



מבוא למערכות לומדות (236756)

סמינר אביב תשפ"א – 15 ביולי 2021

מרצה: ד"ר ניר רחנfeld

מבחן מסכם מועד א'

הנחיות הבחינה:

- **משך הבחינה:** 3 שעות.
- **חומר עזר:** המבחן בחומר סגור (לא ספרים, מחברות, דפי נוסחאות).
- אין צורך במחשבון.
- מותר לכתוב בעט או בעיפרון, כל עוד הכתב קרייא וברור.
- יש לכתוב את תשובה תיכום **על גבי שאלון זה** בכתב יד קרייא. תשובה בכתב יד שאינו קרייא לא תיבדק.
- במבחן 12 עמודים ממושפרים מה"כ, כולל שער זה שמספרו 1 ושלושה עמודי טיווח בסוף הגילון.
- אין בחרה בין השאלות. יש בסה"כ 104 נקודות והציון המירבי הוא 100.
- נא לכתוב רק את המבוקש ולצער הסברים קצרים **עפ"י** הנחיות.
- **בתום המבחן יש להגיש את שאלון זה בלבד.**
- **בצהוב:** הבחרות שפורסמו בזמן הבחינה.

מבנה הבחינה:

- **חלק א' [72 נק']:** 4 שאלות פתוחות [כל אחת 18 נק']
- **חלק ב' [32 נק']:** 8 שאלות סגורות (אמריקאיות) [כל אחת 4 נק']

בהצלחה!

חלק א' – שאלות פתוחות [72 נק']

שאלה 1 [18 נק']

נתון סט אימון עם תיוגים ביןaries ולומדים עליו מסוג. בסט האימון אין שתי דוגמאות זהות. בעת, בוחרים באקראי 3 דוגמאות אימון שונות וMSCPs אותם (כל אחת עם התיוג שלה) כך שיופיעו פעמיים בסט האימון. לבסוף, מאמנים מסוג חדש על סט האימון המעודכן.

לכל אחד מאלגוריתמי הלמידה הבאים סמן האם גבולות ההחלטה (decision boundaries) של המסוג החדש לאחר השכפול זהים בהכרח לא בהכרח המקורי. רק אם סימנתם שהגבולות לא בהכרח זהים, הסבירו בקצרה מדוע. הניחו שאין צעדים סטטיסטיים (אקראים) בritch האלגוריתמים.

א. 3ID המשמש באנתרופופיה ובונה עץ עמוק מירבי 4 גבולות ההחלטה: **זהים בהכרח / לא בהכרח זהים**

הסבר (אם סימנתם "לא בהכרח"): היו ירים כטבוקם צפויים מרגע אחד כל אחד מהם יתגלו או לא, אבל לא יתגלו כל אחד בוודאות

ב. NN-k כאשר $k = 1$ גבולות ההחלטה: **זהים בהכרח / לא בהכרח זהים**

הסבר (אם סימנתם "לא בהכרח"):

ג. NN-k כאשר $k = 3$ גבולות ההחלטה: **זהים בהכרח / לא בהכרח זהים**

הסבר (אם סימנתם "לא בהכרח"): היו ירים כטבוקם צפויים מרגע אחד כל אחד מהם יתגלו או לא, אבל לא יתגלו כל אחד בוודאות

אנו מודים ב + פלוס מינוס ב - (אילגין)

ד. Perceptron בהנחה שהדעתה פריד ליניארית גבולות ההחלטה: **זהים בהכרח / לא בהכרח זהים**

הסבר (אם סימנתם "לא בהכרח"):

המונר

$$= \oplus =$$

ה. AdaBoost with decision stumps גבולות ההחלטה: **זהים בהכרח / לא בהכרח זהים**

= + \rightarrow \leftarrow - \rightarrow - \leftarrow - הסבר (אם סימנתם "לא בהכרח"):

ו. SVM-Hard בהנחה שהדעתה פריד ליניארית גבולות ההחלטה: **זהים בהכרח / לא בהכרח זהים**

הסבר (אם סימנתם "לא בהכרח"):

שאלה 2 [18 נק']

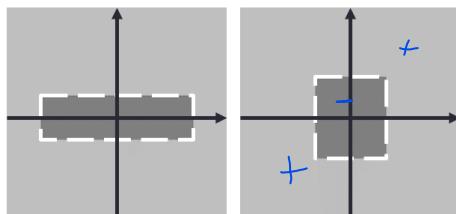
א. [6 נק'] הוכיחו את תכונת המונוטוניות של VC-dimension
 $\text{VCdim}(\mathcal{H}_1) \leq \text{VCdim}(\mathcal{H}_2)$, אם $\mathcal{H}_1 \subseteq \mathcal{H}_2$ אזי

