הטכניון – מכון טכנולוגי לישראל

הפקולטה למדעי המחשב

מבוא למערכות לומדות (236756)

סמסטר חורף תשפ"ב – 07 במרץ 2022

מרצה: ד"ר יונתן בלינקוב

<u>מבחן מסכם מועד ב'</u>

הנחיות הבחינה:

- **משך הבחינה:** 3 שעות.
- **חומר עזר:** המבחן בחומר סגור (ללא ספרים, מחברות, דפי נוסחאות).
- אין צורך במחשבון.
- מותר לכתוב בעט **<u>בלבד</u>**.
- מותר לענות בעברית או באנגלית.
- יש לכתוב את התשובות **על גבי שאלון זה** בכתב יד קריא. תשובה בכתב יד לא קריא – לא תיבדק.
- במבחן 16 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.**

מבנה הבחינה:

- **חלק א' [76 נק']:** 4 שאלות פתוחות.
- **חלק ב' [24 נק']:** 4 שאלות סגורות (אמריקאיות) [כל אחת 6 נק'].

בהצלחה!

חלק א' – שאלות פתוחות [76 נק']

שאלה 1 – Linear regression & Optimization [14 נק']

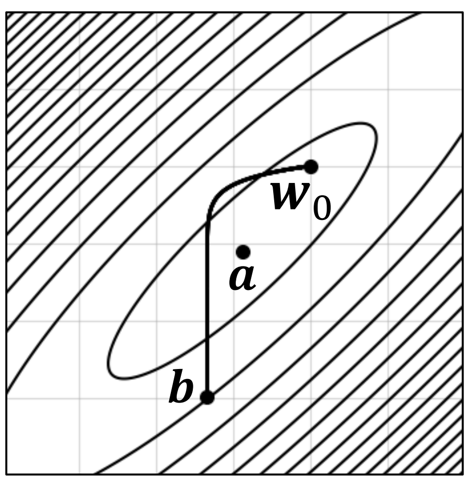
נתונה בעיית רגרסיה ליניארית דו-ממדית (בעיה בשני פרמטרים): $\operatorname{argmin}_{\boldsymbol{w}\in\mathbb{R}^2}\frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{w}^\top\boldsymbol{x}_i-y_i)^2$.

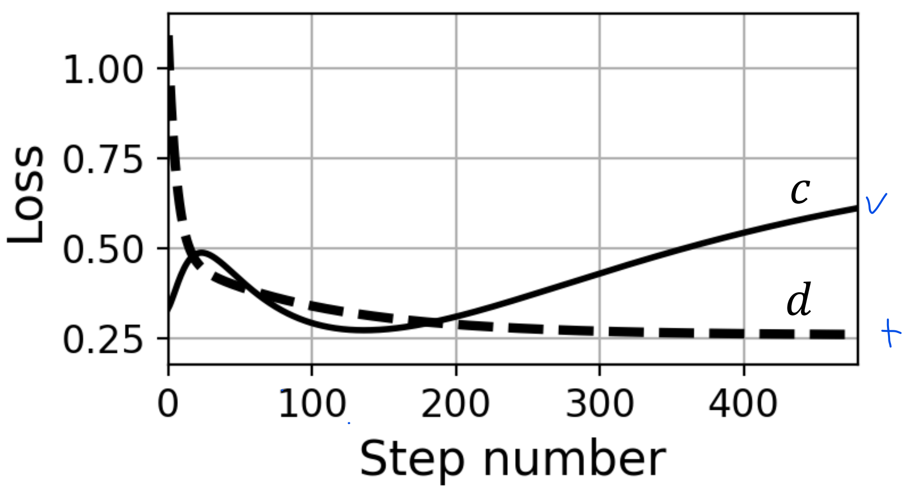
אוספים דאטה S ומחלקים אותו לסט אימון ולסט ואלידציה.

