הטכניון – מכון טכנולוגי לישראל
הפקולטה למדעי המחשב

מבוא למערכות לומדות (236756)

סמסטר אביב תשפ"א – 15 ביולי 2021

מרצה: ד"ר ניר רוזנפלד

מבחן מסכם מועד א'

הנחיות הבחינה:

- **משך הבחינה:** 3 שעות.
- **חומר עזר:** המבחן בחומר סגור (ללא ספרים, מחברות, דפי נוסחאות).
- אין צורך במחשבון.
- מותר לכתוב בעט או בעיפרון, כל עוד הכתב קריא וברור.
- יש לכתוב את תשובותיכם **על גבי שאלון זה** בכתב יד קריא. תשובה בכתב יד שאינו קריא לא תיבדק.
- במבחן 12 עמודים ממוספרים סה"כ, כולל עמוד שער זה שמספרו 1 ושלושה עמודי טיוטה בסוף הגיליון.
- אין בחירה בין השאלות. יש בסה"כ 104 נקודות והציון המירבי הוא 100.
- נא לכתוב רק את המבוקש ולצרף הסברים קצרים עפ"י ההנחיות.
- **בתום המבחן יש להגיש את שאלון זה בלבד.**
- בצהוב: הבהרות 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 בינאריים ולומדים עליו מסווג. בסט האימון אין שתי דוגמאות זהות.

כעת, בוחרים באקראי 3 דוגמאות אימון שונות ומשכפלים אותן (כל אחת עם התיוג שלה) כך שיופיעו פעמיים בסט האימון.

לבסוף, מאמנים מסווג חדש על סט האימון המעודכן.

לכל אחד מאלגוריתמי הלמידה הבאים סמנו האם גבולות ההחלטה (decision boundaries) של המסווג החדש לאחר השכפול זהים בהכרח לאלה של המסווג המקורי. רק אם סימנתם שהגבולות **לא בהכרח זהים**, הסבירו בקצרה מדוע.

הניחו שאין צעדים סטוכסטיים (אקראיים) בריצת האלגוריתמים.

א. ID3 המשתמש באנטרופיה ובונה עץ בעומק מירבי 4 גבולות ההחלטה: **זהים בהכרח** / **לא בהכרח זהים**

הסבר (אם סימנתם "לא בהכרח"): ה entropy מחשב את יחס ה + לעומת הכלל ויחס ה(−) לעומת הכלל, שכפול נק' עלול לגרום לשינוי הentropy ולכן לשינוי במיקום הפיצול

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

הסבר (אם סימנתם "לא בהכרח"):

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

הסבר (אם סימנתם "לא בהכרח"): נקודה (−1) הנמצאת קרוב ל(+1) שהשתכפל ; (−1) שלא השתכפלה, ייתכן (ויותר) כעת תסווג בתור (+1) במקום (−1)

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

הסבר (אם סימנתם "לא בהכרח"): אמנם נגיע למפריד שמפריד, בגלל השכפול ה"סיכוי" של המסווג עלול להידבק ולהוביל בסוף לתוצאה שונה (בגלל השכפולים)

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

הסבר (אם סימנתם "לא בהכרח"): השכפולים עלולים לגרום לתעדף אלגו' שונים ובכך לשנות את המסווג הסופי.

ו. Hard-SVM בהנחה שהדאטה 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:

לכל שתי מחלקות היפותזות, אם $\mathcal{H}_1 \subseteq \mathcal{H}_2$ אזי $\text{VCdim}(\mathcal{H}_1) \leq \text{VCdim}(\mathcal{H}_2)$.

הוכחה:

$VCdim(H_1)=k$ כלומר קיים סט בגודל k כך שכל לייבלינג אפשרי H_1 מסווגת לנכון, נניח בה"כ כי $h_1,h_2,...,h_i \in H_1$

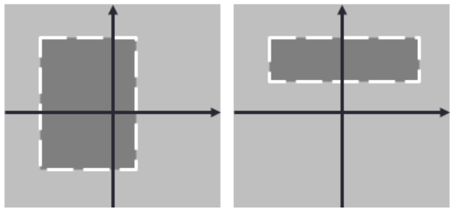
הן ההיפותזות המתאימות את הלייבלינג של הסט, בפרט מכיוון ש $H_1 \subseteq H_2$ אזי $h_1,...,h_i \in H_2$