הוכחה:
 $y_1, \dots, y_n \in \{+, -\}$ $C = \{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$ $\text{VCdim}(H_2) = D$
 $h(x_i) = y_i$ $x_i \in C$ $h \in H_1$ $\text{dim}(H_1) = D$
 $\text{dim}(H_1) = D \leq \text{dim}(H_2)$ כי $H_1 \subseteq H_2$ $h \in H_1$

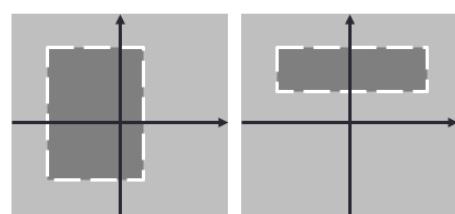
בתרגול הגדנו את מחלקת ההיפותזות $\mathcal{H}_{\text{aligned}}$ של מבנים מקבילים לצירים בדו-מימד והראינו שמתקיים הבירה: האיזור שבתו המלבן מסווג כחיובי והائزור החיצוני כשלילי.

כעת נגדיר את מחלקת $\mathcal{H}_{\text{centered}}$ של מבנים מקבילים לצירים בדו-מימד **שומרץם** בדיק בראשית הצירים.

$\mathcal{H}_{\text{centered}}$ שתי היפותזות מתור



$\mathcal{H}_{\text{aligned}}$ שתי היפותזות מתור



$$\text{VCdim}(\mathcal{H}_{\text{centered}}) = \boxed{2}$$

ב. [6 נק'] מהו מימד ה-VC של המחלוקת החדשה? מלאו.

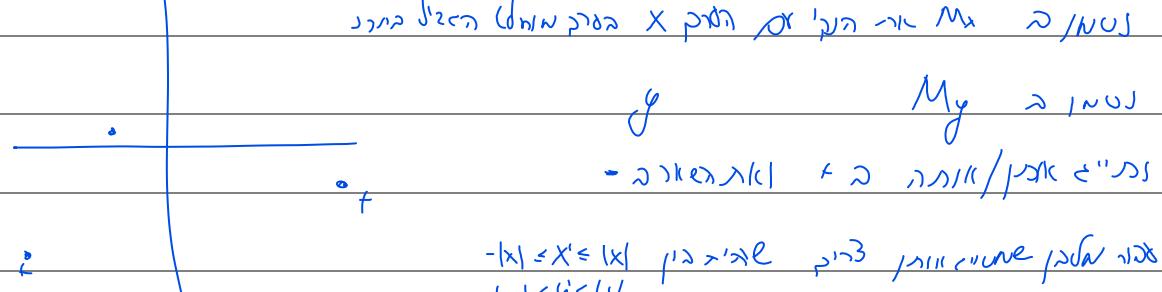
ג. [6 נק'] הוכחו את תשובתכם לסעיף הקודם.

$$\text{הוכחה: } \text{VCdim} \geq 2 \quad (\text{I})$$

$$\text{רנ' } \text{VCdim} = 3 \quad \text{או } \text{הוכחה כפופה ל } \text{VCdim} < 3$$

$$\exists h \forall y_1, y_2 \exists h(x_1) = y_1 \wedge h(x_2) = y_2$$

לעתה נוכיח מה זה אומר ווליאן ה-VC מודול גורן



$$M_y \geq 1$$

ולכן $M_x \geq 1$ כלומר $M_x \geq 1$

$$-|x_1| \leq x_1 \leq |x_1| \quad -|x_2| \leq x_2 \leq |x_2|$$

כך ש- $|x_1| + |x_2| \leq M_x$

$$h(x_1) = + \wedge h(x_2) = - \quad \text{ולכן } h(x_1) \neq h(x_2) \quad \text{ולכן } \text{VCdim} \geq 2$$

$$\text{ולכן } \text{VCdim} \geq 2$$

$$\boxed{\text{VCdim} = 2} \quad \text{ס. ס}$$

שאלה 3 [18 נק']

נתון סט אימון עם דוגמאות חד-ממדיות $\mathbb{R} \ni y_1, \dots, y_m \in \mathbb{R}$, $x_1, \dots, x_m \in \mathbb{R}$ ותיוגים רציפים.

$$\hat{w} = \operatorname{argmin}_{w \in \mathbb{R}} (\sum_{i=1}^m (wx_i - y_i)^2 + R(w))$$

פתרונות בעית גרגסיה ליניארית עם רגולרייזציה (עם $\lambda = 1$):

בפתרון שהתקבל, המשקל שמתאים למאפיין היחיד מקיים $0 \neq \hat{w}$.

$$\widehat{\mathbf{u}} = \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^2} (\underbrace{\sum_{i=1}^m (\mathbf{u}^\top \mathbf{x}'_i - y_i)^2 + R(\mathbf{u})}_{\triangleq \mathcal{L}'(\mathbf{u})})$$

