תעודת זהות:
מספר מחברת:

מבוא ללמידה חישובית - מבחן מועד ב'
סמסטר א' תשע"ו (2015/6)

בית הספר למדעי המחשב, אוניברסיטת תל־אביב

מרצים: פרופ' ליאור וולף, פרופ' ערן הלפרין
מתרגל: רגב שוייגר

8.4.2016

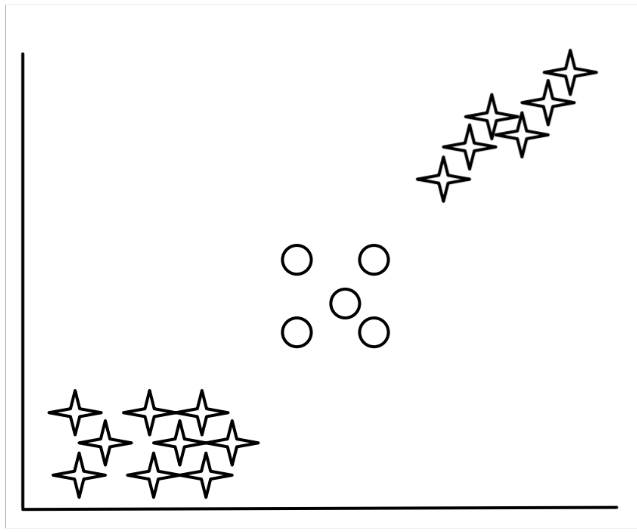
<u>הוראות:</u>

1. מומלץ לקרוא את כל ההנחיות והשאלות בתחילת המבחן לפני תחילת כתיבת התשובות.
2. משך הבחינה – **שלוש שעות**.
3. חומר עזר מותר: דף נוסחאות בגודל A4.
4. **יש לענות על השאלות במקום המיועד לכך בטופס השאלון (טופס זה)**. מחברות הבחינה לא תקראנה, ותשמשנה כטיוטה בלבד.
5. יש למלא בכל דף של השאלון מספר ת.ז. ומספר מחברת.
6. במבחן 5 שאלות:
   - הניקוד לכל שאלה מופיע ליד מספר השאלה.
   - יש לענות תשובות ברורות, ענייניות ותמציתיות.
7. מותר להשתמש בכל טענה שהוכחה בכיתה (בהרצאה, בתרגול, או בתרגיל בית) בתנאי שמצטטים אותה במדויק. טענות אחרות (כאלה שהוכחו בספר, בהרצאות מהסמסטר הקודם, וכו') יש להוכיח.
8. אם לא נאמר אחרת, יש להניח שדגימות במדגם נוצרות באופן בלתי תלוי ומאותה התפלגות (i.i.d).

תעודת זהות:
מספר מחברת:

1 שאלה 1 - 20 נקודות

נתון המדגם הדו־מימדי הבא, בו הנקודות מסווגות לשתי מחלקות:

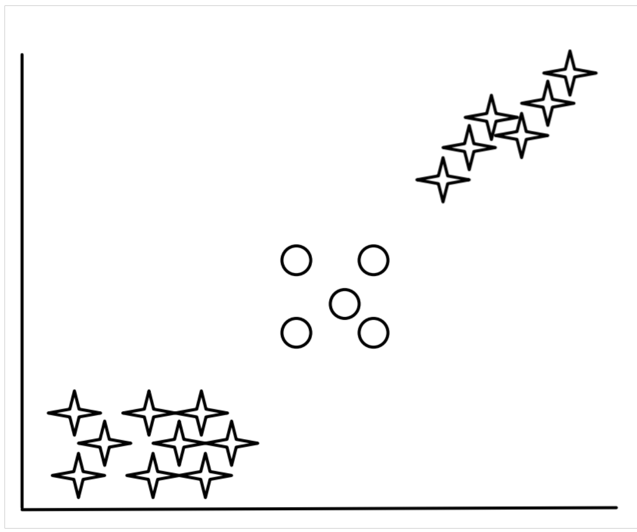


עבור כל אחד מהאלגוריתמים הבאים, קבע/י האם ניתן להריץ אותו עד לקבלת מסווג עם שגיאת למידה אפס, על המדגם הנתון. אם כן, צייר/י קו הפרדה מתאים למסווג המתקבל. אם לא, הסבר/הסבירי מדוע.

