IML (67329) | תרגיל 3

שם: רונאל חרדים | ת"ז:208917641

חלק I

$Theoretical$:

שאלה 1

נוכיח כי בעיית האופטימיזציה הבאה שקולה:

$$\underset{(\mathbf{w},b)}{argmin}\|\mathbf{w}\|^2 \quad \text{s.t.} \quad \forall i y_i\left(\langle\mathbf{w},\mathbf{x}_i\rangle+b\right)\geq 1 = \underset{\mathbf{v}\in\mathbb{R}^n}{argmin}\frac{1}{2}\mathbf{v}^\top Q\mathbf{v}+\mathbf{a}^\top\mathbf{v} \quad \text{s.t.} \quad A\mathbf{v}\leq\mathbf{d}$$

הוכחה:

$$\underset{(\mathbf{w},b)}{argmin}\|\mathbf{w}\|^2 = \underset{(\mathbf{w},b)}{argmin}\left([w,b]^T I\,[w,b]\right) s.t \underbrace{\begin{bmatrix} (y_1,x_1) & . & y_1 \\ . & . & \\ . & . & \\ . & & . \\ (y_m,x_m) & & y_m \end{bmatrix}}_{A} \cdot \begin{bmatrix} w \\ b \end{bmatrix} \geq \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{b} =$$

$$\underset{(\mathbf{w},b)}{argmin}\ 0.5\cdot[w,b]^T\,2I\,[w,b]+O^T\,[w,b]\ \ s.t \begin{bmatrix} (y_1,x_1) & . & y_1 \\ . & . & \\ . & . & \\ . & & . \\ (y_m,x_m) & & y_m \end{bmatrix} \cdot \begin{bmatrix} w \\ b \end{bmatrix} \leq \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

נשים לב כי עבור $[w,b]=V, 2I=Q, O^T=a$ נקבל את הדרוש.

שאלה 2

נוכיח את השקילות הנדרשת:

$$\underset{\mathbf{w},\{\xi_i\}}{argmin}\frac{\lambda}{2}\|\mathbf{w}\|^2+\frac{1}{m}\sum_i\xi_i \quad \text{s.t.} \quad \forall i y_i\langle\mathbf{w},\mathbf{x}_i\rangle\geq 1-\xi_i\wedge\xi_i\geq 0 = \underset{\mathbf{w},\{\xi_i\}}{argmin}\frac{\lambda}{2}\|\mathbf{w}\|^2+\frac{1}{m}\sum_i\ell^{\text{hinge}}\ \left(y_i\langle\mathbf{w},\mathbf{x}_i\rangle\right)$$

נשים לב כי אנו מנסים למצוא את הערך המינימלי שמתקבל לכל ξ_i ולכל w אופטימלי תחת התנאים הנ"ל. ומתקיים:

$$y_i \langle \mathbf{w}, \mathbf{x}_i \rangle \geq 1 \Rightarrow \xi_i = 0$$

אחרת

$$\xi_i = 1 - y_i \langle \mathbf{w}, \mathbf{x}_i \rangle$$

במטרה למנם את ξ_i מאחר ולכל ξ_i מתקיים $\xi_i = l_{hinge}(y_i \langle \mathbf{w}, \mathbf{x}_i \rangle)$ ולכן מתקיים עבור m ערכים:

$$\underset{\mathbf{w},\{\xi_i\}}{argmin} \frac{\lambda}{2}\|\mathbf{w}\|^2 + \frac{1}{m}\sum_i \ell^{\text{hinge}} \left(y_i \langle \mathbf{w}, \mathbf{x}_i \rangle\right)$$

שאלה 3

(א)

נחשב עבור ווקטור עם פיצ'ר בודד:

$$\hat{y}^{MAP} := \underset{k\in[K]}{argmax} f_{Y|X=\mathbf{x}}(k) = \underset{k\in[K]}{argmax} \frac{f_{X|Y=k}(\mathbf{x}) f_Y(k)}{f_X(\mathbf{x})} =$$

$$\underset{k\in[K]}{argmax} \frac{P\left(X=x \mid Y=k\right)\cdot P\left(Y=k\right)}{P\left(X=x\right)} = \underset{k\in[K]}{argmax} P\left(X=x \mid Y=k\right)\cdot P\left(Y=k\right) =$$

$$\underset{k\in[K]}{argmax} \left(\frac{1}{\sigma_k\sqrt{2\pi}}\exp\left(-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}\right)\right)\cdot \pi_k$$

עבור סט אימון $S = \{x_i, y_i\}_{i=0}^m$

$$L(\theta \mid x, y) = f_{x,y|\theta}\left(\{x_i, y_i\}_{i=1}^m\right) \overset{iid}{=} \prod_{i=1}^m f_{x,y|\theta}\left(x_i, y_i\right) = \prod_{i=1}^m f_{x|y=y_i}\left(x_i\right)\cdot f_y\left(y_i\right) =$$

$$= \prod_{i=1}^m N\left(\mu_k, \sigma_k^2\right)\cdot \pi_{y_i}$$

ניתן לחשב $logliklehood$ באופן שקול:

$$\ell(\theta \mid x, y) = \log\left(\prod_{i=1}^m N\left(\mu_k, \sigma_k\right)\cdot \pi_{y_i}\right) = \sum_{i=1}^m \log\left(N\left(\mu_k, \sigma_k^2\right)\cdot \pi_k\right)$$

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

נגדיר משתנה אינדיקטור

$$= l(\theta \mid x, y) = \sum_k n_k \cdot \log\left(\pi_k\right) - \frac{1}{2} n_k \log\left(\sigma_k^2\right) - \frac{1}{2\sigma_k^2} \cdot \sum_{i \mid y_i = k} \left(x_i - \mu_k\right)^2$$