כעת, משכפלים את המאפיין היחיד כך שכל דוגמה מעודכנת היא וקטור \mathbb{R}^2 לבסוף, פותרים את הבעה המעודכנת:

הבהרה: הפונקציה R מקבלת סקלר או וקטור ומחזירה את הנורמה שלו (בסעיפים א'-ב': 2ן בריבוע, בסעיף ג': 1ן).

$$\mathbf{u} = \begin{bmatrix} \hat{w} \\ \omega_2 \end{bmatrix} \text{ כאשר } \|\mathbf{u}\|_2^2 = \sum_j (u_j)^2 \triangleq R(\mathbf{u})$$

א. [6 נק'] נציג כפתרון לבעה המעודכנת את הווקטור $\mathbf{z} \in \mathbb{R}^2$ המקיים $\mathcal{L}'(\mathbf{u}) + R(\mathbf{u}) > \mathcal{L}'(\mathbf{z}) + R(\mathbf{z})$

הוכחו שהו פיתרון לא אופטימלי על ידי הצעת פיתרון אחר $\mathbf{z} \in \mathbb{R}^2$ המקיים $\mathcal{L}'(\mathbf{u}) + R(\mathbf{u}) > \mathcal{L}'(\mathbf{z}) + R(\mathbf{z})$

$$\mathcal{L}'(\mathbf{u}) = \sum_i (z_i - \hat{w}x_i - y_i)^2 \quad \mathbf{z} = \boxed{\begin{bmatrix} \omega_2 \\ \omega_2 \end{bmatrix}}$$

$$R(\mathbf{u}) = \omega^2 > 2 \cdot \frac{\omega^2}{4} = R(2)$$

ב. [6 נק'] נציג כפתרון לבעה המעודכנת את הווקטור $\mathbf{u} \in \mathbb{R}^2$ המקיים $\mathcal{L}'(\mathbf{u}) + R(\mathbf{u}) = \sum_j (u_j)^2 \triangleq R(\mathbf{u})$, סמן את הטענה הנכונה בהכרח והסבירו בקצרה.

א. אחד המשקלים שווה לאפס (משמעות: $\hat{u}_1 = 0 \vee \hat{u}_2 = 0$)

ב. שני המשקלים שונים מאפס (משמעות: $\hat{u}_1 \neq 0 \wedge \hat{u}_2 \neq 0$)

ג. שני המקרים a, b אפשריים

הסבירו בקצרה:

$\mathcal{L}'(\mathbf{u}) = \mathcal{L}'(\mathbf{z}) + R(\mathbf{u}) > R(\mathbf{z})$

ג. [6 נק'] נציג כפתרון לבעה המעודכנת בהכרח והסבירו בקצרה.

א. אחד המשקלים שווה לאפס (משמעות: $\hat{u}_1 = 0 \vee \hat{u}_2 = 0$)

ב. שני המשקלים שונים מאפס (משמעות: $\hat{u}_1 \neq 0 \wedge \hat{u}_2 \neq 0$)

שני המקרים a, b אפשריים

c.

הסבירו בקצרה:

$|u_1| + |u_2| = |\omega_1| + |\omega_2|$

שאלה 4 [18 נק']

הזכירו בבעיות האופטימיזציה של ה-SVM במקרה ללא הומוגני:

Hard SVM

$$\begin{aligned} \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}} & \|\mathbf{w}\|_2^2 \\ \text{s.t. } & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad \forall i \in [m] \end{aligned}$$

Soft SVM

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

א. [4 נק'] כתבו וקטור שמהווה subgradient לפי \mathbf{w} לפונקציית loss של ה-SVM.

תשובה סופית (לרשומכם עמודי טיווה בסוף השאלה):

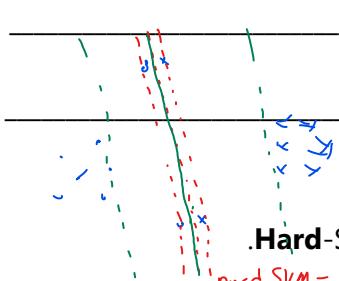
$$\max\{0, 1 - \mathbf{z}\} = \begin{cases} 1 - \mathbf{z} & \mathbf{z} \leq 1 \\ 0 & \mathbf{z} > 1 \end{cases}$$

$$\nabla_{\mathbf{w}} \left(\frac{1}{m} \sum_{i=1}^m \max\{0, 1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b)\} + \lambda \|\mathbf{w}\|_2^2 \right) = \underbrace{\sum_{i=1}^m y_i \mathbf{x}_i}_{2\lambda \mathbf{w}_1} + \underbrace{2\lambda \mathbf{w}_2}_{0 \cdot \mathbf{w}} ; \quad y_i (\mathbf{w}^\top \mathbf{x}_i + b) \leq 1$$

ב. [6 נק'] לכל אחת מהטענות הבאות, סמן אם היא (בהכרח) נכון / שגוייה / לא ניתן לדעת והסבירו בקצרה.
א. עבור סט אימון פריד ליניארית, SVM מגיע ל-**Hard-SVM** גובה מזה של **train accuracy**.

