



## מבוא למערכות לומדות (236756)

סמסטר אביב תשע"ו

### מבחן מסכם מועד א', 6 ביולי 2016

--	--	--	--	--	--	--	--	--

מספר סטודנט:

משך המבחן: 3 שעות.

חומר עזר: אין להשתמש בכל חומר עזר. בעמוד הבא לרשותכם דף נוסחאות והגדרות.

#### הנחיות כלליות:

- המבחן כתוב בלשון זכר ומיועד לנשים ולגברים כאחד.
- מלאו את הפרטים בראש דף זה ובדף השער המצורף, בעט בלבד.
- במבחן 18 דפים ממוספרים סהכ, כולל עמוד זה שמספרו 1. ודאו שיש לכם כל הדפים.
- במבחן 5 שאלות. יש לענות על כל השאלות.
- משך המבחן 3 שעות (180 דקות).
- כל התשובות יכתבו על טופס הבחינה, ויש להחזירו בתום הבחינה.
- אנא כתבו בכתב יד קריא וברור. תשובה בכתב יד שאינו קריא לא תיבדק.
- נא לא לתלוש עמודים ממחברת הבחינה.
- נא לכתוב רק את מה שהתבקשתם ולצרף הסברים קצרים רק כפי שמבוקש בשאלה—אין צורך בהסברים או פרטים נוספים על אלו שהתבקשתם במפורש.

## כל המוסיף גורע

1. המילה העברית ל-feature היא תכונה או מאפיין. המילה העברית ל-label היא תיוג.

**בהצלחה!**



## דף נוסחאות

1. אלא אם כן מצוין אחרת, מרחב התיגים  $\mathcal{Y}$  הוא  $\{-1, +1\}$ .

2.  $\binom{n}{k} \leq n^k$

3.  $L_D^{01}$  = true error = שגיאת הכללה

4.  $L_S^{01}$  = שגיאה אמפירית (ממוצע השגיאות על מדגם)

5. ~~שגיאת אסטימציה~~  $L_D^{01} - L_S^{01}$  (ההגדרה הנכונה תוקנה במבחן על הלוח)

6.  $e \approx 2.72$

7.  $\mathcal{H}$  תהא מחלקת היפוטזות של בעיית למידה כלשהי, ו- $S$  קבוצת אימון שנבחרת באקראי. נסמן

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}} L_D^{01}(h)$$

$$h^* = \operatorname{argmin}_{h \in \mathcal{H}} L_S^{01}(h)$$

אזי, לכל  $\delta > 0$ : בהסתברות של לפחות  $1 - \delta$  מתקיים:

$$L_D^{01}(\hat{h}) \leq L_D^{01}(h^*) + O\left(\sqrt{\frac{VCDIM(\mathcal{H}) + \frac{1}{\log(\delta)}}{|S|}}\right)$$



## חלק א : שאלות קצרות (30 נק')

1. נתונה מחלקת היפותזה סופית  $\mathcal{H}$  שמכילה בידויק 4096 פונקציות סיווג בינאריות. אז מימד ה

VC של  $\mathcal{H}$  הוא בידויק 12. הטענה האחרונה הינה (יש לסמן אפשרות אחת):

אמת ☐

שקר ☒ דוגמא נגדית (חובה לספק במקרה שסימנתם "שקר"):

$$\mathcal{X} = \{1, 2, \dots, 4096\}$$

$$\mathcal{H} = \{h_1, h_2, \dots, h_{4096}\}$$

$$h_i(x) = \begin{cases} +1 & x = i \\ -1 & \text{otherwise} \end{cases}$$

2. משפט סאוור-שלח (Sauer-Shelah) חוסם את פונקציית הגידול (growth function) ☒

כפונקציית של מימד ה VC. הטענה האחרונה הינה (יש לסמן אפשרות אחת):

אמת ☒ (אכן, המשפט חוסם את פונקציית הגידול בפולינום במימד ה VC)

שקר ☐

3. נניח שעבור משפחת היפותזות  $\mathcal{H}$  של מסווגים בינאריים ( $\mathcal{Y} = \{\pm 1\}$ ) מעל מרחב דוגמאות  $\mathcal{X}$

כלשהו מתקיימת התכונה הבאה:

לכל  $\epsilon, \delta \in (0, 1)$  ולכל  $m \leq 2^{\frac{1}{\epsilon} + \frac{1}{\delta}}$  קיימת התפלגות  $\mathcal{D}$  מעל  $\mathcal{X} \times \mathcal{Y}$  כך שכל אלגוריתם

שרץ על קלט (קבוצת אימון)  $S = (x_1, y_1) \dots (x_m, y_m)$  שמוגרל מההתפלגות ומחזיר

היפותזה  $h$  מקיים שבסיכוי לפחות  $\delta$ :

$$L_{\mathcal{D}}(h) > \min_{h' \in \mathcal{H}} L_{\mathcal{D}}(h') + \epsilon$$

אז  $\mathcal{H}$  אינה למידה PAC (Not PAC Learnable). טענה זו הינה (יש לסמן אפשרות אחת):

אמת ☒ (אילו  $\mathcal{H}$  היתה למידה PAC אז היה לה VCdim סופי ואז היה אפשר להשיג

שגיאת הכללה  $\epsilon$  מעל לאופטימום בסיכוי  $1 - \delta$  עם מספר דוגמאות פולינומיאלי ב

$$\frac{1}{\epsilon}, \frac{1}{\delta} \text{ עבור כל התפלגות } \mathcal{D}, \text{ עבור כל } \epsilon, \delta, \text{ וזאת בניגוד להנחה}$$

שקר ☐

4. תהיה  $S$  קבוצת אימון כלשהי בגודל 500, ו  $S'$  תת-קבוצת אימון הכוללת בידויק את 100

הדוגמאות הראשונות של  $S$ . יהי  $h$  פלט של אלגוריתם למידה כלשהו על קלט  $S'$ . אז  $L_S(h)$

הוא דגימה בלתי מוטית (unbiased estimator) של  $L_{\mathcal{D}}(h)$ . (כלומר, התוחלת של  $L_S(h)$  היא

בידויק  $L_{\mathcal{D}}(h)$ ). טענה זו הינה (יש לסמן אפשרות אחת):



אמת ☐

שקר ☒ הסבר קצר (חובה לספק במקרה שסימנתם "שקר"):

כי  $S$  ו  $S'$  לא מדגמים בלתי תלויים

5. הגדירו Improper Efficient PAC Learning, ותנו דוגמא לבעייה שהיא

Improper efficient PAC Learnable אבל לא ידוע אם היא

Efficient PAC Learnable

נאמר שסדרה של מחלקת היפותזות  $\mathcal{H}_n$  מעל  $\mathcal{X}$  היא Improper Efficient PAC Learning אם קיים אלגוריתם  $A$  וקיים פולינום  $p(\cdot, \cdot, \cdot)$  כך שלכל התפלגות  $\mathcal{D}$  מעל  $\mathcal{X} \times \mathcal{Y}$ , ולכל  $\epsilon, \delta$  הפלט של  $A$  על קבוצת אימון  $S = (x_1, y_1) \dots (x_m, y_m)$  שמוגדרת מ  $\mathcal{D}^m$  עבור  $m = \Theta(p(\frac{1}{\epsilon}, \frac{1}{\delta}, n))$  הוא בסיכוי לפחות  $1 - \delta$  היפותזה  $h$  שלא בהכרח שייכת ל  $\mathcal{H}_n$ , המקיימת  $L_{\mathcal{D}}(h) \leq L_{\mathcal{D}}(h^*) + \epsilon$ , כאשר  $h^* = \operatorname{argmin}_{h' \in \mathcal{H}} L_{\mathcal{D}}(h')$ . בנוסף, האלגוריתם  $A$  רץ בזמן פולינומי בקלט שלו, וכן ניתן לחשב את היפותזות הפלט  $h$  בזמן פולינומי בגודל הקלט שלה.

