מבוא למערכות לומדות - תרגיל 3

רועי צרפי 2086481254

חלק תיאורטי:

1) נוכיח כי בעיית ה- Hard-SVM הבעיה הינה בעיית תכנון ריבועי, קרי, נמצא את Q, A ווקטורים a, d כך ש- נוכל להגדיר את הבעיה לצורה

$$\underset{v\in\mathbb{R}^n}{\operatorname{argmin}} \ \frac{1}{2} v^T Q v + a^T v \quad , \quad Av \le d$$

הוכחה: נעזר בריבוע, נשים לב

$$\|w\|^2 = w^T I w = \frac{2}{2} w^T I w + 0 =$$

$$= 2\,\frac{w^T I w}{2} + 0^T w$$

=> נוכל לקבל את הביטוי הנ"ל עם הצבה לבעיית מינימיזציה ע"י בחירת:

$$v = w, \ Q = 2I, \ a = 0_n$$

=> כעת נתבונן באי-השוויון ע"מ לקבל את יתר הפרמטרים הדרושים.

$$\forall i \quad y_i \cdot (\langle w, x_i \rangle + b) \ge \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

תהי x' זה המתקבל מ-x ע"י הכנסת הדיוק b לעוכה כפי שראינו בתרגול.

$$y_i(\langle x_i, w\rangle + b) = y_i(\langle x'_i, w\rangle) =$$

$$= y_i x_i'^T w \geq 1$$

כעת נרצה למצוא את A שיקיים את הנ"ל כך שכאשר נכפול את A ב-w נקבל וקטור עמודה שכל שורה ושורה בו תהיה $y_i x_i'^T w$ ווקטור b שיקיים את הנתון.

נגדיר את Z באופן הבא:

$$Z = \begin{pmatrix} \vdots \\ — \; y_i x'_i \; — \\ \vdots \end{pmatrix}$$

זה x' כך וכפלנו כל שורה ב-y_i (כפל וקטור בסקלר)

באופן זה נקבל כאשר נכפול w Zw יהיה:

$$Zw = \begin{pmatrix} \vdots \\ y_i x_i'^T w \\ \vdots \end{pmatrix}$$ כפי שרצינו.

בנוסף נגדיר את b, ונשים לב כי ע"פ הנתון

$$Zw \geq \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

אך אנו נדרשים לא"ש בכיוון ההפוך, לכן נכפול את האגפים ב-(-1) ונקבל:

$$-Zu \le \begin{pmatrix}-1\\ \vdots\\ -1\end{pmatrix}$$

ולכן נגדיר את $A=-Z$ ו- $b=\begin{pmatrix}-1\\ \vdots\\ -1\end{pmatrix}$
ונקבל את הנדרש.

$\blacksquare$

2) תהי בעיית אופטימיזציה של soft-SVM הבאה:

$$\underset{w,\xi_i}{\operatorname{argmin}} \ \frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_i \xi_i \quad s.t.$$

$$\forall_i \ y_i\langle w, x_i\rangle \geq 1-\xi_i \ \wedge \ \xi_i \geq 0$$

נראה כי בהינתן פונ' loss hinge כך ש-

$$\ell^{hinge}(a) = \max\{0, 1-a\}$$

נראה כי הבעיה שקולה לבעיה

$$\underset{w,\{\xi_i\}}{\operatorname{argmin}} \ \frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_i \ell^{hinge}(y_i\langle x_i, w\rangle)$$

פתרון: נתבונן במקרה בו הגדרנו את ξ_i בתחילה ונראה כי תחת הגדרה זו האילוצים מתקיימים והבעיות אכן שקולות.

$$\xi_i = \begin{cases} 0 & y_i(\langle x_i, w\rangle + b) > 1 \\ 1 - y_i(\langle x_i, w\rangle + b) & else \end{cases}$$

כעת נתבונן במקרה בו

$$\xi_i = \ell^{hinge}(y_i(\langle x_i, w\rangle + b))$$

נשים לב שאכן הצבה זו מגדירה את הבעיה השקולה, וכעת לפתרון כי בכל מקרה שבו הגדרנו את ערך ξ_i מתקיימות האילוצים מהבעיה המקורית, קרי היא שקולה.

(א) עבור $\xi_i = 0$, אזי מתקיים כי

$$y_i(\langle x_i, w\rangle + b) \ge 1 \Rightarrow \ge 1 - \xi_i = 1$$

קרי התנאי הראשון מתקיים.

כמו כן $\xi_i = 0$ ולכן $\xi_i \ge 0$, וקרי גם התנאי השני מתקיים ולכן במקרה זה נקבל שייך לכל ההגבלות.

