

סהב	6	5	4	3	2	1

## מבחן דוגמה – 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

### 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: דף A4 - צד אחד בלבד, עם שם התלמיד ומספר זהות.
4. יש לענות על השאלות במקום המיועד לכך בטופס השאלון (טופס זה). מחברות הבחינה לא ייקראו, וישמשו כטייטה בלבד.
5. יש למלא בכל דף של השאלון מספר ת.ז. ומספר מחברת.
6. במבחן 8 שאלות:
  - הניקוד לכל שאלה מופיע לידיה 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!*

שאלה 1 (10 נקודות).

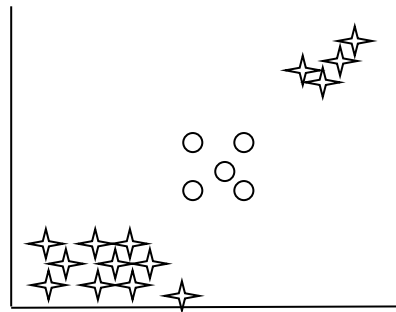
נתונים האלגוריתמים הבאים

א. SVM עם Polynomial kernel

ב. Perceptron

ג. מלבן דו-ממדי

נתון מדגם דו-מימדי הבא, בו שני סיווגים  $\star$  ו- $\circ$

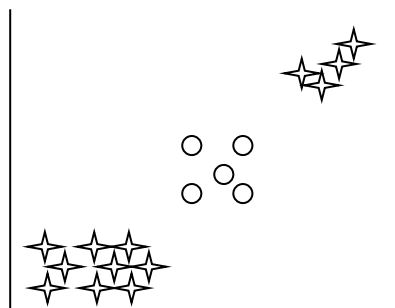


עבור כל אחד מהאלגוריתמים מריצים אותו עד שמקבלים מסווג עם שגיאת למידה אפס (על המדגם) קבע/י האם האלגוריתם יגיע לשגיאה אפס. אם לא הסבר מדוע. אם כן צייר קו הפרדה מתאים למסווג המתקבל:

א. SVM עם Polynomial kernel

לא יכול להגיע לשגיאה אפס. הסבר \_\_\_\_\_

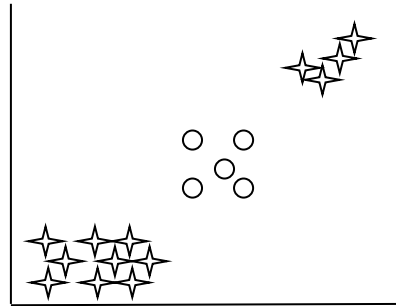
יכול להגיע לשגיאה אפס. קו הפרדה של המסווג:



ב. Perceptron

לא יכול להגיע לשגיאה אפס. הסבר \_

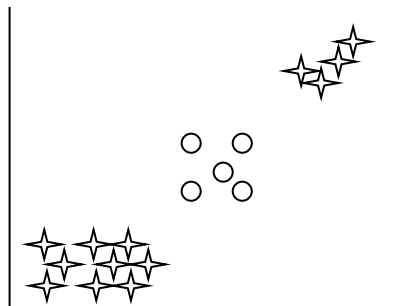
יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



ג. מלבן דו-מימדי (מוגדר ע"י  $(x_1, x_2, y_1, y_2)$  כאשר הסיווג של  $(x, y)$  חיובי אם  $x_1 \leq x \leq x_2$  ו-  $y_1 \leq y \leq y_2$ )

לא יכול להגיע לשגיאה אפס. הסבר \_\_\_\_\_

יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



תעודת זהות:

מספר מחברת:

**שאלה 2 (10 נקודות).**

מוצעים שני השינויים הבאים (כל אחד לחוד) לאלגוריתם ה-Perceptron:

א. במקום לעדכן אחרי כל שגיאה, נעדכן רק כל שגיאה עשירית.  
(כלומר, בשגיאה ה- $t$  של  $(x_t, y_t)$  אזי אם  $1 = t \bmod 10$  אזי  $w_{t+1} = w_t + \gamma x_t$  אחרת  $w_{t+1} = w_t$ ).