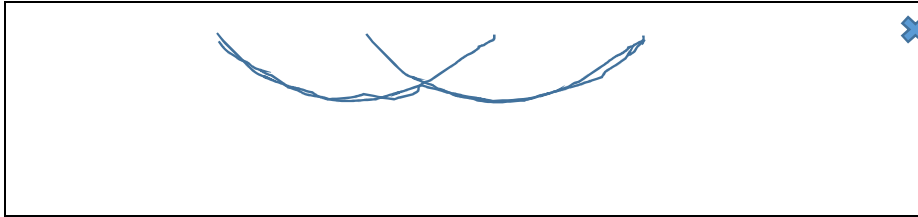
3-Term DNF



6. מינימום של שתי פונקציות קמורות נותן פונקציה קמורה. טענה זו הינה (יש לסמן אפשרות אחת):

☐ אמת

☒ שקר דוגמא נגדית (חובה לספק במקרה שסימנתם "שקר"):



7. כידוע, פונקציית המטרה של Soft-SVM היא  $L_S^{hinge}(w) + \lambda \|w\|^2$

ככל שמגדילים את  $\lambda$ , כך צפוי ששגיאת ההכללה האסטימציה (יש לסמן אפשרות אחת):

☒ תרד (הגדלת  $\lambda$  מעודדת  $\|w\|^2$  קטן יותר, כלומר למרחב היפותזות קטן יותר, ולכן לאסטימציה יותר טובה)

☐ תעלה

☐ תישאר ללא שינוי

8. ככל שמגדילים את מספר האיטרציות של AdaBoost, כך צפוי ששגיאת ההכללה האסטימציה (יש לסמן אפשרות אחת):

☐ תרד



☒ תעלה (כי מחלקת ההיפותזות האפקטיבית, המורכבת מצירופים של היפותזות בסיסיות, גדלה)

☐ תישאר ללא שינוי

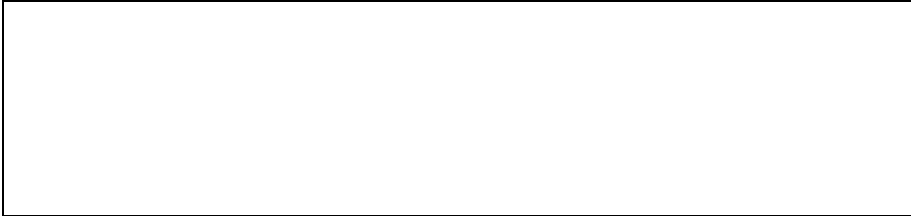
9. נתונה מחלקת היפותזות  $\mathcal{H}$  של מסווגים בינאריים מעל מרחב דוגמאות  $\mathcal{X}$  כלשהו. ידוע שמימד ה VC של  $\mathcal{H}$  הוא לפחות 10. נגדיר מרחב דוגמאות חדש  $\mathcal{X}' = \mathcal{X} \times \mathcal{X}$  של כל הזוגות הסדורים של דוגמאות מ  $\mathcal{X}$ , ומחלקת היפותזות  $\mathcal{H}' = \{h' | h \in \mathcal{H}\}$ , כאשר לכל  $h \in \mathcal{H}$  ההיפותזה  $h'$  מוגדרת כדלקמן:

$$\forall (x, x') \in \mathcal{X} \times \mathcal{X}: \quad h'((x, x')) = \begin{cases} +1 & h(x) = h(x') \\ -1 & \text{otherwise} \end{cases}$$

10. אז מתקיים שמימד ה VC של  $\mathcal{H}'$  הוא לפחות 5. טענה זו הינה (יש לסמן אפשרות אחת):

☒ אמת (ניתן לקחת קבוצה בגודל 10 שניתנת לניתוח ע"י  $\mathcal{H}$  ולחלק אותה לחמישה זוגות)

☐ שקר דוגמא נגדית: (חובה לספק במקרה שסימנתם "שקר"):





## חלק ב: superml.com (26 נק')

לאהרון יש דג באקווריום. אהרון החליט לבדוק אם הדג מגיב לתמונות. הוא הראה לדג תמונות מתוך הסמארטפון שלו, וגילה שהדג מגיב לחלקן ולאחרות הוא לא מגיב בכלל. הוא החליט לבדוק אם אפשר לנבא לאילו תמונות הדג מגיב. אהרון שמע על למידה חישובית, אבל אינו רוצה להיכנס לתחום. הוא מצא באינטרנט שרות מקוון <https://train.superml.com> שמקבל כקלט אוסף תמונות מסוגות כן/לא, לומד משקולות של רשת עמוקה כלשהי (סודית), מחזיר את שגיאת האימון, **שומר בסוד את המשקולות שנלמדו ומחזיר מפתח** כלשהו (למשל, מספר סידורי) ייחודי למשקולות שנלמדו. כדי לעשות פרדיקציה (ניבוי) יש לקרוא לשרות מקוון אחר <https://predict.superml.com> שמקבל כקלט תמונה ומפתח, ומחזיר את הפרדיקציה של התמונה לפי המשקולות שמזוהות עם המפתח.