הטענה: נכון / שגוייה / לא ניתן לדעת.

הסביר: 100% train Accuracy



hard SVM - small Margin - Large Norm
soft SVM - Large Margin - Small Norm

הטענה: נכון / שגוייה / לא ניתן לדעת.

הסביר: soft SVM פולסן על נאום נרחב יותר מאשר הhard SVM

problem of overfitting more hardSVM > more soft SVM

ג. [8 נק'] נתון דатаה פריך ליניארית עם תיוגים בינהירים. פתרים עליו SVM-Hard ומקבלים כפתרון $\mathbf{w}^d \in \mathbb{R}$ ו- $b_1 \in \mathbb{R}$. ידוע שהפתרון שהתקבל הוא הפתרון היחיד לבעה.

עכשו פתרים בעיה SVM-Hard margin=2 מעודכנת עם **margin=2**. משמע, עם אילוצים מעודכנים $2 \geq (\mathbf{w}^\top \mathbf{x}_i + b)$.

הערה בדיעבד: שימו לב שהשאלה לא עוסקת ב- $\text{margin} = \frac{\mathbf{w}^\top \mathbf{x}_i + b}{\|\mathbf{w}\|}$, אלא ב- $(\mathbf{w}^\top \mathbf{x}_i + b)$, כפי שכותוב במפורש בנוסחה.

את הפתרון החדש נסמן ב- $\mathbf{w}_2 \in \mathbb{R}^d$ ו- $b_2 \in \mathbb{R}$.

בכל אחד מהסעיפים הבאים סמןו את התשובה הנכונה ומלאו את החסר (היכן שנדרש).

א. יתכן שלבעיה המעודכנת אין פתרון, למשל כאשר המרחק בין המחלקות לא מספיק גדול לмерות שהדטה פריך ליניארית.

$$\omega^1 = \frac{\omega}{2} \quad b^1 = \frac{b}{2}$$

הטענה: נכונה / שגויה.

ב. אם קיימם פתרון לבעיה המעודכנת אז הפרדייקציות הבינהיריות זהות. משמע, $h_{\mathbf{w}_1, b_1}(\mathbf{x}) = h_{\mathbf{w}_2, b_2}(\mathbf{x})$ $\forall \mathbf{x} \in \mathbb{R}^d$.

הטענה: נכונה / שגויה.

כ. אם קיימם פתרון לבעיה המעודכנת אז קיימים סקלר α עבורו $\mathbf{w}_2 = \alpha \mathbf{w}_1$.

הטענה: נכונה והסקלר שווה ל- 2 / שגויה.

$$g_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$$

ד. אם קיימם פתרון לבעיה המעודכנת אז קיימים סקלר β עבורו $b_2 = \beta b_1$.

הטענה: נכונה והסקלר שווה ל- 2 / שגויה.

חלק ב' – שאלות אמריקאיות [32 נק']

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

. [4 נק'] מזכורת: $\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$, $\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$.

נתון דאטה עם תיוגים ביןaries (P או N, לפחות דוגמה אחת מכל תיוג).

סמננו את כל הטענות השגויה.

a. ערך ה-(AUC) Area under the curve האופטימלי של עקומת ROC הוא 1

b. מתקיים $\text{Precision} + \text{Recall} = 1$

c. תמיד קיים מודל שמשיג $\text{Recall} = 1$

d. ניתן ליצור Confusion matrices רק בעיות עם תיוגים ביןaries (שתי מחלקות)

e. בדאייה מאבחן ($\Pr[y = P]$), מודיע ה- Recall וה- Precision מהים שחוקים למדד ה- Accuracy

. [4 נק'] נרצה להרחיב את מודל ה-NN-k לבעיות רגסיה ($\mathbf{x}_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$).

נגידור את $(\mathbf{x})_{\text{NN}}$ להיות קבוצת האנדקסים של k השכנים הקרובים ביותר ל- \mathbf{x} .

