תרגיל 3

שמ: תומר שבי
ת.ז: 315315523

1) $\arg\min \|w\|^2 \quad \forall i \quad y_i(\langle w, x_i\rangle + b) \geq 1$

צריך למצוא $Q, A \in M_n(\mathbb{R}), a, d$! כך ש $\underset{w,b}{\arg\min} \|w\|^2 \text{ s.t } \forall i \ y_i(\langle w, x_i\rangle + b) \geq 0$
יכול להגדיר כ

$$\underset{v \in \mathbb{R}^n}{\arg\min} \tfrac{1}{2} v^t Q v + a^t v \quad \text{st} \quad Av \leq d$$

הוכחה: נשים לב כי $\quad v^t I w = \|w\|^2$
כמו

$$\forall i \quad y_i(\langle w, x_i\rangle + b) \geq 1 \iff y_i \langle w, x_i\rangle + b y_i \geq 1 \quad \backslash \cdot -1$$

$$-(y_i \langle w, x_i\rangle + b y_i) \leq -1$$

נגדיר $A \in M_{m \times (d+1)}(\mathbb{R})$, $v \in \mathbb{R}^{d+1}$, $d \in \mathbb{R}^{d+1}$

$$\forall i \quad \overset{\text{שורה } i}{[A]_i} = -\begin{bmatrix} y_i x_i \\ \hline y_i \end{bmatrix} \quad v = \begin{bmatrix} w \\ \hline b \end{bmatrix} \quad d = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix}$$

ולכן נקבל

$$\forall i \quad y_i(\langle w, x_i\rangle + b) \geq 1 \iff A_i^t v \leq 1 \iff Av \leq d$$

כעת נרצה למצוא a ! Q כך ש

$$\underset{b, w \in \mathbb{R}^n}{\arg\min} w^t I w = \underset{b, w \in \mathbb{R}^n}{\arg\min} \tfrac{1}{2} v^t Q v + a^t v$$

$$\underset{b, w \in \mathbb{R}^n}{\arg\min} w^t I w = \underset{b, w \in \mathbb{R}^n}{\arg\min} \tfrac{1}{2} [w^t | b] Q \begin{bmatrix} w \\ \hline b \end{bmatrix} + a^t \begin{bmatrix} w \\ \hline b \end{bmatrix}$$

נשים לב כי $a = 0$ ואז Q יצא נגדיר:

$$Q = \begin{bmatrix} \overbrace{\begin{matrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix}}^{d} & \\ & 0 \end{bmatrix}$$

$$\underset{b, w}{\arg\min} \|w\|^t = \underset{b, w \in \mathbb{R}^n}{\arg\min} \tfrac{1}{2} [w^t | b] Q \begin{bmatrix} w \\ \hline b \end{bmatrix} + 0^t \begin{bmatrix} w \\ \hline b \end{bmatrix} = \underset{v \in \mathbb{R}^{n+1}}{\arg\min} \tfrac{1}{2} v^t Q v + a^t v \quad \text{s.t} \quad Av \leq d$$

מצאנו d, a, Q, A לכן.

ב

$$\arg\min \frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_{i=1} \xi_i \quad s.t\ \forall i \quad y_i\langle w,x_i\rangle \geq 1-\xi_i \wedge \xi_i \geq 0$$

נסמן $\ell^{hinge}(a) = \max(0, 1-a)$

נשים שבבעיה המקורית בין s.t לו:

$$\forall i \quad y_i\langle w,x_i\rangle \geq 1-\xi_i \quad \wedge \xi_i \geq 0$$

$$\Longleftrightarrow \xi_i \geq 1-y_i\langle w,x_i\rangle$$

$$\ell^{hinge}(y_i\langle w,x_i\rangle) = \begin{cases} 1-y_i\langle w,x_i\rangle & 1-y_i\langle w,x_i\rangle > 0 \\ 0 & \text{אחרת} \end{cases}$$

נשים שהגדרנו $\xi_i \geq 0$ $\forall i$ נקבל כי

$$\xi_i = \begin{cases} 1-y_i\langle w,x_i\rangle & y_i\langle w,x_i\rangle < 1 \Longleftrightarrow 1-y_i\langle w,x_i\rangle > 0 \\ 0 & \text{אחרת} \end{cases}$$

