



מבוא למערכות לומדות (236756)

סמינר חורף תשפ"ד – 12 באפריל 2024

מרצה: ד"ר יונתן בלינקוב

מבחן מסכם מועד א'

הנחיות הבחינה:

- **משך הבחינה:** שלוש שעות.
- **חומר עזר:** המבחן בחומר סגור (לא ספרים, מחברות, דפי נסחאות).
- מחשבון: מותר.
- כל כתיבה: עט בלבד.
- יש לכתוב את התשובות על גבי שאלון זה.
- מותר לענות בעברית או באנגלית.
- הוכחות והפרוכות צריכות להיות פורמליות.
- קרייאות:
- סימונים לא ברורים בשאלות רב-ברירה ו/או תשבות מילוליות בכתב יד לא קראי יובילו לפסילת התשובה.
- לא יתקבלו ערעורים בנושא.
- בבחן 14 עמודים ממוספרים סה"כ, כולל עמוד שער זה שמספרו 1 ושלושה עמודי טיווח בסוף הגיליון.
- נא לכתוב רק את המבוקש ולצף הסבירים קיצרים עפ"י ההנחיות.
- **בתום המבחן יש להגיש את שאלון זה בלבד.**
- **לזכאים להערכת חלופית מתאפשרת בחירה בין שאלות 3 ו-4.**

בהצלחה!

שאלה 12] Maximum Likelihood Estimation : 1 נק'

נתונות m דגימות חד-ממדיות $\{x_1, x_2, \dots, x_m\}$ עבור $p \in (0,1)$ שנדגמו. $X \sim \text{Geo}(p)$ משתנה מקרי (הסתברות מטבע) בסיכוי p עד להשגת "הצלחה" אחת.

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

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

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

משמעותו של \hat{p} הוא פיתוח ביטוי עבור $\Pr(x_1, x_2, \dots, x_m; p)$.

$$\begin{aligned} \ln L(S; p) &= \ln \Pr(x_1, x_2, \dots, x_m; p) = \ln \prod_{i=1}^m \Pr(x_i; p) = \ln \left(\prod_{i=1}^m (1-p)^{x_i-1} p \right) = \text{תשובה:} \\ &= \ln \sum_{i=1}^m \ln(\Pr(x_i; p)) = \ln \sum_{i=1}^m \ln((1-p)^{x_i-1} p) = \ln \sum_{i=1}^m ((x_i-1) \ln(1-p) + \ln(p)) \\ \frac{\partial}{\partial p} L(S; p) &= \frac{m}{p} + \sum_{i=1}^m \frac{x_i-1}{1-p} = \frac{m}{p} + \frac{m}{1-p} + \sum_{i=1}^m \frac{x_i}{1-p} = 0 \Rightarrow \frac{\frac{m(1-p)+mp}{p(1-p)}}{\cancel{p(1-p)}} = \frac{\sum x_i}{\cancel{p(1-p)}} \\ \Rightarrow \frac{1}{p} &= \frac{1}{m} \sum x_i \Rightarrow \bar{x} = \frac{1}{p} \Rightarrow p = \frac{1}{\bar{x}} \end{aligned}$$

