הטכניון – מכון טכנולוגי לישראל

הפקולטה למדעי המחשב

מבוא למערכות לומדות (236756)

סמסטר אביב תשפ"ב – 12 ביולי 2022

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).
- מותר להשתמש במחשבון.
- יש לכתוב בעט **בלבד**.
- יש לכתוב את התשובות **על גבי שאלון זה**.
- מותר לענות בעברית או באנגלית.
- קריאוּת:
  - תשובה בכתב יד לא קריא – **לא תיבדק**.
  - בשאלות רב-ברירה – הקפידו להקיף את התשובות בבירור. סימונים לא ברורים יביאו לפסילת התשובה.
  - לא יתקבלו ערעורים בנושא.
- 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: Feature selection and Classification models [20 נק']

נתון סֵט אימון עם $d \geq 3$ מאפיינים (features) ועם תיוגים בינאריים, משמע $\boldsymbol{x}_i \in \mathbb{R}^d, y_i \in \{-1,1\}$.

בסט האימון אין שתי דוגמאות זהות.

נתון שהקורלציה בין שני מאפיינים מסוימים A, B בסט האימון היא בדיוק -1.

משמע, לכל דוגמה $\boldsymbol{x}_i$ מתקיים $x_i[B] = a x_i[A] + b$ עבור $a < 0$ וְ-$b \in \mathbb{R}$.

לומדים שני מסווגים:

- בשלב הראשון: לומדים מסווג על סט האימון המקורי ומחשבים עליו את דיוק האימון.
- בשלב השני:
  - מסירים את המאפיין B ומקבלים סט אימון מעודכן (שבו $d-1$ פיצ'רים).
  - מאמנים מסווג חדש על סט האימון המעודכן, ומחשבים עליו את דיוק האימון המעודכן.

עבור כל אלגוריתם למידה, סמנו האם דיוק האימון של המסווג החדש על סט האימון המעודכן זהה בהכרח לזה של המסווג המקורי על סט האימון המקורי.

הסבירו בקצרה את תשובותיכם (2-4 משפטים בכל סעיף).

הניחו שאין צעדים סטוכסטיים (אקראיים) בריצת האלגוריתמים ואין שגיאות נומריות (בפרט, בעיות קמורות מתכנסות לפיתרון האנליטי שלהן במדויק).

א. kNN עם $k = 1$ (דוגמה לא נחשבת שכנה של עצמה). דיוק האימון זהה בהכרח? **כן / לא**

הסבר: ______________________________

____

____

____

ב. Hard-SVM ליניארי לא הומוגני בהנחה שהדאטה המקורי פריד. דיוק האימון זהה בהכרח? **כן / לא**

הסבר: ____________________

____

____

____

ג. Soft-SVM לינארי לא הומוגני עם $\lambda = 1$ בהנחה שהדאטה המקורי פריד. דיוק האימון זהה בהכרח? **כן / לא**

הסבר: ____________________

____

____

____

ד. ID3 המשתמש באנטרופיה ובונה עץ בעומק מירבי 4. דיוק האימון זהה בהכרח? **כן / לא**

הסבר: ____________________

____

____

____

שאלה 2: Kernel SVM [18 נק']

<u>תזכורת</u>: פונקציה מהווה קרנל חוקי אם ניתן לכתוב אותה בתור $K(u,v) = \langle \phi(u), \phi(v) \rangle = \sum_{i=1}^{n} \phi_i(u)\phi_i(v)$ עבור מיפוי כלשהו $\phi: \mathcal{X} \to \mathbb{R}^n$.

נתונות שתי פונקציות קרנל חוקיות $K_1, K_2: (\mathcal{X} \times \mathcal{X}) \to \mathbb{R}$ והמיפויים המתאימים להן $\phi, \psi: \mathcal{X} \to \mathbb{R}^n$. משמע, מתקיים:
$K_1(u,v) = \langle \phi(u), \phi(v) \rangle, \qquad K_2(u,v) = \langle \psi(u), \psi(v) \rangle$.

הוכיחו שהפונקציות הבאות מהוות קרנלים חוקיים (כפי שמוסבר בתזכורת לעיל).