ולכן הן גם מנתצות אותו סט הנ' ולכן $VCdim(H_1) \leq VCdim(H_2)$

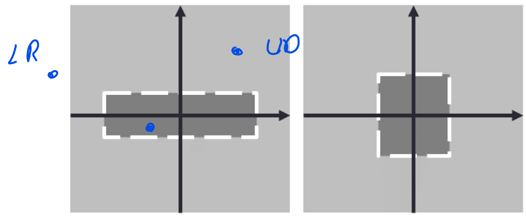
בתרגול הגדרנו את מחלקת ההיפותזות $\mathcal{H}_{\text{aligned}}$ של מלבנים מקבילים לצירים בדו-מימד והראינו שמתקיים $\text{VCdim}(\mathcal{H}_{\text{aligned}}) = 4$. הבהרה: האיזור שבתוך המלבן מסווג כחיובי והאיזור החיצוני כשלילי.

כעת נגדיר את מחלקת $\mathcal{H}_{\text{centered}}$ של מלבנים מקבילים לצירים בדו-מימד **<u>שמרכזם</u>** בדיוק בראשית הצירים.

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



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



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

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

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

הוכחה:

$VCdim(H) \geq 2$: נקח את קב' הנק': $S=\{\underbrace{(0,2)}_{x_1},\underbrace{(1,0)}_{x_2}\}$

ונעבור על כל התיוגים האפשריים:
1. $y_1=y_2=+1$, נבחר ריבוע עם אורך צלע 100
2. $y_1=y_2=-1$ נבחר ריבוע עם אורך צלע 0.001
3. $y_1=+1, y_2=-1$ נבחר מלבן עם רוחב $(-0.01,0.01)$ וגובה $(-100,100)$
4. $y_1=-1, y_2=+1$ נבחר מלבן עם רוחב $(-100,100)$ וגובה $(-0.01,0.01)$

כלומר קיימים נק' כך שלכל תיוג אפשרי קיימת השערה במחלקת ההיפותזות שמסווגת אותן נכון.

$VCdim(H)<3$: נבחר 3 נק' אקראיות. נסמן ב LR את הנקודה עם ה $|x|$ הכי גדול ועם $|y|$ הכי גדולה UD

נתייג אותן ב $(+1)$ (במידה ו $LR=UD$ נתייג אותה רק ב$(+1)$)
את השאר ב (-1) נשים לב כי לכל תיוג כזה יכיל נק' (-1) עם ערך $|y|$ קטן
יותר מה $|y|$ של ה UD ו $|x|$ קטן מה $|x|$ של ה LR ולכן הנק' תהיה תמיד
בתוך ~~ה~~ המלבן של ההיפותזה $(+1)$ ולכן, היא תסווג תמיד לא נכון
כלומר הראנו שעבור כל סידור נק', קיים תיוג כך שH אינה מנתצת אותו
ולכן $VCdim(H)<3$

סה"כ $VCdim(H) \geq 2$ וגם $VCdim(H)<3$ ולכן $VCdim(H)=2$.

שאלה 3 [18 נק']

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

פותרים בעיית רגרסיה ליניארית עם רגולריזציה (עם $\lambda = 1$): $\hat{w} = \operatorname{argmin}_{w\in\mathbb{R}}\left(\sum_{i=1}^{m}(wx_i - y_i)^2 + R(w)\right)$

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

כעת, משכפלים את המאפיין היחיד כך שכל דוגמה מעודכנת היא וקטור $\boldsymbol{x}'_i = \begin{bmatrix} x_i \\ x_i \end{bmatrix} \in \mathbb{R}^2$.

לבסוף, פותרים את הבעיה המעודכנת: $\hat{\boldsymbol{u}} = \operatorname{argmin}_{\boldsymbol{u}\in\mathbb{R}^2}\left(\underbrace{\sum_{i=1}^{m}(\boldsymbol{u}^\top \boldsymbol{x}'_i - y_i)^2}_{\triangleq \mathcal{L}'(\boldsymbol{u})} + R(\boldsymbol{u})\right)$.

הבהרה: הפונקציה R מקבלת סקלר או וקטור ומחזירה את הנורמה שלו (בסעיפים א'-ב': L2 בריבוע, בסעיף ג': L1).

א. [6 נק'] כאשר $R(\boldsymbol{v}) = \|\boldsymbol{v}\|_2^2 \triangleq \sum_j (v_j)^2$, נציע כפיתרון לבעיה המעודכנת את הווקטור $\boldsymbol{u} = \begin{bmatrix} \hat{w} \\ 0 \end{bmatrix}$.

