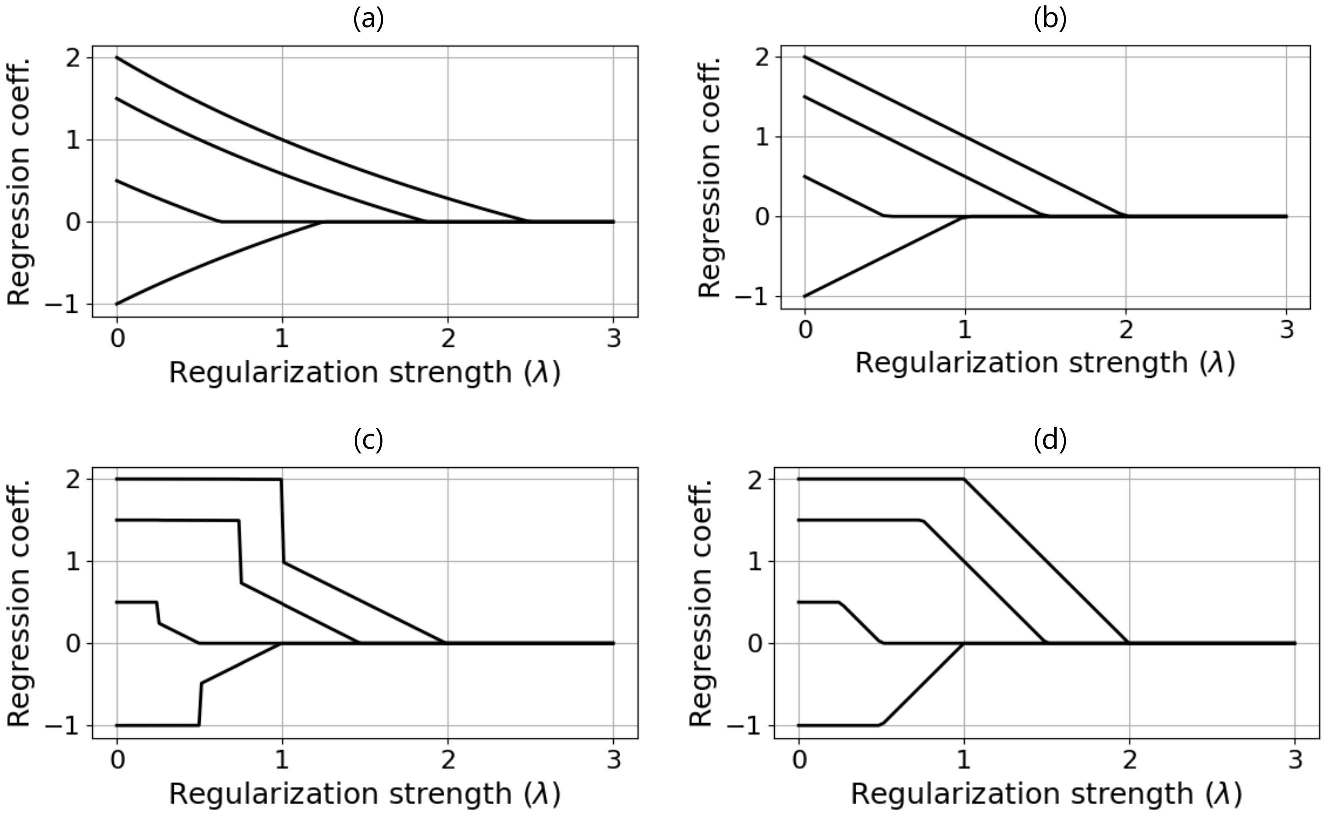
<u>נתונה טענה 2:</u> תחת ההנחה, מתקיים $(\hat{\mathbf{w}}_{\mathrm{CP}})_i = \begin{cases} (\hat{\mathbf{w}}_{\mathrm{LS}})_i, & |(\hat{\mathbf{w}}_{\mathrm{LS}})_i| \ge 2\lambda \\ 2\cdot\mathrm{sign}((\hat{\mathbf{w}}_{\mathrm{LS}})_i)\cdot(|(\hat{\mathbf{w}}_{\mathrm{LS}})_i| - \lambda), & \lambda < |(\hat{\mathbf{w}}_{\mathrm{LS}})_i| < 2\lambda \\ 0, & |(\hat{\mathbf{w}}_{\mathrm{LS}})_i| \le \lambda \end{cases}$.

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

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

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

כעת, פותרים בעיית רגרסיה בארבעה ממדים ($\mathbf{X} \in \mathbb{R}^{m\times 4}$) המקיימת את הנחת האורתוגונליות, משמע $\mathbf{X}^\top\mathbf{X} = m\mathbf{I}_{4\times 4}$.
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.

התרשים שמתאים למקדמים של $\hat{\mathbf{w}}_{\ell 1}$ הוא: (d) (c) (b) (a)

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

הסבר:

צריך להסתכל על הטענות כפונק' של λ בהינתן $(\hat{\boldsymbol{w}}_{\mathrm{LS}})_i$.

רואים מהנוסחאות, שמקדם מתאפס כאשר $\lambda = (\hat{\boldsymbol{w}}_{\mathrm{LS}})_i$. לכן למשל המקדם הגדול (2) צריך להתאפס כאשר $\lambda = 2$, ותרשים a נפסל.

לפי אותן נוסחאות, LASSO לינארי לגמרי בין 0 עד שהמקדם מתאפס (מתאים רק לתרשים b).

CP אמור להיות לינארי ורציף בין החלקים בהם הוא קבוע (מתאים לתרשים d).

