תעודת זהות:
מספר מחברת:

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

בית הספר למדעי המחשב, אוניברסיטת תל־אביב

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

4.9.2015

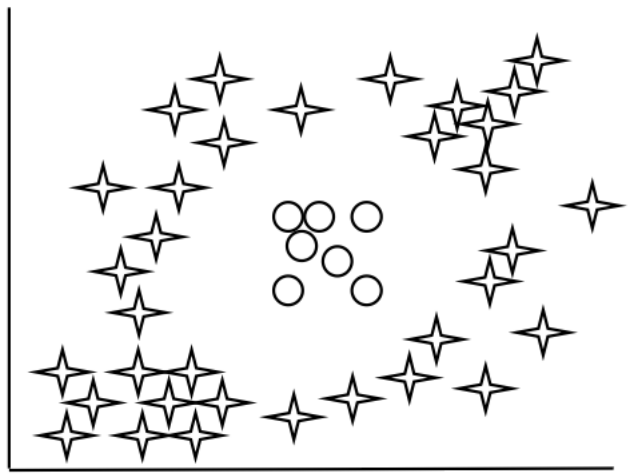
<u>הוראות:</u>

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:

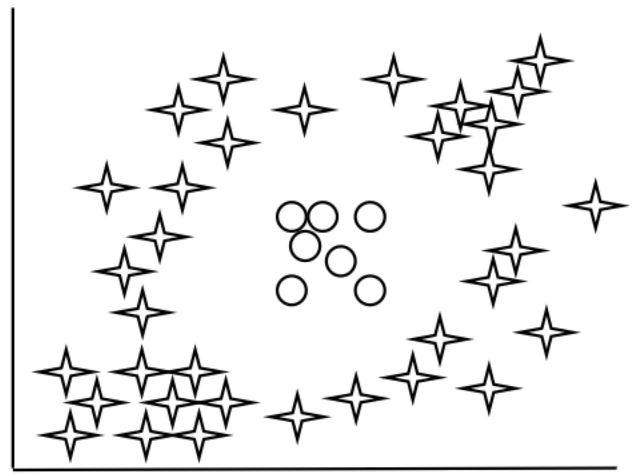


עבור כל אחד מהאלגוריתמים הבאים, קבע/י האם ניתן להריץ אותו עד לקבלת מסווג עם שגיאת למידה אפס, על המדגם הנתון. אם כן, צייר/י קו הפרדה מתאים למסווג המתקבל. אם לא, הסבר/הסבירי מדוע.

1. עצי החלטה, כאשר ה־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:

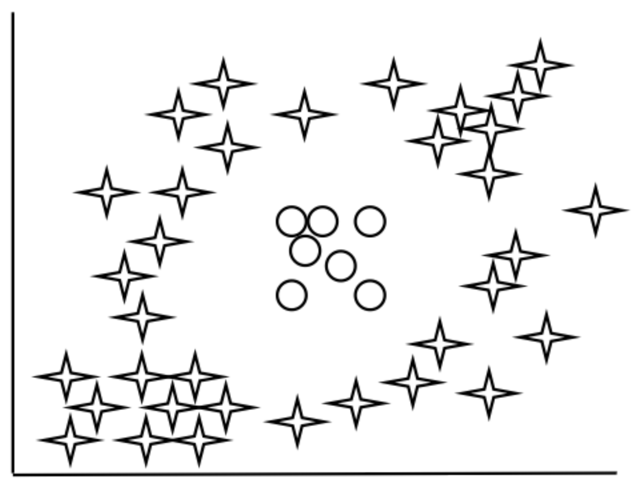
   ☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



2. 3-Nearest Neighbor.

☐ לא יכול להגיע לשגיאה אפס. הסבר:

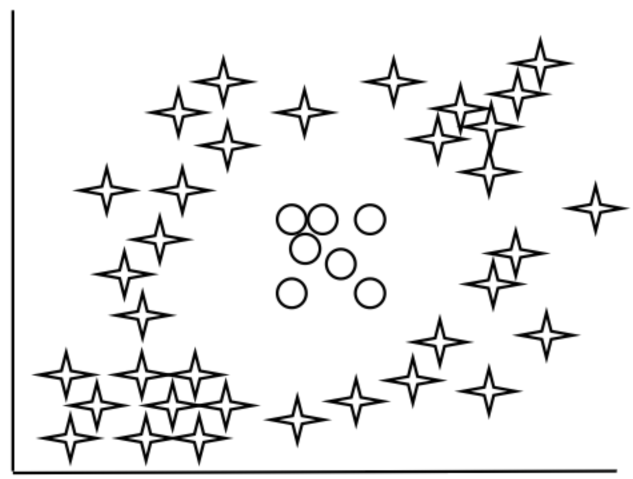
☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