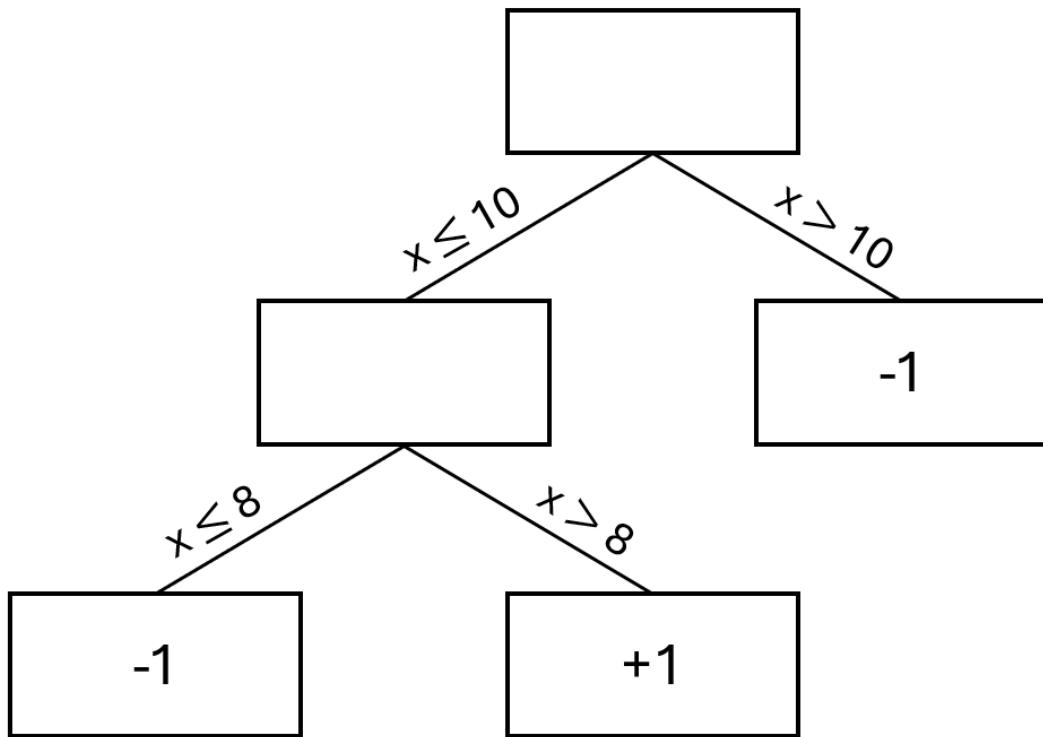
שאלה 2 : VC-dimension of Decision Trees [4 נק']

- א. [4 נק'] להלן הגדירה של "ניתוץ". השמננו מההגדירה את הטעותים.
השלימו את שלושת הטענות החסרים. בכל מקום כתבו בבירור האם חסר בהגדירה A או E.

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

מעתה והלאה בשאלת זו מרחיב הדוגמאות הוא $\mathcal{X} = \mathbb{R}$ ומרחב התוצאות הוא $\mathcal{Y} = \{-1, +1\}$.
מחלקת ההיפותזות \mathcal{H} עליה נשאל הינה עצי החלטה מעומק $1 \leq L$.

תשובות: עצי ההחלטה בקורס הינם עצים בינאריים בהם הצמתים הפנימיים מתפצלים לפי thresholds. העומק של העץ לא כולל את השורש אבל כן את העליים. לדוגמה, להלן מוצג עץ החלטה "חוקי" מעומק 2.



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

הוכחה פורמלית: גורם H מוגדר הולistically כפונקציית גוף הפליגים כפואים

$$x_1 < x_2 < \dots < x_{2^k} \quad \text{e} \quad \text{os} \quad x_1, \dots, x_{2^k} \quad (11)$$

କେବଳ ଏହାରେ ମାତ୍ରାକୁ ପରିଚାରିତ କରିବାକୁ ପାଇଁ ଏହାରେ ଯଦିଏହାରେ ଏହାରେ ଏହାରେ

$(y_1, \dots, y_{2^{k-1}} - \text{Left} : \text{בנוסף ל } y_{2^k}, \dots, y_{2^n})$ ו- $(y_{2^k+1}, \dots, y_{2^n} - \text{Right}$

1. **לעכידת מושג** – מושג אחד או יותר מושכלים על ידי מושג אחר.

1. **תְּמִימָה** (תְּמִימָה) **תְּמִימָה** (תְּמִימָה) **תְּמִימָה** (תְּמִימָה) **תְּמִימָה** (תְּמִימָה)

ת-+---+-+--- 100% מילוי סוף סוף כ- 2%

לפיכך $h(x) = y$ אם ורק אם $x \in h^{-1}(y)$



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

הוכחה פורמלית: $\lim_{n \rightarrow \infty} c_n = c$ מוכיחו ה' $c = \lim_{n \rightarrow \infty} c_n$

$$x_1 < x_2 < \dots < x_{e+1} \quad \forall x_i \in C: \quad h(x_i) = y_i \quad : e \geq 0 \quad h \in H \quad \text{and} \quad \forall i \in I$$

$$\left\{ \begin{array}{l} y_i=1 \mid i:1,2 = 1 \\ y_i=-1 \mid i:1,2 = 0 \end{array} \right\} \quad \text{רְשָׁאָת: ۷۷۷/۱۵۱۲} \quad \text{בָּזֶל}$$

איך ניתן לסייע לילדי מוגבלים?

