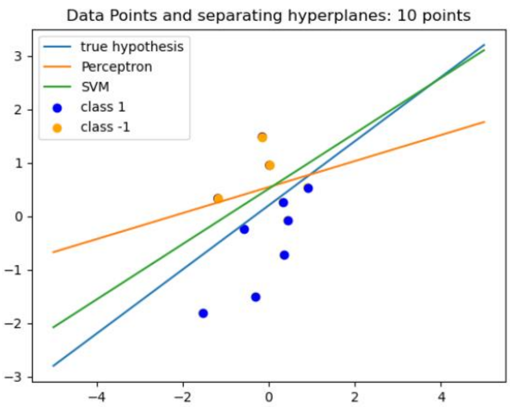
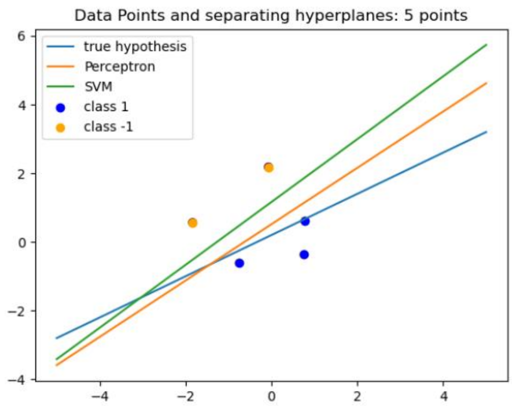
316163260

