



מבוא למערכות לומדות (236756)

סמינר חורף תשפ"ב – 10 בפברואר 2022

מרצה: ד"ר יונתן בלינקוב

מבחן מסכם מועד א'

הנחיות הבוחינה:

- **משך הבדיקה:** 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.
- מותר לענות בעברית או באנגלית.
- יש לכתוב את התשובות **על גבי שאלון זה** בכתב יד קרייא. תשובה בכתב יד לא קרייא – לא תיבדק.
- במבחן 16 עמדים ממוספריםסה"כ, כולל שער זה שמספרו 1 ושלושה עמדים טיווה בסוף הגילוון.
- נא לכתוב רק את המבוקש ולצף הסבירים קיצרים עפ"י הנחיות.
- **בתום המבחן יש להגיש את שאלון זה בלבד.**

מבנה הבדיקה:

- **חלק א' [75 נק']:** 4 שאלות פתוחות.
- **חלק ב' [25 נק']:** 5 שאלות סגורות (אמריקאיות) [כל אחת 5 נק'].

בהצלחה!

חלק א' – שאלות פתוחות [75 נק']

שאלה 1 [16 נק']

חוקרת מהטכניון עבדת על בעית סיווג ביןאי כלשהי. ברשותה dataset שבו $150 = m$ דוגמאות שונות (distinct). החקירה הריצה שלושה מודלים, וכל מודל ביצעה hyperparameter tuning:

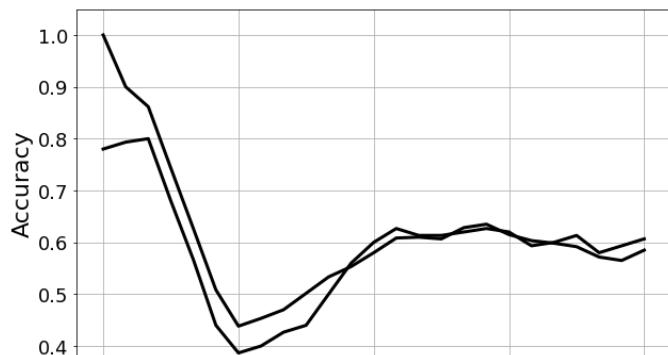
- (a) מודל: kNN (נק' נחשבת שכנה של עצמה), היפרפרמטר: מספר השכנים k , טווח: 1 עד 97.
- (b) מודל: עץ החלטה, היפרפרמטר: עומק מירבי, טווח: 1 עד 40.
- (c) מודל: Kernel SVM, היפרפרמטר: חזק הרגולרייזציה γ , טווח: 10^{-3} עד 10^7 .

לכל מודל, היא ציירה גרף של דיק האימון ודיק ההכללה (בעזרת cross validation 5-fold) בציר y כפונקציה של ערך היפרפרמטר בציר x (הערכים גדלים משמאלי לימין).

בקבוק תקלה, הכתיב על ציר x נמחק מכל הגרפים.

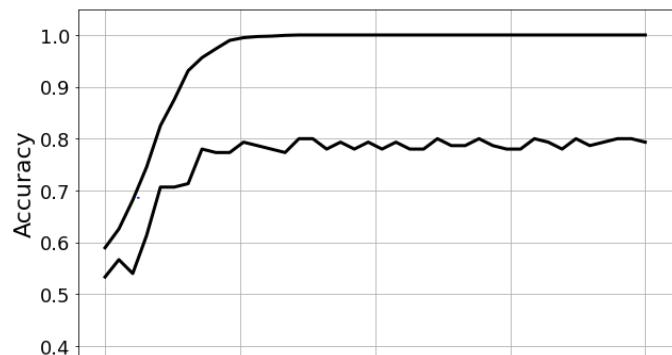
בנוסף, הגרפים נשמרו בטיעות בשחור לבן, כך שלא ניתן להבדיל בקלות בין דיק האימון לדיק ההכללה. א. [8 נק'] במקום המתאים מתחתם, כתבו את אותן שמות שמתאימה למודל ולהיפרפרמטר שיצרו אותו. הערה: **השאלה לא מוגדרת היטב ובדיעבד היו שתי תשובות נכונות לשאלה זו.**

תרשים ב'