הוכיחו שזהו פיתרון לא אופטימלי על ידי הצעת פיתרון אחר $\boldsymbol{z} \in \mathbb{R}^2$ המקיים $\mathcal{L}'(\boldsymbol{u}) + R(\boldsymbol{u}) > \mathcal{L}'(\boldsymbol{z}) + R(\boldsymbol{z})$.

$$\left.\begin{aligned} \mathcal{L}'(u) &= \sum(\hat{w}x_i - y_i)^2 \\ \mathcal{L}'(z) &= \sum(2\cdot\tfrac{1}{2}\hat{w}x_i - y_i)^2 \end{aligned}\right\} =$$

$$\left.\begin{aligned} R(u) &= |w|^2 \\ R(z) &= \tfrac{1}{2}|w|^2 \end{aligned}\right\} R(z) < R(u)$$

$\boldsymbol{z} = \boxed{\left[\tfrac{1}{2}\hat{w}, \tfrac{1}{2}\hat{w}\right]}$

ב. [6 נק'] כאשר $R(\boldsymbol{v}) = \|\boldsymbol{v}\|_2^2 \triangleq \sum_j (v_j)^2$, סמנו את הטענה הנכונה בהכרח והסבירו בקצרה.

a. אחד המשקלים שווה לאפס (משמע: $\hat{u}_1 = 0 \lor \hat{u}_2 = 0$)

b. שני המשקלים שונים מאפס (משמע: $\hat{u}_1 \neq 0 \land \hat{u}_2 \neq 0$)

c. שני המקרים a, b אפשריים

ג) $R(u) = |w|$

$R(z) = \frac{1}{4}|w| + \frac{1}{4}|w| = \frac{1}{2}|w|$

$\mathcal{L}'(u) =$

הסבירו בקצרה:

הוכחנו בסעיף הקודם כי עבור כל פתרון שבו אחד המשקלים שווה לאפס יש פתרון טוב יותר שבו 2 המשקלים שונים מאפס

ג. [6 נק'] כאשר $R(\boldsymbol{v}) = \|\boldsymbol{v}\|_1 \triangleq \sum_j |v_j|$, סמנו את הטענה הנכונה בהכרח והסבירו בקצרה.

a. אחד המשקלים שווה לאפס (משמע: $\hat{u}_1 = 0 \lor \hat{u}_2 = 0$)

b. שני המשקלים שונים מאפס (משמע: $\hat{u}_1 \neq 0 \land \hat{u}_2 \neq 0$)

c. שני המקרים a, b אפשריים

הסבירו בקצרה:

כל עוד $|u_1|+|u_2|=|w|$ אם $u_1+u_2=w$ [illegible]

שאלה 4 [18 נק']

היזכרו בבעיות האופטימיזציה של ה-SVM במקרה הלא הומוגני:

Soft SVM

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

Hard SVM

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

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

$$\max\{\}: \begin{cases} 0 & z \geq 1 \\ 1-z & z<1 \end{cases}$$

תשובה סופית (לרשותכם עמודי טיוטה בסוף השאלון):

$$\nabla_{\boldsymbol{w}}\left(\frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - \overbrace{y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b)}^{z}\} + \lambda\|\boldsymbol{w}\|_2^2\right) = \begin{cases} 2\lambda\|w\|_2 & ; y_i \cdot w^T x_i + b \geq 1 \\ 2\lambda\|w\|_2 - \frac{1}{m}\sum y_i x_i & ; y_i \cdot w^T x_i + b < 1 \end{cases}$$

ב. [6 נק'] לכל אחת מהטענות הבאות, סמנו האם היא (בהכרח) **נכונה / שגויה / לא ניתן לדעת** והסבירו בקצרה.

a. עבור סט אימון פריד ליניארית, **Hard**-SVM מגיע ל-accuracy **train** גבוה מזה של **פרספטרון**.

הטענה: **נכונה / שגויה / לא ניתן לדעת**.

הסבר: הדאטא פריד לינארית ולכן בשניהם הדיוק אימון יגיע ל 100%

b. עבור סט אימון פריד ליניארית, **Soft**-SVM מגיע ל-accuracy **test** גבוה מזה של **Hard**-SVM.

הטענה: **נכונה / שגויה / לא ניתן לדעת**.