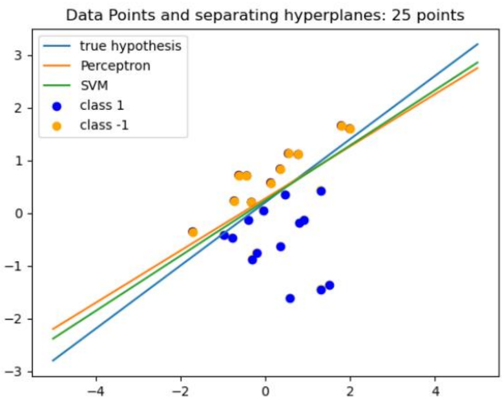
נועה בן דרור

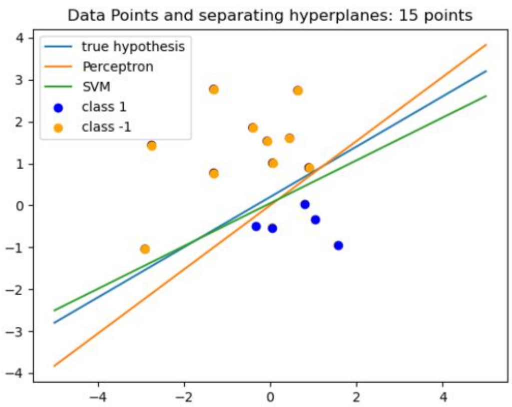
<u>IML – תרגיל 3 – חלק מעשי</u>

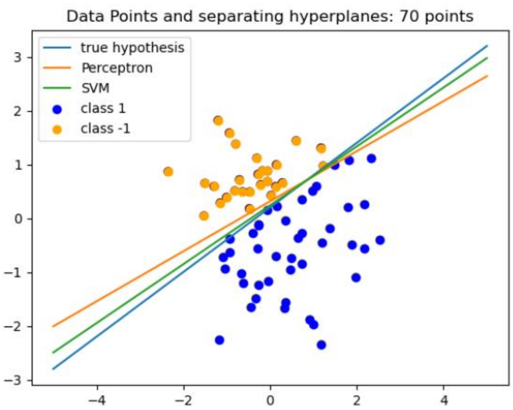
(9



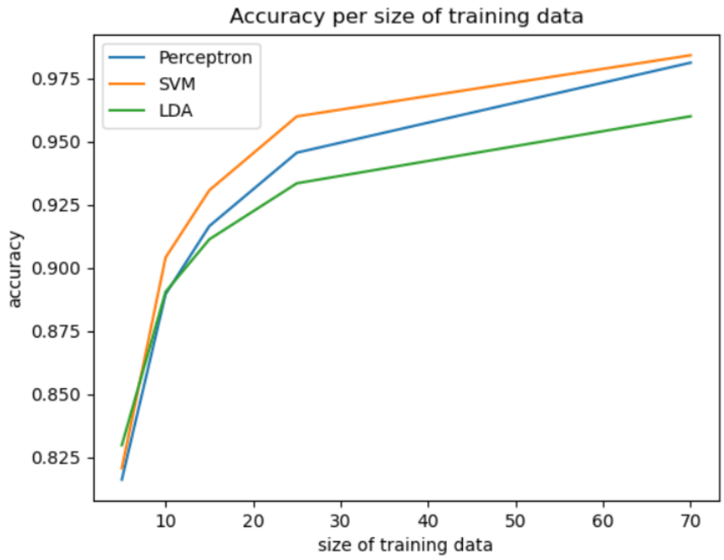






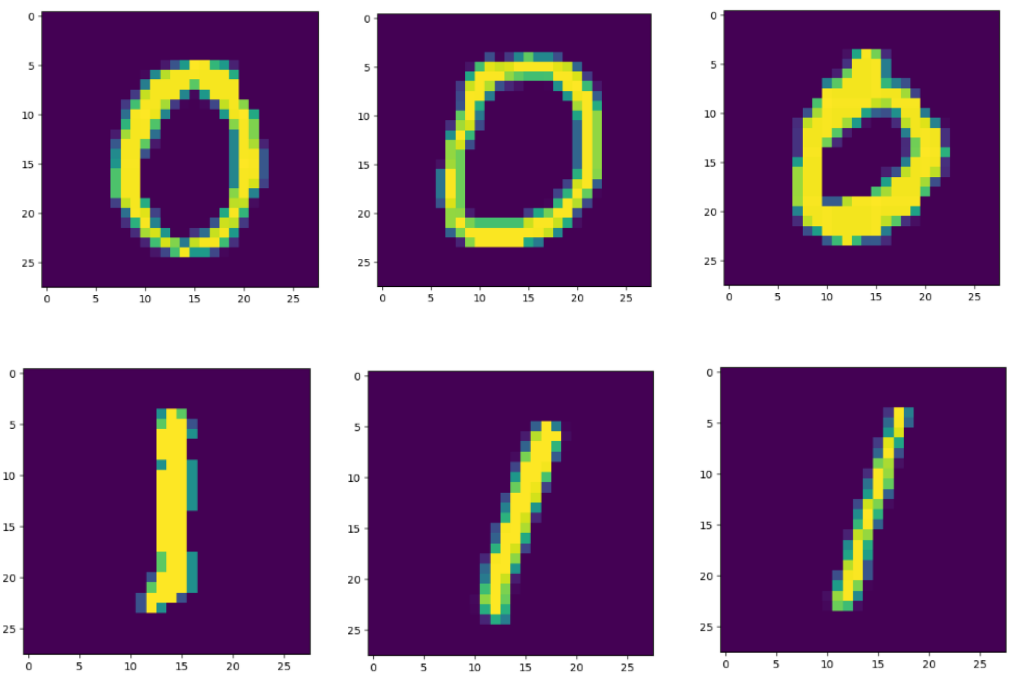


10)

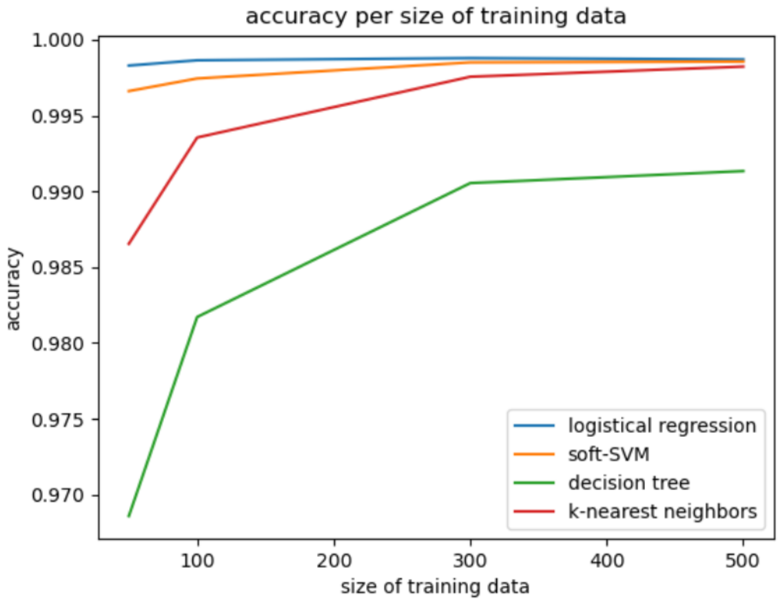


11) ה – classifier שעשה את העבודה הכי טובה הוא ה- SVM. נזכיר שה- hard-SVM ממקסם את השוליים, ולכן מתקיימת רמה נוספת של אנליזה שמודל הפרספטרון לא לוקח בחשבון. LDA עשה את העבודה הכי פחות טובה, כי הוא מניח הנחה שלא מתקיימת בדאטה שלנו, ומשתמש בשערוכים.