בכל סעיף, הגדירו <u>בבירור</u> פונקציית מיפוי $\phi': \mathcal{X} \to \mathbb{R}^{n'}$ מתאימה והראו שמתקיים $K'(u,v) = \langle \phi'(u), \phi'(v) \rangle$ כנדרש.

א. [5 נק'] הפונקציה $K'(u,v) = \left(K_1(u,v)\right)^2$.

הוכחה:

$K'(u,v) = (K_1(u,v))^2 = \sum_{i=1}^{n} \varphi_i(u)\varphi_i(v) \sum_{j=1}^{n} \varphi_j(u)\varphi_j(v) = \sum_{i=1}^{n}\sum_{j=1}^{n} (\varphi_i(u)\varphi_i(v))(\varphi_j(u)\varphi_j(v)) \Rightarrow$

$\varphi'(u) = (\varphi_1(u)\varphi_1(u), \varphi_1(u)\varphi_2(u), \dots, \varphi_1(u)\varphi_n(u), \varphi_2(u)\varphi_1(u), \dots, \varphi_2(u)\varphi_n(u), \dots \varphi_n(u)\varphi_n(u))$

$\varphi': \mathcal{X} \to \mathbb{R}^{n^2}$

ב. [5 נק'] הפונקציה $K'(u,v) = K_1(u,v) + 3 \cdot K_2(u,v) + 1$.

הוכחה:

$K'(u,v) = \sum_{i=1}^{n} \varphi_i(u)\varphi_i(v) + 3\sum_{i=1}^{n} \psi_i(u)\psi_i(v) + 1 =$

$= \sum_{i=1}^{n} [\varphi_i(u)\varphi_i(v) + 3\psi_i(u)\psi_i(v)] + 1 =$

$\varphi'(u) = [\varphi(u), \sqrt{3}\,\psi(u), 1]$

ג. [8 נק'] אילו מהפעולות הבאות אמורות <u>להפחית</u> overfitting ב-Kernel SVM?

סמנו את <u>כל</u> התשובות המתאימות (השאלה אינה עוסקת במקרי קצה אלא במקרה הסביר).

~~a.~~ לעבור מפונקציית הקרנל $K(\boldsymbol{u},\boldsymbol{v}) = (\boldsymbol{u}^\top \boldsymbol{v})^p$ לפונקציית הקרנל $K(\boldsymbol{u},\boldsymbol{v}) = (\boldsymbol{u}^\top \boldsymbol{v} + 1)^p$, כאשר $p \in \mathbb{N}_{\geq 2}$.

$(u^T v + 1)(u^T v + 1)$

$= u^2 v^2 + 2uv + 1$

$\varphi(u) = (u^2, \sqrt{2}u, 1)$

$(uv)^2 = u^2 v^2$ $\quad \varphi(u) = (u^2)$

$\varphi(u) = u^p$

~~b.~~ להקטין את השונות σ^2 של קרנל RBF (שמוגדר בתור $K(\boldsymbol{u},\boldsymbol{v}) = \exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{u}-\boldsymbol{v}\|_2^2\right\}$).

$\sigma^2 \to \infty \quad e^0 = 1 \qquad \sigma^2 \to 0 \quad e^{-\infty \cdot \|u-v\|} = 0$ 1מין

~~c.~~ לפתור (במדויק) את הבעיה ה-primal במקום לפתור (במדויק) את הבעיה ה-dual.

d. להגדיל את מקדם הרגולריזציה λ (בהתאמה: להקטין את C).

e. להגדיל את סט האימון (באופן $i.i.d.$ מאותה התפלגות של הדאטה המקורי).

~~f.~~ להגדיל את סט המבחן (באופן $i.i.d.$ מאותה התפלגות של הדאטה המקורי).