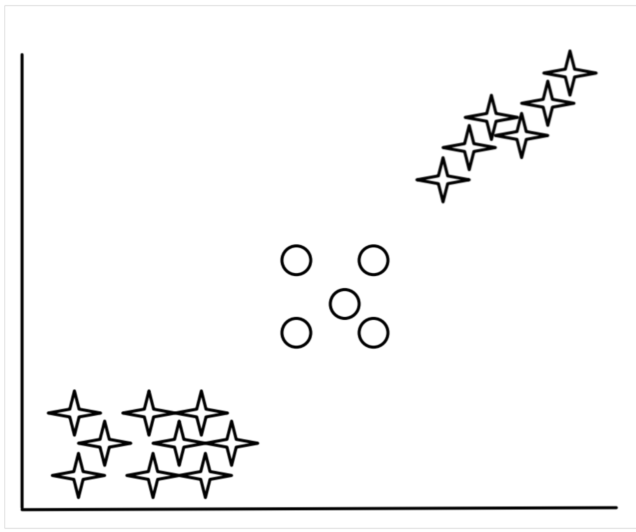
1. אלגוריתם המוצא מלבן דו־מימדי המוגדר על ידי ארבעה מספרים (x_1, x_2, y_1, y_2).
כלומר, נקודה (x, y) תסווג עם סיווג מסוים אם $x_1 \leq x \leq x_2$ וגם $y_1 \leq y \leq y_2$,
ועם הסיווג השני אחרת.

☐ לא יכול להגיע לשגיאה אפס. הסבר:

☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



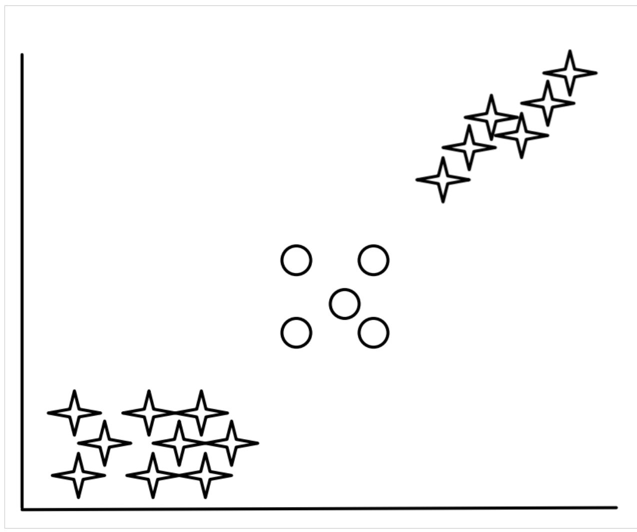
2. רגרסיה לוגיסטית, כאשר הסיווג נקבע לפי העיגול של התוצאה ל־0 או ל־1.
   - ☐ לא יכול להגיע לשגיאה אפס. הסבר:
   - ☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



3. SVM עם Polynomial kernel ריבועי.

☐ לא יכול להגיע לשגיאה אפס. הסבר:

☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



תעודת זהות:
מספר מחברת:

שאלה 2 - 20 נקודות

נתון לנו מדגם מסווג $(\mathbf{x}_i, y_i)$, כאשר $\mathbf{x}_i \in \mathbb{R}^2$ ו- $y_i \in \{+1, -1\}$. נרצה לפתור את בעיית הרגרסיה הבאה:

$$y_i = w^2(\mathbf{x}_i)_1 + (\mathbf{x}_i)_2$$

1. מהי פונקציית האופטימיזציה המתאימה לבעיה החדשה?

2. מהו ה־w שמביא למינימום את פונקציית המטרה שהוגדרה בסעיף 1?

שאלה 3 - 20 נקודות

האלגוריתם שראינו בכיתה ל-Soft Margin SVM מניח שלכל הנקודות יש אותה חשיבות בעינינו. בשאלה זו, נניח כי לכל נקודה $\mathbf{x}_n$ נתונה משקולת $0 \leq v_n \leq 1$ שמגדירה את החשיבות שלה. נרצה כעת להביא למינימום את סכום החריגות מהשוליים, המשוקלל לפי החשיבות. כלומר, נתונה הבעיה הבאה:

$$\begin{aligned} &\min_{\mathbf{w},b,\boldsymbol{\xi}} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{n=1}^{N} v_n\xi_n \\ &\text{s.t.}\ \ y_n(\mathbf{w}^T\mathbf{x}_n + b) \geq 1 - \xi_n \qquad \forall n = 1, \ldots, N \\ &\qquad \xi_n \geq 0 \end{aligned}$$

כאשר $\mathbf{x}_n \in \mathbb{R}^d$ ו- $y_n \in \{+1, -1\}$.