12)



(14



logistical regression time: 0.0320210599899292 m: 50

soft svm time: 0.05595876693725586 on m: 50

decision tree time: 0.011732168197631836 on m: 50

knn time: 0.2588677549362183 on m: 50

logistical regression time: 0.027227387428283692 m: 100

soft svm time: 0.06716924190521241 on m: 100

decision tree time: 0.01156142234802246 on m: 100

knn time: 0.4352590084075928 on m: 100

logistical regression time: 0.039820923805236816 m: 300

soft svm time: 0.1008923101425171 on m: 300

decision tree time: 0.023917679786682126 on m: 300

knn time: 1.298315396308899 on m: 300

logistical regression time: 0.03851727962493897 m: 500

soft svm time: 0.09460421562194823 on m: 500

decision tree time: 0.02191664218902588 on m: 500

knn time: 1.710689787864685 on m: 500

ניתן לראות כי לכל m, לאלגוריתם ה-knn לוקח הכי הרבה זמן לרוץ. אחריו, לכל m, נמצא soft-svm מבחינת זמן ריצה. לאחר מכן, נמצא logistic regression, ואחריו decision tree, שרץ בזמן הקצר ביותר לכל m.

השתמשתי בספריית sk-learn בשביל ארבעת המודלים.

נועה בן דרור
316163260

תרגיל 3 IML - חלק תיאורטי

(1) לפי חוק בייס מתקיים כי: $P(X|Y)\cdot P(Y)=P(Y|X)\cdot P(X)$ וכן

$$h_D = \begin{cases} +1 & P(y=1|x) \geq \frac{1}{2} \\ -1 & \text{otherwise} \end{cases}$$

אז נקבל:

$$P(y=1|x) \overset{\text{בייס}}{=} \frac{P(x|y=1)\cdot P(y=1)}{P(x)}$$

אם הפונקציה h_D מחזירה 1 זה אומר ש- $P(y=1|x) \geq \frac{1}{2}$, וזה אומר:

$$h_D(x)=1$$

$$\Downarrow$$

$$P(y=1|x) \geq \frac{1}{2}$$

$$\Downarrow$$

$$\underset{y\in\{\pm1\}}{\operatorname{argmax}}\{P(y=1|x)\} = \underset{y\in\{\pm1\}}{\operatorname{argmax}}\left\{\frac{P(x|y=1)\cdot P(y=1)}{P(x)}\right\} =$$

(P(X) בלתי תלוי ב- y)

$$\underset{y\in\{\pm1\}}{\operatorname{argmax}}\{P(x|y=1)\cdot P(y=1)\}$$

באותו אופן זה מתקיים אם הפונקציה h_D מחזירה -1 באופן סימטרי. לכן נקבל כי

$$h_D = \underset{y\in\{\pm1\}}{\operatorname{argmax}}\{P(x|y)\cdot P(y)\}$$

② מתקיים:

$$h_D(x) = \underset{y\in\{\pm1\}}{\operatorname{argmax}}\{P(x|y)\cdot P(y)\} = \underset{y\in\{\pm1\}}{\operatorname{argmax}} \ln\{P(x|y)\cdot P(y)\} =$$

(ln פונ' מונ' עולה)

$$\cancel{\underset{y\in\{\pm1\}}{\operatorname{argmax}} \{P(x|y)}$$

$$\underset{y\in\{\pm1\}}{\operatorname{argmax}}\{\ln P(x|y) + \ln P(y)\} =$$

$$\underset{y\in\{\pm1\}}{\operatorname{argmax}}\left\{\ln\left(\frac{1}{(2\pi)^d \det(\Sigma)} \exp\left(-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)\right)\right) + \ln P(y)\right\} =$$

$$\underset{y\in\{\pm1\}}{\operatorname{argmax}}\left\{\ln\left(\exp\left(-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)\right)\right) + \ln P(y)\right\} =$$

$$\underset{y\in\{\pm1\}}{\operatorname{argmax}}\left\{-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y) + \ln P(y)\right\} =$$

$$\underset{y\in\{\pm1\}}{\operatorname{argmax}}\left\{-\frac{1}{2}x^T\Sigma^{-1}x + \frac{1}{2}\mu_y^T\Sigma^{-1}x + \frac{1}{2}x^T\Sigma^{-1}\mu_y - \frac{1}{2}\mu_y^T\Sigma^{-1}\mu_y + \ln P(y)\right\} =$$

$$\underset{y\in\{\pm1\}}{\operatorname{argmax}}\left\{x^T\Sigma^{-1}\mu_y - \frac{1}{2}\mu_y^T\Sigma^{-1}\mu_y + \ln P(y)\right\} =$$

נגדיר. $$\underset{y\in\{\pm1\}}{\operatorname{argmax}}\{\delta_y(x)\}$$

(3) עבור כל נקודה שעבורה $y=1$, ונניח שיש $x_n,\dots,x_1$ נקודות כאלה.