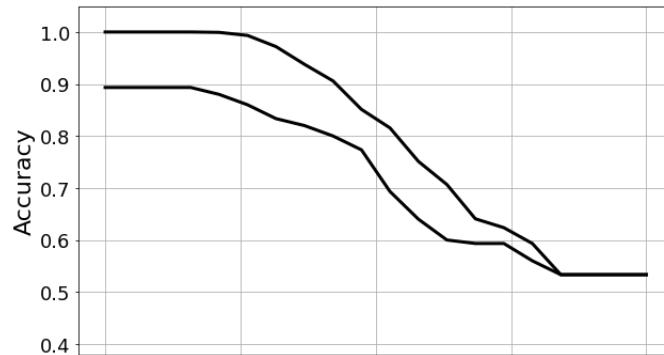
a

תרשים א'



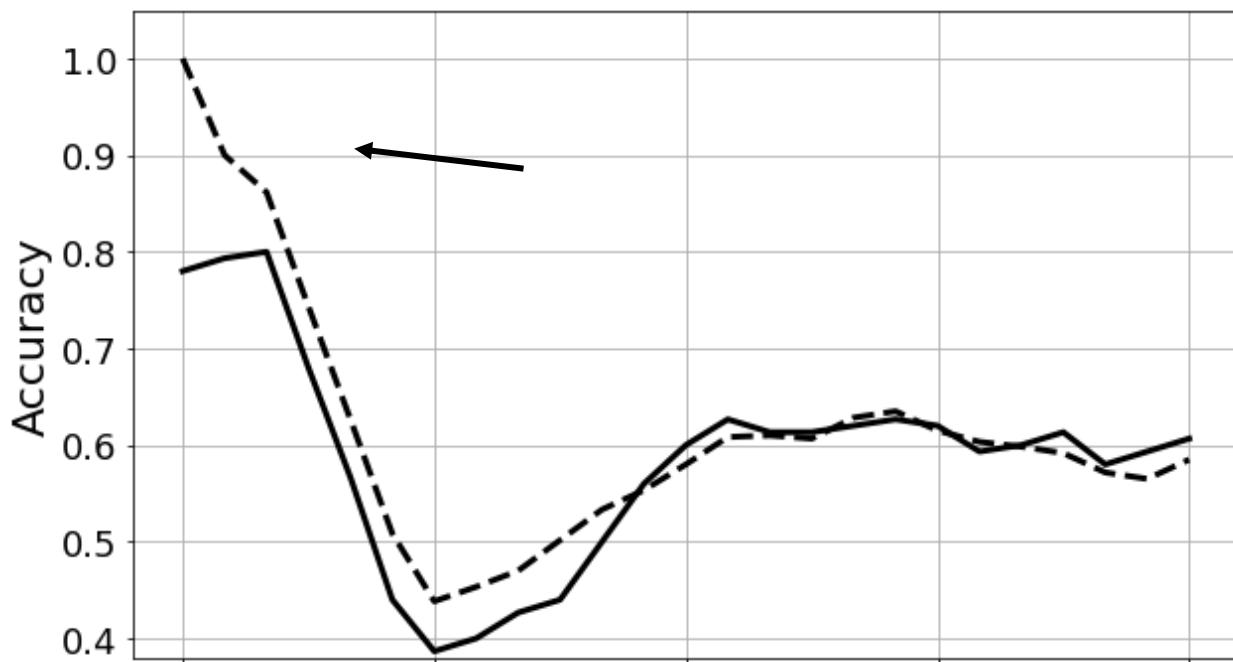
b

תרשים ג'



c

ב. [8 נק'] להלן תרשימים ב' מוגדל.



הסתכלו על העקומה המקבוקות שבתרשים (מסומנת בחצ'). האם העקומה מתארת את דיק האימון או את דיק ההקללה? הסבירו בקצרה. התבססו על התרשים ועל מאפייני המודל שיצר את עקומה זו (מבין שלושת המודלים).