קיבלנו שהגדרת ξ_i האפשרית הקטנה של ξ_i זהה בדיוק כמו הפונקציה: $\ell^{hinge}(y_i\langle w,x_i\rangle)$

ולכן הבעיה

$$\arg\min_{w,\{\xi_i\}} \frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_{i=1} \xi_i \quad s.t\ \forall i \quad y_i\langle w,x_i\rangle \geq 1-\xi_i \wedge \xi_i \geq 0$$

שקולה לבעיה

$$\arg\min_{w,\{\xi_i\}} \frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_{i=1} \ell^{hinge}(y_i\langle w,x_i\rangle)$$

כנדרש $\boxtimes$

3) ב) $X\in\mathbb{R}$ בנוסף $\{(X_i,y_i)\}_{i=1}^m$ בת״ל הוא יהיה

$$\mathcal{L}(\Theta|X,y)=f_{X,y|\Theta}(\{X_i,y_i\}_{i=1}^m)$$

$$\overset{iid}{=}\prod_{i=1}^m f_{X,y|\Theta}(X_i,y_i)$$

$$=\prod_{i=1}^m f_{X,y|y=y_i}(X_i)\cdot f_{y|\Theta}(y_i)$$

$$=\prod_{i=1}^m \mathcal{N}(x_i|\mu_{y_i},\sigma_{y_i}^2)\cdot Mult(y_i|\pi)$$

כעת נ' לוג לייקליהוד

$$\ell(\Theta|X,y)=\log\left(\prod_{i=1}^m \mathcal{N}(x_i|\mu_{y_i},\sigma_{y_i}^2)\cdot Mult(y_i|\pi)\right)$$

$$=\sum_{i=1}^n \log(\mathcal{N}(x_i|\mu_{y_i},\sigma_{y_i}^2))+\log(Mult(y_i|\pi))$$

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

$$=\sum_{i=1}^m \log\left(\frac{1}{\sqrt{2\pi\sigma_{y_i}^2}}\right)-\frac{(X_i-\mu_{y_i})^2}{2\sigma_{y_i}^2}+\log(\pi_{y_i})$$

$$=\sum_{i=1}^n \log\left((2\pi\sigma_{y_i}^2)^{-\frac{1}{2}}\right)-\frac{(X_i-\mu_{y_i})^2}{2\sigma_{y_i}^2}+\log(\pi_{y_i})$$

$$=\sum_{i=1}^m -\frac{1}{2}\log(2\pi)-\frac{1}{2}\log(\sigma_{y_i}^2)-\frac{(X_i-\mu_{y_i})^2}{2\sigma_{y_i}^2}+\log(\pi_{y_i})$$

$$=-\frac{m}{2}\log(2\pi)+\sum_k n_k\log(\pi_k)-\frac{1}{2}n_k\log(\sigma_k^2)-\frac{1}{2\sigma_k^2}\sum_{\{i:y_i=k\}}(X_i-\mu_k)^2$$

כאשר $n_k=\sum_{i=1}^n \mathbb{1}_{[y_i=k]}$

כאשר נזכור כי $\sum_k \pi_k=1 \iff \sum_k \pi_k-1=0$, כאשר נשתמש בכופלי לגרנז' עם $g(\pi)=\sum_k \pi_k-1$

לכן קיים λ המקיים את המשוואה הבאה

$$\mathcal{L}=\ell(\Theta|X,y)-\lambda g(\pi)$$

כעת נגזור לפי π_k ונקבל

$$\frac{\partial\mathcal{L}}{\partial\pi_k}=\frac{\partial}{\partial\pi_k}\ell(\Theta|X,y)-\lambda\frac{\partial}{\partial\pi_k}g(\pi)$$

$$=\frac{n_k}{\pi_k}-\lambda=0$$

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

ולכן

$$g(\pi)=\sum_k \pi_k-1=0$$

$$\Rightarrow\sum_k\frac{n_k}{\lambda}=1$$

$$\frac{1}{\lambda}\cdot m=1$$

$$m=\lambda$$

ולכן $\pi_k^{MLE}=\frac{n_k}{m}$ ואין לנו צורך במקסימום לפי (S) בעוד נרצה למצוא את μ_k שיביא לנו לוג-לייקליהוד מקסימלי