שאלה 3: VC-dimension [16 נק']

נגדיר את מחלקת ההיפותזות של מסווגים ליניאריים הומוגניים ב-d ממדים:

$$\mathcal{X} = \mathbb{R}^d,\ \ \mathcal{H}_{\text{lin}}^d = \{\boldsymbol{x} \mapsto \text{sign}(\boldsymbol{w}^\top \boldsymbol{x}):\ \boldsymbol{w} \in \mathbb{R}^d\}.$$

כדי למנוע טעויות, נקבע שלאורך כל השאלה $\text{sign}(0) = 0$ (ואחרת הפונקציה מחזירה ± 1 בהתאם לסימן).

א. [7 נק'] הוכיחו שמתקיים $\text{VCdim}(\mathcal{H}_{\text{lin}}^d) \geq d$.

הוכחה:

האינדקס ה i

יהיו $\forall i \in [d]:\ X_i = [0,\dots,0,1,0,\dots,0]$

לכל תיוג אפשרי שלהן $y_i \in \{\pm 1\}$ נגדיר $\forall i \in [d]:\ w_i := y_i$

ונקבל $w = (w_1, w_2, \dots, w_d)$ נשים לב כי $\forall i \in [d]\ \ y_i w^T x_i = y_i (y_1 y_2 \dots y_i \dots y_d) \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = y_i^2 > 0$

i-th index

ובפרט $\sum_{i=1}^{d} (1 - y_i w^T x_i) = 0$

כלומר ה loss הוא 0, סיווגנו את כולם נכון ולכן עבור כל תיוג קיים מסווג לינארי הומוגני

אשר שייך למחלקה אשר מסווגת את כל הנק' נכון $\Rightarrow$ ניתצנו את המחלקה $\Rightarrow$ $VCdim(H^d_{lin}) \geq d$

ב. [9 נק'] הוכיחו שממד ה-VC הוא בדיוק d על ידי כך שתוכיחו שמתקיים $\text{VCdim}(\mathcal{H}_{\text{lin}}^d) < d + 1$.

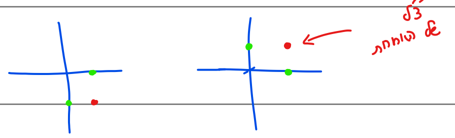
<u>רמז</u>: כל אוסף $\boldsymbol{x}_1, \dots, \boldsymbol{x}_d, \boldsymbol{x}_{d+1}$ של $d + 1$ וקטורים כלשהם ב-$\mathbb{R}^d$ הינו תלוי ליניארית, ובהכרח אחד הווקטורים באוסף (בה"כ $\boldsymbol{x}_{d+1}$) מקיים $\boldsymbol{x}_{d+1} = \sum_{i=1}^{d} z_i \boldsymbol{x}_i$ עבור אוסף סקלרים $z_1, \dots, z_d \in \mathbb{R}$ שלפחות אחד מהם שונה מאפס.

הוכחה (לרשותכם דפי טיוטה בסוף הגיליון):

(לכל סידור נק׳ כלשהו, יש תיוג כך ש[illegible] לא ניתן לנתץ)

יהי $x_1, \dots, x_d, x_{d+1}$ נק׳ במרחב $\mathbb{R}^d$, לפי הרמז ובה״כ $x_{d+1} = \sum_{i=1}^{d} z_i x_i$ ו-z_i כמתוארים

$$y_i := \begin{cases} sign(z_i) & i \le d \wedge z_i \ne 0 \\ 1 & i \le d \wedge z_i = 0 \\ -1 & i = d+1 \end{cases}$$



נניח כי קיים $w \in \mathbb{R}^d$ המתייג נכון את כל הנק׳ כלומר $\forall i: sign(w^T x_i) = y_i$

$$sign(w^T x_{d+1}) = sign\left(w^T \sum_{i=1}^{d} z_i x_i\right) = sign\left(\sum_{i=1}^{d} \underbrace{z_i w^T x_i}_{\ge 0}\right) = 1 \ne -1 = y_{d+1}$$

כלומר הגענו לסתירה להגדרה של y_{d+1} ולכן לא קיים w כזה

כלומר עבור כל $x_1, \dots, x_{d+1}$ כלשהם ניתן למצוא תיוג כך שלא ניתן לנתצו באמצעות מחלקת ההיפותזות שלנו. ולכן $VCdim(\mathcal{H}^d_{lin}) < d+1$

שאלה 4: Bagging and Boosting [22 נק']

משמאל מופיע אלגוריתם AdaBoost.