תשובה תמציתית:

דיק האימון/accurace כפולה (K=1) מהתוצאות שצרכנו להציג בפונקציית loss, וכך ה"מגזר" (רכירן) הוא הוליגומטric accuracy (כגיא 100%)

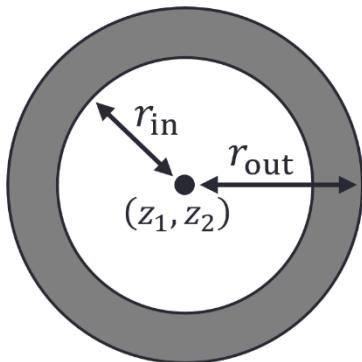
שאלה 2 – 21] PAC learning – נק' 21]

תהי \mathcal{H} מחלקת היפותזות של Bagels/donuts בדו-ממד:

$$\mathcal{H} = \{h_\theta: \mathbb{R}^2 \rightarrow \{\pm 1\} \mid \theta = (z_1, z_2, r_{\text{out}}, r_{\text{in}}), r_{\text{out}} > r_{\text{in}} \geq 0\}$$

כאשר היפותזה בודדת מוגדרת באופן הבא:

$$h_\theta(x) = \begin{cases} +1, & r_{\text{out}} \geq \sqrt{(x_1 - z_1)^2 + (x_2 - z_2)^2} \geq r_{\text{in}} \\ -1, & \text{otherwise} \end{cases}$$



דגשים לגבי בל היפותזה $\mathcal{H} \in h_\theta$:

o המרכזים של המעגלים משותפים ולא בהכרח בראשית הצירים.

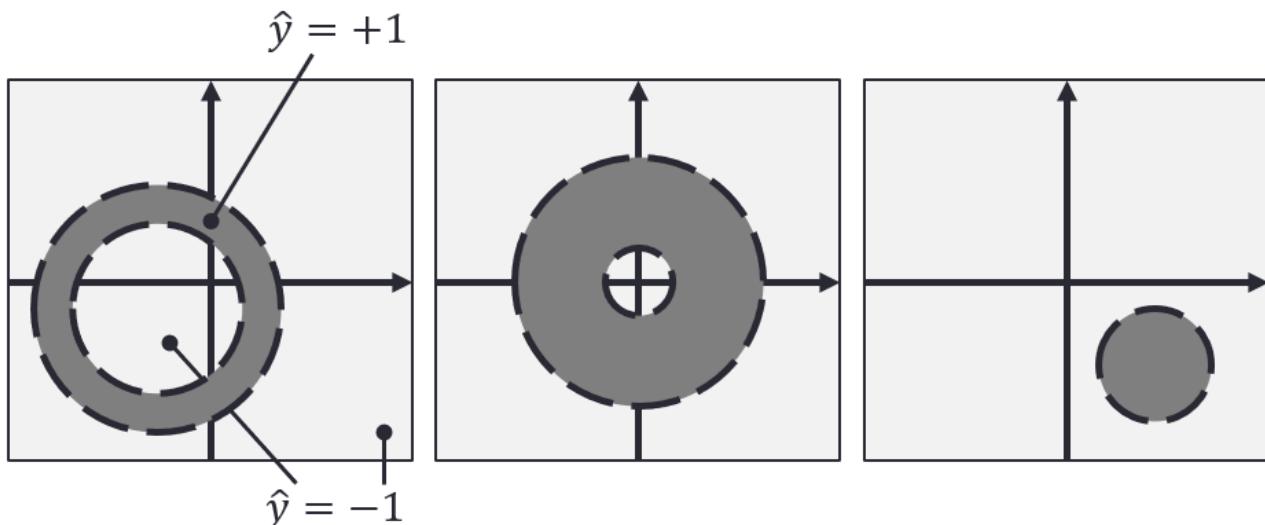
o הרדיוס של המעגל הפנימי יכול להיות אפס.

o השטח שבתוך ה-donut לא יכול להיות אפס.

o האזור בין שני המעגלים מסוג כחיווי, והאזורים האחרים כשליליים.

o מדובר אך ורק במעגלים ולא באלייפסות.

דוגמה לשולוש היפותזות מתוך \mathcal{H} :



א. [3 נק'] להלן ההגדרה של "ניתוץ". השמננו מההגדרה את הכפניות.

השלימו את שלושת הטענות החסרים. בכל מקום כתבו האם חסר בהגדירה A או E.