[illegible]

המשך

$$l(\Theta|x,y) = -\frac{m}{2}\log(2\pi) + \sum_k n_k \log(\pi_k) - \frac{1}{2} n_k \log(\sigma_k^2) - \frac{1}{2\sigma_k^2}\sum_{\{i:y_i=k\}}(x_i-\mu_k)^2$$

$$\frac{\partial l}{\partial \mu_k} = +\frac{1}{\cancel{2}\sigma_k^2}\sum_{\{i|y_i=k\}} -\cancel{2}(x_i-\mu_k) = 0$$

$$\Rightarrow \sum_{\{i|y_i=k\}} x_i - \mu_k = 0$$

$$\Rightarrow \underbrace{\frac{1}{n_k}\sum_{i|y_i=k} x_i}_{\text{נסמן ב } x'_k} = \mu_k^{MLE}$$

נמצא באופן דומה את σ_k

$$\frac{\partial l}{\partial \sigma_k^2} = -\frac{1}{2} n_k \cdot \frac{1}{\sigma_k^2} + \frac{1}{2(\sigma_k^2)^2}\sum_{i:y_i=k}(x_i-\mu_k)^2$$

נציב את x'_k

$$= -\frac{n_k}{2\sigma_k^2} + \frac{1}{2\sigma_k^4}\sum_{\{i|y_i=k\}}(x_i - x'_k)^2 = 0 \quad / \cdot 2\sigma_k^2$$

$$= \frac{1}{\sigma_k^2}\sum_{\{i|y_i=k\}}(x_i-x'_k)^2 = n_k$$

$$\Rightarrow \sigma_k^{2\,MLE} = \frac{1}{n_k}\sum_{\{i|y_i=k\}}(x_i-x'_k)^2$$

3) ב)

מש' בתרגיל הקודם i

$$\mathcal{L}(\Theta|x,y) = \prod_{i=1}^{m} \mathcal{N}(x_i|\mu_{y_i}, \sigma_{y_i}^2)\cdot Mult(y_i|\pi)$$

כי פיצ'רים ב"ת

$$= \prod_{i=1}^{m}\prod_{j=1}^{d}\mathcal{N}(x_{ij}|\mu_{jy_i}, \sigma_{y_ij}^2)\cdot Mul(y_i|\pi)$$

כמו מקודם נגזור

$$l(\Theta|x,y) = \sum_{i=1}^{m}\left(\log(Mul(y|\pi)) + \sum_{j=1}^{d}\log(\mathcal{N}(x_{ij}|\mu_{y_ij}, \sigma_{y_ij}^2))\right)$$

$$= \sum_k n_k \log(\pi_k) + \sum_k \sum_{j=1}^{d} \mathbb{1}_{[y_i=k]}\log(\mathcal{N}(x_{ij}|\mu_{kj}, \sigma_{kj}^2))$$

נשים לב שכל פיצ'ר הוא בעצם גם פיצ'ר x_j הוא נורמלי שאלו התוצאות לכל נורמל בנפרד

ו- לפי הנ"ל

$$\pi_k^{MLE} = \frac{n_k}{m} \qquad \mu_{kj}^{MLE} = \frac{1}{n_k}\sum_{\{i|y_i=k\}} x_{ij} \qquad \sigma_{ij}^{2\,MLE} = \frac{1}{n_k}\sum_{\{i|y_i=k\}}(x_{ij}-\mu_{kj})^2$$

4) א) כאן $\quad y \sim Mul(\pi) \qquad x_j | y = k \overset{ind}{\sim} Poi(\lambda_{kj})$

$$\mathcal{L}(\Theta | X, y) = f_{X,y|\Theta}(\{x_i, y_i\}_{i=1}^{m})$$

$$\overset{\text{ב״ת}}{=} \prod_{i=1}^{m} f_{x,y|\Theta}(x_i, y_i)$$

$$= \prod_{i=1}^{m} f_{x|y=y_i}(x_i) \cdot f_{y|\Theta}(y_i)$$

$$= \prod_{i=1}^{m} Poi(x_i | \lambda_{y_i}) \cdot Mul(y_i | \pi)$$

בדומה קודם נגזור את הלוג לייקליהוד

$$\ell(\Theta | X, y) = \sum_{i=1}^{m} \log\left(\frac{e^{-\lambda_{y_i}} \lambda_{y_i}^{x_i}}{x_i!}\right) + \log(\pi_{y_i})$$