נחשב:

$$\mu_{+1}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\mu_{-1}=\frac{1}{m-n}\sum_{j=1}^{m-n}x_j$$

$$\Sigma=\frac{1}{m}\sum_{i=1}^{m}(x_i-\mu_i)(x_i-\mu_i)^T$$

(δ זו הדלתא של קרונקר)

$$P(y=1)=\frac{1}{m}\sum_{i=1}^{m}\delta(y_i,1)$$

$$P(y=0)=\frac{1}{m}\sum_{i=1}^{m}\delta(y_i,0)$$

(4) יש 2 סוגי שגיאות:

false positive - סיווג של לא-ספאם כספאם.

false negative " " ספאם כלא-ספאם.

השגיאה הראשונה היא הגרועה יותר, כי אז מייל שהוא לא-ספאם, שייתכן והוא חשוב, יגיע לתיקיית הספאם, שבה המשתמש לא יסתכל, ולכן יפספס את המייל החשוב.

(5) בעיית ה- hard-SVM: $\underset{(w,b)}{\operatorname{argmin}}\ \|w\|^2 \text{ s.t. } \forall i,\ y_i(\langle w,x_i\rangle+b)\geq 1$

צריך למצוא $Q\in\mathbb{R}^{n\times n}$, $A\in\mathbb{R}^{m\times n}$, $a\in\mathbb{R}^n$, $d\in\mathbb{R}^m$, כדי להמיר את הבעיה

ל- $\underset{v\in\mathbb{R}^n}{\operatorname{argmin}}\left\{\frac{1}{2}v^TQv+a^Tv \mid Av\leq d\right\}$

נבחר ראשית $Q = I_n$ ו- $a = v \cdot \frac{1}{2}$ ונקבל:

$$\frac{1}{2}v^TQv + a^Tv = \frac{1}{2}v^TI_nv + \frac{1}{2}v^Tv = v^Tv = ||v||^2$$

ואז נצטרך להמיר $y_i(\langle w,x_i\rangle + b) \geq 1$ ל- $Av \leq d$. אם כך:

$\langle w,x_i\rangle \geq \frac{1-by_i}{y_i}$ ולכן, עבור $A \in R^{m\times n}$ תהי A_i השורה ה-i של A, אז:

$$A_i = -x_i, 1 \quad s.t \quad A = -\begin{bmatrix} x_1^T, 1 \\ \vdots \\ x_m^T, 1 \end{bmatrix}$$

$$v = (w,b)$$

$$d \in R^m \quad s.t \quad d_i = -\frac{1}{y_i}$$

(6) יהי w ונבדוק של ϵ_i יהיה מינימלי, ואז הממוצע שלהם יהיה מינימלי.

כמו כן, אנו יודעים שפתרון חוקי צריך לקיים $\epsilon_i \geq 1-y_i\langle w,x_i\rangle$

$$\epsilon_i \geq \ell^{hinge}(y_i\langle w,x_i\rangle)$$

נראה שכדי ש- ϵ_i יהיה מינימלי צריך להתקיים שיוויון ממש.

נניח בשלילה כי יש פתרון $\epsilon_1,\dots,\epsilon_m$ כך שקיים $1\leq i \leq m$ כך ש-

$\epsilon_i > \ell^{hinge}(y_i\langle w,x_i\rangle)$ אז נסמן $\psi = \ell^{hinge}(y_i\langle w,x_i\rangle)$ ונשים לב שפתרון שמכיל את ψ במקום ϵ_i הוא עדיין חוקי, מאחר ש- $\psi \geq 0$ וכן $\psi \geq 1-y_i\langle w,x_i\rangle$.

אז בנינו פתרון עם $\epsilon_1,\dots,\epsilon_{i-1},\psi,\epsilon_{i+1},\dots,\epsilon_m$ שהוא קטן יותר מהפתרון המינימלי. לכן לכל $1\leq i$

$\varepsilon_i = \ell^{hinge}(y_i \langle w_i, x \rangle)$ ונקבל.