(ב) $\xi_i = 1 - y_i(\langle x_i, w\rangle + b)$, נציב.

$$y_i(\langle x_i, w\rangle + b) \ge \cancel{1} - (\cancel{1} + y_i(\langle x_i, w\rangle + b))$$

קרי נקבל שוויון ולכן הא"ש מתקיים.

כמו כן מכיוון שלפי הנתון אזי $y_i(\langle x_i, w\rangle + b) \le 1$

ולכן אילו ועבור נקבל כי

$$0 \le 1 - y_i(\langle x_i, w\rangle + b) = \xi_i$$

קרי $0 \le \xi_i$ ולכן גם תנאי השני מתקיים.

$\Leftarrow$ לסיכום, הראנו כי בהינתן הגדרה זו התנאים מתקיימים באופן מוכלל ולכן הבעיות זהות כנדרש!

∎

ג. כמתואר בו נניח כי

$$y \sim \text{Multinomial}(\Pi)$$

$$x_j \mid y=k \overset{iid}{\sim} \mathcal{N}(\mu_{kj}, \sigma^2_{kj})$$

ד. נניח כי $x \in \mathbb{R}$, דהיינו קיים פיצ'ר יחיד, נבצע fit ל-Θ הנתון $\{(x_i, y_i)\}$ על מנת שנכתיר את כלל המקסימיזציה הנדונה.

פתרון: נרצה למצוא λ המקסם את הביטוי

$$\underset{\lambda \in \mathbb{N}}{\operatorname{argmax}} \frac{f_{X|y=\lambda}(x) f_y(\lambda)}{f_X(x)}$$

ניתן לראות כי $f_X(x)$ אינו רלוונטי לחישוב ה-argmax ולכן נתעלם ממנו ונקבל

$$\underset{\lambda \in \mathbb{N}}{\operatorname{argmax}} f_{X|y=\lambda}(x) f_y(\lambda)$$

נעזר ב-MLE ע"מ לקבל את הערך המקסימלי של ה״פ:

$$L(S \mid X, y) \overset{iid}{=} \prod_{i=1}^{m} f_{X|y=\lambda}(x) \cdot f_y(\lambda) =$$

$$= \prod_{i=1}^{m} \mathcal{N}(x \mid \mu_k, \sigma_k^2) \cdot \text{Mult}(\lambda \mid \Pi) =$$

$$= \prod_{i=1}^{m} \frac{1}{\Sigma_{y_i}\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\Sigma_{y_i}}\right)^2} \cdot \pi_{y_i} \quad / \log \quad =$$

נסמן פה
מכיוון ש- $j=1$ [illegible]!

$$= \log\left(\frac{1}{\Sigma_{y_i}\sqrt{2\pi}}\right) \sum_{i=1}^{m} -\frac{1}{2}\left(\frac{x_i-\mu_{y_i}}{\Sigma_{y_i}}\right)^2 \cdot \log(\pi_{y_i}) =$$

$$= \sum_{i=1}^{m} \log(\pi_k) - \frac{1}{2}\log(2\pi) - \log(\Sigma_k) - \frac{1}{2}\left(\frac{x_i-\mu_k}{\Sigma_k}\right)^2 =$$

$$= m\log(\pi_{y_i}) - \frac{m}{2}\log(2\pi) - m\log(\Sigma_{y_i}) - \frac{1}{2}\sum_i\left(\frac{x_i-\mu_{y_i}}{\Sigma_{y_i}}\right)^2$$

$$= m\left(\log(\pi_{y_i}) - \log(\Sigma_{y_i}) - \frac{1}{2}\log(2\pi)\right) - \frac{1}{2}\sum_i\left(\frac{x_i-\mu_{y_i}}{\Sigma_{y_i}}\right)^2$$

פיתוח והצבת n_k y_i

$$= -\frac{m}{2}\log(2\pi) + \sum_{k\in[K]} -\frac{1}{2}n_k\log(\Sigma_k) - \sum_{i=1, y_i=k}^{m}\frac{(x_i-\mu_k)^2}{2\Sigma_k} + n_k\log(\pi_k)$$

כעת פונק' ה-log-likelyhood שלנו שאותה נרצה
למקסם, נחזור על תהליך הגזירה שראינו בהרצאה
עד שנמצא את ערכי המקסימום של הפונק'.

$\Leftarrow$ ראינו כי נדרוש ש- $\sum_{k\in[K]}\pi_k = 1$

$\Leftarrow$ $g(\pi) = \sum_{k\in[K]}\pi_k - 1 = 0$ ״

$$L = \ell(\Theta \mid \{x_i, y_i\}_{i=1}^{n}) - \lambda g(\pi)$$ נגזור ⇐

ונמצא את הפרמטרים השונים!