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

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

$$= \sum_{k} \left(n_k \log(\pi_k) - n_k \lambda_k + \log(\lambda_k) \sum_{\{i|y_i=k\}} x_i\right) + \sum_{i=1}^{m} \log(x_i!)$$

נגזור לפי λ_k

$$\frac{\partial \ell}{\partial \lambda_k} = -n_k + \frac{1}{\lambda_k} \sum_{\{i|y_i=k\}} x_i = 0$$

$$\Rightarrow \lambda_k^{MLE} = \frac{1}{n_k} \sum_{\{i|y_i=k\}} x_i$$

ולגבי פרמטרי שתלוי ב π_k כדי לגזור דקדוק ונזכור ($\sum_k \pi_k = 1$ כשמשתמשים בכופלי לגרנג'
לאחר נקבל שוב ש

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

כנדרש ∎

ב) בדומה לסעיף 3.1, נניח שכל פיצ'ר בת"ל נקבל

$$\mathcal{L}(\Theta | X, y) = \prod_{i=1}^{m} Poi(x_i | \lambda_{y_i}) \cdot Mul(y_i | \pi)$$

$$= \prod_{i=1}^{m} \prod_{j=1}^{d} Poi(x_{ij} | \lambda_{y_i j}) \cdot Mul(y_i | \pi)$$

ולכן הלוג לייקליהוד הוא:

$$\ell(\Theta | X, y) = \sum_{i=1}^{m} \log(\pi_{y_i}) + \sum_{j=1}^{d} \log(Poi(x_{ij} | \lambda_{y_i j}))$$

$$= \sum_{k} n_k \log(\pi_k) + \sum_{\{i|y_i=k\}} \sum_{j=1}^{d} \log(Poi(x_{ij} | \lambda_{kj}))$$

ושוב כמו ב 3.2 נניח שכל פיצ'ר X_j בלתי תלוי ; נגזור באותו אופן ונקבל את ה λ_k שלנו
בסעיף א' ל j הקבוע