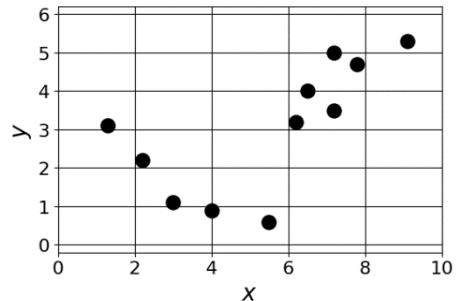
לפיכך $h(x_i) = h(x_{i+1}) = -1$

$V(\text{dom}(H)) \leq 2^{\aleph_0}$ \Leftrightarrow $\exists S \subseteq \mathcal{P}(\omega) \text{ such that } H \text{ is well-founded and } \forall x \in S \exists y \in S \text{ such that } x \neq y \text{ and } x \in y$

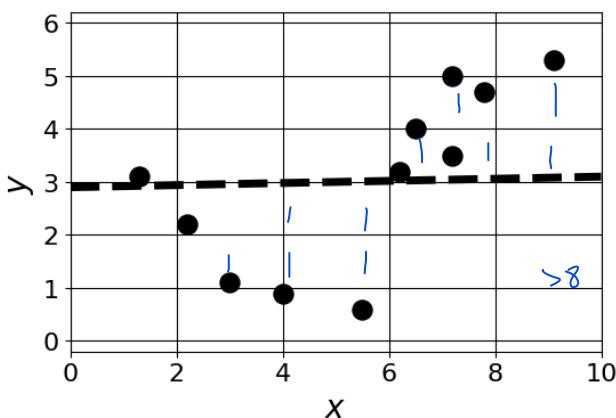
$$V(\dim(H)) \leq 2^k$$

שאלה 3 : 32] Linear Regression [ק']

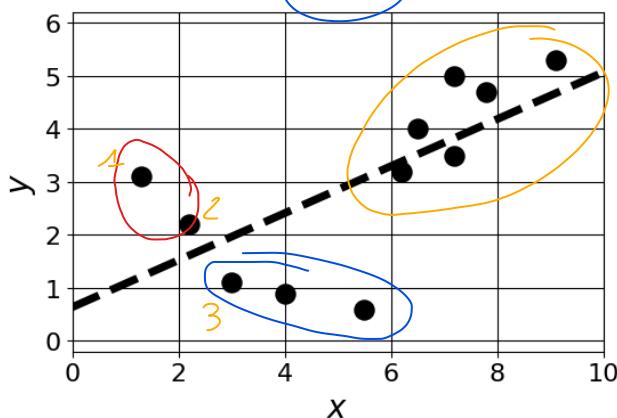
לזקאים להערכתה חלופית **בלבד** (כפי שהוגדרו באתר הקורס): סמננו את התيبة הזו אם ברצונכם **לדלג** על שאלת זו. המשקל של שאלות 1,2,4 יתפזר באופן ייחסי על פני 100 נקודות. ניתן לדלג רק על שאלת אחת מתוך שאלות 3,4.



(A)



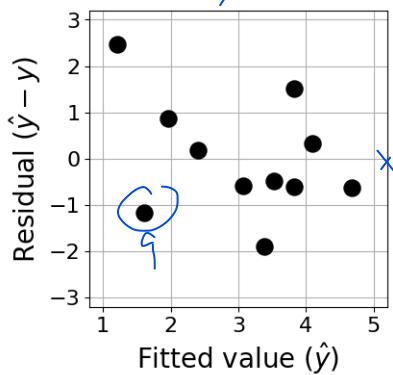
(B)



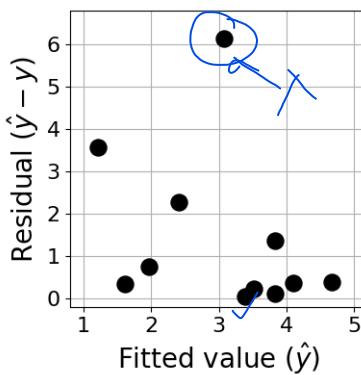
ב. [נק'] להלן תרגימי Residual analysis מהסוג שהוצג בהרצאה (שים לב לציר ה- y השוניים).

הקיפו את האות של התרשים היחיד שמתאים לנוטנים ולקו הרגסיה מהסעיף הקודם.