Initialize $D^{(1)} = \left(\frac{1}{m}, \dots, \frac{1}{m}\right)$

For t=1, …, T:

$$h_t = \mathcal{A}\left(S, D^{(t)}\right)$$

$$\epsilon_t = \sum_i D_i^{(t)} \cdot \mathbf{1}_{h_t(x_i) \neq y_i}$$

$$\alpha_t = \frac{1}{2} \log\left(\frac{1}{\epsilon_t} - 1\right)$$

$$D_i^{(t+1)} = \frac{1}{Z_t} D_i^{(t)} \exp\left(-\alpha_t y_i h_t(x_i)\right)$$

$$h_s(x) = \operatorname{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

א. [5 נק'] שימו לב שבכל איטרציה האלגוריתם מעדכן את ההתפלגות באופן $D_i^{(t+1)} = \frac{1}{Z_t} D_i^{(t)} \exp\left(-\alpha_t y_i h_t(x_i)\right)$ כאשר $Z_t \triangleq \sum_{j=1}^{m} D_j^{(t)} \exp\left(-\alpha_t y_j h_t(x_j)\right)$ הוא גורם נירמול.

הוכיחו שמתקיים $D_i^{(t+1)} = c \cdot \exp\left(-y_i \sum_{k=1}^{t} \alpha_k h_k(x_i)\right)$ עבור קבוע כלשהו $c > 0$ (שתלוי ב-t).

הוכחה (לרשותכם דפי טיוטה בסוף הגיליון):

$$D_i^{(t+1)} = \frac{1}{Z_t} D_i^{(t)} e^{-\alpha_t y_i h_t(x_i)} = \frac{1}{Z_t Z_{t-1}} D_i^{(t-1)} e^{-y_i \alpha_t h_t(x_i) - y_i \alpha_{t-1} h_{t-1}(x_i)} = \dots$$

$$= \frac{1}{\prod_{i=1}^{t} Z_i} e^{-y_i \sum_{k=1}^{t} \alpha_k h_k(x_i)}$$

בסעיף הבא נוכיח שהבחירה של האלגוריתם ב-$\frac{1}{2}\log\left(\frac{1}{\epsilon_t}-1\right)$ $\alpha_t =$ הינה אופטימלית.

<u>תזכורת</u>: הגדרנו בתרגול את ההיפותזה ה-unthresholded ה"חזקה" באיטרציה t בתור: $h_s^{(t)} = \sum_{k=1}^{t}\alpha_k h_k(x_i)$,

והראינו שקיים קבוע חיובי C כך שערך ה-exp. loss של ההיפותזה הוא:

$$\mathcal{L}_{\exp}\left(h_s^{(t)}\right) \triangleq \frac{1}{m}\sum_{i=1}^{m}\exp\{-y_i\sum_{k=1}^{t}\alpha_k h_k(x_i)\} = \underbrace{C}_{>0}\cdot\left(e^{-\alpha_t}+(e^{\alpha_t}-e^{-\alpha_t})\overbrace{\sum_i D_i^{(t)}1_{h_t(x_i)\neq y_i}}^{\varepsilon_t}\right),$$

כאשר C אינו תלוי בבחירה של המקדם α_t שנבחר בזמן t.

ב. [5 נק'] הוכיחו שבהינתן h_t וְ-$\left\{D_i^{(t)}\right\}_{i\in[m]}$, הבחירה בְּ-$\alpha_t = \frac{1}{2}\log\left(\frac{1}{\epsilon_t}-1\right)$ ממזערת את $\mathcal{L}_{\exp}\left(h_s^{(t)}\right)$.

<u>הערה</u>: שימו לב ש-$\epsilon_t \triangleq \sum_i D_i^{(t)}\cdot\mathbf{1}_{h_t(x_i)\neq y_i}$ והוא קבוע בהינתן h_t וְ-$\left\{D_i^{(t)}\right\}_{i\in[m]}$.