$$\lambda_{kj}^{MLE} = \frac{1}{n_k} \sum_{\{i|y_i=k\}} x_{ij}$$

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

ושוב הגענו לכל הנגזרות לכן

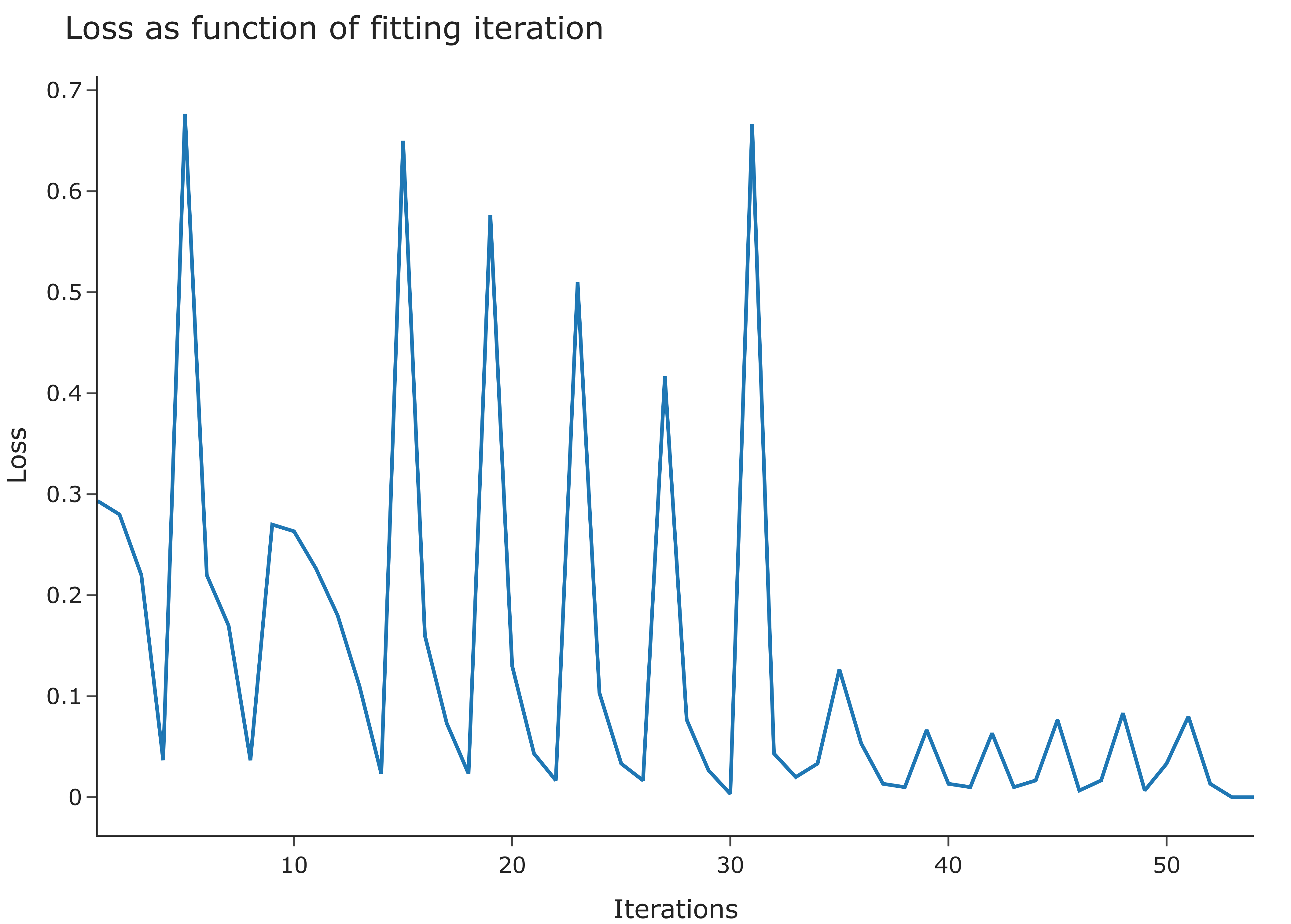
