תעודת זהות: ____________________

מספר מחברת: ____________________

1	2	3	4	5	6	7	8	סהב

מבחן מועד א' – מבוא ללמידה חישובית
סמסטר א' תשע"ד (2013)

בית הספר למדעי המחשב, אוניברסיטת תל-אביב

מרצים: פרופ' ערן הלפרין, פרופ' ליאור וולף, פרופ' ישי מנצור,

מתרגל: מריאנו שיין

30.1.2014

הוראות

1. מומלץ לקרא את כל ההנחיות והשאלות בתחילת המבחן, לפני תחילת כתיבת התשובות.
2. משך הבחינה – **שלוש שעות**. לא תינתן כל הארכה נוספת.
3. חומר עזר מותר: **ללא**.
4. **יש לענות על השאלות במקום המיועד לכך בטופס השאלון (טופס זה).** מחברות הבחינה לא ייקראו, וישמשו כטיוטה בלבד.
5. יש למלא בכל דף של השאלון מספר ת.ז. ומספר מחברת.
6. במבחן 8 שאלות:
   - הניקוד לכל שאלה מופיע לידה 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)

בהצלחה!

תעודת זהות: ______________________________

מספר מחברת: ______________________________

<u>שאלה 1 (10 נקודות).</u>

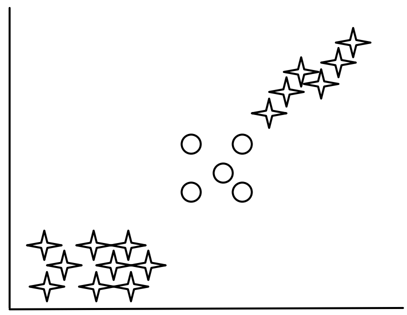
נתונים האלגוריתמים הבאים

א. עץ החלטה עם decision stubs (כלומר $x_i > a$ עבור כל מימד i וקבוע a)

ב. SVM עם Gaussian kernel

ג. מסווג מסוג 1-NN (NN=Nearest Neighbor)

נתון מדגם דו-מימדי הבא, בו שני סיווגים ו-



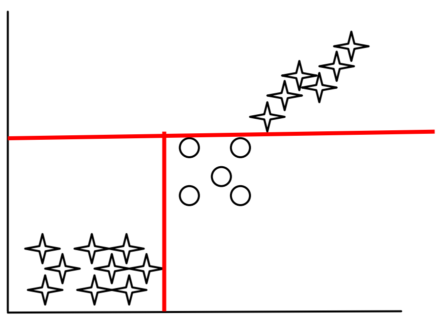
עבור כל אחד מהאלגוריתמים מריצים אותו עד שמקבלים מסווג עם שגיאת למידה אפס (על המדגם)

קבע/י האם האלגוריתם יגיע לשגיאה אפס. אם לא הסבר מדוע. אם כן צייר קו הפרדה מתאים למסווג המתקבל:

א. עץ החלטה עם decision stubs (כלומר $x_i > a$ עבור כל מימד i וקבוע a)

☐ לא יכול להגיע לשגיאה אפס. הסבר ______________________________

☒ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



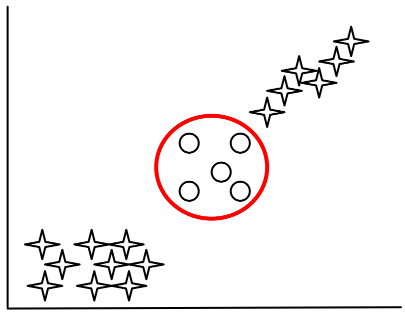
תעודת זהות: ______________________________

מספר מחברת: ______________________________

ב. תוכנית SVM עם Gaussian kernel

☐ לא יכול להגיע לשגיאה אפס. הסבר ________________________________

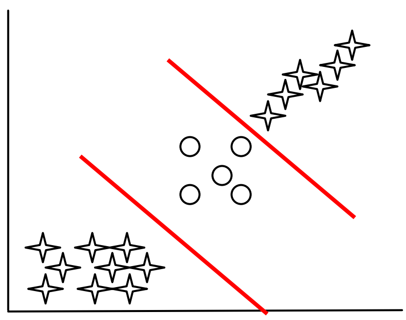
☒ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



ג. מסווג מסוג 1-NN (NN=Nearest Neighbor)

☐ לא יכול להגיע לשגיאה אפס. הסבר ________________________________

☒ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



תעודת זהות: ____________________________

מספר מחברת: ____________________________

שאלה 2 (10 נקודות).

נתונה בעיה עם קלט של 3 משתנים בינאריים, $x_1, x_2, x_3\ \varepsilon\{0,1\}$, ופלט בינארי $y\ \varepsilon\{0,1\}$

1. עבור מסווג המניח מודל הסתברותי Naïve Bayes כמה פרמטרים צריך לשערך מהנתונים? (רק את המספר)

$\boxed{7}$

2. עבור מסווג במודל הסתברותי כללי (ללא הנחת אי התלות של Naïve Bayes) כמה פרמטרים צריך לשערך מהנתונים? (רק את המספר)

$\boxed{16}$

תעודת זהות: ______________________

מספר מחברת: ______________________

שאלה 3 (10 נקודות).