מתחילים מווקטור $\boldsymbol{w}_0=\boldsymbol{0}$ ופותרים את הבעיה (עבור סט האימון) בעזרת gradient descent (לא SGD) עם גודל צעד η.

בתרשים השמאלי: המסלול המלא שנוצר מאימון עם GD החל מ-$\boldsymbol{w}_0$ (המסלול מתואר ע"י עקומה במרחב $\mathbb{R}^2$, ומראה את כל הפתרונות $\boldsymbol{w}_0=\boldsymbol{0},\boldsymbol{w}_1,\boldsymbol{w}_2,\dots,\boldsymbol{w}_{480}=\boldsymbol{b}$). קווי המתאר (ה-level sets) מתארים loss landscape שהמינימום שלו הוא בנקודה a. עליכם להבין האם מדובר ב-loss על ה-training או ה-validation.

בתרשים הימני: מופיע גרף ההתכנסות המראה את ה-training loss ואת ה-validation loss.





א. [4 נק'] התאימו בין הפתרונות $\boldsymbol{a},\boldsymbol{b}$ שבתרשים השמאלי לבין הפיתרון האופטימלי על סט האימון $\boldsymbol{w}^*_{\text{train}}$ והפיתרון האופטימלי על סט הואלידציה $\boldsymbol{w}^*_{\text{val}}$. התאימו בין העקומות c,d לבין ה-training loss וה-validation loss.

מלאו את המקומות הריקים באותיות a,b,c,d כנדרש.

$\boldsymbol{w}^*_{\text{train}}$ matches $\underbrace{\text{b}}_{\text{השלימו}}$, $\boldsymbol{w}^*_{\text{val}}$ matches $\underbrace{\text{a}}_{\text{השלימו}}$,

training loss matches $\underbrace{\text{d}}_{\text{השלימו}}$, validation loss matches $\underbrace{\text{c}}_{\text{השלימו}}$.

שימו לב כיצד שני ה-losses (ובפרט ה-training loss) לא יורדים מתחת 0.25.

ב. [5 נק'] אילו מהפתרונות הבאים עשויים לשפר את ה-loss ה-training בסוף האימון (ביחס לתוצאות המוצגות לעיל)?

סמנו את כָּל התשובות המתאימות.

a. הוספת רגולריזציית ℓ^2.

~~b.~~ שימוש במדיניות early stopping (עצירת ה-GD לפני התכנסות, לפי קריטריון כלשהו).

~~c.~~ אימון עם SGD (עם batch_size=1) במקום GD, במשך מספר צעדים זהה (480).

d. מיפוי של שני ה-features המקוריים ל-feature mapping פולינומיאלי.

~~e.~~ סיבוב מערכת הצירים של ה-features המקוריים (ב-dataset S כולו) ב-$45°$ סביב ראשית הצירים ($\boldsymbol{w}_0 = \underline{\mathbf{0}}$).

ג. [5 נק'] אילו מהפתרונות הבאים עשויים לשפר את ה-loss ה-validation בסוף האימון (ביחס לתוצאות המוצגות לעיל)?

סמנו את כָּל התשובות המתאימות.

a. הוספת רגולריזציית ℓ^2.

b. שימוש במדיניות early stopping (עצירת ה-GD לפני התכנסות, לפי קריטריון כלשהו).

c. אימון עם SGD (עם batch_size=1) במקום GD, במשך מספר צעדים זהה (480).

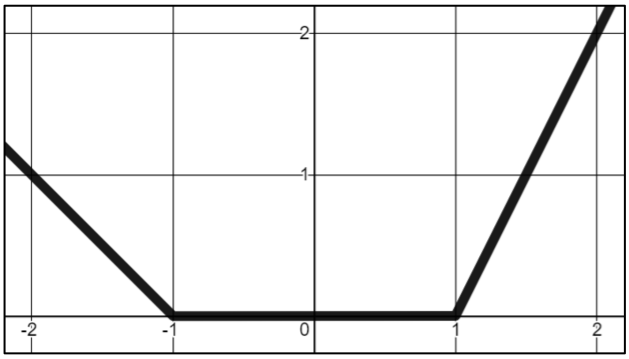
d. מיפוי של שני ה-features המקוריים ל-feature mapping פולינומיאלי.

e. סיבוב מערכת הצירים של ה-features המקוריים (ב-dataset S כולו) ב-$45°$ סביב ראשית הצירים ($\boldsymbol{w}_0 = \underline{\mathbf{0}}$).

