1	2	3	4	5	6	סהכ

תעודת זהות: _____

מספר מחברת: _____

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 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: _____

שאלה 1 (10 נקודות).

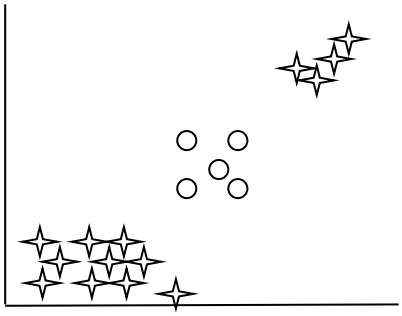
נתונים האלגוריתמים הבאים

א. SVM עם Polynomial kernel

ב. Perceptron

ג. מלבן דו-ממימדי

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



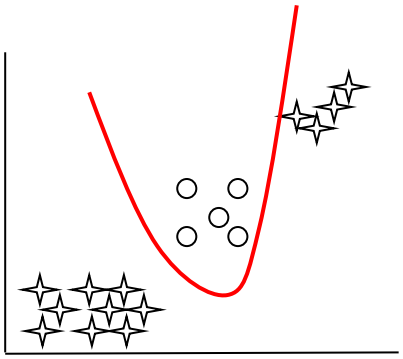
עבור כל אחד מהאלגוריתמים מריצים אותו עד שמקבלים מסווג עם שגיאת למידה אפס (על המדגם)

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

א. SVM עם Polynomial kernel

☐ לא יכול להגיע לשגיאה אפס. הסבר _____

☒ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



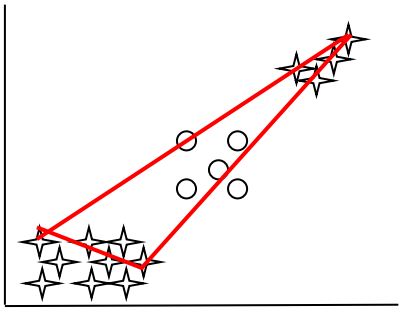
תעודת זהות: _____

מספר מחברת: _____

ב. Perceptron

☒ לא יכול להגיע לשגיאה אפס. הסבר _לא ניתן להפרדה ליניארית. ישנו עיגול שהוא בקמור של הכוכבים. ראה בציור.

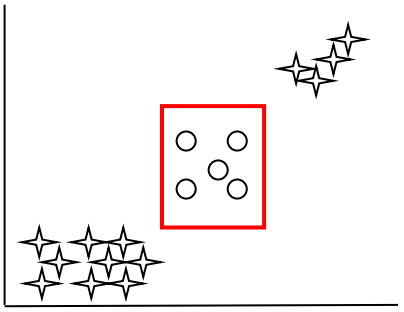
☐ יכול להגיע לשגיאה אפס. קו ההפרדה של המסווג:



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

☐ 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 $w_{t+1}=w_t+y_tx_t$ אחרת $w_{t+1}=w_t$.)

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

_נתבונן רק בתת-הסדרה של השגיאות עם עדכון משקולות. לפי משפט ה-perceptron מספר השגיאות בסדרה זאת חסום ע"י M. מכן שמספר השגיאות בסדרה המקורית חסום ע"י 10M+9

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 $x\neq 0$:

נניח ש- (x,y)שגיאה ו-w אלו המשקולות לפני העדכון ו-w' אחרי העדכון

אם $y=+1$ אזי $x^Tw < 0$ כיוון שזאת שגיאה . מכאן $x^Tw'=x^T(-w/2) >0$

אם $y=-1$ אזי $x^Tw > 0$ כיוון שזאת שגיאה . מכאן $x^Tw'=x^T(-w/2) <0$

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

על הסדרה $x_1=(1,0)$, $y_1=-1$, $x_2=(0,1)$, $y_2=+1$ וחוזרים עליה אין סוף פעמים, בכל צעד עושים שגיאה.
הסדרה ניתנת להפרדה לניארית ע"י $w=(-1,+1)$

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) $1\le i\le m$

$$\max_{z,b} \Sigma_i w_i (z^t x_i + b - y_i)^2$$

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

כן ☒ לא ☐

אם כן תן דוגמא אם לא הוכח:

עבור הנקודות (-1,1) (0,0) ו- (1,1) רגרסיה רגילה תתן z=0 ו- b=2/3 ורגרסיה עם משקולות w_1=1000, w_2=2 ו- w_3=1000 תתן z=0 ו- b=2000/2002

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)?

נשתמש באי שוויון לכל נקודת אימון:

$y(w^t x+b) \geq 1+\xi$

ואז (היות וייבחר ξ מינימלי) נקבל $b=(1+\xi)y - w^t x$

תעודת זהות:

מספר מחברת:

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 ($I=U^tU$) . 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 θ 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	☐ תשתנה ☒ לא תשתנה הסבר: במקום לבחור $x_i < b$ נבחר $x_i < b+x_{0,i}$	☐ תשתנה ☒ לא תשתנה הסבר: במקום לבחור $x_i < b$ נבחר $x_i < b/a$	☐ תשתנה ☒ לא תשתנה הסבר: במקום לבחור $x_i < b$ נבחר $x_i < b/d_i$	☒ תשתנה ☐ לא תשתנה הסבר: כיוון שיש לנו רק קווים מקבילים לצירים אם U מבצעת סיבוב של המימדים לא נוכל לבצע את אותו סיווג ע"י decision stumps בצירים החדשים	☒ תשתנה ☐ לא תשתנה הסבר: כיוון שיש לנו רק קווים מקבילים לצירים אם A מבצעת סיבוב של המימדים לא נוכל לבצע את אותו סיווג ע"י decision stumps בצירים החדשים