הוכחה:

$$\frac{\partial \mathcal{L}_{\exp}(h_s^{(t)})}{\partial \alpha_t} = \cancel{C}\cdot\left(-e^{-\alpha_t}+\varepsilon_t e^{\alpha_t}+\varepsilon_t e^{-\alpha_t}\right) = 0 \quad /\cdot e^{\alpha_t}$$

$$-1+\varepsilon_t e^{2\alpha_t}+\varepsilon_t = 0 \Rightarrow e^{2\alpha_t} = \frac{1-\varepsilon_t}{\varepsilon_t} \Rightarrow 2\alpha_t = \log\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \Rightarrow$$

$$\Rightarrow \alpha_t = \frac{1}{2}\log\left(\frac{1}{\varepsilon_t}-1\right)$$

הסעיף הבא לא קשור לאלגוריתם AdaBoost אלא לשיטת Ensemble שממשקלת היפותזות באופן כללי.

ג. [12 נק'] נתון סֶט אימון $\{(x_i, y_i)\}_{i=1}^{m}$ עם סיווגים בינאריים ($y_i \in \{-1, +1\}$).

צוות של 50 מומחים ומומחיות הגדיר ידנית 50 היפותזות בינאריות $h_1, \dots, h_{50} : \mathcal{X} \to \{-1, +1\}$.

המטרה שלנו היא למשקל אותן, משמע, ליצור היפותזה $h_\alpha(x) = \operatorname{sign}\left(\sum_{k=1}^{50} \alpha_k h_k(x)\right)$.

כדי למנוע טעויות, נקבע שלאורך כל השאלה $\operatorname{sign}(0) = 0$ (ואחרת הפונקציה מחזירה ± 1 בהתאם לסימן).

<u>נתון</u>: קיימים משקלים $\alpha_1^*, \dots, \alpha_{50}^* \in \mathbb{R}$ כך שהשגיאה האמפירית (שגיאת האימון) היא אפס.

כעת, נרצה <u>ללמוד</u> אוסף משקלים $\alpha_1, \dots, \alpha_{50} \in \mathbb{R}$.

(i) נסחו בעיית אופטימיזציה מתאימה **וקמורה** (ביחס למשקלים הנלמדים $\alpha_1, \dots, \alpha_{50}$) כך שהפיתרון שלה ממזער את השגיאה האמפירית.

תשובה (לרשותכם טיוטה בסוף הגיליון):

$h(x) = [h_1(x), \dots, h_{50}(x)]^T \in \{\pm 1\}^{50}$ $\alpha = [\alpha_1, \dots, \alpha_{50}] \in \mathbb{R}^{50}$

$\cancel{\underset{\alpha \in \mathbb{R}^{50}}{\arg\min} \sum_{i=1}^{m} \max\{0, 1 - y_i \sum_{k=1}^{50} \alpha_k h_k(x_i)\}} = \underset{\alpha \in \mathbb{R}^{50}}{\arg\min} \sum_{i=1}^{m} \max\{0, 1 - y_i \alpha^T h(x_i)\}$

(ii) נמקו <u>בקצרה</u> מדוע הבעיה שהגדרתם קמורה (מומלץ להשתמש בטענות מההרצאה ומהתרגול).

תשובה (לרשותכם טיוטה בסוף הגיליון):

כיוון הביטויים בתוך ה max קמורים, & max קמורים מבוין קמור

סכימה של קמורים מבוין קמור

(iii) יהי אוסף המשקלים $\alpha_1, \dots, \alpha_{50} \in \mathbb{R}$ פיתרון (אופטימלי) של הבעיה הקמורה שהגדרתם לעיל.

הוכיחו שהשגיאה האמפירית של פיתרון זה היא מינימלית.