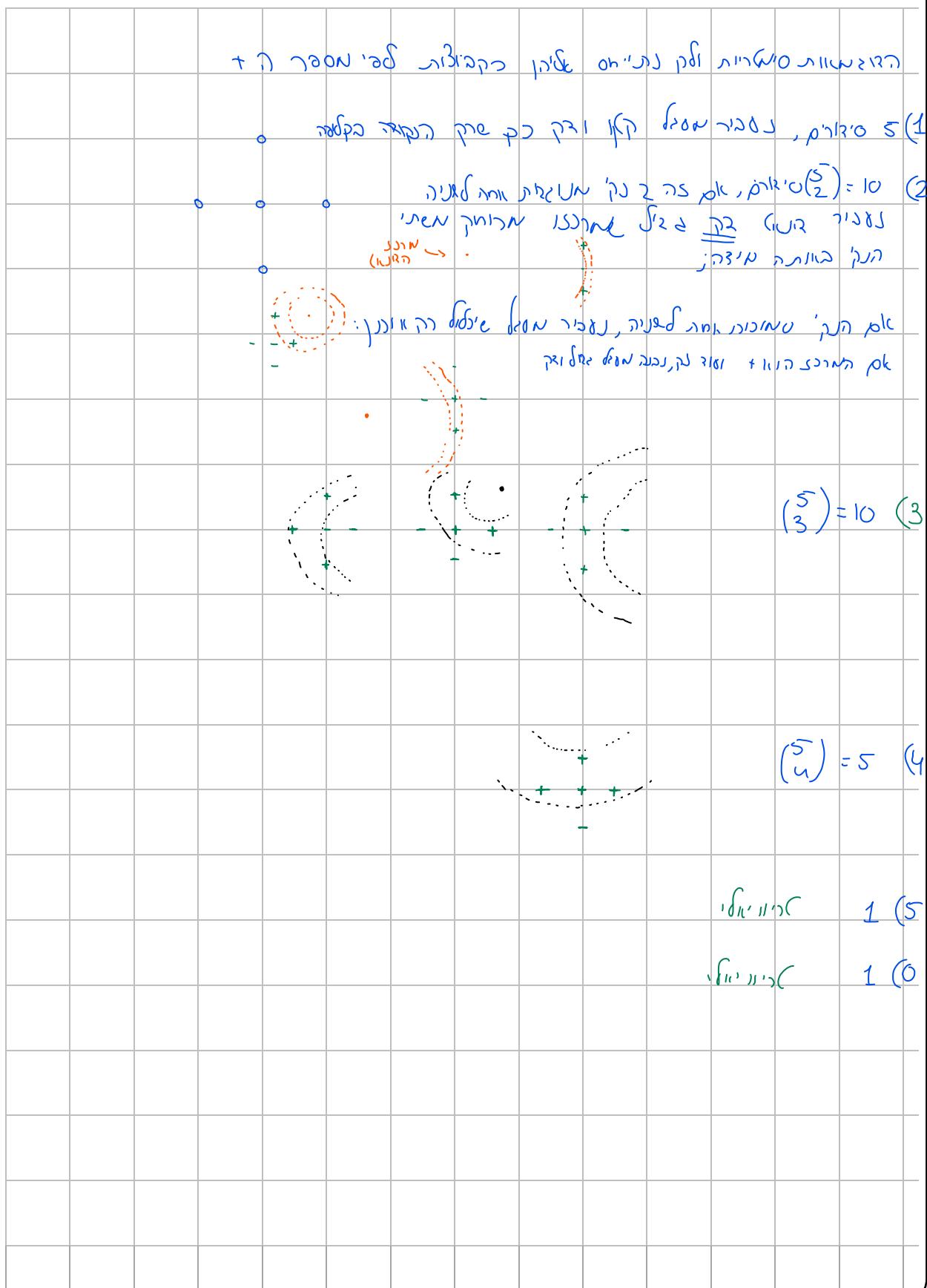
$$\mathcal{H} \text{ shatters } C \Leftrightarrow \underbrace{\bigwedge_{\substack{\text{השלימו}}} y_1, \dots, y_{|C|} \in \mathcal{Y}:}_{\substack{\text{השלימו}}} \underbrace{\bigvee_{\substack{\text{השלימו}}} h \in \mathcal{H}:}_{\substack{\text{השלימו}}} \underbrace{\bigwedge_{\substack{\text{השלימו}}}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

$$\text{VCdim}(\mathcal{H}) \geq \boxed{5}$$

ב. [13 נק'] כתבו את החסם התחתון **ההduk ביוור** שתוכלו למצוא לממד ה-VC:

הוכיחו את החסם התחתון שכתבתם (אין להוכיח שוויזן). יש לכתבו הסבר מיולי תמציתי ולצרכו תרשימים נדרשים.