עלות שרות הלמידה המקוון (זמן וכסף) היא לינארית בגודל קבוצת האימון. עלות שרות הפרדיקציה היא יחידה אחת של זמן/כסף. קבלת סיווג אמיתי (מהדג) לוקח לאהרון (ולדג) יחידת זמן אחת.

היעד של אהרון הוא שגיאת הכללה  $\epsilon_0$  בסיכוי  $1 - \delta_0$  עבור שני פרמטרים  $\epsilon_0, \delta_0$  קבועים כלשהם. יש לו תקציב של מיליון יחידות זמן/כסף. במידה ש [superml.com](https://superml.com) יכול להשיג ביצועים אלה, אהרון כמובן מעוניין "לא לצאת פראיר" ולשלם לא הרבה יותר מהמינימום הנדרש.

רישמו (בפסאודוקוד, או כרשימה של הוראות פשוטות וברורות בעברית) תכנית עבודה מפורטת עבור אהרון, המבוססת על עקרונות שנלמדו בקורס. (דוגמא לעיקרון שלא נלמד בקורס: "לחפש בפייסבוק מה אנשים כותבים על [superml.com](https://superml.com), ולפעול בהתאם").  
שימו לב: מותר להשתמש בסימונים  $O(\dots)$ ,  $\Theta(\dots)$  וכו' כדי להתייחס לביטויים שידועים לכם רק עד כדי קבוע אוניברסלי...



הנקודה כאן היא שלא ידוע מספר הדוגמאות  $m$  הדרושות לצורך השגת המטרה. לכן צריך "לחפש" את  $m$ . כדי לשמור על עלות לכל היותר קבוע כפול המינימום הדרוש, ננסה את  $m$  בסדרה גיאומטרית (למשל  $1, 2, 4, 8, 16, \dots$  עד שנגמר התקציב). כדי לדעת מתי אפשר לעצור, יש לעשות ולידציה.

Foreach  $i=1,2,3,\dots$  until end of budget

Let  $m = 2^i$

Generate fresh training set of size  $m$  (Note: doesn't have to be fresh, can reuse)

Call training service, obtain key

// now validate

Draw  $\Theta\left(\frac{\log\left(\frac{2^i}{\delta_0}\right)}{\epsilon_0^2}\right)$  fresh examples, denote them  $S$  (Note: Here *has* to be fresh)

Compute empirical error on  $S$  (using last key)

If the error is at most  $\epsilon_0/2$  then stop and return 'Success' with key

End

Return 'Fail'

Note: The reason we need  $2^i$  in the log of the validation set size is because we want the sum of the validation failure probabilities to be order of  $\delta_0$  (here we took validation failure probability sequence  $\frac{\delta_0}{2}, \frac{\delta_0}{4}, \frac{\delta_0}{8}, \dots$ )

Note: The total cost is at most  $O(1)$  times the cost required to train using the "correct" number  $m$  of samples.





### חלק ג : גרעינים (Kernels) 12 נק'

יהיו  $K_1, K_2$  ו-  $K_3$  גרעינים תקינים. הגרעין  $K_1$  מתאים למיפוי מאפיינים

$\phi(x) \in \mathbb{R}^{75}$  (feature map). הגרעין  $K_2$  מתאים למיפוי מאפיינים  $\phi(x) \in \mathbb{R}^{75}$

$\mathbb{R}^{100}$ . הגרעין  $K_3$  מתאים למיפוי מאפיינים  $\psi(x) \in \mathbb{R}^{20}$  המקיים:

$$\forall i = 1..20, \forall x: \psi_i(x) = 0.5 \phi_i(x)$$

סמנו את התשובה הנכונה בעיגול. בכל מקרה שבו סימנתם "כן", עליכם לתאר את מיפוי

המאפיינים המתאים לגרעין התקין שהתקבל.