הסבר: יכול להיות שהמשתנים של ה hard וביולידציה יהיו טובים יותר בעולם האמיתי
ויכול להיות שה soft יביא לביצועים טובים יותר ויתכן למשל באימון היה הפרדה לגדול
ולא ממש יכול להיות דומה ל hard)

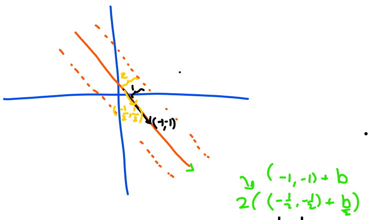
ג. [8 נק'] נתון דאטה פריד ליניארית עם תיוגים בינאריים.

פותרים עליו <u>Hard</u>-SVM ומקבלים כפיתרון $\boldsymbol{w}_1 \in \mathbb{R}^d$ ו-$b_1 \in \mathbb{R}$. ידוע שהפיתרון שהתקבל הוא הפיתרון <u>היחיד</u> לבעיה.

עכשיו פותרים בעיית <u>Hard</u>-SVM מעודכנת עם **margin=2**. משמע, עם אילוצים מעודכנים $y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 2$.

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

smaller norm



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

a. ייתכן שלבעיה המעודכנת אין פיתרון, למשל כאשר המרחק בין המחלקות לא מספיק גדול למרות שהדאטה פריד ליניארית.

לא קשור

הטענה: **נכונה / שגויה.**

b. אם קיים פיתרון לבעיה המעודכנת אזי הפרדיקציות הבינאריות זהות. משמע, $h_{\boldsymbol{w}_1,b_1}(\boldsymbol{x}) = h_{\boldsymbol{w}_2,b_2}(\boldsymbol{x})$ $\forall \boldsymbol{x} \in \mathbb{R}^d$.

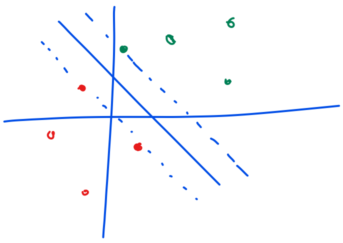
הטענה: **נכונה / שגויה.**

$y_i(w_1^\top x_i + b_1) \geq 1 \quad / \cdot 2$

$y_i((2w_1^\top)x_i + 2b_1) \geq 2$

c. אם קיים פיתרון לבעיה המעודכנת אזי קיים סקלר α עבורו $\boldsymbol{w}_2 = \alpha \boldsymbol{w}_1$.

הטענה: **נכונה והסקלר שווה ל-**2 / **שגויה.**



d. אם קיים פיתרון לבעיה המעודכנת אזי קיים סקלר β עבורו $b_2 = \beta b_1$.

הטענה: **נכונה והסקלר שווה ל-**2 / **שגויה.**

חלק ב' – שאלות אמריקאיות [32 נק']

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

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

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

סמנו את **כל** הטענות **השגויות**.

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

b. מתקיים Precision+Recall=1

c. תמיד קיים מודל שמשיג Recall=1

d. ניתן ליצור Confusion matrices רק בבעיות עם תיוגים בינאריים (שתי מחלקות)

e. בדאטה מאוזן ($\Pr[y = \text{P}] \approx 0.5$), מדדי ה-Precision וה-Recall נהיים שקולים למדד ה-Accuracy

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

נגדיר את $\text{NN}(\boldsymbol{x})$ להיות קבוצת האינדקסים של k השכנים הקרובים ביותר ל-$\boldsymbol{x}$.

סמנו את **שתי** פונקציות החיזוי (של y בהינתן $\boldsymbol{x}$) המתאימות ביותר לבעיה.

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

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

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

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

קרוב $\bar{x}$ ⇐ הרבה חשיבות

רחוק $\bar{x}$ ⇐ קצת חשיבות

e. $h(\boldsymbol{x}) = \sum_{i\in\text{NN}(\boldsymbol{x})}(\max\{0,\ \|\boldsymbol{x}-\boldsymbol{x}_i\|_2^2\}\, y_i)$

same

ג. [4 נק'] נתונים סט אימון וסט מבחן הנדגמים מהתפלגות כלשהי $\mathcal{D}$ ומחלקת היפותזות כלשהי $\mathcal{H}$.

בוחרים היפותזה מתוך המחלקה ע"י ERM.

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

לכל סוג שגיאה **סמנו** את האפשרות המתאימה בהכרה.