סמננו את שתן פונקציות החיזוי (של u בהינתן \mathbf{x}) המתאימות ביותר לעביה.

$$h(\mathbf{x}) = \frac{1}{k} \sum_{i \in \text{NN}(\mathbf{x})} \mathbf{x}_i$$

$$h(\mathbf{x}) = \frac{1}{k} \sum_{i \in \text{NN}(\mathbf{x})} y_i$$
 ↗
.b

$$h(\mathbf{x}) = \sum_{i \in \text{NN}(\mathbf{x})} (\|\mathbf{x} - \mathbf{x}_i\|_2 y_i)$$
 .c

$$h(\mathbf{x}) = \sum_{i \in \text{NN}(\mathbf{x})} \left(\frac{\exp\{-\|\mathbf{x} - \mathbf{x}_i\|_2\}}{\sum_{j \in \text{NN}(\mathbf{x})} \exp\{-\|\mathbf{x} - \mathbf{x}_j\|_2\}} y_i \right)$$
 .d

$$h(\mathbf{x}) = \sum_{i \in \text{NN}(\mathbf{x})} (\max\{\dots, \|\mathbf{x} - \mathbf{x}_i\|_2^2\}, y_i)$$
 .e

. [4 נק'] נתונים סט אימון וסט מבחן הנדגמים מהתפלגות כלשהי \mathcal{D} ומחלקה היפותזית כלשהי \mathcal{H} .

בוחרים היפותזה מתרון המחלקה ע"י ERM.

עתה מוסיפים לשט האימון מספר דוגמאות חדשות (הנדגמות מ- \mathcal{D}), ולומדים היפותזה חדשה ע"י MRM על \mathcal{H} .

כל סוג שגיאה סמננו את האפשרות המתאימה בהכרח.

a. שגיאת האימון: תעלה / תרד / לא תשתנה / לא ניתן לדעת

b. שגיאת המבחן: תעלה / תרד / לא תשתנה / לא ניתן לדעת

. [4 נק'] מה התפקיד של פונקציית h-sigmoid במסוג regression? סמננו את התשובה הנכונה.

א. להכניס non-linearity ולהפוך את ה- margin להסתברות לא ליניארים

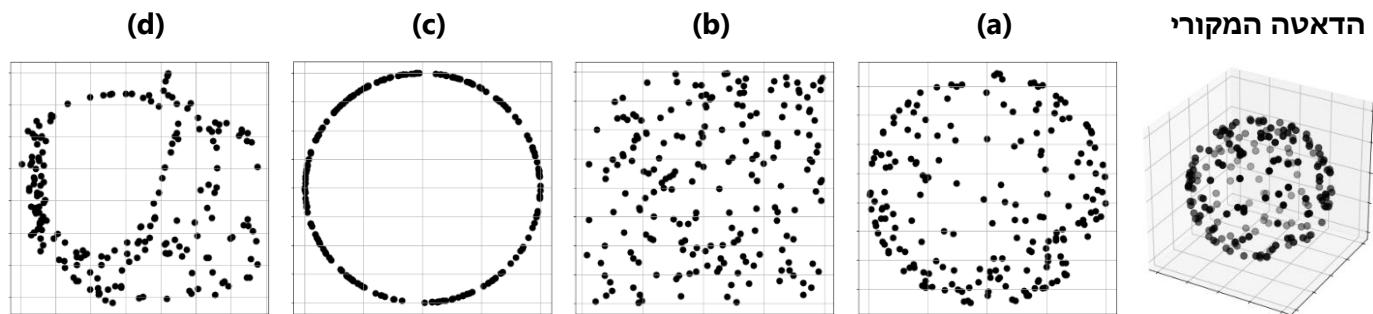
b. להפוך את ה- margin להסתברות

א. לעשות גרסה פולינומיאלית בין $\mathbf{x} \in \mathbb{R}^d$ ל- $y \in \mathbb{R}$

d. להוסיף רגולריזציה למסוג שנלמד

e. למקסם את האנטרופיה

[4 נק'] נתון דата שנדגים מהתפלגות אחידה על ספירה (sphere) תלת-ממדית סביב ראשית הצירים. מפעילים PCA כדי להוריד ממד מ-3 ל-2. הקיפו את האות שמתאימה לתרשים של הדאטה לאחר הורדת הממד.



.[4 נק'] נתונה בעית `multiclass`.

אימנו מסוג All vs. One (שיטת א') ומסוג `multinomial logistic regression` (שיטת ב').

לאחר האימון, נוספה לבעה מחלוקת חדשה עם דוגמאות חדשות.

על מנת להתמודד עם המחלוקת החדשה (סמננו את הטענה הנכונה ב'וטר):

a. בשיטה א' אנחנו צריכים לאמן רק מסוג אחד נוספת ואין שינוי במסוגים שכבר אומנו

b. בשיטה ב' אנחנו צריכים לאמן רק מסוג אחד נוספת ואין שינוי במסוגים שכבר אומנו

c. שתי הטענות a+b נוכנות

d. כל הטענות הקודמות שגויות

.[4 נק'] נסתכל על רשת Multi-Layer Perceptron לרגרסיה שלומדת פונקציה $h: \mathbb{R}^d \rightarrow \mathbb{R}$.

כפונקציית אקטיבציה נשתמש בפונקציית הזוזות $z = z(\sigma)$. לסיום, נשתמש ב-loss ריבועי (MSE) ללא רגולריזציה.