לא	כן	דוגמא: האם $K_1(x, x')$ גרעין תקין באופן כללי? מיפוי: $\theta(x)$
לא	כן	האם $K_1(x, x') + K_2(x, x')$ גרעין תקין באופן כללי? מיפוי: $(\theta(x), \phi(x))$
לא	כן	האם $K_1(x, x') + 10 \cdot K_2(x, x')$ גרעין תקין באופן כללי? מיפוי: $(\theta(x), \sqrt{10}\phi(x))$
לא	כן	האם $K_1(x, x') - K_2(x, x')$ גרעין תקין באופן כללי? מיפוי:
לא	כן	האם $K_2(x, x') - K_3(x, x')$ גרעין תקין באופן כללי? מיפוי:
לא	כן	האם $K_3(x, x') - K_2(x, x')$ גרעין תקין באופן כללי? מיפוי:
לא	כן	האם $(K_2(x, x') - 1)^3$ גרעין תקין באופן כללי? מיפוי:



הפקולטה למדעי המחשב  
סמסטר אביב תשע"ו 2015/16

הטכניון, מכון טכנולוגי לישראל  
מבוא למערכות לומדות

--	--	--



## חלק ד: מודלים הסתברותיים (22 נק')

דני רוצה לאמן מודל הסתברותי כלשהו בשביל לחזות את הנימה של ביקורות לסרטים (חיובית / שלילית). לצורך כך הוא משתמש בקבוצת אימון מתויגת  $S$  המכילה מספר רב של ביקורות לסרטים. בנוסף בכדי לאמן את המודל לקח דני מילון ענק  $V$  המכיל חלק ניכר מהמילים בשפה העברית. לצורך אימון המודל הצטרך דני לשערך את ההסתברויות הבאות  $\{P(x) = \text{prob } x \text{ appears in review} | x \in V\}$ .  
דני בחר לשערך זאת באופן הבא:

$$\tilde{P}(x) = \frac{\# \text{training examples containing } x}{\# \text{ training examples}}$$

1. מהי השיטה הסטטיסטית בה דני השתמש?

MLE

2. מה הבעייתיות בשיטה שלו?

עבור  $x$  כך ש  $P(x)$  קטן יחסית, בסיכוי לא מבוטל נקבל  $\tilde{P}(x) = 0$ . אם תגיע עכשיו דוגמא שמכילה את המילה  $x$  אז נקבל  $\text{Likelihood}=0$ .  
(באופן סימטרי, גם אם  $P(x)$  קרוב מדי ל 1).

3. הצע פתרון לדני בכדי להתמודד עם בעיה זו?



MAP שנותן סיכוי  $< 0$  לכל אפשרות של הפרמטר  $\Theta_x = P(x)$  ב  $[0,1]$  (לדוגמא, התפלגות אחידה על התחום, ששקולה לפיתרון הידוע בשם Laplace Smoothing).