הוכחה (לרשוטכם דפי טויטה בסוף הגילון):



המשך סעיף ב':

ג'. [5 נק'] חוקרת וחוקר רוצחים לאמן מודל סיוג בינהרי.

החוקרת משתמשת בחלוקת היפותזות H שהגדנו.

החוקק משתמש בחלוקת היפותזות של donuts שמרכזם בראשית הצירים, משמעו:

$$\mathcal{H}' = \{h_\theta \mid \theta = (0, 0, r_{\text{out}}, r_{\text{in}}), r_{\text{out}} > r_{\text{in}} \geq 0\} \subset \mathcal{H}$$

מי צפוי להזדקק לפחות דוגמאות בתרגיל הלמידה ע"מ להבעית (במוני ח' PAC) שגייאת הכללה $0.1 = \epsilon$? נמקו בקצרה.

תשובה תמציתית:

$$m_H(\varepsilon, \delta) = \Theta\left(\frac{VC(H) \log\left(\frac{1}{\delta}\right) + \log\left(\frac{1}{\varepsilon}\right)}{\varepsilon}\right)$$

~~שאלה 3 – רגרסיה ליניארית ו-*s*-models [21 נק']~~

נתון דאטה $S = \{(x_i, y_i)\}_{i=1}^m$ שהגיא ממודל ליניארי $\epsilon_i \sim N(0, 1)$ נורמלי. רעש אקראי מפilog. i.i.d. עם $y_i = \mathbf{w}^\top \mathbf{x}_i + \epsilon_i$.

שימנו לב: הדוגמאות $\mathbb{R}^d \in \mathbf{x}_i$ והתיוגים $\mathbb{R} \in y_i$ נתונים. וקטור המשקלים $\mathbb{R}^d \in \mathbf{w}$ לא ידוע והוא רצוי למדוד.

תזכורת: הוכחנו שתחת הנחות אלה ה-likelihood שווה ל:

$$L(\mathbf{w} | \{\mathbf{x}_i, y_i\}_i) = \Pr(\{\mathbf{x}_i, y_i\}_i | \mathbf{w}) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(\mathbf{w}^\top \mathbf{x}_i - y_i)^2\right\}$$

א. [5 נק'] הוכחו שתחת הנחות השאלה, בעית ה-LS ללא רגולריזציה, משמע $\arg\min_{\mathbf{w}} \frac{1}{m} \sum_{i=1}^m (\mathbf{w}^\top \mathbf{x}_i - y_i)^2$

שיטה לבועית ה-MLP הבאה: $\arg\max_{\mathbf{w}} L(\mathbf{w} | \{\mathbf{x}_i, y_i\}_i)$

הוכחה ע"י פיתוח תמציתי מנומך:

שימו לב: עדין מניחים שהרשות ϵ מתפלגת גאוסיאנית ככתוב בתחילת השאלה.
 משמע, כל משקל נדגם בז' באופן הבא: $w_k \sim \text{Laplace}(0, b)$, עבור $0 < b$ נתון (משותף לכל המשקלים).
 בעוד, נניח בנוסף **שוקטור המשקלים** הלא ידוע w הגע מהתפלגות לפלאס.

$$Z \sim \text{Laplace}(\mu, b) \Rightarrow f(z) = \frac{1}{2b} \exp\left\{-\frac{1}{b}|z - \mu|\right\}$$