האלגוריתם עדיין מבצע רק מספר סופי של שגיאות. נכון  לא נכון  הסבר:

---

ב. לאחר כל שגיאה נעדכן  $w_{t+1} = -w_t/2$  כאשר  $w_1 = (1, \dots, 1)$

הראה שלעולם לא נבצע אותה שגיאה פעמיים רצוף עבור  $x \neq 0$ :

---

---

---

האלגוריתם עדיין מבצע רק מספר סופי של שגיאות. נכון  לא נכון  הסבר:

תעודת זהות:

מספר מחברת:

**שאלה 3 (10 נקודות).**

משנים את אלגוריתם הרגרסיה הלינארית כך שלכל נקודה  $i$  יש משקל  $w_i > 1$

א. כתוב את תוכנית האופטימיזציה המתאימה:

קלט  $(x_i, y_i)$   $1 \leq i \leq m$



ב. האם המשקולות ישנו את הפתרון (ביחס לפתרון שימצא אלגוריתם ללא משקלים לנקודות)?

כן  לא

אם כן תן דוגמא אם לא הוכח:

---

תעודת זהות:

מספר מחברת:

**שאלה 4 (10 נקודות)**

לאחר למידת KERNEL SVM נקבל כלל החלטה מהצורה  $f(x) = \sum_i \alpha_i y_i K(x_i, x) + b$   
נניח שפתרנו את בעיית האופטימיזציה הפרימאלית כדי ללמוד SOFT MARGIN SVM אבל שכחנו  
לרשום את הערך של  $b$ . איך נוכל לחשב באמצעות נקודה שמפירה את ה MARGIN את הערך של  $b$   
(כפונקציה של הפתרון הפרימאלי ללא שימוש בערך של  $b$ )?

**שאלה 5 (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.  $D$  היא מטריצה ריבועית אלכסונית, ללא אפס על האלכסון.  $U$  מטריצה יוניטרית ( $U^T U = I$ ).  $A$  היא מטריצה ריבועית מדרגה מלאה. (הכל במספרים ממשיים).

**AdaBoost over decision stumps:** the weak learners are obtained by considering for every feature  $i$  and for every possible threshold  $\theta$  weak learners of the form:  $f_{i,\theta}(x) = (x_i > \theta)$

	$T(x) = x + x_0$	$T(x) = ax$	$T(x) = Dx$	$T(x) = Ux$	$T(x) = Ax$
AdaBoost over decision stumps	<input type="checkbox"/> תשתנה <input type="checkbox"/> לא תשתנה הסבר:	<input type="checkbox"/> תשתנה <input type="checkbox"/> לא תשתנה הסבר:	<input type="checkbox"/> תשתנה <input type="checkbox"/> לא תשתנה הסבר:	<input type="checkbox"/> תשתנה <input type="checkbox"/> לא תשתנה הסבר:	<input type="checkbox"/> תשתנה <input type="checkbox"/> לא תשתנה הסבר:

תיבה רב מימדית: פרמטרים  $(c_1, \dots, c_d)$  ו-  $(b_1, \dots, b_d)$  מסווג דוגמא כחיובית אם לכל  $i$  מתקיים  $c_i \leq x_i \leq b_i$  בוחרים את התיבה הקטנה ביותר העקבית עם הדוגמאות

	$T(x) = x + x_0$	$T(x) = ax$	$T(x) = Dx$	$T(x) = Ux$	$T(x) = Ax$
תיבה רב מימדית	<input type="checkbox"/> תשתנה <input type="checkbox"/> לא תשתנה הסבר:	<input type="checkbox"/> תשתנה <input type="checkbox"/> לא תשתנה הסבר:	<input type="checkbox"/> תשתנה <input type="checkbox"/> לא תשתנה הסבר:	<input type="checkbox"/> תשתנה <input type="checkbox"/> לא תשתנה הסבר:	<input type="checkbox"/> תשתנה <input type="checkbox"/> לא תשתנה הסבר:

**שאלה 6 (15 נקודות)**

עבור בעיית קלסיפיקציה, נתונה מחלקת השערות סופית  $H$  מעל מרחב  $X$   
נתונה התפלגות  $P$  מעל  $H$  (התפלגות prior)  
נתונה התפלגות  $D$  מעל  $X$   
נבחרת פונקציית מטרה  $f$  לפי ההתפלגות  $P$   
נבחר מדגם של  $m=100$  דוגמאות  $S$  לפי  $D$  ומסווג לפי  $f$

לפי Maximum Likelihood (ML) נבחרת ההשערה  $h_{ml}$   
לפי Maximum A Posterior (MAP) נבחרת ההשערה  $h_{map}$   
לפי Bayes Inference מוגדרת ההשערה  $H_{bayes}$  הממצעת את השגיאה של כל  
ההשערות לפי ה-posterior. פורמלית:

$$H_{bayes}(x) = 1 \Leftrightarrow \sum_{h \in H} h(x) \Pr(h|S) > 0.5$$

נגדיר  $error(h) = \Pr_D [f(x) \neq h(x)]$

ממוצעים בביטויים הבאים הם על בחירת המדגם  $S$ .  
עבור כל טענה, קבע אם היא נכונה או לא.  
אם נכונה הוכח. אם לא נכונה, תן דוגמא נגדית או הסבר.

א: תמיד  $E[error(h_{ml})] < 2E[error(H_{bayes})]$  כן  לא

ב: יתכן כי  $E[error(h_{ml})] < E[error(h_{map})]$  כן  לא

ג: לכל מדגם  $S$  מתקיים  $error(H_{bayes}) \leq error(h_{ml})$  כן  לא



**שאלה 7 (15 נקודות)**

נתון מדגם ובו דגימות מהצורה  $(x, y)$  כך ש  $x$  הוא וקטור בעל מספר features.

עבור כל אחד מהאלגוריתמי הרגרסיה הבאים יש לבחור את ההשפעה של נרמול ה features ע"י הפעלת טרנספורמציה ליניארית במדגם (ללא נרמול ה  $y$ ) ביחס להרצת האלגוריתם ללא נרמול, מתוך האפשרויות הבאות:

1. קו הרגרסיה המתקבל לא משתנה
2. ייתכן שקו הרגרסיה משתנה אך השגיאה על המדגם לא משתנה
3. ייתכן שהשגיאה על המדגם משתנה

א. רגרסיה ליניארית  1  2  3

הסבר: \_\_

---

---

---

ב. Ridge Regression  1  2  3

הסבר: \_\_

---

---

---

ג. רגרסיה ליניארית Online (Stochastic Gradient Descent) המשתמשת בקבוע למידה  $\alpha=0.01$  לאחר 10000000 צעדים

1  2  3

הסבר: \_\_

**שאלה 8 (15 נקודות)**

נתונות דגימות  $x$  הנוצרות ע"פ פונקציית צפיפות של התפלגות אקספוננציאלית:

$$f(x; s) = se^{-sx}$$

כאשר  $s$  נבחר תחילה (עבור כל דגימה) מתוך שלוש אפשרויות  $s_1, s_2, s_3$  (איננו יודעים את ההסתברות לבחירת כל אפשרות,  $p_1, p_2, p_3$  בהתאמה).

נרצה להעריך את  $s_1, s_2, s_3$  ואת ההסתברויות  $p_1, p_2, p_3$ :

א. יש לכתוב את פונקציית ה  $\log$ -likelihood המתאימה (אין צורך לחשב ML)

ב. נרצה להשתמש באלגוריתם EM כדי לחשב עבור כל דגימה מאיזה משלושת המקורות נוצרה:  
יש לפתח ולנסח את נוסחת העדכון לשלב ה E:

ג. יש לפתח ולנסח את נוסחת העדכון לשלב ה M:

---

---