1. מהו הלגרנז'יאן בבעיה זו?

תעודת זהות:
מספר מחברת:

2. מהי הבעיה הדואלית במקרה זה? יש להציג את צורתה הסופית, שלא תלויה במשתנים $\mathbf{w}, b, \boldsymbol{\xi}$.

תעודת זהות:
מספר מחברת:

שאלה 4 - 20 נקודות

באלגוריתם PCA נתון לנו מדגם $\mathbf{x}_i$, כאשר $\mathbf{x}_i \in \mathbb{R}^d$, עבור $i = 1, \ldots, n$. בשאלה זו נרצה להרחיב את האלגוריתם ל-Kernel PCA.

קיימת פונקציית מיפוי $\phi : \mathbb{R}^d \to H$ (כאשר H הוא מרחב כלשהו אליו ממופות הנקודות) וקיים קרנל K, כך שמתקיים, לכל שתי נקודות $\mathbf{x}, \mathbf{x}'$:

$$K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$$

פונקציית המיפוי לא בהכרח נתונה, אך הקרנל נתון לנו במפורש. כלומר, לכל שתי נקודות $\mathbf{x}, \mathbf{x}'$, אנחנו יכולים לחשב את $K(\mathbf{x}, \mathbf{x}')$, אבל לא את $\phi(\mathbf{x}), \phi(\mathbf{x}')$.

נזכיר, כי באלגוריתם PCA אנו דורשים כי המדגם יהיה ממורכז, כלומר ש-

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}$$

כדי להשיג מטרה זו, מבצעים שלב עיבוד מקדים בו מחסרים מכל קואורדינטה את הממוצע שלה. כיוון שכעת המרחב בו נפעל הוא H, נדרוש במקום זאת ש-

$$\frac{1}{n} \sum_{i=1}^{n} \phi(\mathbf{x}_i) = \mathbf{0}$$

אך נשים לב שכעת לא ניתן לחסר במפורש את הממוצע במרחב H.

(המשך בעמוד הבא)

1. במהלך האלגוריתם, משתמשים במטריצה $\overline{\overline{K}}$ המוגדרת על ידי

$$\overline{\overline{K}}_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j)$$

הראה/י כיצד ניתן להחליף אותה במטריצה חדשה $\overline{\overline{K'}}$ כך שהאיבר המתאים לכל שתי נקודות מהמדגם $\mathbf{x}_i, \mathbf{x}_j$, יהיה שקול להפעלה של הקרנל המקורי על אותן שתי נקודות, לאחר שהמדגם עבר מרכוז במרחב H. אין להשתמש ישירות במיפויים $\phi(\mathbf{x}_i)$. מה הסיבוכיות של פעולה זו?

2. נניח שהמדגם ממורכז כנדרש. אנו רוצים להפעיל את אלגוריתם PCA על המיפויים $\phi(\mathbf{x}_i)$. מכיוון שהטווח H של ϕ יכול להיות מרחב ממימד אינסופי או גבוה, לא נחפש את הצירים הראשיים (ה־Principal Components) עצמם, אלא נסתפק באפשרות לבצע מכפלה פנימית של כל ציר עם המיפוי $\phi(\mathbf{x})$ של נקודה חדשה כלשהי, $\mathbf{x}$.

נסמן ב־$\mathbf{u}_1, \ldots, \mathbf{u}_k$ את k הצירים הראשיים הראשונים במרחב H המתאימים לנקודות $\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_n)$, ותהא $\mathbf{x}$ נקודה כלשהי חדשה. הראה/י כיצד ניתן לחשב את

$$\langle \mathbf{u}_j, \phi(\mathbf{x}) \rangle$$

עבור $j = 1, \ldots, k$. מהי הסיבוכיות של הפתרון?

תעודת זהות:
מספר מחברת:

שאלה 5 ־ 20 נקודות

נתון לנו מדגם $(\mathbf{x}_i, y_i)$, כאשר $\mathbf{x}_i \in \mathbb{R}^d$ ו־ $y_i \in \mathbb{R}$. נרצה לשנות את בעיית הרגרסיה הלינארית כדי שתתאים ל־ L1 loss.

1. מהי פונקציית האופטימיזציה המתאימה לבעיה החדשה?

2. כיצד ניתן לפתור את הבעיה החדשה באמצעות Linear Programming?