בכל אחת מהשאלות הבאות יש לבחור את היחס המתאים בין הכמות המתוארת במשפט הראשון (א) לכמות המתוארת במשפט השני (ב). למשל, אם (א) גדול מ (ב) יש לבחור באפשרות הראשונה גדול ☐). כמו כן יש להסביר בקצרה את הבחירה.

בתוחלת על המדגם שנבחר:

א. ממוצע השגיאה על המדגם שעליו נבחרת ההשערה ע"י ERM

ב. ממוצע השגיאה על דוגמאות שלא ראינו של ההשערה שנבחרת ע"י ERM.

גדול ☐   קטן ☒   שווה ☐   לא ניתן לדעת ☐

הסבר:

כיוון שביצענו מינימיזציה של השגיאה ב- ERM גרמנו ל- overfitting.

כיוון שמדובר בתוחלת על המדגם אין "מדגם רע".

בהינתן מדגם S יש השערה h(S) ש- ERM מחזיר.

על מדגם חדש T תוחלת השגיאה היא בדיוק השגיאה של T.

על S כיוון שבחרנו את ההשערה שהביאה למינימום, התוחלת קטנה יותר.

מידת ה- overfitting זה ההפרש בין ממוצע השגיאה על המדגם והסתברות השגיאה:

א. מידת ה-overfitting כשלאלגוריתם הלמידה ERM יש 100 השערות לבחירה

ב. מידת ה-overfitting כשלאלגוריתם הלמידה ERM יש 200 השערות לבחירה

גדול ☐   קטן ☒   שווה ☐   לא ניתן לדעת ☐

הסבר:

ככול שיש יותר השערות מידת ה- overfiting תגדל

דרך את לראות את זה, זה דרך חסמי ההכללה למחלקה סופית.

לכן אם m קבוע, אם מגדילים את H אזי ε צריך לגדול.

תעודת זהות: ____________________

מספר מחברת: ____________________

שאלה 4 (15 נקודות).

נתון מדגם בגודל m מהצורה $(x_{i,1}, x_{i,2}, y_i)$, ובו נתונים דו-מימדיים (x_1,x_2) ופלט y.

רוצים לבצע רגרסיה כך: $y=w^2 x_1+x_2$

א. יש להגדיר את פונקציית המטרה של הרגרסיה (עבור שגיאה ריבועית):

ב. יש חשב את w שמביא למינימום את פונקציית המטרה שהוגדרה בסעיף א:

$$W^{2*} = \max\left\{0, \frac{\sum_{\{i=1\}}^{m} x_{i,1}(y_i - x_{i,2})}{\sum_{\{i=1\}}^{m} {x_{i,1}}^2}\right\}$$

החישוב:

$$\frac{d}{dw}\sum_{\{i=1\}}^{m}\left(w^2 x_{i,1} + x_{\{i,2\}} - y_i\right)^2 = \sum_{\{i=1\}}^{m} 2w x_{i,1}\left(w^2 x_{i,1} + x_{i,2} - y_i\right) = 0$$

$$\boldsymbol{w = 0 \text{ or } w \neq 0 \text{ and }} \sum_{\{i=1\}}^{m} x_{i,1}\left(w^2 x_{i,1} + x_{i,2} - y_i\right) = 0$$

תעודת זהות: ______________________________

מספר מחברת: ______________________________

<u>שאלה 5 (12 נקודות).</u>

מריצים SVM על בעייה מסויימת, ואז מריצים פעם שנייה, כאשר את הנקודה הראשונה לוקחים פעם אחת, על הנקודה השנייה חוזרים פעמיים, על השלישית שלוש פעמים וכן הלאה.

האם הפתרון בפעם השנייה יהיה שונה מזה של הפעם הראשונה:

א. עבור SVM ליניארי עם HARD MARGIN

כן ☐ לא ☒ הסבר:

כיוון שחייבים לקיים את כל האילוצים, ובודקים את ה-margin לפי המקרה הגרוע לכן אין משמעות לשכפול הנקודות.

ב. עבור SVM עם HARD MARGIN עם קרנל גאוסיני

כן ☐ לא ☒ הסבר:

כיוון שחייבים לקיים את כל האילוצים, ובודקים את ה-margin לפי המקרה הגרוע לכן אין משמעות לשכפול הנקודות.

ג. עבור SVM ליניארי עם SOFT MARGIN

כן ☒ לא ☐ הסבר:

כיוון שלא חייבים לקיים את כל האילוצים, שגיאה בנקודה ראשונה שווה חצי משגיאה בנקודה השנייה, ולכן פתרון האופטימיזציה יכול להשתנות.