3. SVM עם Polynomial kernel, d=2.

☐ לא יכול להגיע לשגיאה אפס. הסבר:

☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



2 שאלה 2 - 15 נקודות

1. מפעילים את אלגוריתם PCA, על מדגם אימון מסוים, לקבלת מערכת צירים בעלת k וקטורים (כאשר אנחנו כוללים שלב מירכוז מקדים, בו אנו מזיזים את כל נקודות המדגם באותו וקטור, כך שממוצע כל קואורדינטה הוא 0). כעת מפעילים על כל נקודה במדגם טרנספורמציה נתונה T, ועל המדגם החדש מפעילים PCA בשנית לקבלת k וקטורים נוספים. האם שתי מערכות הצירים זהות (עד כדי החלפת סימן אפשרית של כל וקטור עצמי)? נמק/י את תשובתך בקצרה.

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

   ☐ מערכות זהות

   ☐ לא בהכרח זהות

   הסבר:

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

   ☐ מערכות זהות

   ☐ לא בהכרח זהות

   הסבר:

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

   ☐ מערכות זהות

   ☐ לא בהכרח זהות

   הסבר:

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

   ☐ מערכות זהות

   ☐ לא בהכרח זהות

   הסבר:

2. מריצים אלגוריתם EM ללמידת הפרמטרים של מודל Gaussian Mixture Model על מדגם נתון. האלגוריתם מאותחל כך: וקטור ההתפלגות המגדיר את ההסתברויות לבחור בכל גאוסיאן מאותחל להתפלגות האחידה. השונות ההתחלתית של של גאוסיאן היא 1, והתוחלת ההתחלתית של הגאוסיאן ה-i היא הנקודה ה-i במדגם (יש פחות צבירים מאשר נקודות). לוקחים נקודה קבועה $\mathbf{z}$ ומחשבים את ההסתברות האפוסטריורית שלה להשתייך לכל אחד מהצבירים, לאחר הרצת האלגוריתם. כעת מפעילים על כל נקודה במדגם טרנספורמציה נתונה T, ועל המדגם החדש מפעילים את האלגוריתם בשנית. כעת מחשבים את ההסתברות האפוסטריורית של $T(\mathbf{z})$ להשתייך לכל אחד מהצבירים. בשתי הפעמים מריצים את האלגוריתם אותו מספר איטרציות.

האם וקטור ההסתברויות עבור $\mathbf{z}$ זהה לוקטור ההסתברויות עבור $T(\mathbf{z})$?

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

☐ בהכרח זהה

☐ לא בהכרח זהה

הסבר:

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

☐ בהכרח זהה

☐ לא בהכרח זהה

הסבר:

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

☐ בהכרח זהה

☐ לא בהכרח זהה

הסבר:

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

☐ בהכרח זהה

☐ לא בהכרח זהה

הסבר:

תעודת זהות:
מספר מחברת:

3 שאלה 3 - 20 נקודות

1. ברגרסיה לינארית סטנדרטית, נתון המודל הבא: יהיו $\mathbf{x}_1, \ldots, \mathbf{x}_p \in \mathbb{R}^n$ וקטורים נתונים, ותהא X המטריצה שעמודותיה הם $\mathbf{x}_1, \ldots, \mathbf{x}_p$. יהי $\mathbf{e} \in \mathbb{R}^n$ וקטור של משתנים מקריים שעבורו לכל $i = 1, \ldots, n$, $e_i \sim \mathrm{N}(0, \sigma^2)$. כאשר σ^2 היא שונות קבועה ונתונה. יהיו $\mathbf{a} = (a_1, \ldots, a_p)^T$ וקטור פרמטרים, ונגדיר:

$$\begin{aligned}\mathbf{y} &= X\mathbf{a} + \mathbf{e} \\ &= a_1 \cdot \mathbf{x}_1 + \ldots a_p \cdot \mathbf{x}_p + \mathbf{e}\end{aligned}$$

מהי ההתפלגות הרב מימדית של הוקטור $\mathbf{y}$?

2. כעת נכליל את המודל. בנוסף להגדרות סעיף 1, יהיו $\mathbf{z}_1 \ldots, \mathbf{z}_q \in \mathbb{R}^n$ וקטורים נתונים, ותהא Z המטריצה שעמודותיה הם $\mathbf{z}_1, \ldots, \mathbf{z}_q$.