1.) $$\frac{\partial L}{\partial \pi_k} \underset{\text{רק בסכימה מופיע } \pi_k}{=} n_k \cdot 1/\pi_k = \frac{n_k}{\pi_k}$$

$$\frac{\partial L}{\partial \pi_k} = \frac{n_k}{\pi_k} - \lambda = 0 \Rightarrow \pi_k = \frac{n_k}{\lambda}$$ ⇐

כעת נציב בביטוי כי $\sum_k \pi_k = 1$

$$\sum_{k \in \{K\}} \frac{n_k}{\lambda} = 1 \Rightarrow \frac{n}{\lambda} = 1 \Rightarrow \lambda = n$$

$$\pi_k^{MLE} = \frac{n_k}{n}$$

2.) $$\frac{\partial L}{\partial \mu_k} = \frac{\partial}{\partial \mu_k}\left(-\sum_{i=1, y_i=k}^{n} \frac{(x_i - \mu_k)^2}{2\Sigma_k}\right) =$$

$$= -\frac{1}{2\Sigma_k} \cdot \sum_{i=1, y_i=k}^{n} -2x_i + 2\mu_k =$$

$$= \frac{1}{\Sigma_k} \sum_{i=1, y_i=k}^{m} (x_i - \mu_k) = \frac{1}{\Sigma_k}\left(\left(\sum_{i=1, y_i=k}^{m} x_i\right) - n_k \mu_k\right)$$

נשווה ל-0 ונעביר אגף

$$\frac{1}{\Sigma_k}\left(\sum_{i=1, y_i=k}^{m} x_i - n_k \mu_k\right) = 0$$

$$\frac{1}{\cancel{\Sigma_k}} \sum_{i=1, y_i=k}^{m} x_i = \frac{n_k \mu_k}{\cancel{\Sigma_k}}$$

$$\Downarrow$$

$$\mu_k = \frac{1}{n_k} \sum_{i=1, y_i=k}^{m} x_i = \mu_k^{MLE}$$

3.) $$\frac{dL}{d\Sigma_k} = \frac{d}{d\Sigma_k}\left(-\frac{1}{2} n_k \cdot \log(\Sigma_k) - \sum_{i=1, y_i=k}^{m} \frac{(x_i - \mu_k)^2}{2\Sigma_k}\right)$$

$$= -\frac{n_k}{2\Sigma_k} + \sum_{i=1, y_i=k}^{m} \frac{(x_i - \mu_k)^2}{+2\Sigma_k^2} = 0$$

$$\Downarrow$$

$$\sum_{i=1, y_i=k}^{m} \frac{(x_i - \mu_k)^2}{2\Sigma_k^2} = \frac{n_k}{2\Sigma_k} \quad / \cdot 2\Sigma_k^2$$

$$\sum_{i=1, y_i=k}^{m} (x_i-\mu_k)^2 = n_k \Sigma_k$$

$$\Sigma_k^{MLE} = \frac{1}{n_k} \sum_{i=1, y_i=k}^{m} (x_i-\mu_k)^T (x_i-\mu_k) \Leftarrow$$

הביטוי מהריבוע שקול
סתם במקרה הרב-מימדי!

6) נכתוב פונקציה דומה לסעיף א' אך הפעם אלו פרמטרים
עם פיצ'ר בודד, קרי

$$X = x_{i,j} \qquad \begin{matrix} 1 \le i \le m \\ 1 \le j \le d \end{matrix}$$

$$L(\Theta | S) = \prod_{i=1}^{m} N(x_{i,j} | \mu_{y_i}, \Sigma_{y_i}) \cdot mult(y_i | \pi)$$

פיצ'רים ב"ת

$$= \prod_{i=1}^{m} \left(\prod_{j=1}^{d} N(x_{i,j} | \mu_{y_i}, \Sigma_{y_i}) \right) \pi_{y_i}$$

נפעיל log ונפתח $\Downarrow$

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{d} \log(N(x_{i,j} | \mu_{y_i}, \Sigma_{y_i}) + \log(\pi_{y_i}) \right) =$$

$$= -\frac{md}{2} \log(2\pi) + \sum_{k \in \{k\}} \sum_{j=1}^{d} \left(-\frac{1}{2} n_k \log(\sigma^2_{y_i,j}) - \sum_{i=1, y_i=k}^{m} \frac{(x_{i,j}-\mu_{k,j})^2}{\sigma^2_{k,j}} + \frac{n_k}{d} \cdot \log(\pi_k) \right)$$