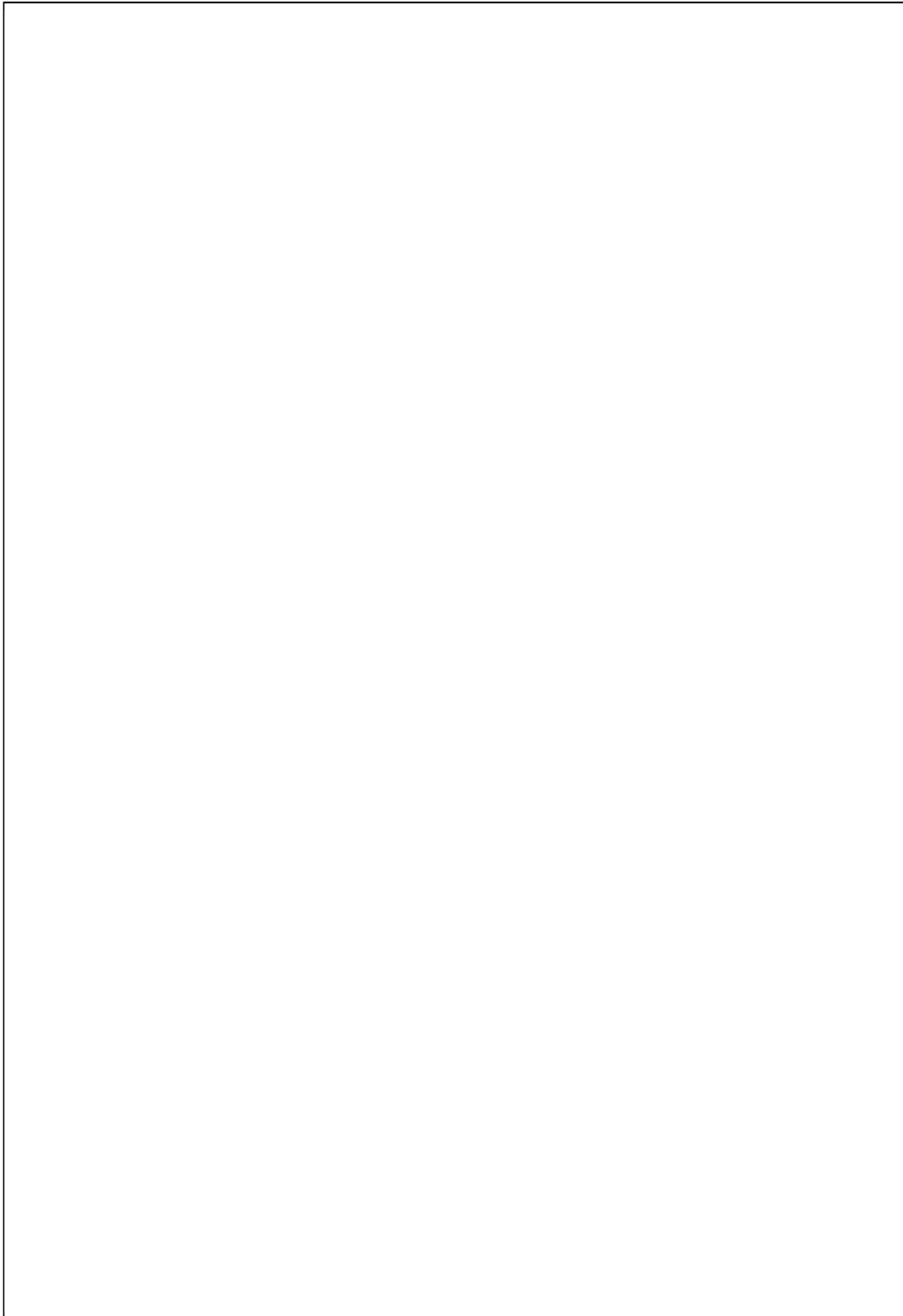
4. התפלגות "יש עתיד" מוגדרת כהתפלגות עם פונקציית הצפיפות הבאה:

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

נניח כי  $x_1, \dots, x_n$  נדגמו i.i.d מהתפלגות "יש עתיד" עם פרמטרים  $\mu, \sigma^2$  לא ידועים. מהם המשערכים ל  $\mu, \sigma^2$  לפי MLE? (אין קשר בין סעיף זה לסעיפים הקודמים). (אין צורך לרשום הוכחה)

$$\hat{\mu} = \frac{\sum_{i=1}^n \ln(X_i)}{n};$$

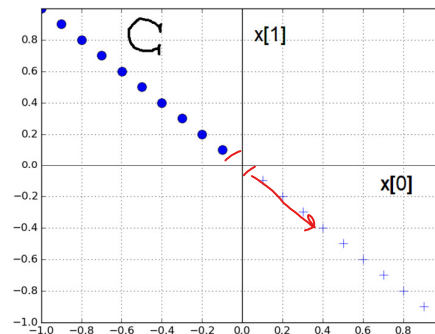
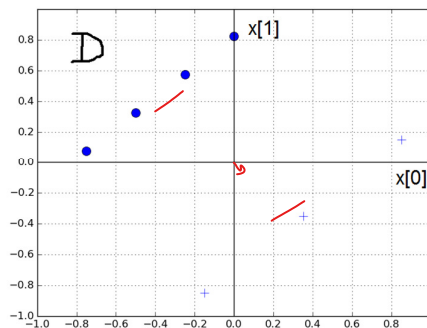
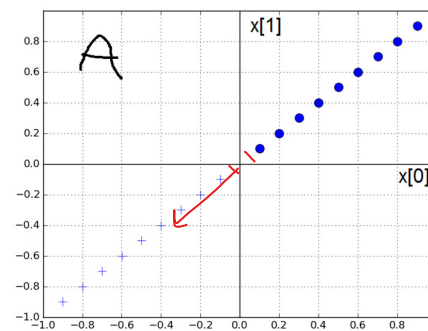
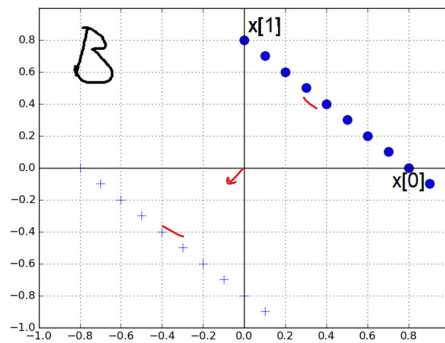
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \left( \ln(X_i) - \frac{\sum_{i=1}^n \ln(X_i)}{n} \right)^2}{n}$$