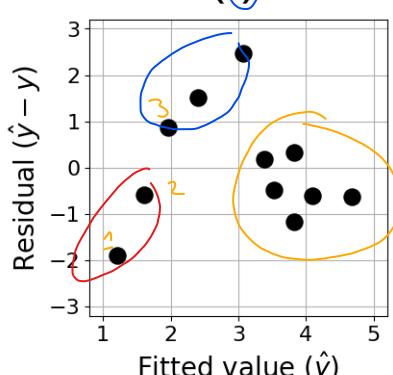
(A)



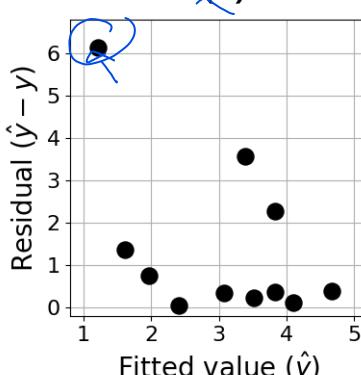
~~(B)~~



(c)



~~(D)~~

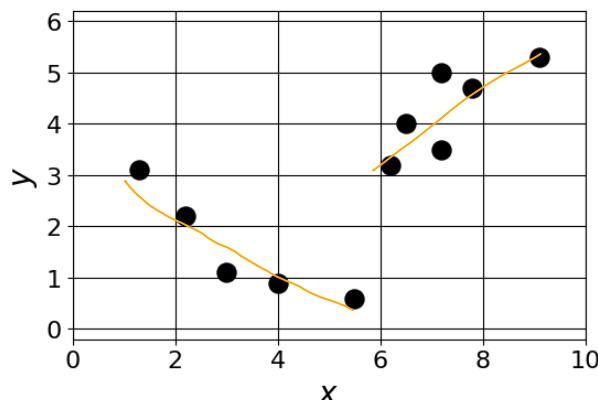


הַיְלָדִים

נגדיר מחלוקת Thresholded linear regression המאפשרת התאמה של שני קווים וגרסיה נפרדים:

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

כמו במקרה הילינרי, נגדיר היפותזה אופטימלית ב- \mathcal{H}_1 בתור היפותזה שمبיאה למינימום את MSE על הדאטה.



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

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

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

$\tau_{mn} = x_{mn} - 1$ $\tau_{mn} = x_{mn} + 1$	$T = \{\tau_m, \tau_1, \dots, \tau_n, \tau_{max}\}$ תשובה: 1.1
פונקציית MSE מינימלית כפולה, כנראה מינימום כפוי	
פונקציית MSE מינימלית כפולה, כנראה מינימום כפוי	
(בג') פונקציית MSE מינימלית כפולה, כנראה מינימום כפוי	

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

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

$$h_{w_1, b_1, w_2, b_2, \gamma} = \underbrace{w_1 x + b_1}_{w_2 x + b_2} + \underbrace{\gamma}_{\gamma \in \mathbb{T}}$$

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

הוכחה: $\exists h \in \mathcal{H}_2$

$$h_{u_1, c_1 u_2, c_2}(x) = u_1 \max\{0, x - c_1\} + u_2 x + c_2$$

$$\begin{aligned} x \geq c_1: h(x) &= u_1(x - c_1) + u_2 x + c_2 = (\underbrace{u_1 + u_2}_{u}) x + \underbrace{c_2 - u_1 c_1}_{c} = u x + c = w_1 x + b, \\ x < c_1: h(x) &= u_1 \cdot 0 + u_2 x + c_2 = u_2 x + c_2 = w_2 x + b_2 \end{aligned}$$

$$h_{u_1, c_1 u_2, c_2}(x) = h_{w_1, b_1, w_2, b_2, \gamma}(x) \in \mathcal{H}_1$$

$$\mathcal{H}_2 \subseteq \mathcal{H}_1$$

ר' 18

ג. [8 נק'] נתון דאטה חד-ממדי עם תיוגים רציפים ($x_i \in \mathbb{R}$, $y_i \in \mathbb{R}$) מחולק באופן אקריא 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-MSE).

כל שורה בטבלה מתארת תוצאות מתהליך הלמידה המתואר על דאטה נתון כלשהו.

חלק מהשורות מתארות תוצאות אפשריות וחלק מתארות תוצאות בלתי אפשריות.

כל אחת מרבע השורות, סמן אם היא אפשרית או לא.

