תעודת זהות:
מספר מחברת:

מבוא ללמידה חישובית - מבחן מועד א'
סמסטר א' תשע"ה (2014/5)

בית הספר למדעי המחשב, אוניברסיטת תל־אביב

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

3.2.2015

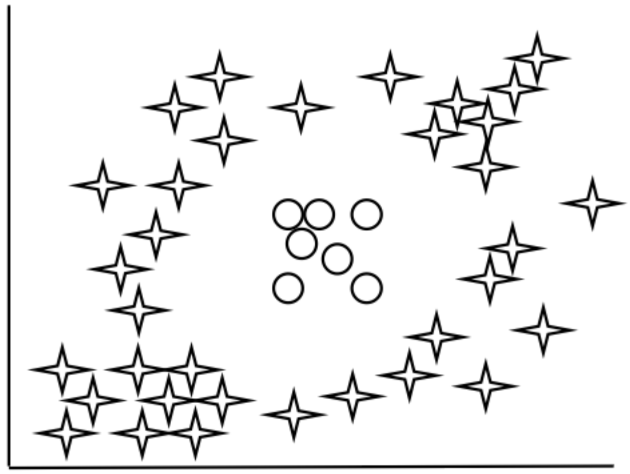
<u>הוראות:</u>

1. מומלץ לקרוא את כל ההנחיות והשאלות בתחילת המבחן לפני תחילת כתיבת התשובות.
2. משך הבחינה – **שלוש שעות**. לא תינתן כל הארכה נוספת.
3. חומר עזר מותר: דף נוסחאות בגודל A4.
4. **יש לענות על השאלות במקום המיועד לכך בטופס השאלון (טופס זה).** מחברות הבחינה לא ייקראו, ותשמשנה כטיוטה בלבד.
5. יש למלא בכל דף של השאלון מספר ת.ז. ומספר מחברת.
6. במבחן 6 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 ־ 15 נקודות

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

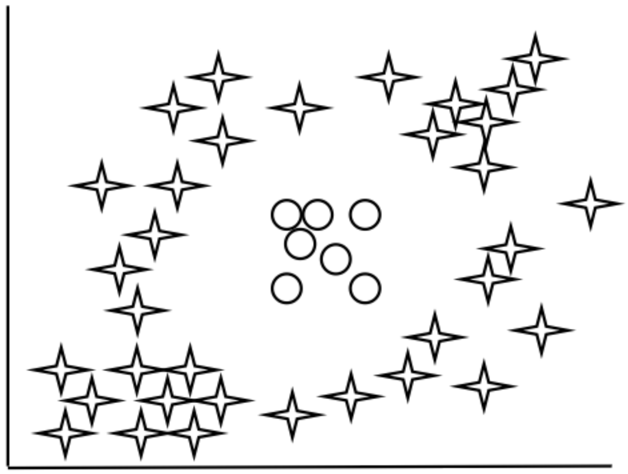


עבור כל אחד מהאלגוריתמים הבאים, קבע/י האם ניתן להריץ אותו עד לקבלת מסווג עם שגיאת למידה אפס, על המדגם הנתון. אם כן, צייר/י קו הפרדה מתאים למסווג המתקבל. אם לא, הסבר/הסבירי מדוע.

1. AdaBoost, כאשר ה־weak classifiers מתוכם האלגוריתם בוחר הם כל ה־decision stumps המקבילים לצירים (כלומר, עבור כל מימד i, וסף a, ניתן לבחור classifier שמסווג על פי תוצאת ההשוואה $x_i < a$).

☐ לא יכול להגיע לשגיאה אפס. הסבר:

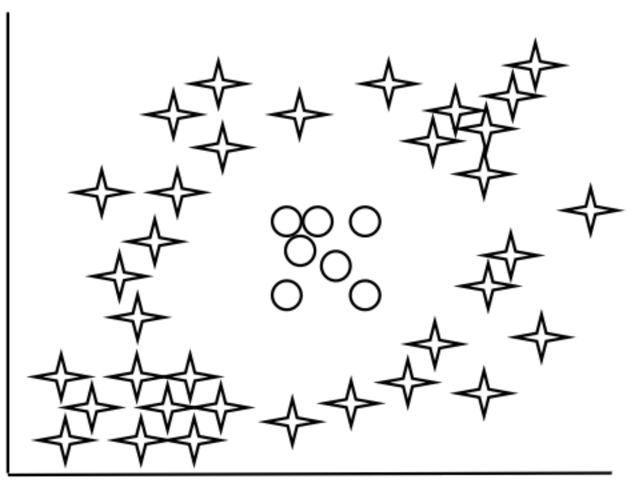
☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