סמננו את הטענה הנכונה.

a. כמחלקת מודלים, ה-capacity שלו של מודלים ליניארים (linear regression)

b. השכבה האחורה ברשות צריכה להיות שכבת Softmax

c. לביעית האופטימיזציה יש מינימום גלובלי יחיד

d. כל הטענות הקודמות שגויות

.[4 נק'] נתונות דוגמאות $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathbb{R}^d$, $y_1, \dots, y_m \in \{-1, +1\}$ מתוויות באופן ביןארי, משמע $\{-1, +1\}$.

רוצים למדוד הורדת ממד ליניארית (בעזרת מטריצה \mathbf{W}) ל- k ממדים כאשר $d \gg k$.

רוצים לשאלח הטללה, דוגמאות מתוויות זהים תהיה קרובות אחת לשניה ורחוקות מהדוגמאות של התיאוג השני.

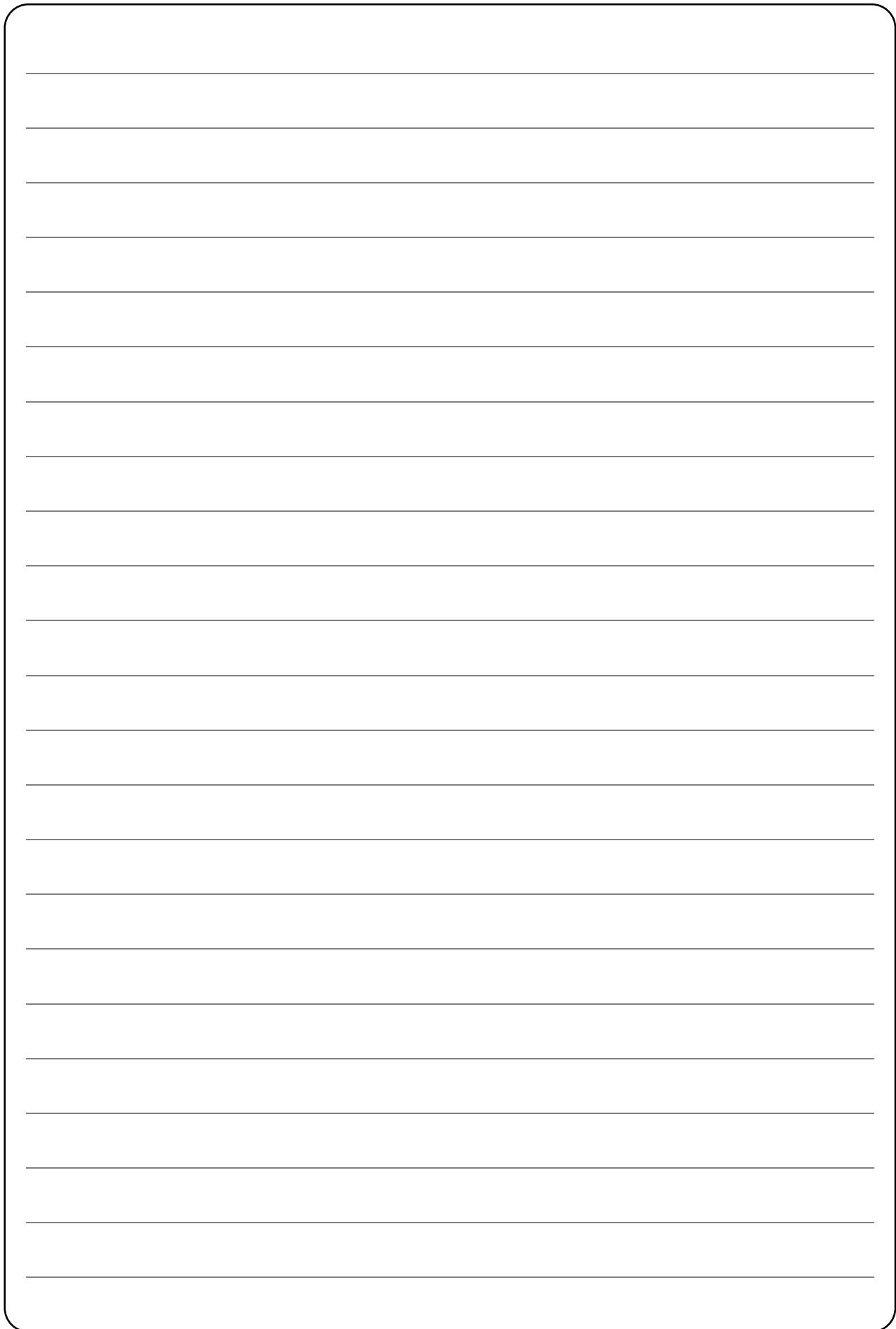
סמננו את בעית האופטימיזציה (היחידה) שמתאימה לפיתרון הבעה.