$\mathcal{H}_2 \subseteq \mathcal{H}_1$

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	X 15	37	כן / לא
17	35	42	51	כן / לא
0	11	0	9	כן / לא

Overfitting
on \mathcal{H}_1 , \mathcal{H}_2 is better

\mathcal{H}_1 more complex

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

לזקאים להערכת חלופית בלבד (כפי שהוגדרו באתר הקורס): סמן את התيبة הזו אם ברצונכם **לדלג** על שאלה זו. המشكل של שאלות 1,2,3 יתפזר באופן ייחסי על פני 100 נקודות. ניתן לדלג רק על שאלה אחת מתוך שאלות 4,3.

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

```

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              y_i = sign(w^T x_i)      y_i.w^T x_i < 0 ? מהר
9
10         if y_i != y_i:   y_i.w^T x_i < 1
11             w = w + y_i x_i
12             errorFound = True
13
14         if not errorFound:
15             break

```

ראינו בתרגול שפרספטרון שקל ל-SGD על $\sum_i \max\{0, -y_i \mathbf{w}^\top \mathbf{x}_i\}$. בעת נראה תכונה דומה ל- Margin Perceptron שקיים שפה נוקציה $\ell(\mathbf{w}; \mathbf{x}_i, y_i) = \max\{0, 1 - y_i \mathbf{w}^\top \mathbf{x}_i\}$. נזכיר שהפה נוקציה $\ell(\mathbf{w}; \mathbf{x}_i, y_i) = \sum_i \ell(\mathbf{w}; \mathbf{x}_i, y_i)$ קמורה ב- \mathbf{w} .

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

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

ב. [9 נק'] הצעו שינוי נקודתי לאלגוריתם הפרספטרון (ציינו את מספרי השורות שמתעדכנות ואת השינויים הנדרשים) כך שהאלגוריתם המעודכן יהיה שקל ל-SGD על $\ell(\mathbf{w})$ שהגדכנו. יש להסביר בקצרה כיצד השיקולות מתקיימת.

תשובה: $\ell(\mathbf{w}; \mathbf{x}_i, y_i) = \max(0, 1 - y_i \mathbf{w}^\top \mathbf{x}_i)$ $\ell(\mathbf{w}; \mathbf{x}_i, y_i) = \max(0, 1 - y_i \mathbf{w}^\top \mathbf{x}_i)$ $\text{margin} \sim$	

ללא תלות בעיפויות הקודמים, נחזר לעסוק באלגוריתם הפרסptrון הרגיל כפי ש谟יע בתחילת השאלה.

משפט התחנשות של הפרספטרון:

זה דאטה פריד לינארית הומוגנית $\{(\mathbf{x}_i, y_i)\}_{i=1}^m$. נסמן את הנורמה המרבית של דוגמה $y_i \mathbf{x}_i$ ב- R (סקלר).
 $\mathbf{w}_{\text{SVM}} \triangleq \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^d} \|\mathbf{w}\|_2 \text{ s.t. } \forall i: y_i \mathbf{w}^\top \mathbf{x}_i \geq 1$,
 $\|\mathbf{w}_{\text{SVM}}\|_2 = R$ טיעיות חייזרי בזמן האימון שלו.

ג. [9 נק'] נתון דатаה $S = \{(x_i, y_i)\}_{i=1}^{100}$ פריך לינארית הומוגנית ש"הבטחת" המשפט עבورو הינה $10 = 10$.
 יוצרים דטהה חדש ע"י הכפלת כל הדוגמאות ב-2, משמע $S' = \{(2x_i, y_i)\}_{i=1}^{100}$
 מהי הבטחת המשפט לדטהה החדש? משמע, כמה טעויות חיזוי (לכל היותר) יעשה אלגוריתם הפרספטורן על הדטהה
 החדש? כתבו ערך מספרי מפורש והסבירו בקצרה את תשובתכם.