חלק ב

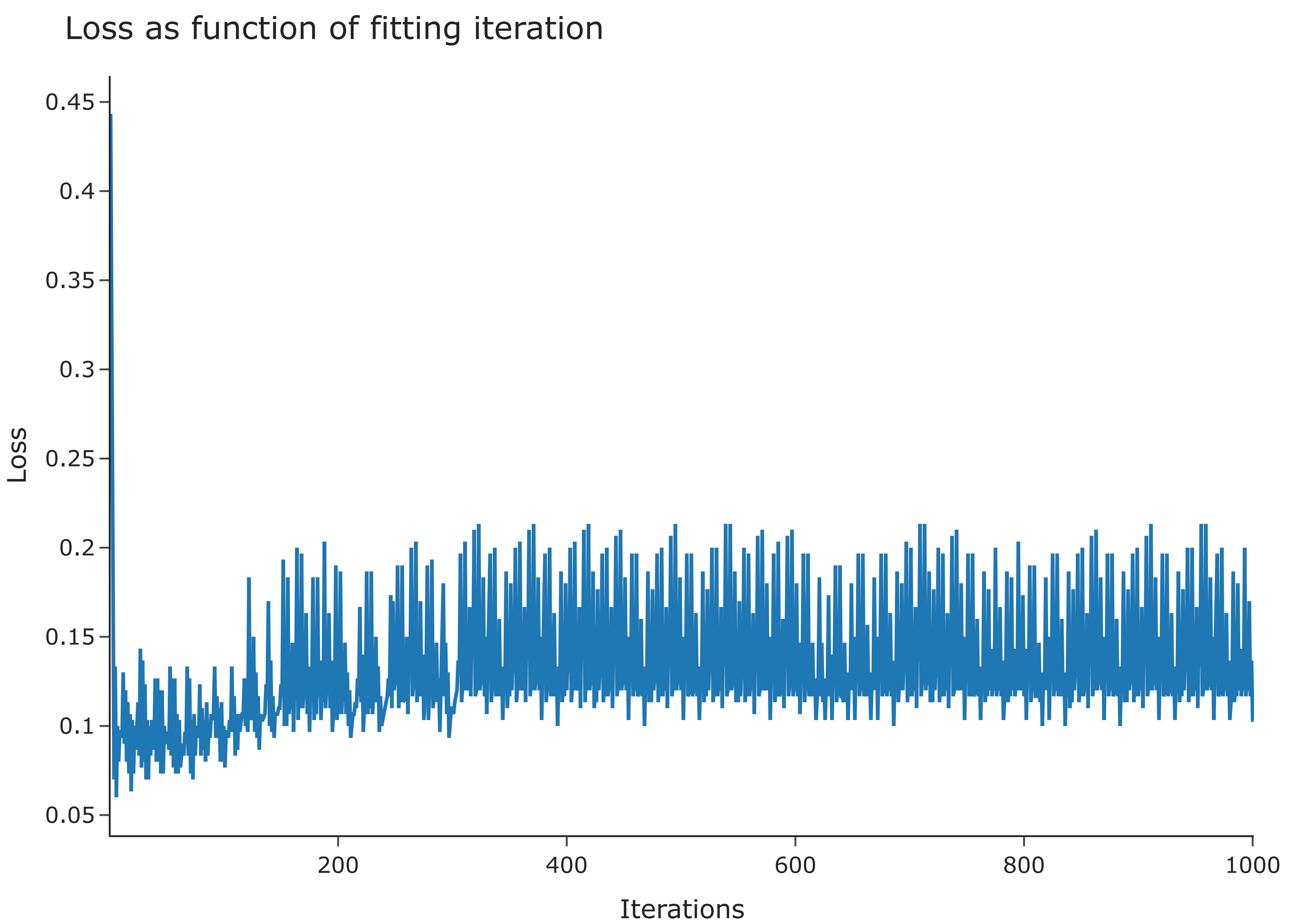
1) למדנו כי לאחר 54 איטרציות הצלחנו למצוא קו מפריד בהצלחה את הדאטה שלנו
כי השגיאה של הגרף עוברת ירידה קבועה עד לאפס הצלחנו