$$\begin{aligned} & \underset{\mathbf{W} \in \mathbb{R}^{k \times d}}{\operatorname{argmin}} \left(\sum_{i,j} \|\mathbf{W}\mathbf{x}_i - \mathbf{W}\mathbf{x}_j\|_2^2 \right) \\ & \underset{\mathbf{W} \in \mathbb{R}^{k \times d}}{\operatorname{argmin}} \left(\sum_{i,j} \|\mathbf{W}\mathbf{x}_i - \mathbf{W}\mathbf{x}_j\|_2^2 + \sum_{i,j:y_i \neq y_j} \|y_i - y_j\|_2^2 \right) \end{aligned} \quad \boxed{\text{Same}}$$

$$\underset{\mathbf{W} \in \mathbb{R}^{k \times d}}{\operatorname{argmin}} \left(\sum_{i,j:y_i=y_j} \|\mathbf{W}\mathbf{x}_i - \mathbf{W}\mathbf{x}_j\|_2^2 + \sum_{i,j:y_i \neq y_j} \max \{0, 1 - \|\mathbf{W}\mathbf{x}_i - \mathbf{W}\mathbf{x}_j\|_2^2\} \right) \quad \boxed{c}$$

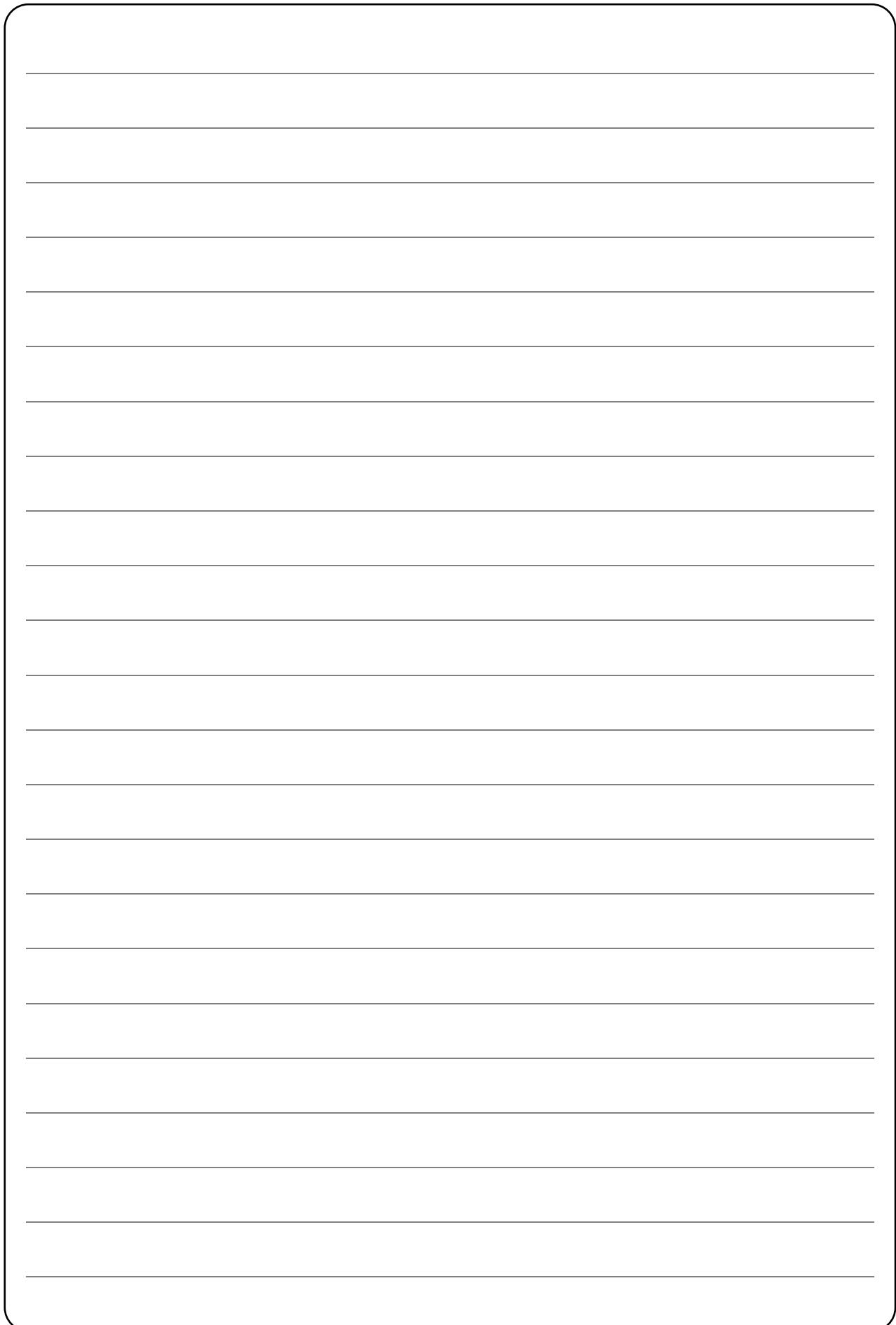
$$\underset{\mathbf{W} \in \mathbb{R}^{k \times d}}{\operatorname{argmin}} \left(\sum_{i,j:y_i \neq y_j} \ln \left(1 + \|\mathbf{W}\mathbf{x}_i - \mathbf{W}\mathbf{x}_j\|_2^2 \right) + \sum_{i,j} (W_{i,j})^2 \right) \quad \boxed{d}$$