תיבה רב מימדית: פרמטרים $(c_1, \dots, c_d)$ ו- $(b_1, \dots, b_d)$
מסווג דוגמא כחיובית אם לכל i מתקיים $c_i \le x_i \le b_i$
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	☐ תשתנה ☒ לא תשתנה הסבר: נוסיף x_0 ל-c ו-b ונקבל את התיבה העקבית הקטנה ביותר. הסיווג יהיה זהה.	☐ תשתנה ☒ לא תשתנה הסבר: נכפיל את c ו-b ב- a ונקבל את התיבה העקבית הקטנה ביותר. הסיווג יהיה זהה.	☐ תשתנה ☒ לא תשתנה הסבר: נכפיל את c ו-b את המימד ה-i ב- d_i ונקבל את התיבה העקבית הקטנה ביותר. הסיווג יהיה זהה.	☒ תשתנה ☐ לא תשתנה הסבר: עבור נקודות בתוך מלבן מקביל לצירים בדו מימד, אחרי סיבוב ב45 מעלות נקבל מלבן מינימלי שמכיל נקודות נוספות. (1,1)=	☒ תשתנה ☐ לא תשתנה הסבר: חיובי מקרה פרטי של הסעיף Ux

תעודת זהות:
מספר מחברת:

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 פונקצית 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 ☒

יתכן שהשגיאה של ML הרבה יותר גרועה. לדוגמא אם ישנה השערה בודדת h שלאחר המדגם תמיד נשארת עם הסתברות גבוהה יותר לאחר 100 דוגמאות, וישנן הרבה מאוד השערות אחרות H עם משקל קצת יותר קטן ומאוד שונות מ-h. אזי לרוב תבחר השערה מתוך H והשגיאה של ML תהייה מאוד גבוהה, אך השגיאה של bayes מאוד נמוכה

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

יתכן ויש מספר רב מאוד של השערות. התפלגות D שמה משקל מאוד גדול על אחת (אבל זה עדיין זניח). MAP יבחר תמיד את ההשערה הזאת, ויספוג לרוב שגיאה גדולה. ML יתעלם מ-D ויבחר לרוב השערה די מדויקת.

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

יתכן שעבור דגימה מסויימת S ה- ML בחר את ההשערה הנכונה ויש לו שגיאה אפס. H_{bayes} יש שגיאה לא אפס כיוון שהוא ממצע גם על השערות עם שגיאה לא אפס.

תעודת זהות:
מספר מחברת:

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 לניארית 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 לינארית ☐1 ☒ 2 ☐ 3

הסבר:__לדוגמה, אם נכפיל כל משתנה ב-2 אז מקדם הרגרסיה שלו יוכפל ב-0.5 השגיאה הריבועית לא משתנה כי התחזיות זהות

_____
_____
_____

ב. Ridge Regression ☐1 ☐ 2 ☒ 3

הסבר:_אם נכפיל משתנה ב-2 ואת המקדם נחלק ב-2 אזי התחזית לא תשתנה אבל הנורמה בריבוע של המקדמים ירדה ל-0.25 וזה ישנה את ה-tradeoff בין השגיאה לנורמה של המקדמים ו"יאפשר" לנו לקחת מקדמים יותר גדולים "אפקטיבית".

_____
_____
_____

ג. רגרסיה לינארית Online (Stochastic Gradient Descent) המשתמשת בקבוע למידה α=0.01 לאחר 10000000 צעדים

☐1 ☐ 2 ☒ 3

הסבר:__הכפלת משתנים גוררת שינוי בקבוע הלמידה וכתוצאה מכך בקצב ההתכנסות או בהתכנסות בכלל, לפיכך ייתכן שהשינוי יגרור שינוי בקו הרגרסיה במתקבל (וכתוצאה מכך לשינוי בשגיאה

תעודת זהות:

מספר מחברת:

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 s1, s2, s3 (איננו יודעים את ההסתברות לבחירת כל אפשרות, p1, p2, p3 בהתאמה).

נרצה להעריך את s1, s2, s3 ואת ההסתברויות p1, p2, p3:

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

log-likelihood = $\sum_{i=1}^{m} \log[\sum_{j=1}^{3}(p_j s_j e^{-s_j x_i}$)]

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

$a_{ij} = \Pr(s = s_j|x_i; p_1,p_2,p_3,s_1,s_2,s_3,\}$ = $\Pr(s = s_j; p_{j,}\}\Pr(x = x_i|s = s_j\ ;\ s_{j,}\}$/ A_i

$=p_j\ s_j\ \exp(-s_j\ x_i)/A_i$

כאשר

$A_i= \Sigma_r\ p_r\ s_r\ \exp(-s_r\ x_i)$

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

$$Q(\theta|\theta^t) = Q(p_{1,}p_2,p_{3,}s_{1,}s_{2,}s_{3,}|\{a_{ij}\})$$

$$= \sum_i \sum_j a_{ij} \log \Pr(x_i, s_j|p_{1,}p_2,p_{3,}s_{1,}s_{2,}s_{3,})$$

$$= \sum_i \sum_j a_{ij}\ log p_j s_j e^{-s_j x_i}$$

בעיית האופטימיזציה המתאימה היא לפיכך:

$\max_{p,s} \Sigma_i\ \Sigma_j\ a_{i,j}\ (\log\ p_j + \log\ s_j - s_j x_i)$

such that $\Sigma_j\ p_j=1,\ p_j \geq 0$

ופתרונה

$p_j= \Sigma_i\ a_{i,j}/m$

$s_j= (\Sigma_i\ a_{i,j})/(\ \Sigma_i\ x_i\ a_{i,j})$

תעודת זהות: _____

מספר מחברת: _____