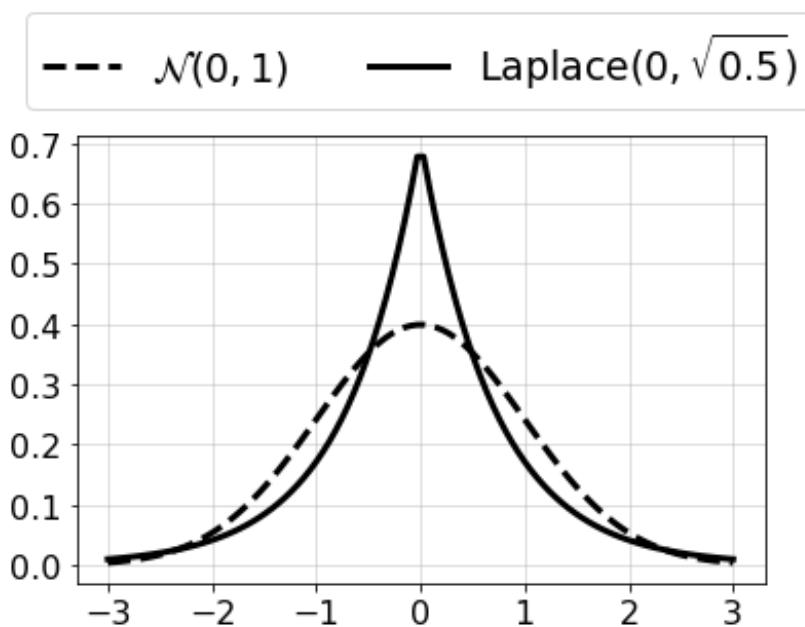
פונקציית הצפיפות של התפלגות לפלאס הינה:

ב. [11 נק'] הוכחו שתחת כל ההנחהות, בעיית LS עם רגוליזציה ℓ^1 , משמע $\lambda \|\mathbf{w}\|_1$, שקיים מינימום ייחודי. שקולה לביעית MAP תחת ה-*prior* על המשקלים, משמע $\Pr_{\mathbf{w}}(\mathbf{w} \mid \{(x_i, y_i)\}_i, \mu = 0, b)$. ניתן להשתמש בחישובים שכבר הוכחו בסעיף הקודם.

הוכחה ע"י פיתוח תמציתי מנומך:

השונות של התפלגות לפלאס נתונה ע"י $\text{Var}[w_k] = 2b^2$.

התרשימים משווה בין התפלגות לפלאס להתפלגות נורמלית שהשונות שלהן היא 1.



תזכורת: הוכחנו בתרגול שאם מניחים $(0,1) \sim \mathcal{N}_k \omega$, בעיית h-MAP שקולה לבעיית LS עם רגולריזציה ℓ^2 .

ג. [5 נק'] מתוך הסתכילות בתרישים, מתוך המזכירות ומתוך מה שהוכחתם בסעיף הקודם, הסבירו בקצרה ובאופן אינטואיטיבי (לא פורמלי) הבדל שלמדנו בין אופי הפיתרון שמתקבלים ע"י רגולריזציה ℓ^2 ללאה המתקבלים ע"י רגולריזציה ℓ^1 .

הסביר קצר:

שאלה 4 – Kernel SVM – נק' 17 [Kernel SVM]

עבור פרמטר $\gamma > 0$, נגדיר את ה-Gaussian kernel לקלט חד-ממדי באופן הבא:

$$K: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, \quad K(a, b) = \exp(-\gamma(a - b)^2) = \varphi(a)^\top \varphi(b)$$

א. [12 נק'] הצעו פונקציית מיפוי $\varphi: \mathbb{R}^p \rightarrow \mathbb{R}^m$ וhoevoico בערותה שהפונקציה K מהויה קרNEL חוקי (בחד ממד).

שימנו לב: עליכם לבחור $\{\infty\} \cup N \in m$ מתאים, סופי או אינסופי.

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$\left[1, \frac{2ab}{1!}, \frac{(2ab)^2 a^2 b^2}{2!}, \dots \right]$$

$$-r^2 + 2ab + b^2$$

]

תזכורת: פירוק טור טילור של e^x לכל x נתון ע"י:

$$e^{-\gamma(a-b)^2} = e^{-\gamma a^2} \cdot e^{2ab} \cdot e^{-\gamma b^2} = e^{-\gamma a^2} \cdot \sum_{n=0}^{\infty} \frac{(2ab)^n}{n!} \cdot e^{-\gamma b^2} =$$

$$= e^{-\gamma a^2} \cdot \sum_{n=0}^{\infty} \frac{(2ab)^n a^n b^n}{n!} e^{-\gamma b^2} = e^{-\gamma a^2} \cdot \sum_{n=0}^{\infty} \left(\frac{(2ab)^{\frac{n}{2}} a^n}{\sqrt{n!}} \cdot \frac{(2ab)^{\frac{n}{2}} b^n}{\sqrt{n!}} \right) e^{-\gamma b^2}$$