2. Perceptron (כאשר הדגימות מוזנות לו אחת אחרי השניה בסדר שרירותי כלשהו).

☐ לא יכול להגיע לשגיאה אפס. הסבר:

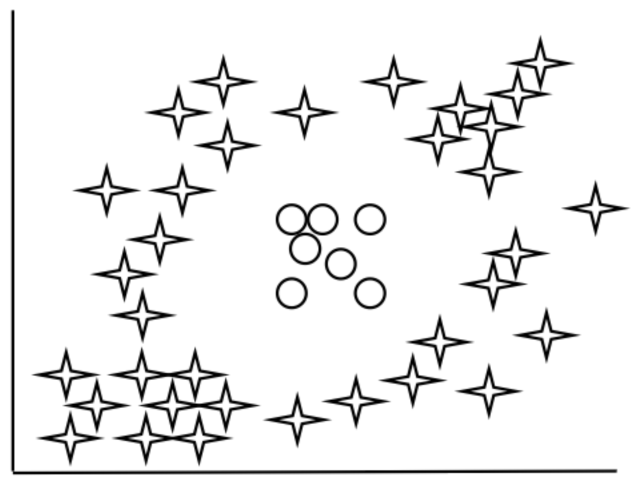
☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



3. SVM עם Gaussian kernel.

☐ לא יכול להגיע לשגיאה אפס. הסבר:

☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



2 שאלה 2 - 15 נקודות

מאמנים מסווג א' (classifier) באמצעות אלגוריתם A ללימוד מסווגים, על מדגם אימון מסוים. כעת מפעילים על כל נקודה במדגם טרנספורמציה נתונה T, ועל המדגם החדש מאמנים מסווג ב' באמצעות אותו האלגוריתם. תהי $\mathbf{x}$ נקודה חדשה. מפעילים את מסווג א' על $\mathbf{x}$, ומפעילים את מסווג ב' על $T(\mathbf{x})$. האם הם בהכרח יחזירו את אותו הסיווג? נמק/י את תשובתך בקצרה.

1. A הוא Hard Margin SVM.

(א) $T(\mathbf{x}) = \mathbf{x} + \mathbf{x}_0$, כאשר $\mathbf{x}_0$ הוא וקטור.
☐ בהכרח אותו סיווג
☐ לא בהכרח אותו סיווג
הסבר:

(ב) $T(\mathbf{x}) = a\mathbf{x}$, כאשר $a \neq 0$ הוא סקלר.
☐ בהכרח אותו סיווג
☐ לא בהכרח אותו סיווג
הסבר:

(ג) $T(\mathbf{x}) = D\mathbf{x}$, כאשר D היא מטריצה אלכסונית, ללא 0 על האלכסון.
☐ בהכרח אותו סיווג
☐ לא בהכרח אותו סיווג
הסבר:

(ד) $T(\mathbf{x}) = U\mathbf{x}$, כאשר U היא מטריצה אורתונורמלית.
☐ בהכרח אותו סיווג
☐ לא בהכרח אותו סיווג
הסבר:

תעודת זהות:
מספר מחברת:

2. A הוא עץ החלטה עם Decision stumps.

(א) $T(\mathbf{x}) = \mathbf{x} + \mathbf{x}_0$, כאשר $\mathbf{x}_0$ הוא וקטור.

☐ בהכרח אותו סיווג

☐ לא בהכרח אותו סיווג

הסבר:

(ב) $T(\mathbf{x}) = a\mathbf{x}$, כאשר $a \neq 0$ הוא סקלר.

☐ בהכרח אותו סיווג

☐ לא בהכרח אותו סיווג

הסבר:

(ג) $T(\mathbf{x}) = D\mathbf{x}$, כאשר D היא מטריצה אלכסונית, ללא 0 על האלכסון.

☐ בהכרח אותו סיווג

☐ לא בהכרח אותו סיווג

הסבר:

(ד) $T(\mathbf{x}) = U\mathbf{x}$, כאשר U היא מטריצה אורתונורמלית.