שאלה 2 – Deep learning [20 נק']

נתונות שתי הפונקציות הרציפות $f,g:\mathbb{R}\to\mathbb{R}$ שבתרשימים הבאים:

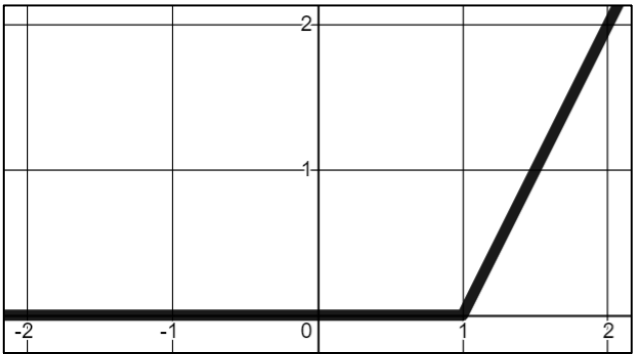
$f(x)$



בתחום $[-1,1]$ הפונקציה היא אפס.

בתחום $(1,\infty)$ השיפוע הוא $\mathbf{2}$ ובתחום $(-\infty,-1)$ הוא $\mathbf{-1}$.

$g(x)$



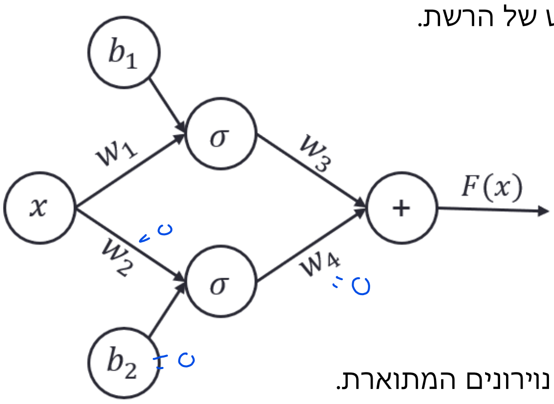
בתחום $(-\infty,\mathbf{1}]$ הפונקציה היא אפס.

בתחום $(\mathbf{1},\infty)$ השיפוע של $\boldsymbol{g}$ הוא $\mathbf{2}$.

נרצה ללמוד את הפונקציות f,g בעזרת רשת הנוירונים הבאה: $F(x)=w_3\cdot\sigma(w_1\cdot x+b_1)+w_4\cdot\sigma(w_2\cdot x+b_2)$.

$f(x) = 2\ \sigma(x-1)+\sigma(-x-1)$

כאשר $w_1,w_2,w_3,w_4,b_1,b_2\in\mathbb{R}$ הם פרמטרים סקלריים ו-$x\in\mathbb{R}$ הוא הקלט של הרשת.



פונקציית **האקטיבציה** σ יכולה להיות אחת מהשתיים:

1. סיגמואיד, משמע $\sigma(z)=\frac{1}{1+e^{-z}}$.
2. ReLU, משמע $\sigma(z)=\max\{0,z\}$.

$w_3\max\{0,w_1x+b_1\}$

בשני הסעיפים הבאים נראה שניתן לממש את הפונקציות f,g בעזרת רשת הנוירונים המתוארת.

א. [4 נק'] נבחר $w_2=b_2=w_4=0$. כתבו את ערכי w_1,w_3,b_1 ואת הבחירה של σ שמקיימים: $\forall x\in\mathbb{R}:\ F(x)=g(x)$.