$$\varphi_n(a) = e^{-\gamma a^2} \frac{a^n \cdot (2ab)^{\frac{n}{2}}}{\sqrt{n!}}$$

$$\therefore \left[1, 2ab, \frac{4a^2 b^2}{2}, \frac{8a^3 b^3}{6}, \dots \right]$$

$$\varphi: \mathbb{R} \rightarrow \mathbb{R}^m \quad (\varphi(a)) = e^{-\gamma a^2} \frac{(2ab)^{\frac{n}{2}} a^n}{\sqrt{n!}}$$

$$\varphi(a) = \left[1, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{6}}, \dots \right]$$

$$\frac{1}{\sqrt{0}}, \frac{1}{\sqrt{1}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{3}}, \dots$$

$$(\varphi(a)) = \left[(\varphi_0(a), \varphi_1(a), \varphi_2(a), \dots) \right] \quad e^{-\gamma(a-b)^2} = \left[a_0, a_1, a_2, \dots \right]$$

$$(\varphi(a))^\top (\varphi(b)) = e^{-\gamma(a-b)^2} = k(a, b)$$

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} a_0, b_0, a_1 b_0 + a_0 b_1, a_2 b_0 + a_1 b_1 + a_0 b_2, \dots \end{bmatrix}$$

המשך לסעיף ב':

- ב. [5 נק'] נתון dataset עם $m = 1000$ דוגמאות חד-ממדיות. נרצה לפתור את הבעיה עם ה-Gaussian kernel שהגדנו. מבחינת יעילות, האם עדיף לפתור את ה-primal problem עם ה-feature mapping שמצאתם, או שעדיין לפתור את ה-dual problem עם פונקציית ה-kernel שהגדה? ענו והסבירו בקצרה.

משמעות:

במקרה של מינימיזציה, מינימום פונקציית האנרגיה מושג בנקודת קיטול.

ਪਾਨੀ ਕਿਸੇ ਵੀ ਜ਼ਰੂਰੀ ਪ੍ਰਣਾਲੀ ਨਹੀਂ ਹੈ।

האם Fe^{2+} יתפרק ל- Fe^{3+} ו- Fe^{2+} ?

חלק ב' – שאלות אמריקאיות [25 נק']

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

א. [5 נק'] סמננו את כל התשובות הנכונות ביחס לאלגוריתמי (1v1) ו-(1vA).

(הנחנו שיש 10 מחלקות ומעלה).

$\binom{10}{2}$ vs 10

ל-Av1 סיבוכיות מקומ נמוכה יותר מאשר 1v1 בזמן האימון.

a.

ל-Av1 סיבוכיות מקומ נמוכה יותר מאשר 1v1 בזמן המבחן (אחר שהאימון הושלם).

b.

ר~~c~~ רק אחד משני האלגוריתמים ניתן למייקבול (parallelization).

1v1 נוטה יותר ליצור בעיות לא מאחנות (imbalanced).

c.

1vALL

ב. [5 נק'] סמננו את כל הטענות שמשלימות בצורה הגיונית את הטענה הבאה.

באופן כללי, ככל שה-complexity של מחלוקת היפותזות עולה:

$$\begin{matrix} C \rightarrow \infty \\ \times \rightarrow 0 \end{matrix}$$

bus - how well the avg model fits the avg

a. ~~the-bias~~ עוללה.

c.

the-variance עוללה.

b.

c. צריך פחות נתונים על מנת להקליל כראוי.

d?

d. יש יותר נטייה ל-overfitting.

e.

תהליך האימון של מסוג בוודד דורש זמן רב יותר.

$$\mathcal{L}(z) = (\max\{0, 1-z\})^2$$

$$\mathcal{L}(z) = \begin{cases} 0 & z \geq 1 \\ (1-z)^2 & z < 1 \end{cases}$$

$$z > 1 \rightarrow 0$$



ג. [5 נק'] נגידר את פונקציית hinge loss:squared hinge loss

סמננו את כל הטענות הנכונות ביחס לפונקציה זו.

a. הפונקציה קמורה ביחס ל-z.

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

c.

הfonktsia חוסמת מלמעלה את hinge loss-0-1 loss בכל מקום.

d.