## חלק ה: SVM (10 נק')

בציורים שלפניכם משורטטות קבוצות אימון (training sets) ניתנות להפרדה, עבור 4 מקרים A,B,C,D. בכל המקרים, מרחב הדוגמאות הוא  $\mathbb{R}^2$ , העיגולים מסמנים דוגמאות שליליות והפלוסים '+' מסמנים דוגמאות חיוביות.



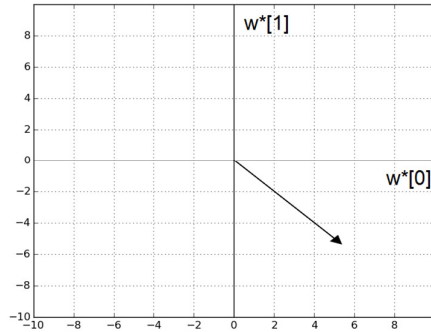
עבור כל אחד מהמקרים A,B,C,D הריצו Hard-SVM. להזכירכם, זו ההגדרה של

Hard-SVM:

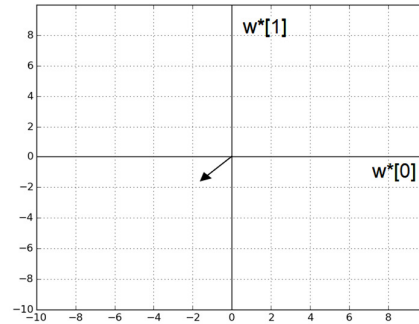
$$(w^*, b^*) = \underset{\{w \in \mathbb{R}^2, b \in \mathbb{R} : \forall i=1..m \langle w, x_i \rangle y_i - b \geq 1\}}{\operatorname{argmin}} \|w\|^2$$



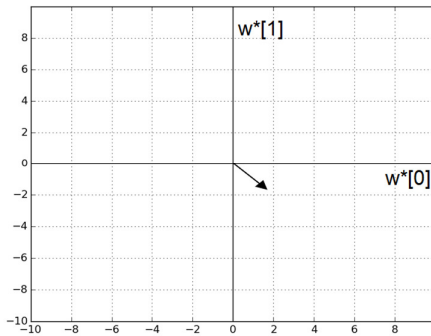
לפניכם שירטוט של הווקטור האופטימלי  $w^*$  המוחזר בכל אחד מהמקרים. בגלל בילבול, סדר השירטוטים השתבש. עליכם לרשום במסגרות המתאימות לאיזה מארבעת המקרים (A,B,C,D) מתאים כל שירטוט של  $w^*$ .



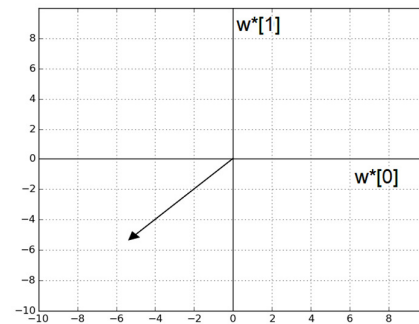
C



B



D



A



עמוד לטיוטה





עמוד לטיוטה



## עמוד לטיוטה



## עמוד לטיוטה



## עמוד לטיוטה