תשובה סופית (לרשותכם דפי טיוטה בסוף הגיליון):

First layer: $w_1=$ ~~2~~ 1 (השלימו); $b_1=$ ~~-2~~ -1 (השלימו); **Second layer:** $w_3=$ ~~1~~ 2 (השלימו).

Activation: Sigmoid or ReLU (circle your choice).

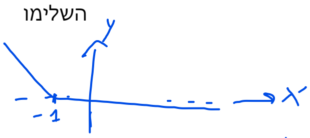
ב. [5 נק'] כתבו את ערכי w_1,w_2,w_3,w_4,b_1,b_2 ואת הבחירה של σ שמקיימים: $\forall x\in\mathbb{R}:\ F(x)=f(x)$.

תשובה סופית:

First layer: $w_1=$ 1 (השלימו); $w_2=$ -1 (השלימו); $b_1=$ -1 (השלימו); $b_2=$ -1 (השלימו);

Second layer: $w_3=$ 2 (השלימו); $w_4=$ 1 (השלימו).

Activation: Sigmoid or ReLU (circle your choice).



$w_4\max\{0,w_2x+b_2\}$

$\max\{0,-x-1\}$

הערה: סעיף ג' לא תלוי בסעיפים הקודמים.

ג. [4 נק'] הציעו פונקציית loss שמתאימה לבעיית הרגרסיה שהוגדרה, כך שמזעור של $\sum_{i=1}^{m} \ell(f(x_i),\ F(x_i))$ יביא ללמידת פרמטרים מתאימים שיקיימו $F(x) = f(x)$.

Answer: $\underbrace{\boldsymbol{\ell}(a,b)}_{\ell:\mathbb{R}\times\mathbb{R}\to\mathbb{R}_{\geq 0}} = \underbrace{(a-b)^2}_{\text{השלימו}}$.

הערה: סעיף ד' תלוי בסעיף ב' רק דרך הבחירה של פונקציית האקטיבציה.

$2\sigma(x-1)$

$w_3=2 \quad w_1=1 \quad b=-1$

בסעיף הבא נבחן את הקמירות של הבעיה שנוצרה.

$\sigma(z)=\max\{0,z\}$

תזכורת: הפונקציה $g: \mathbb{R} \to \mathbb{R}$ נקראת פונקציה <u>קמורה</u> אם מתקיים
$$\forall z_1, z_2 \in \mathbb{R},\ \ \forall t \in [0,1]: \quad tg(z_1) + (1-t)g(z_2) \geq g(t\, z_1 + (1-t) z_2)$$

ד. [7 נק'] בסעיף זה נניח שוב $w_2 = b_2 = w_4 = 0$.

הוכיחו / הפריכו: הפונקציה $\ell(a,\ F(x))$ קמורה ביחס לפרמטר w_1 (בהינתן כל בחירה של b_1, w_3, a, x).

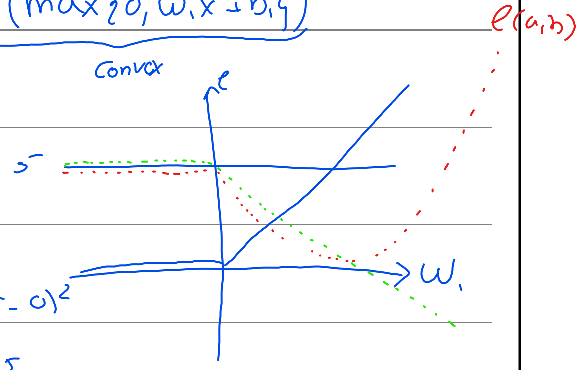
ניתן לענות לפי הגדרת הקמירות או לפי מאפיינים שלמדנו (אך יש לציין אותם במפורש).