יהיו $b_1 \ldots, b_q \sim \mathrm{N}(0, \tau^2)$ משתנים מקריים, ונסמן $\mathbf{b} = (b_1, \ldots, b_q)^T$. השונות τ^2 נתונה. כעת:

$$\begin{aligned}\mathbf{y} &= X\mathbf{a} + Z\mathbf{b} + \mathbf{e} \\ &= a_1 \cdot \mathbf{x}_1 + \ldots a_p \cdot \mathbf{x}_p + b_1 \cdot \mathbf{z}_1 + \ldots b_p \cdot \mathbf{z}_p + \mathbf{e}\end{aligned}$$

מודל כזה נקרא Linear Mixed Model (LMM). כעת מהי ההתפלגות של $\mathbf{y}$?

תעודת זהות:
מספר מחברת:

4 שאלה 4 ־ 20 נקודות

באלגוריתם SVM, נתון לנו מדגם מסווג $(\mathbf{x}_n, y_n)$, כאשר $\mathbf{x}_n \in \mathbb{R}^d$ ו־ $y_n \in \{+1, -1\}$. אנו פותרים את הבעיה הבאה:

$$\min_{\mathbf{w},b} \frac{1}{2}\mathbf{w}^T\mathbf{w}$$
$$\text{s.t. } y_n(\mathbf{w}^T\mathbf{x}_n + b) \geq 1 \qquad \forall n = 1, \ldots, N$$

כאשר $\mathbf{w} \in \mathbb{R}^d$ הוא וקטור המשקולות, $b \in \mathbb{R}$ הוא קבוע ההזזה. כזכור, את הבעיה הזו ניתן להמיר לבעיה דואלית ולפתור אותה באופן שקול. נסמן ב־ α_i את כופלי לגרנז' של הלגרנז'יאן בבעיה הדואלית.

1. הראה/י כי מתקיימות שתי המשוואות:

$$\sum_{n=1}^{N} \alpha_n y_n = 0, \qquad \mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n$$

2. נניח כי עבור i מסוים, $\mathbf{x}_i$ הוא support vector. הרא/י כיצד ניתן לחשב את b כפונקציה של
$\mathbf{x}_n, y_n, \alpha_n$.

3. נזכיר כי עבור i מסוים, $\mathbf{x}_i$ הוא support vector אם $\alpha_i > 0$. השתמש/י בשני הסעיפים הקודמים כדי להוכיח כי:

$$\sum_{n=1}^{N} \alpha_n = \|\mathbf{w}\|^2$$

תעודת זהות:
מספר מחברת:

5 שאלה 5 - 15 נקודות

1. נתונים $x_1, \ldots, x_n \in \mathbb{R}$. הוכח/הוכיחי כי המספר μ המביא למינימום את:

$$\sum_{i=1}^{n} |x_i - \mu|$$

הוא החציון של המדגם.

תעודת זהות:
מספר מחברת:

2. פונקצית הצפיפות של התפלגות לפלס החד מימדית $Laplace(\mu,\sigma)$ היא

$$f(x)=\frac{1}{2\sigma}e^{-\frac{|x-\mu|}{\sigma}}$$

נתון מדגם $x_1,\ldots,x_n\sim Laplace(\mu,\sigma)$. מהו אומד הנראות המירבית של הפרמטרים μ,σ?

3. נתונים $x_1, \ldots, x_n \in \mathbb{R}$. בנוסף נתונות משקולות חיוביות $w_1 \ldots, w_n$. נגדיר את $\mu_{(w_1,\ldots,w_n)}(x_1, \ldots, x_n)$ בתור המספר μ המביא למינימום את:

$$\sum_{i=1}^{n} w_i |x_i - \mu|$$

מהו $\mu_{(w_1,\ldots,w_n)}(x_1, \ldots, x_n)$?

תעודת זהות:
מספר מחברת:

6 שאלה 6 - 20 נקודות

פונקצית הצפיפות של התפלגות לפלס החד מימדית $Laplace(\mu, \sigma)$ היא

$$f(x) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}}$$

נתון מדגם הנלקח מתוך תערובת של התפלגויות לפלס חד מימדיות. כלומר, נתונות k התפלגויות לפלס $f_1 = Laplace(\mu_1, 1), \ldots, f_k = Laplace(\mu_k, 1)$. בנוסף נתון וקטור הסתברות $\mathbf{p} = (p_1, \ldots, p_k)$ הקובע את ההסתברות לדגום מתוך כל אחת מההתפלגויות. עבור כל נקודת מדגם $i = 1, \ldots, n$, תחילה מגרילים ערך z_i לפי ההתפלגות $\mathbf{p}$ כדי לבחור מאיזו התפלגות לפלס לדגום. ואז, הערך x_i נדגם מתוך ההתפלגות $Laplace(\mu_{z_i}, 1)$.

מהו אלגוריתם EM המתאים לאמידת הפרמטרים μ_i, p_i?

אפשר להשתמש בתוצאות שאלה 5 ללא הוכחה.