פונקציה של קבועים בנפרד

נעבור עכ״ש לפתרון בעיית מקסימום:

נבצע שוב פיתוח בעזרת מכפלי לה-גרנז' באופן

זהה!

$$L = \ell(\Theta|S) - \lambda g(\pi) \quad , \quad g(\pi) = \sum \pi_k - 1$$

$$\frac{dL}{d\pi_k} = \frac{d}{d\pi_k}\left(\sum_{j=1}^{d} \frac{n_k}{d}\log(\pi_k)\right) - \lambda = \qquad (1)$$

$$= \cancel{\sum_{j=1}^{d}} \frac{n_k}{\cancel{d}} \cdot {}^{1}/_{\pi_k} - \lambda = 0$$

$$\Rightarrow \frac{n_k}{\pi_k} = \lambda \Rightarrow \pi_k = \frac{n_k}{\lambda}$$

ראינו ב-ex כי $\lambda = m$ לפי אילוץ $\sum_k \pi_k = 1$

$$\Leftarrow \quad \pi_k^{MLE} = \frac{n_k}{m}$$

הבהרה חשובה!

(2) נעבור לפי $\mu_{k,j}$

$$\frac{dL}{d\mu_{k,j}} \overset{?}{=} \frac{d}{d\mu_{k,j}}\left(-\sum_{i=1, y_i=k}^{m} \frac{(x_{i,j} - \mu_{k,j})^2}{2\sigma_{k,j}^2}\right) = \dots =$$

$$= +\frac{1}{\cancel{2}\sigma_{k,j}^2} \sum_{i=1, y_i=k}^{m} +\cancel{2}x_{i,j} \mp \cancel{2}\mu_{k,j} =$$

$$= \frac{1}{\sigma_{k,j}^2} \sum_{i=1, y_i=k}^{m} x_{i,j} - \frac{n_k \mu_{k,j}}{\sigma_{k,j}^2} = 0$$

$$\frac{1}{\sigma_{k,j}^2} \sum_{i=1, y_i=k}^{m} x_{i,j} = \frac{n_k \mu_{k,j}}{\sigma_{k,j}^2} \qquad \Leftarrow$$

$$\Rightarrow \quad \hat{\mu}_{k,j}^{MLE} = 1/n_k \sum_{i=1, y_i=k}^{m} x_{i,j}$$

[illegible]

$$\frac{dL}{d\sigma_{k,j}^2} = \frac{d\left(-\frac{1}{2} n_k \log(\sigma_{k,j}^2) - \sum \frac{(x_{i,j}-\mu_{k,j})^2}{2\sigma_{i,j}^2}\right)^{(3)}}{d\sigma_{k,j}^2} =$$

$$= -\frac{n_k}{2\sigma_{k,j}^2} + \sum_i \frac{1}{+2\sigma_{k,j}^4} \cdot (x_{i,j} - \mu_{k,j})^2 = 0$$

$$\frac{n_k}{2\sigma_{k,j}^2} = \frac{1}{2\sigma_{k,j}^4} \sum_i (x_{i,j} - \mu_{k,j})^2 \quad / \cdot 2\sigma_{k,j}^4$$

$$\sigma_{k,j}^2 n_k = \sum_i (x_{i,j} - \mu_{k,j})^2 \quad /: n_k$$

1 MLE

$$\sigma_{k,j}^2 = \frac{1}{n_k} \sum_i \left(x_{i,j} - \mu_{k,j}\right)^2$$

* זהו אומד מוטה ולכן עדיף לבצע וכפול ב-

כפי שלמדנו בהסתברות כדי לקבל אומד $\frac{1}{n_k - 1}$

לא - מוטה!

4) נרצה למצוא את השערה נקודתית אך שהפעם הרכיבים
הם התפלגות Pois, דהיי

$$x_j \mid y=k \overset{\text{iid}}{\sim} Pois(\lambda_{k,j})$$

א. נראה את חוק ה- likelyhood ונציג

$$L(\Theta \mid S) = f_{X \mid Y \mid \Theta}(S) \overset{\text{iid}}{=} \prod_{i=1}^{n} f_{X,Y \mid \Theta}(x_i, y_i) =$$

כלל ב.ס.

$$\overset{ב}{=} \prod_{i=1}^{n} \frac{f_{X \mid Y = y_i}(x) \cdot f_Y(y_i)}{f_X(x)} =$$

