תעודת זהות:
מספר מחברת:

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

בית הספר למדעי המחשב, אוניברסיטת תל־אביב

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

26.1.2016

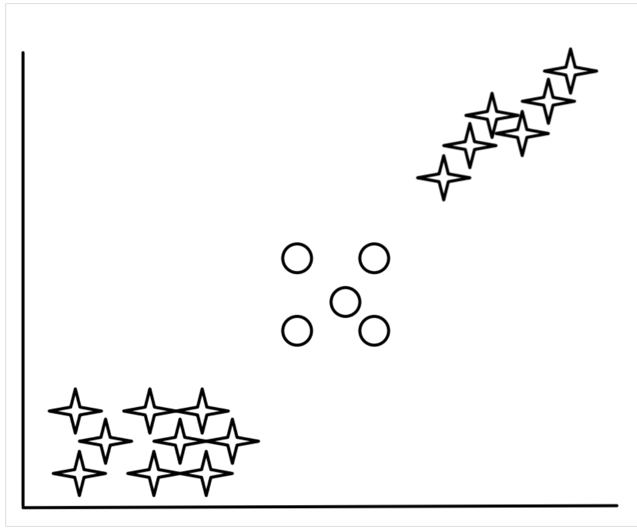
<u>הוראות:</u>

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:

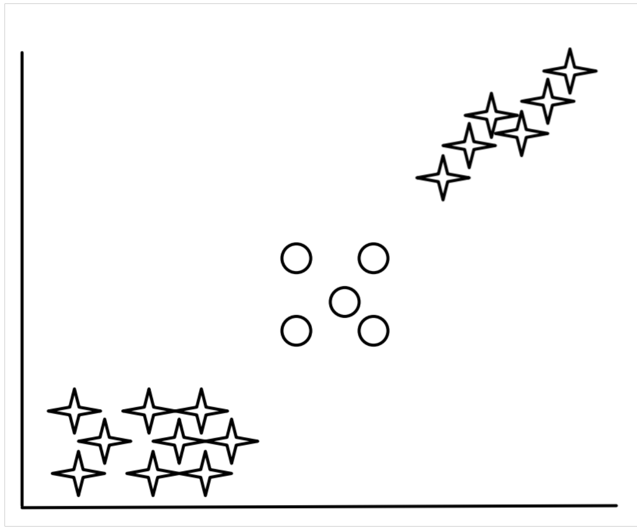


עבור כל אחד מהאלגוריתמים הבאים, קבע/י האם ניתן להריץ אותו עד לקבלת מסווג עם שגיאת למידה אפס, על המדגם הנתון. אם כן, צייר/י קו הפרדה מתאים למסווג המתקבל. אם לא, הסבר/הסבירי מדוע.

1. עץ 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:

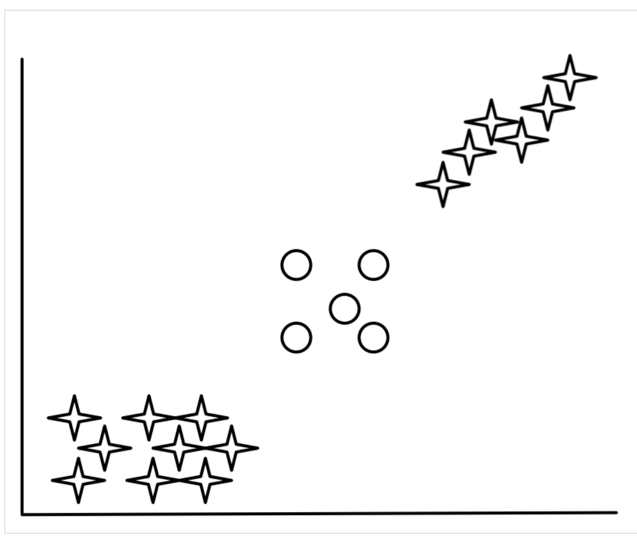
   ☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



2. 1-Nearest Neighbor, כאשר הנקודה הנבדקת לא נכללת ברשימת הנקודות השכנות.

☐ לא יכול להגיע לשגיאה אפס. הסבר:

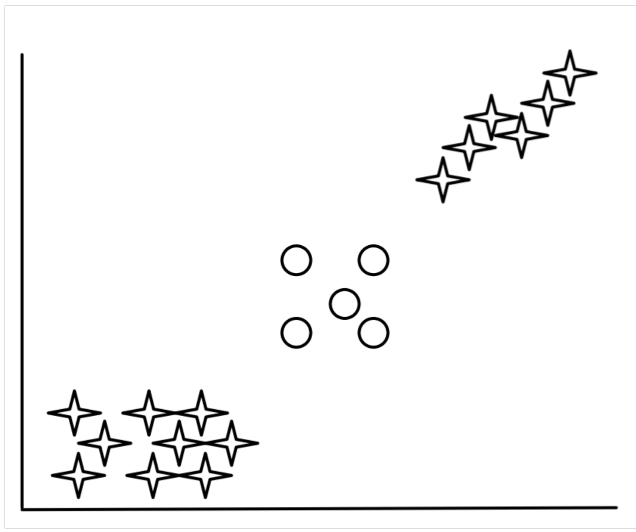
☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