<u>שימו לב</u>: עליכם להשתמש בבחירה של σ מסעיף ב' ובבחירה של ℓ מסעיף ג'. אין להציב ערכים מסעיף א'.

תשובה (לרשותכם דפי טיוטה בסוף הגיליון):

$\ell(a,b) = (a - w_3 \max\{0, w_1 x + b_1\})^2 =$

$= \underbrace{a^2}_{\text{const}} \underbrace{- 2a w_3 \underbrace{\max\{0, w_1 x + b_1\}}_{\text{convex}}}_{\text{not convex}} + w_3^2 \underbrace{(\max\{0, w_1 x + b_1\})^2}_{\text{convex}}$



for example: $a=5 \quad w_3=1 \quad b_1=0$

$x=1 \qquad (5-\max\{0,w_1\})^2$

$z_1=-5\ ;\ z_2=5\ ;\ t=\frac{1}{2}\ ;\ \frac{1}{2}(5-0)^2 + \frac{1}{2}(5-5)^2 \overset{?}{\geq} (5-0)^2$

$\frac{1}{2}(-5) + \frac{1}{2}5 = 0 \qquad 12.5 \overset{?}{\geq} 25$

$\Downarrow$

Not Convex

~~שאלה 3~~ – Naïve Bayes [18 נק']

בשאלה זו נראה ש-Gaussian Naïve Bayes הינו מסווג לינארי.

נתון דאטה $S=\{(\boldsymbol{x}_i,y_i)\}_{i=1}^{m}$ עם דוגמאות $\boldsymbol{x}_i\in\mathbb{R}^d$ ותיוגים $y_i\in\{0,1\}$.

לפי מידול Gaussian Naïve Bayes, נניח שההתפלגות המותנה של כל כניסה $k=1,\ldots,d$ הינה:

$$(\mathrm{X}[k]\mid Y=y)\sim\mathcal{N}(\mu_y[k],\sigma[k]^2),$$

כאשר $\mathrm{X}[k]$ הוא המשתנה המקרי המתאים לכניסה ה-k ב-x , המסומנת $x[k]$.

תזכורת: פונקציית הצפיפות של התפלגות גאוסיאנית $\mathcal{N}(\mu,\sigma^2)$ נתונה ע"י: $P(z)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right)$.

א. [4 נק'] הוכיחו שמתקיים: $P(X=x\mid Y=1)=\left(\prod_{k=1}^{d}\frac{1}{\sigma[k]\sqrt{2\pi}}\right)\exp\left(-\sum_{k=1}^{d}\frac{1}{2\sigma[k]^2}(x[k]-\mu_1[k])^2\right)$.

הוכחה <u>מנומקת</u>:

טענה (ללא הוכחה): בעזרת חוק בייס ונוסחת ההסתברות השלמה, ניתן להראות:

$$P(Y=1 \mid X=x) = \frac{1}{1+\frac{P(Y=0)P(X=x \mid Y=0)}{P(Y=1)P(X=x \mid Y=1)}}$$

טענה (ללא הוכחה): נסמן $p \triangleq P(Y=1)$. תחת הנחות השאלה ניתן להראות:

$$\frac{P(Y=0)P(X=x \mid Y=0)}{P(Y=1)P(X=x \mid Y=1)} = \frac{1-p}{p} \cdot \exp\left(\sum_{k=1}^{d}\left(\frac{\mu_0[k]-\mu_1[k]}{\sigma[k]^2}x[k] + \frac{\mu_1[k]^2-\mu_0[k]^2}{2\sigma[k]^2}\right)\right)$$

ב. [7 נק'] בעזרת האמור לעיל, הוכיחו שמתקיים: $P(Y=1 \mid X=x) = \frac{1}{1+\exp(\boldsymbol{w}^\top \boldsymbol{x}+b)}$.

בסוף ההוכחה, ציינו במפורש את ערכי $\boldsymbol{w} \in \mathbb{R}^d, \boldsymbol{b} \in \mathbb{R}$ המקיימים זאת.