ניתן להתעלם כי הוא הקבוע

$$= \prod_{i=1}^{n} pois(x_i \mid \lambda_k) \cdot mult(y_i \mid \pi) =$$

j=1 2) ולכן נתייחס ל-λ כהגדרה.

$$= \prod_{i=1}^{n} \frac{\lambda_{y_i} e^{-\lambda y_i}}{x_i!} \cdot \pi_{y_i}$$

⇐ נפעיל log ונקבל!

$$\ell(\Theta \mid S) = \log\left(\prod_{i=1}^{n}\left(\frac{\lambda_{y_i}^{x_i} \cdot e^{-\lambda y_i}}{x_i!} \cdot \pi_{y_i}\right)\right) =$$

$$= \sum_{i=1}^{m} \left(\log\left(\frac{\lambda_{y_i}^{x_i} \cdot e^{-\lambda_{y_i}}}{x_i!} \right) + \log(\pi_{y_i}) \right) =$$

$$= \sum_{i=1}^{m} \left(\log(\lambda_{y_i}^{x_i}) - \log(x_i!) - \lambda_{y_i} + \log(\pi_{y_i}) \right)$$

$$= \sum_{i=1}^{m} \left(x_i \cdot \log(\lambda_{y_i}) - \log(x_i!) - \lambda_{y_i} + \log(\pi_{y_i}) \right)$$

$\Leftarrow$ נבצע מהלך זהה עד כדי פירוק הסכום לפי
ההתפלגות השונות:

$$\sum_{k} \left(\sum_{i, y_i = k} \left(x_i \log(\lambda_{y_i}) - \log(x_i!) \right) - n_k \lambda_{y_i} + n_k \log(\pi_k) \right)$$

נעזר כעת שנית בכופלי לגרנג':

כאשר נגדיר $g(\pi) = \sum_{k} \pi_k - 1$

ונגדיר $L = \ell(\theta \mid S) - \lambda' g(\pi)$

(ℓ)

$$\frac{dL}{d\pi_k} = \frac{d\left(\sum_{k} n_k \log(\pi_k) \right)}{d\pi_k} =$$

פה
ויקר
א.פר
היה
בסיכום

$$= n_k \cdot \frac{1}{\pi_k} - \lambda' = 0 \Rightarrow \pi_k = \frac{n_k}{\lambda'}$$

⇐ לפי האילוץ נקבל כי

$$\sum_k \pi_k = \sum_k \frac{n_k}{\lambda'} = n/\lambda' = 1 \Rightarrow \lambda' = n$$

$$\hat{\pi}_k^{MLE} = \frac{n_k}{n}$$

(2)

$$\frac{\partial L}{\partial \lambda_k} = \frac{\partial}{\partial \lambda_k}\left(\sum_i x_i \log(\lambda_k) - n_k \lambda_k \right) =$$

$$= \sum_i x_i \cdot 1/\lambda_k - n_k = 0$$

$$\hat{\lambda}_k^{MLE} = 1/n_k \sum_i x_i \quad ⇐$$

מצאנו את הערכים של האומדים לפי עקרון MLE

כנדרש!

6) כעת יש לנו d פוזיציות וצריך את הביטוי שיופיע
בו שינוי בקריטריון 3.

$$L(\Theta|S) \overset{?}{=} \prod_{i=1}^{n}\left(\prod_{j=1}^{d} Pois(x_{i,j}|\lambda_{y_i,j}) \cdot mult(\pi_{y_i})\right) =$$

$$\overset{\log}{=} \sum_{i=1}^{n}\left(\sum_{j=1}^{d}\log\left(Pois(x_{i,j}|\lambda_{y_i,j})\right) + \log(\pi_{y_i})\right) =$$

$$= \sum_{i=1}^{n}\left(\sum_{j=1}^{d}\log\left(\frac{\lambda_{y_i,j}^{x_{i,j}}\cdot e^{-\lambda_{y_i,j}}}{x_{i,j}!}\right) + \log(\pi_{y_i})\right) =$$

$$= \sum_{i}\left(\sum_{j}\left(x_{i,j}\cdot\log(\lambda_{y_i,j}) - \log(x_{i,j}!) - \lambda_{y_i,j}\right) + \log(\pi_{y_i})\right) =$$

$$\overset{\text{נעבור לפי k}}{=} \sum_{k\in[K]}\left(\sum_{j}\left(\sum_{i}\left(x_{i,j}\log(\lambda_k) - \log(x_{i,j}!)\right)\right) - n_k\lambda_{k,j} + n_k\log(\pi_k)\right)$$