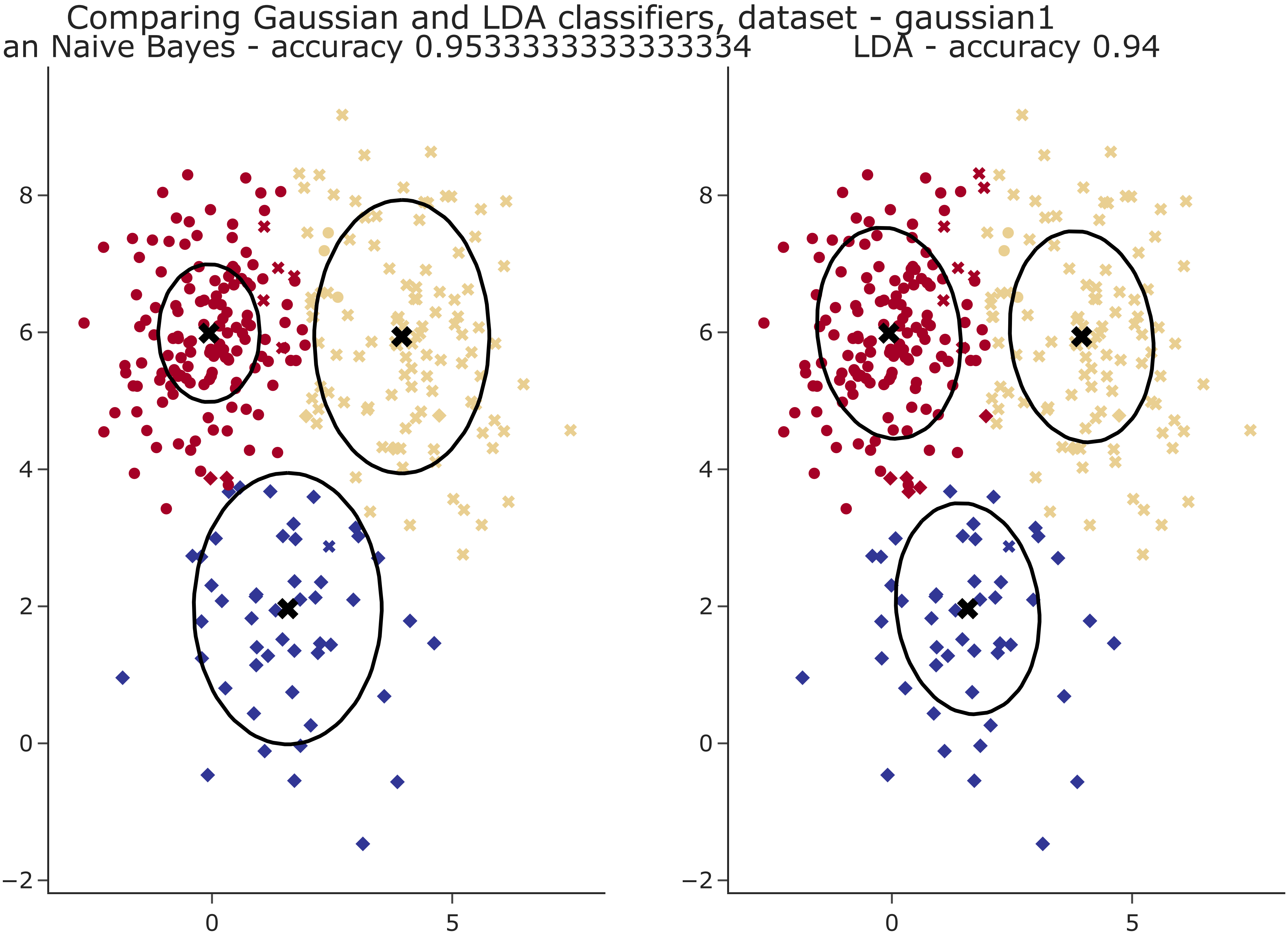
ב) נשים לב כי התוצאה אינה מתכנסת אלא קופצת הרבה לאורך כל ה 1000 איטרציות, אשר לא הצלחנו למצוא
קו מפריד בין הדאטה שבו כל נקודה בדאטה נכונה לכל אחד מהסימנים. בגרף הצבע יש קפיצות שגיאות גדולות
לאורך הקווים אבל הרבה יותר מאן. הסיבה לכך היא שבניגוד לגרף א הקו המפריד לפי הנקודה הקודמת בימין אבל
בגלל שהדאטה אינו ניתן להפרדה בכל פעם שביצענו גלגול, גרמנו לנקודה אחת לפחות לא נכונה ולכן לא הצלחנו להתכנס
לאחר כל האיטרציות