הוכחה:

ג. [7 נק'] בהסתמך על האמור לעיל, הוכיחו כי מתקבל כלל החלטה ליניארי.

משמע, ההיפותזה $h(x) = \mathrm{argmax}_{y\in\{0,1\}} P(Y = y \mid X = x)$ מַשְׁרָה גבול החלטה ליניארי (decision boundary).

הוכחה תמציתית:

שאלה 4 – SVM [24 נק']

היזכרו בבעיות ה-SVM במקרה ההומוגני (נניח שמתקיים $\lambda = 1$ בבעיה ה-Soft):

Soft SVM

$$\underset{\boldsymbol{w}\in\mathbb{R}^d}{\operatorname{argmin}} \left(\frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y_i \cdot \boldsymbol{w}^\top \boldsymbol{x}_i\} + \|\boldsymbol{w}\|_2^2\right)$$

Hard SVM

$$\underset{\boldsymbol{w}\in\mathbb{R}^d}{\operatorname{argmin}} \|\boldsymbol{w}\|_2^2 \quad \text{s.t. } y_i \cdot \boldsymbol{w}^\top \boldsymbol{x}_i \geq 1, \ \forall i \in [m]$$

נתון דאטה d-ממדי $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}$ עם סיווגים בינאריים (± 1). ידוע שהדאטה פריד ליניארית ע"י מפריד הומוגני.
צוות מחקר פתר את שתי בעיות האופטימיזציה שלמעלה, וקיבל את $\boldsymbol{w}_{\text{hard}}, \boldsymbol{w}_{\text{soft}} \in \mathbb{R}^d$ כפתרונות.

א. [8 נק'] האם ניתן לומר שאחד מהמקרים $\|\boldsymbol{w}_{\text{hard}}\|_2 \leq \|\boldsymbol{w}_{\text{soft}}\|_2$ או $\|\boldsymbol{w}_{\text{hard}}\|_2 \geq \|\boldsymbol{w}_{\text{soft}}\|_2$ מתקיים בהכרח?
אם כן, איזה מהם? בכל מקרה – הסבירו בקצרה.

הסבר תמציתי:

$\|w_{hard}\|_2 \geq \|w_{soft}\|_2$: כי hard המטרה היא למזער את הנורמה
וכתוצאה אנו מוצאות את ה w שמפריד את הדאטא
עם הנורמה הקטנה ביותר
לעומת זאת, soft מנסה למזער את הסכום של הנורמה עם ה
מחיר של ה margin violation של הסיווגים הלא נכונים, ולכן היא עלולה
לצאת פתרון עם נורמה גבוהה יותר אבל עם פחות סיווגים לא נכונים
ובמקרה הכי גרוע, הם ימצאו את ה w של ה hard ולכן

ב. [8 נק'] נוסף עוד feature ממקור לא ידוע. הצוות פתר את בעיית ה-Hard-SVM המעודכנת וקיבל את $\boldsymbol{w}'_{\text{hard}} \in \mathbb{R}^{d+1}$.
האם ניתן לומר שאחד מהמקרים $\|\boldsymbol{w}'_{\text{hard}}\|_2 \le \|\boldsymbol{w}_{\text{hard}}\|_2$ או $\|\boldsymbol{w}'_{\text{hard}}\|_2 \ge \|\boldsymbol{w}_{\text{hard}}\|_2$ מתקיים בהכרח?
אם כן, איזה מהם? בכל מקרה – הסבירו בקצרה.

הסבר תמציתי:

מכיוון שהפיצ'ר יכול רק לעזור ולא לפגוע בנו
במידה והפיצ'ר מיותר, לא נשתמש בו ונשים 0 במקומו בוקטור ש
נקבל $\|w_{hard}\| = \|w'_{hard}\|$ אם במידה והפיצ'ר לא פוגע – הוא יכול
רק לעזור לנו למצוא פתרון נורמה קטנה יותר (בעיה המציאת שניתן
לפתור בדרך לא אנליטית ויעילה יותר)

טוב ברק: בכ"מ $\|w'\| \le \|w\|$ ונראה כי $\|w\| = \|w'\|$ אם נבחר ואת הפיצ'ר
היומחן לוקטור 0 ונקבל פתירה.