⇐ ונעזר בנגזרת לפי MLE, לפי הגדרת $g(\pi)$ כפי שלמדנו

$$\hat{\pi}_k^{MLE} = \frac{n_k}{n}$$

(4) מתקבל באופן ישיר כי
(הביטוי כפונקציה של השונה ב-ee)

$$\frac{\partial L}{\partial\lambda_{k,j}} = \frac{d}{d\lambda_{k,j}}\left(\sum_{j}\left(\sum_{i} x_{i,j}\log(\lambda_{k,j}) - n_k\cdot\lambda_{k,j}\right)\right) =$$

$$= \sum_j \left(\left(\sum_i \frac{x_{i,j}}{\lambda_{k,j}} \right) - k_k \right) \rightarrow$$

רק λ תלוי
מתאפס
לכן ניתן
להעביר מן
הסכימה

$$= \sum_i \frac{x_{i,j}}{\lambda_{k,j}} - k_k = 0$$

$$\hat{\lambda}^{MLE}_{k,j} = \frac{1}{k_k} \sum_{i=1, y_i=k}^{n} x_{i,j}$$

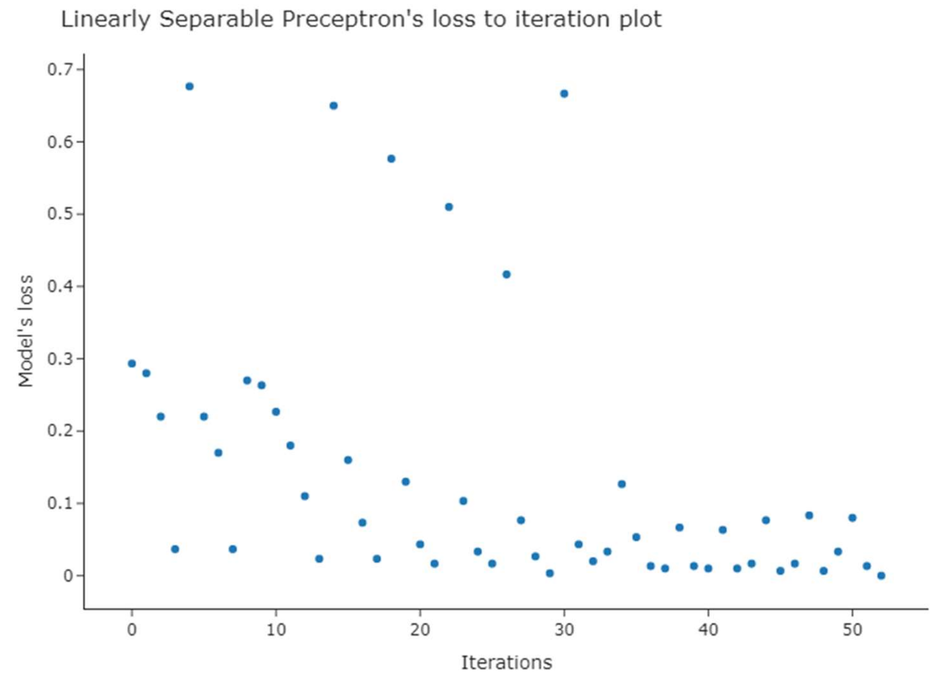
⇐ בכך מצאנו את הערכות הממקסמות את ה"ע ונודא
למצב כאשר נעשה Fit, כנדרש!