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



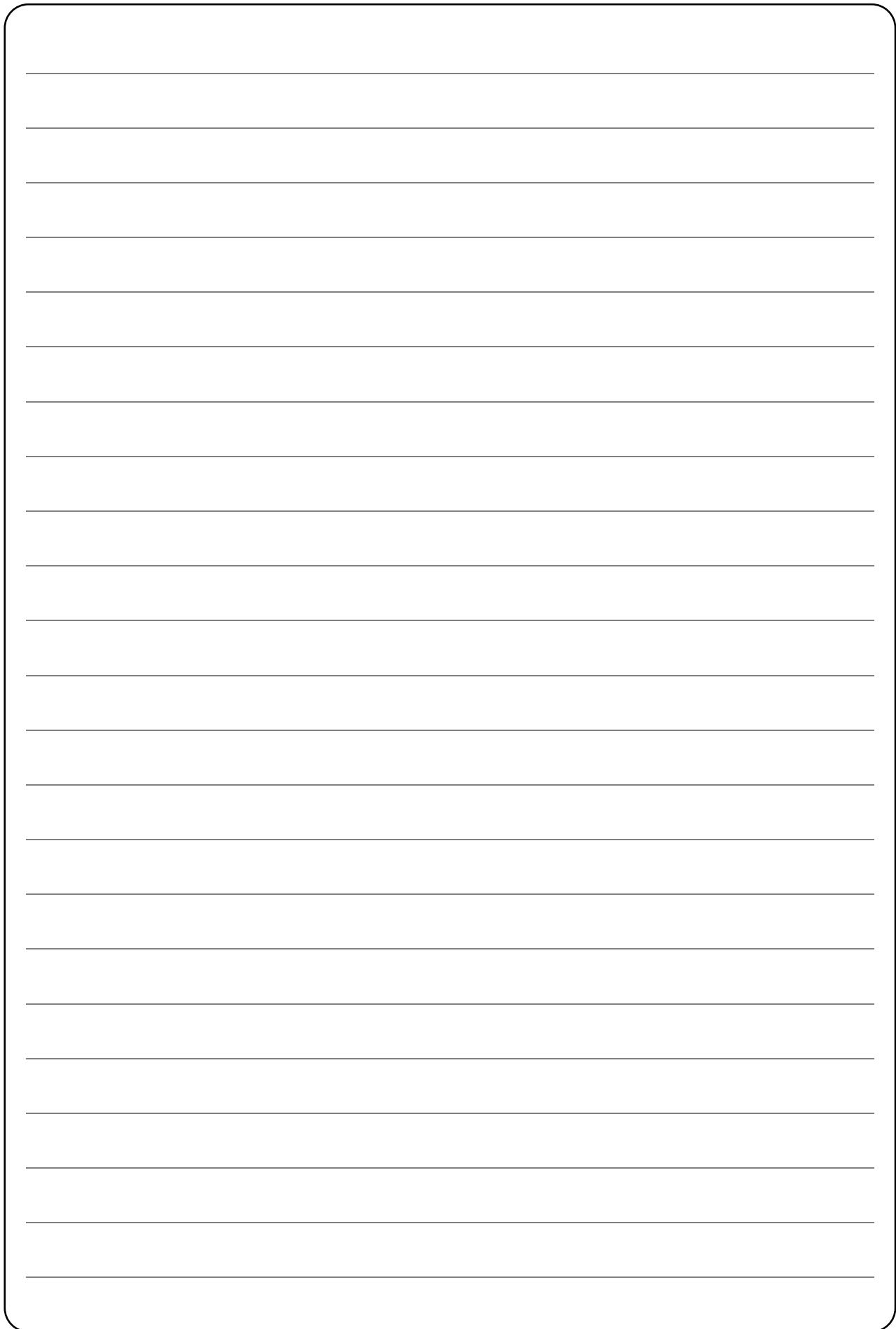
A large rectangular frame with rounded corners, containing 20 horizontal lines spaced evenly apart, intended for handwritten responses.

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



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

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



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