ג. [8 נק'] יהי $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}$ סט אימון כלשהו (לאו דווקא פריד ליניארית) עבורו $\boldsymbol{x}_i \in \mathbb{R}^d, y_i \in \{\pm 1\}$
ויהי $\boldsymbol{w}_{\text{soft}} \in \mathbb{R}^d \setminus \{\underline{\mathbf{0}}\}$ הפתרון של Soft-SVM על סט זה.

עליכם להוכיח **<u>אחת</u>** מבין שתי הטענות הבאות (השנייה מזכה בניקוד חלקי <u>בלבד</u>):

(i) [8 מתוך 8 נק'] לכל סט כזה ניתן להוסיף feature חדש (בלי לשנות את ה-features האחרים ואת התיוגים הנתונים),
כך שמתקיים $\|\boldsymbol{w}_{\text{soft}}\|_2 > \|\boldsymbol{w}'_{\text{soft}}\|_2$, כאשר $\boldsymbol{w}'_{\text{soft}} \in \mathbb{R}^{d+1}$ אופטימלי עבור הבעיה המעודכנת (עם ה-feature החדש).

<u>או</u>

(ii) [4 מתוך 8 נק'] לכל סט כזה ניתן להוסיף feature חדש (בלי לשנות את ה-features האחרים ואת התיוגים הנתונים),
כך שקיים $\boldsymbol{w}' \in \mathbb{R}^{d+1}$ עבורו $\|\boldsymbol{w}_{\text{soft}}\|_2 > \|\boldsymbol{w}'\|_2$ וגם $\mathcal{L}(\boldsymbol{w}_{\text{soft}}) \geq \mathcal{L}'(\boldsymbol{w}')$, כאשר מגדירים

$$\mathcal{L}(\boldsymbol{w}) \triangleq \frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y_i \cdot \boldsymbol{w}^\top \boldsymbol{x}_i\}, \quad \mathcal{L}'(\boldsymbol{w}) \triangleq \frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y_i \cdot \boldsymbol{w}^\top \underbrace{\boldsymbol{x}'_i}_{\substack{\text{הדוגמאות המעודכנות} \\ \text{עם ה-feature החדש}}}\}.$$

הוכחה (יש לציין איזו טענה מוכיחים):

חלק ב' – שאלות רב-ברירה [24 נק']

בשאלות הבאות סמנו את התשובות המתאימות (לפי ההוראות). בחלק זה אין צורך לכתוב הסברים.

א. [6 נק'] היזכרו בבעיות רגרסיה ליניארית עם Least squares (LS), וסמנו את <u>כָּל</u> התשובות הנכונות.

שימו לב: בסעיף זה, הגזירוּת מתייחסת להגדרת ה-gradients <u>המסורתית</u>, ולא ל-subgradients.

a. כאשר פותרים LS עם הפיצ'רים המקוריים, הבעיה קמורה ביחס ל-$\boldsymbol{w}$.

b. כאשר פותרים LS עם feature mapping שנקבע מראש (למשל פולינומיאלי), הבעיה קמורה ביחס ל-$\boldsymbol{w}$.

c. Ridge regression (ℓ^2) היא בעיה קמורה אך לא גזירה ביחס ל-$\boldsymbol{w}$.

d. Lasso (ℓ^1) היא בעיה קמורה אך לא גזירה ביחס ל-$\boldsymbol{w}$.

ב. [6 נק'] הטענות הבאות עוסקות במודלים מסוג Linear Soft SVM, Perceptron, and Logistic Regression.