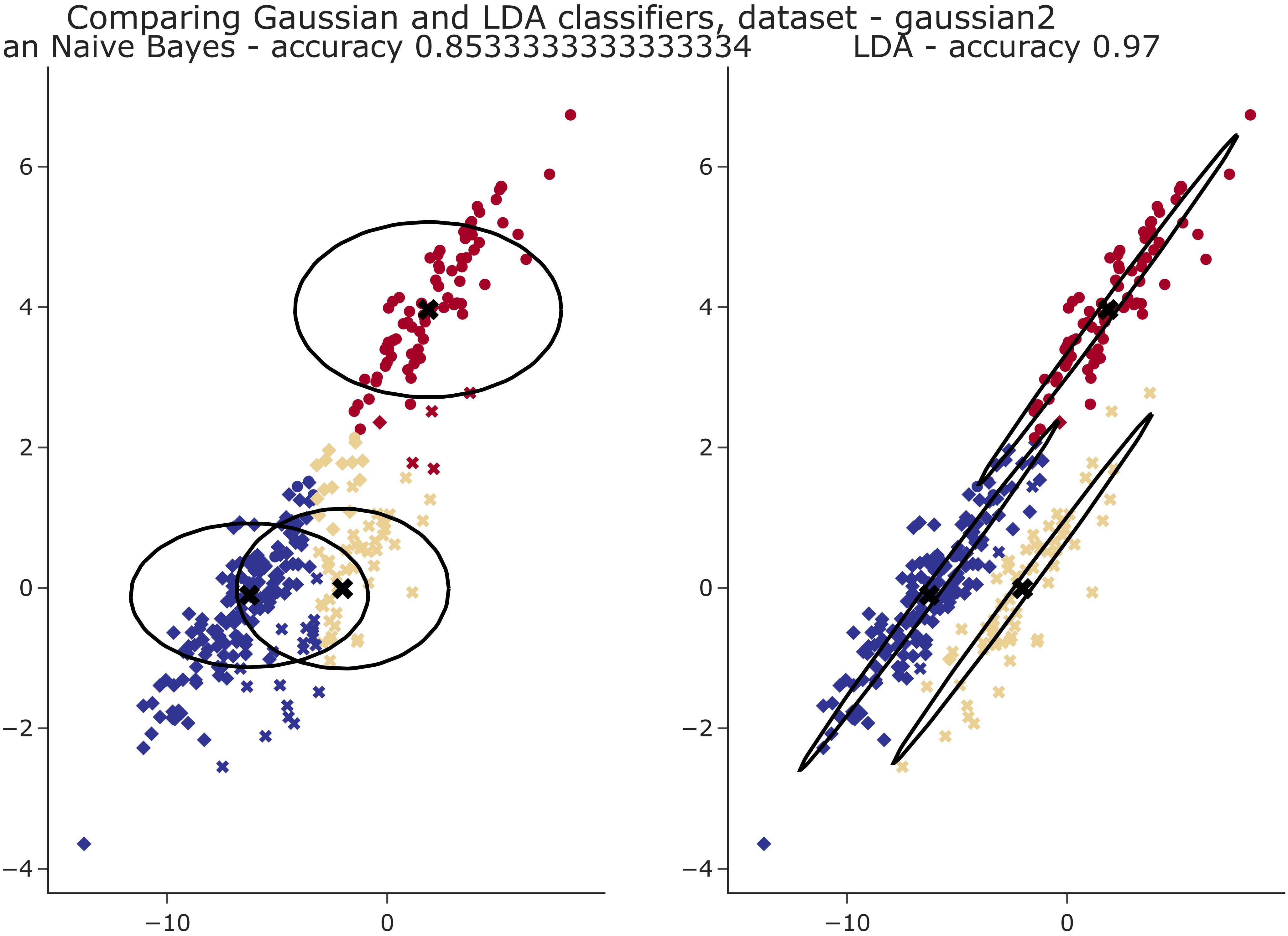
3) א) ניתן לראות כי 2 המודלים מצליחים להשיג תוצאות דומות. ניתן לראות מהצורה של
העיגולים סביב ה "X" כי השונות בין 2 הפיצ'רים היא נמוכה מאוד ולכן 2 המודלים משיגים תוצאה טובה.
שכן ה cov נמוך ולא GNB מניח אי תלות בין הפיצ'רים

ב) עבור הדאטה של הנתונים, השונות בין 2 הפיצ'רים מצליחה להתפס ע"י מטריצת ה cov של LDA
ולכן המודל מצליח למצוא את הסיווג יותר טוב מאשר ה Gaussian Naive Bayes שמניח שהשונות בין 2 הפיצ'רים
בכל אחד מהקלאסים היא 0 שכן האלגוריתם מניח אי תלות בין הפיצ'רים


Loss as function of fitting iteration
Loss
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10
20
30
40
50
Iterations


Loss as function of fitting iteration
Loss
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
200
400
600
800
1000
Iterations


Comparing Gaussian and LDA classifiers, dataset - gaussian1
an Naive Bayes - accuracy 0.9533333333333334
LDA - accuracy 0.94
8
6
4
2
0
−2
0
5
8
6
4
2
0
−2
0
5



Comparing Gaussian and LDA classifiers, dataset - gaussian2
an Naive Bayes - accuracy 0.8533333333333334
LDA - accuracy 0.97
6
4
2
0
−2
−4
−10
0
6
4
2
0
−2
−4
−10
0