הוכחה:

קיים α^* עבור $sign(\alpha^{*T}h(x_i)) = y_i$ ולכן $y_i\alpha^{*T}h(x_i) > 0$

מובטח שקיים פתרון $\beta\alpha^*$ עבור ערך β כלשהו כך ש

$\forall i: y_i \beta\alpha^{*T}h(x_i) > 1$ שמה שמאפס את ה Objective

כלומר ובפרט גם השגיאה האמפירית תתאפס ($\frac{1}{m}\sum \mathbb{I}(h(x) \neq y_i)$)

חלק ב' – שאלות אמריקאיות [24 נק']

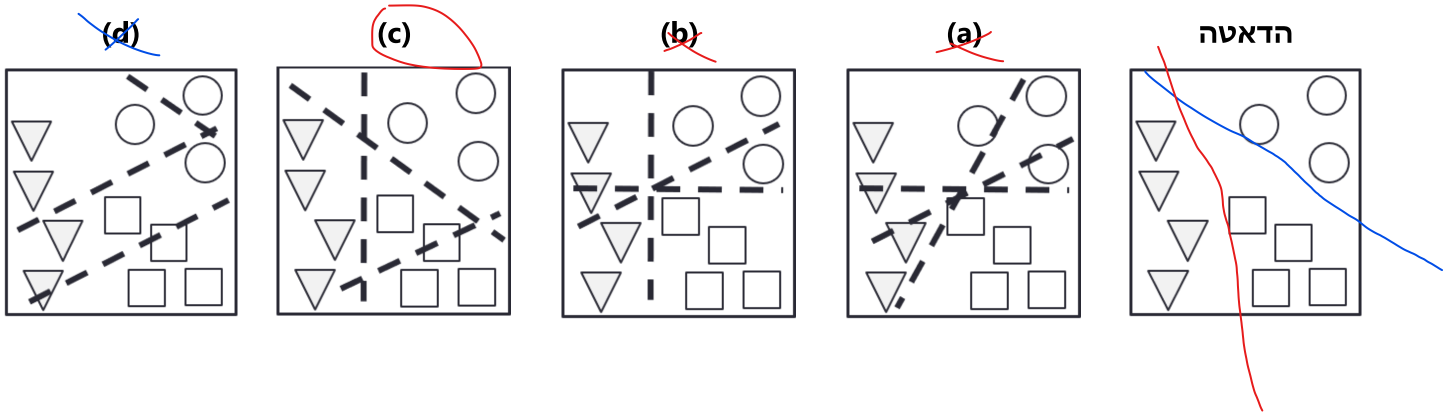
בשאלות הבאות סמנו את התשובות המתאימות (לפי ההוראות). בחלק זה אין צורך לכתוב הסברים.

א. [6 נק'] נתון דאטה דו-ממדי עם שלושה תיוגים אפשריים – משולש, מעגל וריבוע.

מפעילים One-vs-All (OVA) עם פרספטרון בתור מסווג בסיס. לכל בעיה בינארית מריצים את הפרספטרון מספיק איטרציות כך שאם הבעיה פרידה ליניארית – הפרספטרון יתכנס (חישבו מה קורה אחרת).

בכל תרשים מצוירים שלושה גבולות החלטה שאמורים לתאר את הגבולות שהתקבלו ע"י מסווגי הבסיס (הבינאריים).

הקיפו את האות שמתאימה לתרשים היחיד שמתאר גבולות החלטה שיכולים להתקבל ע"י השיטה שתוארה לעיל.



ב. [6 נק'] נתון דאטה d-ממדי $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}$ עם סיווגים בינאריים (±1).

סמנו את כל הטענות הנכונות **בהכרח** ביחס **לשיטות להורדת ממד** לממד נמוך כלשהו.

הבהרה: זו שאלת חשיבה ולא שאלת טריוויה.

a. הדאטה המקורי פריד ליניארית $\Longleftarrow$ הדאטה פריד ליניארית לאחר הורדת ממד ליניארית עם PCA.

b. הדאטה המקורי פריד ליניארית $\Longleftarrow$ בהסתברות גבוהה, הדאטה פריד ליניארית לאחר הורדת ממד ליניארית עם Random Projections.