שאלה 4: Support Vector Regression [30 נק']

שאלה זו עוסקת ברגרסיה 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.

$$\begin{aligned}&\underset{\mathbf{w}\in\mathbb{R}^d,\ b\in\mathbb{R}}{\operatorname{argmin}}\ \|\mathbf{w}\|_2^2\\ &\text{s.t.}\quad \mathbf{w}^\top\mathbf{x}_i + b \le y_i + \epsilon,\ \ \forall i \in [m]\\ &\qquad\ \ \mathbf{w}^\top\mathbf{x}_i + b \ge y_i - \epsilon,\ \ \forall i \in [m]\end{aligned}$$

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

תשובה והסבר קצר: רק עבור דאטה לינארי (ללא ϵ אין מרווח לשגיאה).

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 $\epsilon > 0$:

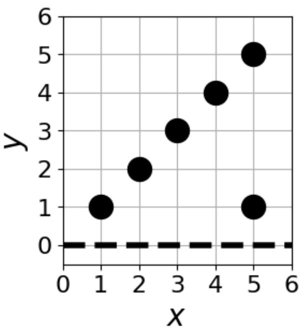
$$\begin{aligned}&\underset{\substack{\mathbf{w}\in\mathbb{R}^d,\ b\in\mathbb{R}\\ \forall i\in[m]:\ \xi_i,\xi_i^*\ge 0}}{\operatorname{argmin}}\ \|\mathbf{w}\|_2^2 + \sum_{i=1}^{m}(\xi_i + \xi_i^*)\\ &\text{s.t.}\quad \mathbf{w}^\top\mathbf{x}_i + b \le y_i + \epsilon + \xi_i,\ \ \forall i \in [m]\\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

בסעיפים הבאים נתונים תרשימים של דאטה חד-ממדי $(x_i, y_i \in \mathbb{R})$ וקווי רגרסיה שונים.

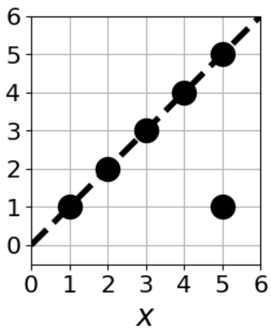
בכל סעיף כתוב ערך של ההיפר-פרמטר ϵ. הקיפו בבירור את האות שמתאימה לקו הרגרסיה שנלמד על ידי Soft-SVR.

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

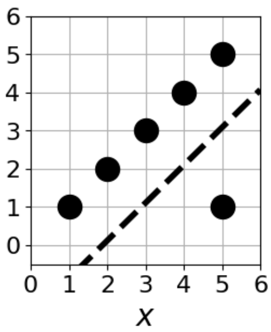
(a)



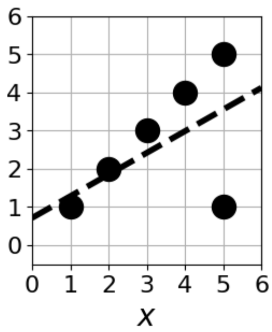
(b)



(c)

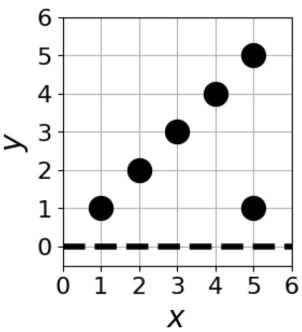


(d)

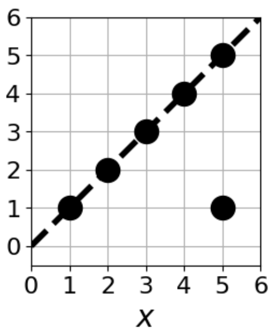


ג. [6 נק'] מהו קו הרגרסיה שנלמד כאשר $\epsilon \to 0$? (תשובה: שימו לב שמתקבל least absolute deviation).

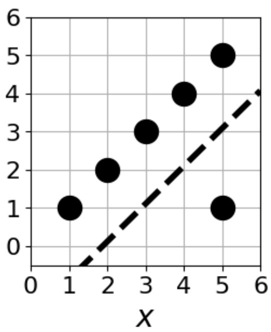
(a)



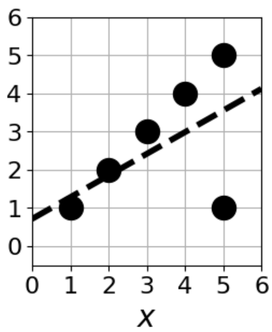
(b)



(c)



(d)



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

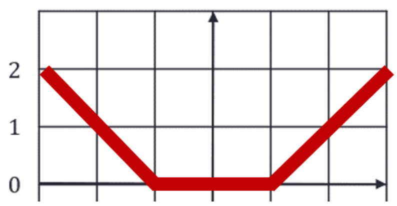
עשינו זאת ע"י הגדרה של $\ell_{\text{hinge}}(\mathbf{w}, b; \mathbf{x}_i, y_i) = \max\{0, 1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b)\}$ ופיתרון הבעיה הבאה:

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

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

תשובה והסבר קצר: רוצים להעניש רק על חריגה גדולה מדי מהערך הרצוי.

$$\ell(\boldsymbol{w}, b; \boldsymbol{x}_i, y_i) = \begin{cases} |y_i - (\mathbf{w}^\top \mathbf{x}_i + b)| - 1, & |y_i - (\mathbf{w}^\top \mathbf{x}_i + b)| > 1 \\ 0, & |y_i - (\mathbf{w}^\top \mathbf{x}_i + b)| \in [-1, 1] \end{cases}$$



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

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

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:

הבעיה מתאימה להפעלת הטריק, כי הדוגמאות מופיעות בה רק בתור מכפלה פנימית.

צריך להחליף את המכפלה $\mathbf{x}_i^\top \mathbf{x}_j$ בפונקציה $K(\mathbf{x}_i, \mathbf{x}_j)$.