תשובה:

2) מינימיזציה של פונקציית האמצעים כפונקציה של R

$\|W\|^2 + \frac{1}{2} \|wsvm\|^2 \leq 2 \cdot \|w\| \cdot \|wsvm\| + \frac{1}{2} \|wsvm\|^2$

$(R \cdot \frac{1}{2} \|wsvm\|^2)^2 = 10$ $R^2 = 2R$ $R = 10$

$\Rightarrow (2R \cdot \frac{1}{2} \|wsvm\|)^2 = 10$

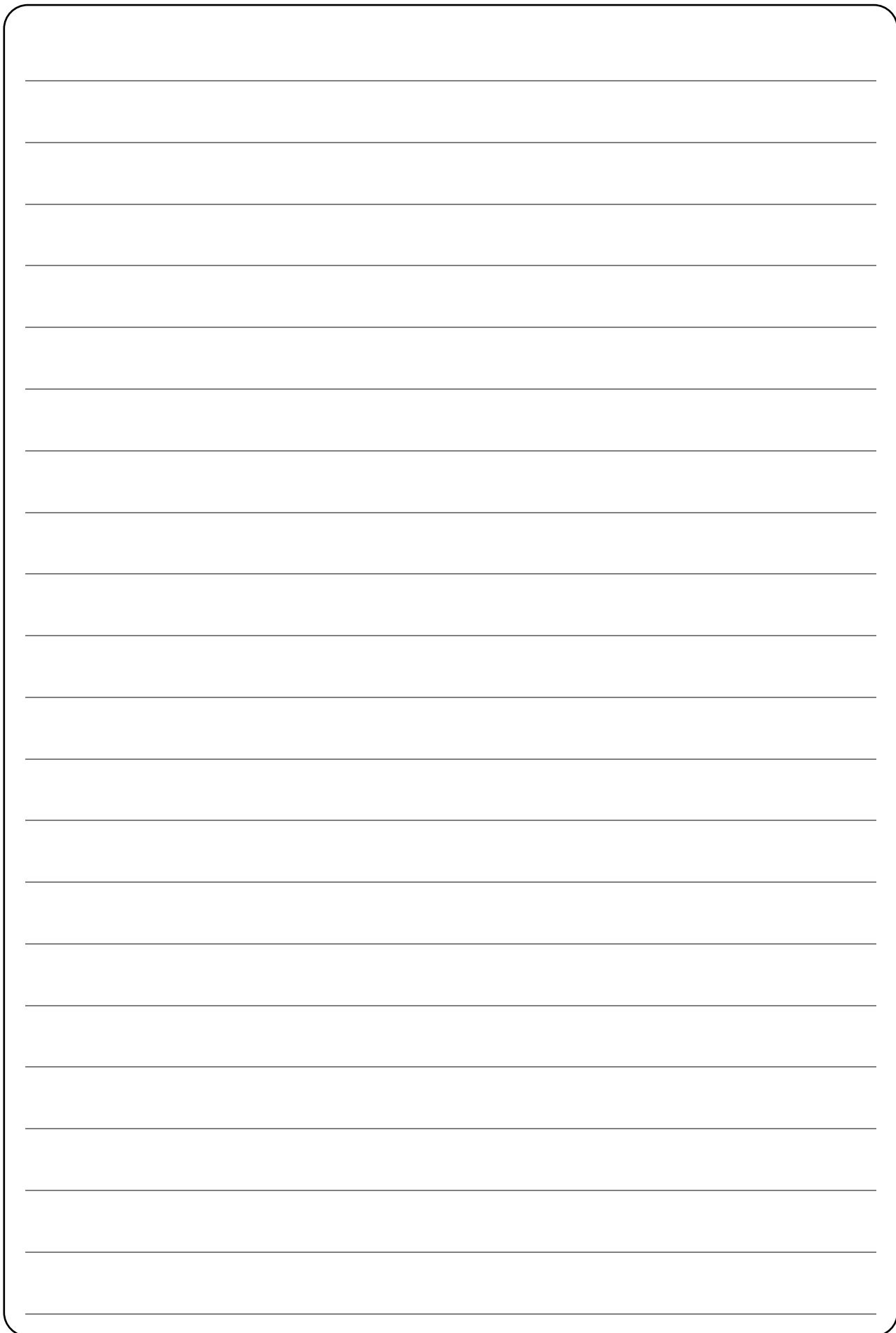
$(R \|wsvm\|)^2 = 10$

השאלה מבקשת $R = 10$

ד. [9 נק'] נתון דата $\{(\mathbf{x}_i, y_i)\}_{i=1}^{100}$ פריד לינארית הומוגנית ש"הבטחת" המשפט עבورو הינה $10 =$
 נסיר את הדוגמה האחרונה ונקבל דטא חדש $\{(\mathbf{x}_i, y_i)\}_{i=1}^{99} = S'$. נסמן את החסם המעודכן בעזרת
 $(R' \cdot \|\mathbf{w}'_{SVM}\|_2)^2$ בטענה שלפניכם, כתבו את היחס ($=, \leq, \geq$) המדויק ביותך כך שתהייה נכון בהכרח: $10 \leq \frac{1}{\|\mathbf{w}'_{SVM}\|_2}$