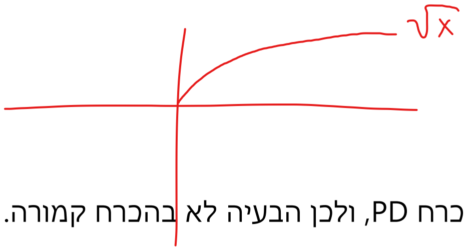
c. הדאטה המקורי פריד ליניארית $\Longleftarrow$ הדאטה פריד ליניארית לאחר הורדת ממד עם Kernel-PCA (עם קרנל לא ליניארי).

d. הדאטה פריד ליניארית לאחר הורדת ממד ליניארית עם PCA $\Longleftarrow$ הדאטה המקורי פריד ליניארית.

e. PCA היא שיטה Supervised בעוד ש-Random Projections היא שיטה Unsupervised.

ג. [6 נק'] נתונה בעיית רגרסיה לינארית (עבור $\lambda \geq 0,\ p \in \{0.5, 1, 2\}$):

$$\text{argmin}_{\boldsymbol{w}}\left(\frac{1}{m}\sum_{i=1}^{m}(y_i - \boldsymbol{w}^\top \boldsymbol{x}_i)^2 + \lambda\|\boldsymbol{w}\|_p^p\right) = \text{argmin}_{\boldsymbol{w}}\left(\frac{1}{m}\|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{w}\|_p^p\right).$$



סמנו את <u>כל</u> הטענות הנכונות.

<u>הערה</u>: בכל הסעיפים, שואלים על קמירות ביחס ל-$\boldsymbol{w}$.

<u>תזכורת</u>: $\|\boldsymbol{w}\|_p^p \triangleq \sum_{k=1}^{d}|w_k|^p$.

a. כאשר $\lambda = 0$: ההסיאן הוא $\frac{2}{m}\boldsymbol{X}^\top\boldsymbol{X}$ שהיא מטריצה PSD אבל לא בהכרח PD, ולכן הבעיה לא בהכרח קמורה.

b. כאשר $\lambda > 0$ וגם $p = 2$: הבעיה קמורה בהכרח.

c. כאשר $\lambda > 0$ וגם $p = 1$: הבעיה קמורה בהכרח.

d. כאשר $\lambda > 0$ וגם $p = 0.5$: הבעיה קמורה בהכרח.

e. כאשר $\lambda > 0$ וגם $p = 1$: הבעיה אינה גזירה ביחס ל-$\boldsymbol{w}$ ולכן לא ניתן להפעיל שיטות גרדיינט (או סאב-גרדיינט).

ד. [6 נק'] רוצים לאמן רשת נוירונים בתצורת Autoencoder.

נסמן את אוסף כל הפרמטרים של הרשת בתור Θ. לכל דוגמה $\boldsymbol{x}_i \in \mathbb{R}^d$ נסמן את הפלט של הרשת בתור $\hat{\boldsymbol{x}}_i \in \mathbb{R}^d$.

בהרצאה ראינו שניתן לאמן את הרשת ע"י פיתרון הבעיה הבאה: $\text{argmin}_\Theta\left(\lambda\|\Theta\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|_2^2\right)$.

אילו מבין הבעיות הבאות עשויות גם כן להתאים לאימון רשת Autoencoder (עבור $\lambda > 0$)?

סמנו את <u>כל</u> התשובות המתאימות.

a. $\text{argmin}_\Theta\left(\lambda\|\Theta\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i + \hat{\boldsymbol{x}}_i\|_2^2\right)$

b. $\text{argmin}_\Theta\left(\lambda\|\Theta\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\exp\{\|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|_2^2\}\right)$

c. $\text{argmin}_\Theta\left(\lambda\|\Theta\|_2^2 - \frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|_2^2\right)$

d. $\text{argmin}_\Theta\left(\lambda\|\Theta\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\max\{0, 1 - \text{sign}(\boldsymbol{x}_i^\top\hat{\boldsymbol{x}}_i)\}\right)$

e. $\text{argmin}_\Theta\left(-\lambda\|\Theta\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\frac{\boldsymbol{x}_i^\top\hat{\boldsymbol{x}}_i}{\|\boldsymbol{x}_i\|_2\|\hat{\boldsymbol{x}}_i\|_2}\right)$

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):