$\boxtimes$

מבוא למערכות לומדות 67577
רועי זהבי 208648154

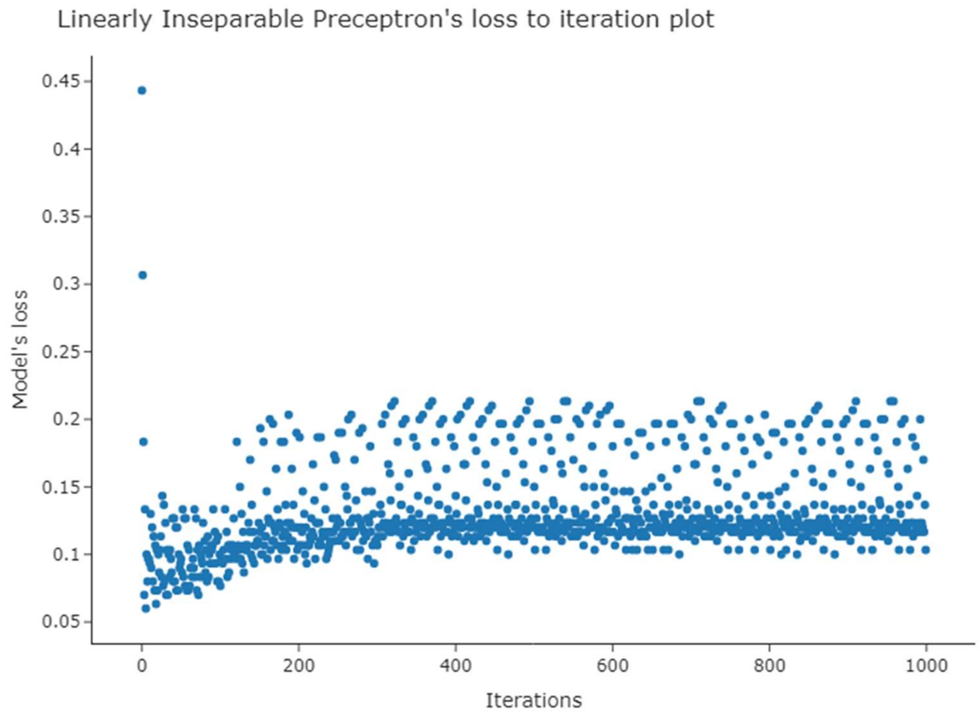
Perceptron Classifier

1. להלן גרף הפלט עבור סט האימון שמופרד ליניארית



ניתן ללמוד מפלט זה כי במקרה של דאטא המפורד ליניארית, אלגוריתם Perceptron מצליח למצוא את העל מישור המפריד, יחסית בקלות. קרי לאחר 52 איטרציות, האלגוריתם עוצר מכיוון שהוא הצליח למצוא על מישור מפריד בעל ERM אפס תחת סט אימון זה.

2. להלן גרף הפלט במקרה הלא מופרד.

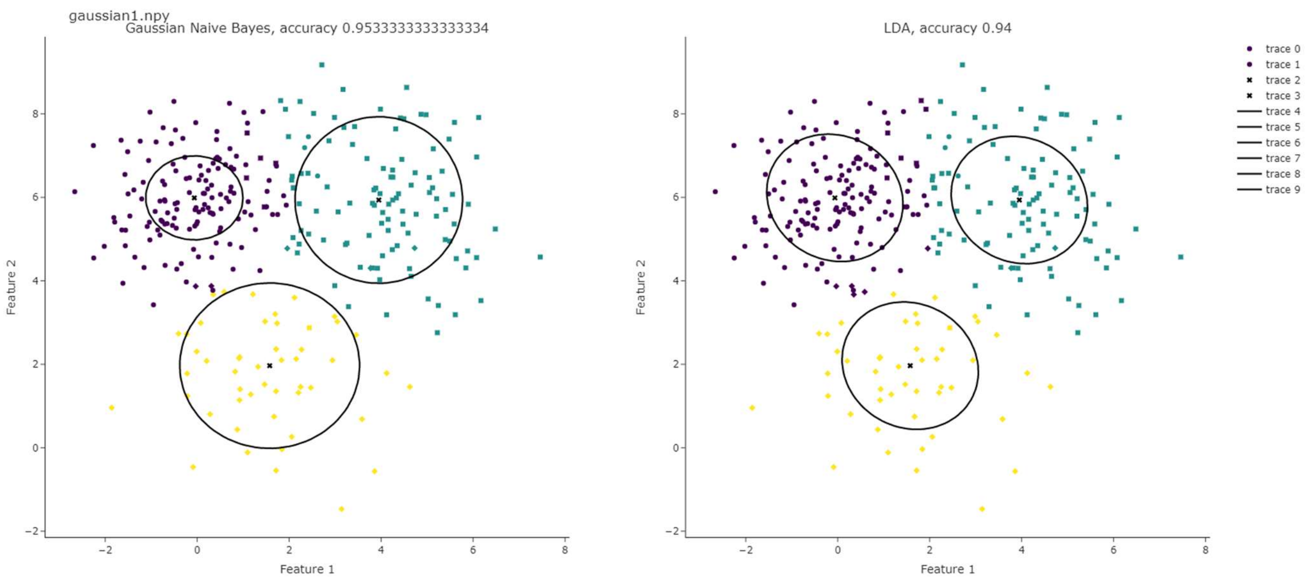


עבור מידע זה נוכל לראות כי ה-Perceptron לא מצליח לעצור. זאת מכיוון שהמידע לא מופרד, אז כל תיקון בעצם פוגע בתיקון הקודם – לכן הניתן לראות כי האלגוריתם נתקע בלולאה אינסופית עד שהוא מגיע לכמות האיטרציות שהגדרנו לו.

ניתן לראות כי השוני מהשאלה הראשונה נובע מכך שעבור מידע מופרד ליניארית, האלגוריתם פועל אך אחרת לא נוכל ללמוד מידע זה ע״י שימוש perceptron.