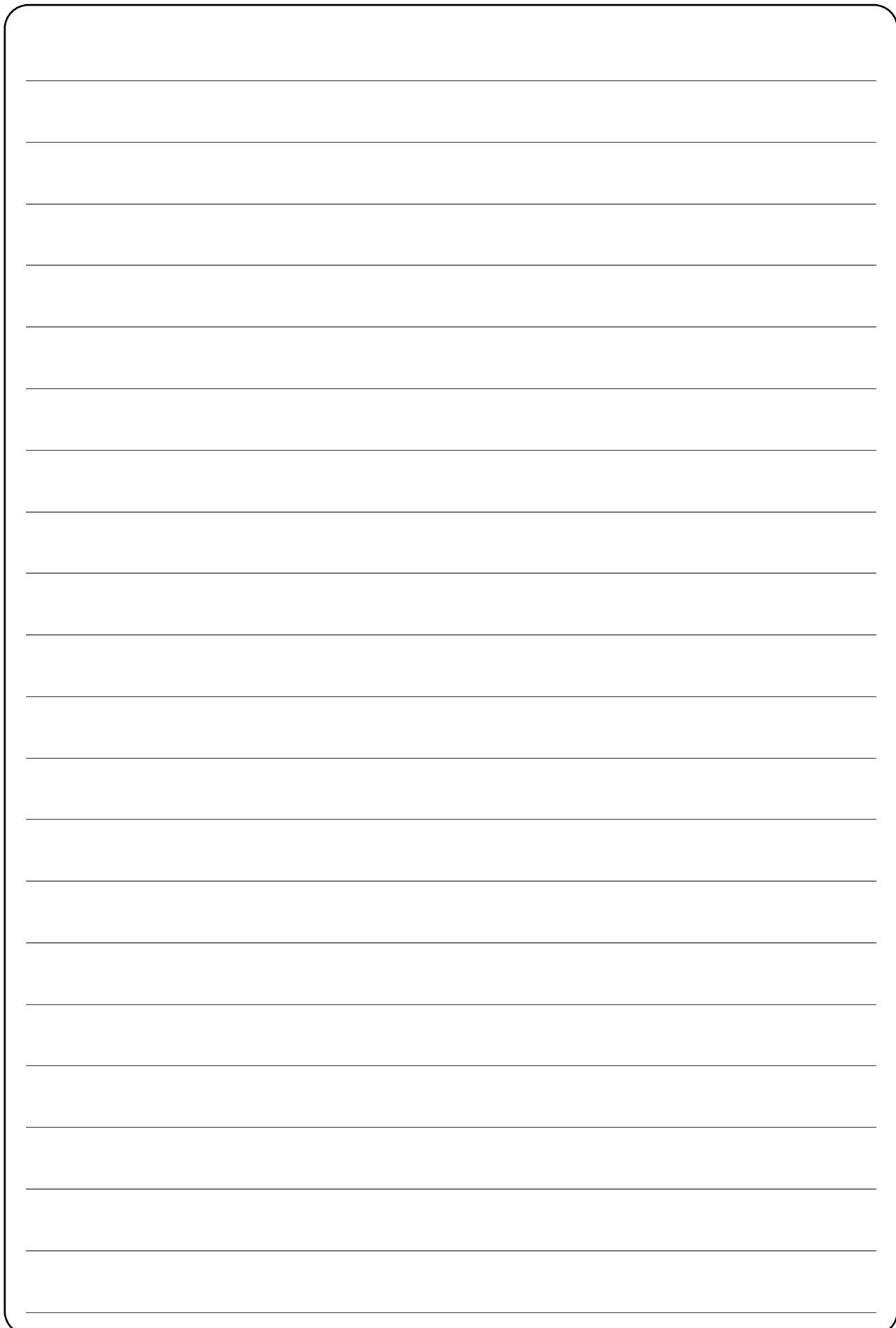
הסברו בקצרה: הוכח רצויו כי אם $\sum_{i=1}^n a_i \geq 0$ אז $\sum_{i=1}^n |a_i| \leq \sum_{i=1}^n a_i^2$