ד. עבור SVM עם SOFT MARGIN עם קרנל פולינומיאלי מדרגה שנייה

כן ☒ לא ☐ הסבר:

כיוון שלא חייבים לקיים את כל האילוצים, שגיאה בנקודה ראשונה שווה חצי משגיאה בנקודה השנייה, ולכן פתרון האופטימיזציה יכול להשתנות.

תעודת זהות: ________________________

מספר מחברת: ________________________

שאלה 6 (15 נקודות).

עבור תוכנית AdaBoost הנח שההשערה החלשה שנלמדת h_t בעלת שגיאה ε_t

עבור כל אחד מהערכים של ε_t קבע איזה משקל תקבל ההשערה h_t והסבר מדוע זה סביר

א. $\varepsilon_t = 0.5$

$\boxed{\alpha_t=0}$

הסבר (למה זאת בחירה הגיונית מעבר לכך שזה מה שהאלגוריתם מבצע):

אם השגיאה חצי אזי אנו לא רוצים להתחשב בהשערה

____

ב. $\varepsilon_t = 0$

$\boxed{\alpha_t =\infty}$

הסבר (למה זאת בחירה הגיונית מעבר לכך שזה מה שהאלגוריתם מבצע):

אם להשערה אין שגיאה בכלל אנו רוצים תמיד להשתמש רק בה

____

____

ג. $0.5 < \varepsilon_t < 1.0$

$\boxed{\alpha_t=ln\frac{1-\epsilon_t}{\epsilon_t}}$

הסבר מה מיוחד בתחום הזה (ולמה זאת בחירה הגיונית מעבר לכך שזה מה שהאלגוריתם מבצע):

המקדם יוצא שלילי וזה הגיוני כיוון שהמינוס של ההערה יש שגיאה טובה מחצי.

____

____

____

תעודת זהות: ______________________

מספר מחברת: ______________________

שאלה 7 (15 נקודות)

בכל תא בטבלה למטה ציינו האם תוצאת האלגוריתם תשתנה כתוצאה מהפעלת הטרנספורמציה המצוינת. נמקו את תשובתכם בקצרה. תוצאת האלגוריתם – ניבוי התוויות של דוגמאות חדשות שעוברות כמובן את אותה טרנספורמציה.

x_0 הוא וקטור. a הוא סקלר שונה מאפס. U מטריצה יוניטרית ($U^t U=I$). A היא מטריצה ריבועית מדרגה מלאה. (הכל במספרים ממשיים).

	ז			
Hard margin linear SVM	☐ תשתנה ☐ לא תשתנה הסבר: אם הפתרון היה w ו- b לאחר הטרנספורמציה נקבל w b-w^t x_0 אותה נורמה של w	☐ תשתנה ☐ לא תשתנה הסבר: אם הפתרון היה w לאחר הטרנספורמציה נקבל w/a בתור פתרון לכל w אנו מחלקים אתהנורמה ב- a^2	☐ תשתנה ☐ לא תשתנה הסבר: הקרנל הליניארי נותר ללא שינוי ולכן הבעיה הדואלית זהה.	☐ תשתנה ☐ לא תשתנה הסבר: אם יש פחות נקודות מהמימד, ניתן עם A להגיע מכל בעיה לכל בעיה אחרת באותו מספר של נקודות חיוביות ושליליות
K Nearest Neighbor, k=3	☐ תשתנה ☐ לא תשתנה הסבר: המרחקים לא משתנים	☐ תשתנה ☐ לא תשתנה הסבר: המרחקים יוכפלו, אבל שכנויות לא ישתנו	☐ תשתנה ☐ לא תשתנה הסבר: המרחקים לא משתנים	☐ תשתנה ☐ לא תשתנה הסבר: אם יש פחות נקודות מהמימד ניתן להגיע ע"י A מכל בעיה לכל בעיה אחרת עם אותו פיזור של תוויות נקודות) בקונפיגורציה כללית)

תעודת זהות: ____________________

מספר מחברת: ____________________

שאלה 8 (13 נקודות).

עבור $\theta \sim U[0,1]$, $X|\theta \sim N(\theta, 1)$ נתון מדגם מ X בעל דגימה בודדת x

א. מהו משערך ML עבור θ

$$\theta^*_{ML} = \begin{cases} 1 & \text{for } x > 1 \\ x & for\ x \in [0,1] \\ 0 & \text{for } x < 0 \end{cases}$$

הסבר:

$$L = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2}\right)$$

$$\ell = -(x-\theta)^2/2 + const$$

$$\frac{d\ell}{d\theta} = -(x-\theta) = 0 \ \rightarrow \ \theta = x$$

צריך לשמור את $\theta \in [0,1]$

ב. מהו משערך MAP עבור θ

$\theta^*_{MAP} =$

הסבר: עבור ההתפלגות האחידה הפתרון זהה לסעיף א כיוון שהצפיפות של ה-prior היא 1

צריך לשמור את $\theta \in [0,1]$