3. SVM עם Gaussian kernel.

☐ לא יכול להגיע לשגיאה אפס. הסבר:

☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



תעודת זהות:
מספר מחברת:

שאלה 2 ־ 20 נקודות

ראינו שהאלגוריתם של בעיית הרגרסיה הלינארית, מוצא פתרון אופטימלי ביחס לפונקציית ה־L2 loss. כעת נגדיר פונקציית loss חדשה, הנותנת לכל נקודה משקל שונה. כלומר, ה־loss על תחזית $\hat{y}_i$ של הנקודה ה־i יהיה:

$$L_i(\hat{y}_i, y_i) = w_i \cdot (\hat{y}_i - y_i)^2$$

1. מהי פונקציית האופטימיזציה המתאימה לבעיה החדשה?

2. האם המשקולות יכולות לשנות את הפתרון הנמצא, ביחס לבעיה המקורית ללא המשקולות? אם כן, תנ/י דוגמה. אם לא, הוכיח/י.

שאלה 3 ־ 20 נקודות

האלגוריתם שראינו בכיתה ל־Soft Margin SVM מניח שאנחנו רוצים להביא למינימום את סכום החריגות מהשוליים, כלומר את $\sum_{i=1}^{n} \xi_i$. בשאלה זו, נרצה להביא למינימום את ריבועי החריגות מהשוליים. כלומר, נתונה הבעיה הבאה:

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

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

1. שימ/י לב שהשמטנו את האילוצים $\xi_n \geq 0$ בבעיה זו. הראה/י שאפשר להוריד את אילוצי אי־השליליות הללו ־ כלומר, שהערך האופטימלי לא משתנה אם האילוצים מתווספים או לא.

2. מהו הלגרנז'יאן בבעיה זו?

3. מהי הבעיה הדואלית במקרה זה?

תעודת זהות:
מספר מחברת:

שאלה 4 - 20 נקודות

1. נתון לנו מדגם מסווג $(\mathbf{x}_i, y_i)$, כאשר $\mathbf{x}_i \in \mathbb{R}^d$ ו- $y_i \in \{+1, -1\}$.

מריצים את אלגוריתם Perceptron על מדגם אימון במשך T איטרציות, עד לקבלת התוצאה הסופית $\mathbf{w}^*$. הראה/י ש-$\mathbf{w}^*$ הוא צירוף לינארי של נקודות המדגם $\mathbf{x}_i$.

2. השתמש/י בתוצאות הסעיפים הקודמים כדי לשנות את Perceptron כך שישתמש ב־kernels. האלגוריתם לא ישמור במפורש את הערך של $\mathbf{w}^*$, אלא ישמור ויעדכן בכל איטרציה את מקדמי הצירוף הלינארי שיוצר אותו. כתב/י פסאודו־קוד שמתאר את האלגוריתם החדש, והדגימ/י כיצד אפשר להשתמש בתוצאה שלו כדי לסווג נקודת מדגם חדשה $\mathbf{x}$, מבלי לשמור במפורש את $\mathbf{w}^*$.

תעודת זהות:
מספר מחברת:

שאלה 5 ־ 20 נקודות

נתון לנו מדגם מסווג $(\mathbf{x}_i, y_i)$, כאשר $\mathbf{x}_i \in \mathbb{R}^d$ ו־ $y_i \in \{+1, -1\}$. נזכיר כי באיטרציה ה־t של אלגוריתם AdaBoost, מעדכנים את ההתפלגות D_t על ידי:

$$\begin{aligned} h_t &= \arg\min_{h \in H} \Pr_{\mathbf{x} \sim D_t}[h(\mathbf{x}) \neq y] \\ \epsilon_t &= \Pr_{\mathbf{x} \sim D_t}[h_t(\mathbf{x}) \neq y] \\ \alpha_t &= \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t} \\ D_{t+1}(\mathbf{x}_i) &= \frac{D_t(\mathbf{x}_i) e^{-\alpha_t y_i h_t(\mathbf{x}_i)}}{Z_t} \end{aligned}$$

כאשר H היא מחלקה של השערות.

1. הראה/י כי

$$\epsilon_t e^{\alpha_t} = \sqrt{\epsilon_t(1 - \epsilon_t)} = (1 - \epsilon_t) e^{-\alpha_t}$$

תעודת זהות:
מספר מחברת:

2. הראה/י כי גורם הנורמליזציה Z_t הוא

$$Z_t = 2\sqrt{\epsilon_t(1-\epsilon_t)}$$

3. הרא/ה כי השגיאה של ההשערה הקודמת ביחס להתפלגות החדשה היא בדיוק חצי, כלומר:

$$\Pr_{\mathbf{x}\sim D_{t+1}}[h_t(\mathbf{x}) \neq y] = \frac{1}{2}$$

תעודת זהות:
מספר מחברת:

4. האם בהכרח $h_{t+1} \neq h_t$?