☐ בהכרח אותו סיווג

☐ לא בהכרח אותו סיווג

הסבר:

3 שאלה 3 - 20 נקודות

נסמן ב־$\mathbf{e}_i \in \mathbb{R}^d$ את וקטור היחידה ה־i של הבסיס הסטנדרטי (כלומר, וקטור בן d איברים שכולם 0 מלבד האיבר ה־i שהוא 1). נגדיר את מדגם האימון הבא:

$$S = \cup_{i=1}^{d} \{\langle \mathbf{e}_i, 1 \rangle , \langle -\mathbf{e}_i, -1 \rangle\}$$

1. כתב/י את בעיית האופטימיזציה הפרימאלית של SVM, עבור המקרה $d = 2$.
מהו פתרון הבעיה, ומהם ה־$\mathbf{w}, b$ האופטימליים? כמה Support Vectors ישנם?

2. במקרה הכללי, עבור d כלשהו, מהם $\mathbf{w}, b$ האופטימליים? כמה Support Vectors ישנם?

4 שאלה 4 ־ 20 נקודות

מדגם $\mathbb{R}^n \in \mathbf{x_1},\ldots,\mathbf{x_m}$ מיוצר כך: נתונה מטריצה W בת n שורות ו־d עמודות, מדרגה מלאה d, כאשר $d < n$. עבור כל $i = 1,\ldots,m$, מגרילים וקטור מקדמים $\mathbf{z}_i \in \mathbb{R}^d$ מההתפלגות הרב נורמלית:

$$\mathbf{z}_i \sim N(\mathbf{0}, \sigma^2 I_d)$$

בנוסף, מגרילים וקטור רעש $\mathbf{e_i} \in \mathbb{R}^n$ מההתפלגות הרב נורמלית:

$$\mathbf{e}_i \sim N(\mathbf{0}, \tau^2 I_n)$$

לבסוף, יוצרים את הנקודה ה־i כך:

$$\mathbf{x}_i = W\mathbf{z}_i + \mathbf{e}_i$$

במודל שתיארנו לעיל, נתייחס ל־$\mathbf{z}_i$ בתור הפרמטרים, בעוד ש־W, σ, τ קבועים וידועים. זהו מודל בייסיאני, כאשר $\mathbf{z}_i \sim N(\mathbf{0}, \sigma^2 I_d)$ מתאר את ה־Prior.

1. כתב/י את פונקציית ה־Log Prior של הפרמטרים, כלומר את:

$$\log \Pr\left(\mathbf{z_1},\ldots,\mathbf{z_m}\right)$$

2. כתב/י את הפונקצית ה־Log Likelihood של המדגם עם ה־Prior המתאים, כלומר את:

$$\log\left(\Pr\left(\mathbf{x_1},\ldots,\mathbf{x_m};\mathbf{z_1},\ldots,\mathbf{z_m}\right)\cdot\Pr\left(\mathbf{z_1},\ldots,\mathbf{z_m}\right)\right)$$

3. מהו אומד ה־MAP של $\mathbf{z_i}$?

5 שאלה 5 - 15 נקודות

נתון מדגם הכולל את הנקודות (x, y) הבאות:

$$(-2, 10), (-1, 5), (0, 0), (1, 5), (2, 10)$$

1. מהו השיפוע של הקו המתאים לרגרסיה לינארית? נמק/י. אין הכרח לחשב את הקו.
2. מהו השיפוע של הקו המתאים ל־PCA של מימד אחד? נמק/י. אין הכרח לחשב את הקו.

תעודת זהות:
מספר מחברת:

6 שאלה 6 - 20 נקודות

נסמן ב־$\mathbb{R}^m$ $\mathbf{e}_i \in$ את וקטור היחידה ה־i של הבסיס הסטנדרטי (כלומר, וקטור בן m איברים שכולם 0 מלבד האיבר ה־i שהוא 1). נגדיר את מדגם האימון הבא:

$$S = \{\langle -\mathbf{e}_1, -1\rangle, \ldots, \langle -\mathbf{e}_m, -1\rangle\}$$

נזכיר שבהקשר של אלגוריתם Perceptron, ה־margin של המפריד המוגדר על ידי וקטור $\mathbf{w}^*$ הוא:

$$\gamma = \min_{\mathbf{x}\in S} \frac{|\mathbf{x}\cdot\mathbf{w}^*|}{\|\mathbf{x}\|}$$

1. נגדיר:

$$\mathbf{w}^* = \frac{1}{\sqrt{m}} \cdot (1, 1, \ldots, 1)$$

מהו ה־Margin של המדגם ביחס למפריד המוגדר על ידי $\mathbf{w}^*$?

תעודת זהות:
מספר מחברת:

2. מפעילים את אלגוריתם Perceptron על המדגם. מהו החסם העליון התיאורטי על כמות השגיאות שהאלגוריתם יבצע, כפונקציה של m?

3. כמה שגיאות יבצע האלגוריתם בפועל על המדגם?