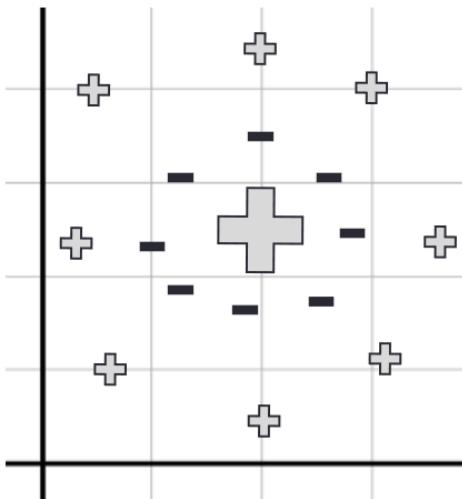
הfonktsia חוסמת מלמעלה את hinge loss בכל מקום.

e.

עבור בעיות סיוג תחת מודל ליניארי, משמע $x_i^T w = z$, הפונק' מעודדת margin מהמפריד.

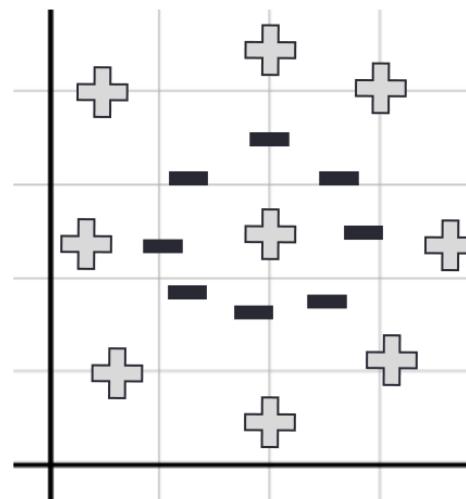
ד. [5 נק'] נתון דатаה עם תיוגים בינהרים ("+" או "-"). מರיצים AdaBoost עם מסוג בסיס לא ידוע. גDAL הוצאות בתרשימים מסמלים את ההסתברויות שהאלגוריתם מקצה (הסתברות גבוהה = צורה גדולה). מריצים את האלגוריתם איטרציה אחת ומקבלים את התרשימים הבאים:

התפלגות לאחר איטרציה אחת



(הנק' האמצעית גדלה והיתר קטן)

התפלגות האחידה ההתחלתית



איזה סוג של מסות בסיס יכול להסביר את התרשים השמאלי שהתקבל? סמן את התשובה הנכונה.

- א. עז החלטה בעומק 1 (decision stump). משמע, שורש ושני עלים.
 - ב. עז החלטה בעומק 2. משמע, שורש, רמת ביןים ועד ארבעה עלים.
 - ג. מסווג שאומר על כל המרחב "שקר" או "אמת".
 - ד. SVM עם קרגל פולינומייאלי ממעלה 2. d.
 - א. כל התשובות הקודומות לא נכוןות.

ה. [5 נק'] היזכרו בפונקציית Softmax שימושה כשכבה האחורה של רשת נוירונים לSieog- K מחלקות:

$$\text{softmax}(f_1(x), \dots, f_K(x); \beta) = \left[\frac{\exp\{\beta f_1(x)\}}{\sum_{i \in [K]} \exp\{\beta f_i(x)\}}, \dots, \frac{\exp\{\beta f_K(x)\}}{\sum_{i \in [K]} \exp\{\beta f_i(x)\}} \right]^\top$$

בשער זה אנו לא מתייחסים כלל לאפשרות ש- β ומণיחים של- β אותו סימן בזמן האימון ובזמן המבחן.

סמננו את בל הטענות הנכונות ביחס לפונקציה זו.

- א. בזמן מבחן (לאחר האימון), כאשר $\beta \rightarrow \infty$, התפלגות הפלט הולכת להתפלגות אחידה.

ב. כאשר משנים את β לאחר האימון בזמן המבחן, כל עוד β שומר על הסימן, אין לו השפעה על הדיק של הרשת.

בזמן אימון, כל עוד הפרמטר β חיובי, אין לו השפעה על מהלך האימון.

בזמן אימון, אם הפרמטר β שלילי, לא ניתן ללמידה את הרשת בעזרת שיטות gradient.

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



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