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



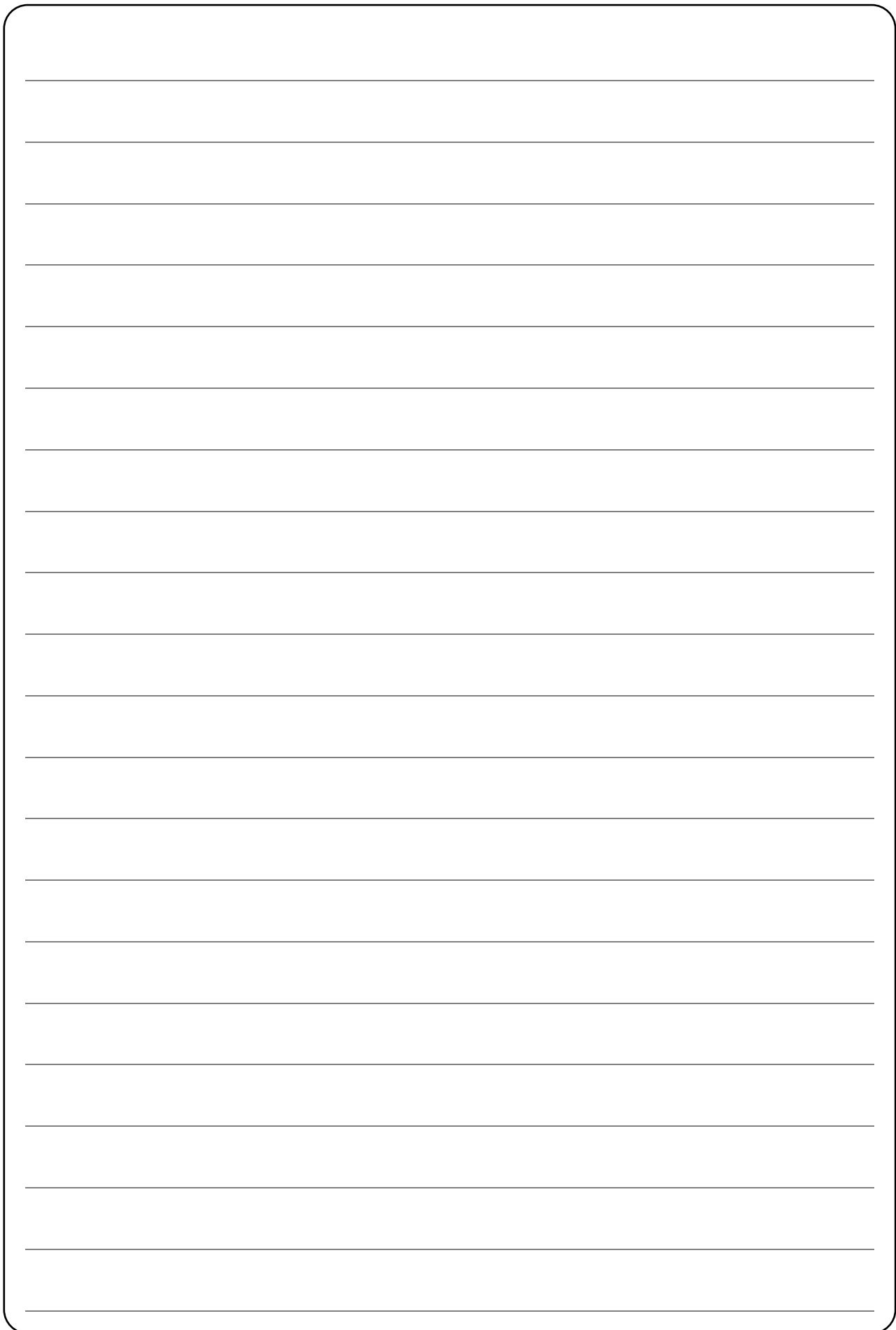
A large rectangular frame with rounded corners, containing 20 horizontal lines spaced evenly apart, intended for handwritten responses.

מסגרת נוספת (יש לציין אם מדובר בטעיטה או בהמשך לשובה אחרת):



A large rectangular frame with rounded corners, containing 20 horizontal lines spaced evenly apart, intended for handwritten responses.

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



A large rectangular frame with rounded corners, containing 20 horizontal lines spaced evenly apart, intended for handwritten responses.