סמנו את <u>כָּל</u> הטענות הנכונות.

a. Logistic Regression יכול ללמוד גם מפרידים לא ליניאריים בגלל שה-sigmoid מכניס non-linearity.

b. ניתן להכליל Logistic Regression לבעיות multiclass בעזרת פונקציית Softmax.

c. כל עוד הדאטה פריד ליניארית, Soft SVM ופרספטרון מחזירים את אותו המפריד.

d. בשלושת האלגוריתמים ניתן להשתמש ב-feature mapping כדי ללמוד מפרידים לא ליניאריים.

e. פרספטרון לומד בעזרת GD (לא stochastic), ואילו Soft SVM יש ללמוד באמצעות SGD.

ג. [6 נק'] נגדיר את פונקציית ה-squared loss הבאה: $\mathcal{L}(z)=(1-z)^2$.

$-2(1-z)$

2

$(1-y_i \cdot w^T x_i)^2$

סמנו את <u>כָּל</u> הטענות הנכונות ביחס לפונקציה זו.

a. הפונקציה קמורה (convex) ביחס ל-z.

b. הנגזרת של הפונקציה היא $\frac{\partial}{\partial z}\mathcal{L}=2z-2$.

c. עבור בעיות <u>סיווג</u> ליניארי (ה-loss) מחושב ע"י $z=y_i\boldsymbol{w}^\top\boldsymbol{x}_i$ והפרדיקציות ניתנות ע"י $(h(\boldsymbol{x}_i)=\operatorname{sgn}(\boldsymbol{w}^\top\boldsymbol{x}_i))$, כאשר ה-training loss הוא 0, גם ה-training error הוא 0.

d. עבור בעיות <u>סיווג</u> ליניארי (ה-loss) מחושב ע"י $z=y_i\boldsymbol{w}^\top\boldsymbol{x}_i$ והפרדיקציות ניתנות ע"י $(h(\boldsymbol{x}_i)=\operatorname{sgn}(\boldsymbol{w}^\top\boldsymbol{x}_i))$, כאשר ה-training error הוא 0, גם ה-training loss הוא 0.

ד. [6 נק'] נגדיר מחלקת היפותזות שמכלילה את המסווגים שמחזיר Adaboost לאחר T צעדים עם מחלקת היפותזות $\mathcal{B}$ בתור מסווגי בסיס. משמע:

$$\mathcal{H}_{\mathcal{B},T} = \left\{ h_{\text{strong}}(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \,\middle|\, \boldsymbol{\alpha} \in \mathbb{R}^T,\ h_t \in \mathcal{B} \right\}$$

לפניכם מספר טענות על ההשפעה של T ו-$\text{VCdim}(\mathcal{B})$ על $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$.

בחרו בטענה היחידה הנכונה (השאלה אינה עוסקת במקרי קצה אלא במקרה הסביר).

a. T גָּדֵל $\Leftarrow$ $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$ גָּדֵל. $\text{VCdim}(\mathcal{B})$ גָּדֵל $\Leftarrow$ $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$ גָּדֵל.

b. T ~~קָטֵן~~ $\Leftarrow$ $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$ גָּדֵל. $\text{VCdim}(\mathcal{B})$ גָּדֵל $\Leftarrow$ $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$ גָּדֵל.

c. T גָּדֵל $\Leftarrow$ $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$ גָּדֵל. $\text{VCdim}(\mathcal{B})$ ~~קָטֵן~~ $\Leftarrow$ $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$ גָּדֵל.

d. T קָטֵן $\Leftarrow$ $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$ גָּדֵל. $\text{VCdim}(\mathcal{B})$ ~~קָטֵן~~ $\Leftarrow$ $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$ גָּדֵל.

e. רק $\text{VCdim}(\mathcal{B})$ משפיע על $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$.

f. רק T משפיע על $\text{VCdim}(\mathcal{H}_{\mathcal{B},T})$.

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

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

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