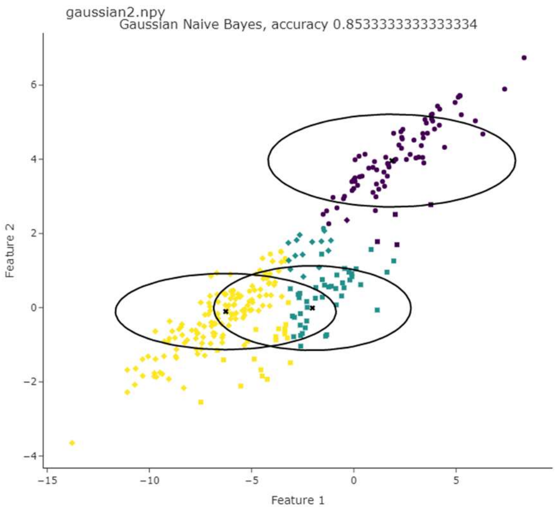
Bayes Classifiers

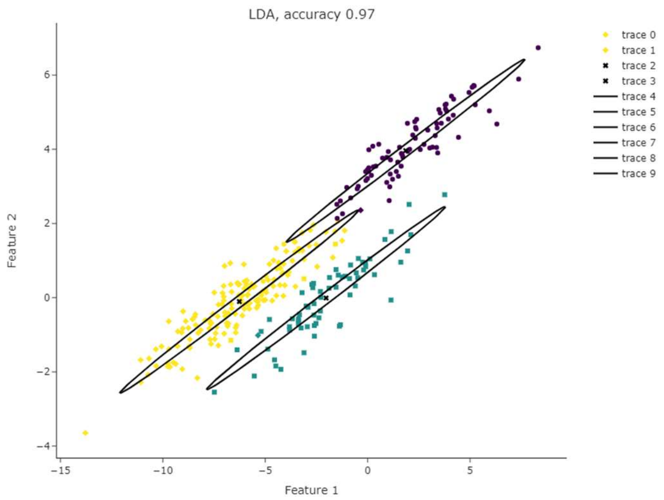
1. להלן הפלטים עבור סט הנתונים הראשון :



מניתוח הפלט ניתן לראות כי האלגוריתמים השונים הצליחו לחזות בצורה טובה דאטא שחולק ל-3 מחלקות שונות, כאשר אלגוריתם ה-$naive\ bayes$ מצליח לחזות בצורה טובה יותר, מכיוון שהוא מיצר מטריצת שונות לכל מחלקה, ובמקרה בו המידע מופרד באופן זה הנ״ל מאפשר לנות לחזות טוב יותר.
כמו כן, ניתן ללמוד כי ככל שהאליפסה שנוצרת ע״י מטריצת השונות מכסה יותר מן הדגימות, אזי האלגוריתם מצליח לחזות בצורה טובה יותר את הדאטא – הנ״ל נובע מן העבודה ככל הגאוסיאן עשינו לו fit מתחקה בצורה טובה יותר להתפלגות המקור ולכן המודלים שלנו המבוססים על התפלגות זו מתפקידים בצורה טובה.

2. להלן הפלטים עבור הסט המידע השני





ניתן לראות כי תחת סט האימון הנ״ל ה-$naive\ bayes$ מתקשה לחזות ביחס לאלגו׳ LDA, הנ״ל נובע מכך שניתן להבין כי ככה״נ המידע שקיבלנו תלוי ליניארית זאת בניגוד להנחת העבודה שאנו מאמצים בעת העבודה עם מודל $naive\ bayes$ (הצדקה לך ניתן לראות גם בצפיפות הדאטא). לכן, הנ״ל מתקשה לחזות את המידע ומצוא מטריצות שונות מתאימות לכל מחלקה (בניגוד למקרה הקודם, שההפרדה הקלה אפשרה בחירת שונות שהטיבה עם הפרדיקציה). זאת מנגד לפעולת ה-LDA שאינה מניחה דבר זה ועל כן היא מצליחה לחזות בצורה טובה את המידע.

טיעון זה ניתן לחיזוק ע״י שימוש במטריצות השונות, ה-LDA מחשב את השונות על כל הדאטא ללא קשר למחלקה, דבר שמתחקה באופן יותר טוב אחר הדאטא (כפי שאנו רואים בתרגול) בניגוד ל-$naive\ bayes$ אשר מייצר מטריצות שונות לכל מחלקה אשר אינם מכסות בצורה טובה את התפלגות הדאטא – ועל כן דיוקו ירד.