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

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

\+ / –

ד. [4 נק'] מה התפקיד של פונקציית ה-sigmoid במסווג logistic regression? סמנו את התשובה הנכונה.

a. להכניס non-linearity ולאפשר ללמוד גבולות החלטה לא ליניאריים

b. להפוך את ה-margin להסתברות

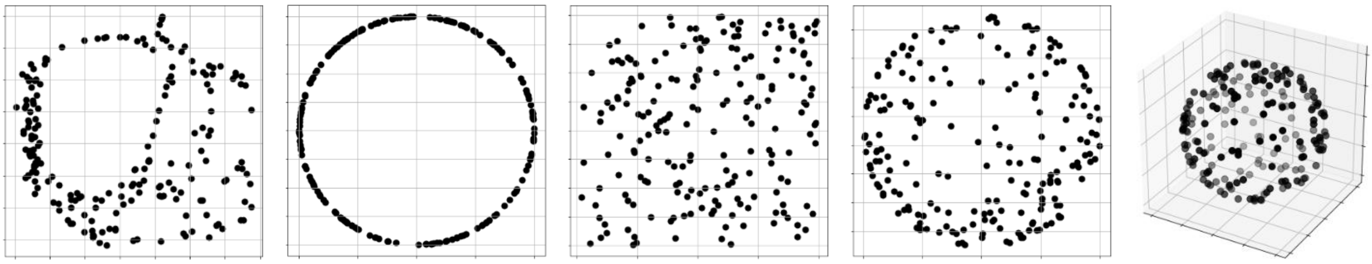
c. לעשות רגרסיה פולינומיאלית בין $\boldsymbol{x} \in \mathbb{R}^d$ ל-$y \in \mathbb{R}$

d. להוסיף רגולריזציה למסווג שנלמד

e. למקסם את האנטרופיה

ה. [4 נק'] נתון דאטה שנדגם מהתפלגות אחידה על ספֵירה (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.

הדאטה המקורי **(a)** **(b)** **(c)** **(d)**



ו. [4 נק'] נתונה בעיית multiclass.
אימנו מסווג One vs. All (**שיטה א'**) ומסווג multinomial logistic regression (**שיטה ב'**).
לאחר האימון, נוספה לבעיה מחלקה חדשה עם דוגמאות חדשות.
על מנת להתמודד עם המחלקה החדשה (סמנו את הטענה הנכונה <u>ביותר</u>):

a. בשיטה א' אנחנו נדרשים לאמן רק מסווג אחד נוסף ואין שינוי במסווגים שכבר אומנו
b. בשיטה ב' אנחנו נדרשים לאמן רק מסווג אחד נוסף ואין שינוי במסווגים שכבר אומנו
c. שתי הטענות a+b נכונות
d. כל הטענות הקודמות שגויות

ז. [4 נק'] נסתכל על רשת Multi-Layer Perceptron ל<u>רגרסיה</u> שלומדת פונקציה $h: \mathbb{R}^d \to \mathbb{R}$.
כפונקציית אקטיבציה נשתמש בפונקציית הזהות $\sigma(z) = z$. לסיום, נשתמש ב-loss ריבועי (MSE) ללא רגולריזציה.
סמנו את הטענה הנכונה.

a. כמחלקת מודלים, ה-capacity שקול לזה של מודלים ליניאריים (linear regression) !
b. השכבה האחרונה ברשת צריכה להיות שכבת Softmax
c. לבעיית האופטימיזציה יש מינימום גלובאלי <u>יחיד</u>
d. כל הטענות הקודמות שגויות

ח. [4 נק'] נתונות דוגמאות $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^d$ מתוייגות באופן בינארי, משמע $y_1, \dots, y_m \in \{-1, +1\}$.
רוצים ללמוד הורדת ממד ליניארית (בעזרת מטריצה $\mathbf{W}$) ל-k מימדים כאשר $k \ll d$.
רוצים שלאחר ההטלה, דוגמאות מתיוגים זהים תהיינה קרובות אחת לשניה ורחוקות מהדוגמאות של התיוג השני.
סמנו את בעיית האופטימיזציה (היחידה) שמתאימה לפיתרון הבעיה.

a. $\underset{\mathbf{W} \in \mathbb{R}^{k \times d}}{\operatorname{argmin}} \left(\sum_{i,j} \left\| \mathbf{W}\boldsymbol{x}_i - \mathbf{W}\boldsymbol{x}_j \right\|_2^2 \right)$

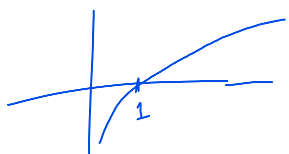
עונש אם שונים

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

Sum 0

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

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



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

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

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