נגזור:

$$\frac{\partial l\left(\theta \mid x_1 y\right)}{\partial \mu_k} = 2\left(\sum_k n_k \cdot \log\left(\pi_k\right) - \frac{1}{2} n_k \log\left(\sigma_k^2\right) - \frac{1}{2\sigma_k^2} \cdot \sum_{i \mid y_i = k} \left(x_i - \mu_k\right)^2\right) 2\partial \mu_k$$

$$= 2 \sum_{i \mid y_i = k} \left(x_i - \mu_k\right)^2 / 2\mu_k = \sum_{i \mid y_i = k} 2\left(x_i - \mu_k\right) \cdot (-1) = 0$$

$$\Rightarrow -n_k \cdot \mu_k + \sum_{i : y_i = k} x_i = 0$$

וקיבלנו כי:

$$\hat{\mu}_k^{\mu LE} = \frac{1}{n_k} \sum_{i \mid y_i = k} x_i$$

נמשיך:

$$\frac{\partial l\left(\theta \mid X_1 y\right)}{\partial \sigma_k^2} = \partial\left(\sum_k n_k \cdot \log\left(\pi_k\right) - \frac{1}{2} n_k \log\left(\sigma_k^2\right) - \frac{1}{2\sigma_k^2} \cdot \left[\sum_{i \mid y_i = k} \left(x_i - \mu_k\right)^2\right) / \partial \sigma_k^2$$

$$= -n_k \cdot \frac{1}{\sigma_k} + \frac{1}{\sigma_n^3} \sum_{i \mid y_1 = k} \left(x_i - \mu_k\right)^2 = 0$$

קיבלנו כי:

$$\Sigma_k^2 = \frac{1}{n_k} \sum_{i \mid y_1 = k} \left(x_i - \mu_k\right)^2$$

וגם $\pi_k = \frac{n_k}{m}$

(ב)

נחשב עבור ווקטור עם d פיצ'רים:

פונקציית $liklehood$ תיראה כך:

$$l(\theta \mid x, y) = \sum_k n_k \cdot \log\left(\pi_k\right) + n_k \cdot \log\left(\frac{1}{\left|\Sigma_k\right|}\right) - \frac{1}{2} \cdot \sum_{i \mid y_i = k} \left(x_i - \mu_k\right) \sum_k^{-1} \left(x_i - \mu_k\right)^\top$$

נגזור:

$$\frac{\partial l(\theta \mid x, y)}{\partial \mu_k} = \sum_{il_{y:}=h} (x_i - \mu_k) \cdot \sum_k^{-1} = 0$$

ונקבל כי:

$$\hat{\mu}_k^{\mu LE} = \frac{1}{n_k} \sum_{i|y_i=k} x_i$$

נשים לב כי:

$$\frac{2 - n_k \log\left(|\Sigma_k|\right)}{2\Sigma_n^{-1}} = -\frac{1}{2} n_k \Sigma_k$$

לכן:

$$\frac{\partial l\left(\theta \mid x_1 y\right)}{\partial \sum_k^{-1}} = -\frac{1}{2} n_k \Sigma_k - \frac{1}{2} \sum_{iI_{y_i=h}} \left(x_i - \mu_k\right)\left(x_i - \mu_k\right)^{\top}$$

ונקבל כי:

$$\Sigma_k^{MLE} = \frac{1}{n_k} \sum_{i|y_1=k} \left(x_i - \mu_k\right)\left(x_i - \mu_k\right)^T$$

שאלה 4

(א)

נחשב עבור ווקטור עם פיצ'ר בודד:

$$\hat{y}_i^{MAP} := \underset{k \in [K]}{argmax} f_{Y|X=\mathbf{x}}(k) = \underset{k \in [K]}{argmax} \frac{f_{X|Y=k}(\mathbf{x}) f_Y(k)}{f_X(\mathbf{x})} =$$

$$\underset{k \in [K]}{argmax} \frac{P\left(X = x \mid Y = k\right) \cdot P\left(Y = k\right)}{P\left(X = x\right)} = \underset{k \in [K]}{argmax} \prod_{i=0}^{d} \frac{\left(\frac{e^{-\lambda_{kj}} \lambda_{kj}^{x_i}}{x_i!}\right) \cdot \pi_k}{P\left(X = x\right)} =$$

$$\underset{k \in [K]}{argmax} \prod_{i=0}^{d} \left(\frac{e^{-\lambda_{kj}} \lambda_{kj}^{x_i}}{x_i!}\right) \cdot \pi_k$$

נוציא log ונקבל:

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

$$= \sum_{k} -\lambda_k \cdot n_k + \left[\sum_{i|y_i=k} x_i \cdot \log\left(\partial_k\right) - \log\left(x_i\ |\right) \right] + n_k \cdot \log\left(\pi_k\right)$$

נגזור לפי λ_k ונקבל:

$$\frac{\partial l(\theta \mid x, y)}{\partial \lambda_k} = -n_k + \sum_{i|y,-k} x_i \left(\frac{1}{\lambda_k}\right) = 0 \Rightarrow \lambda_k^{MLE} = \frac{1}{n_k} \sum_{i|y_i=k} x_i$$

(ב)

באופן דומה לסעיף קודם ומכך שהתפלגות הדגימות פואסונית כך ש $\lambda_k = \sum_{j=1}^{n} \lambda_{kj}$ נובע כי $\lambda_k^{MLE} = \frac{1}{n_k} \sum_{i|y_i=k} x_i$

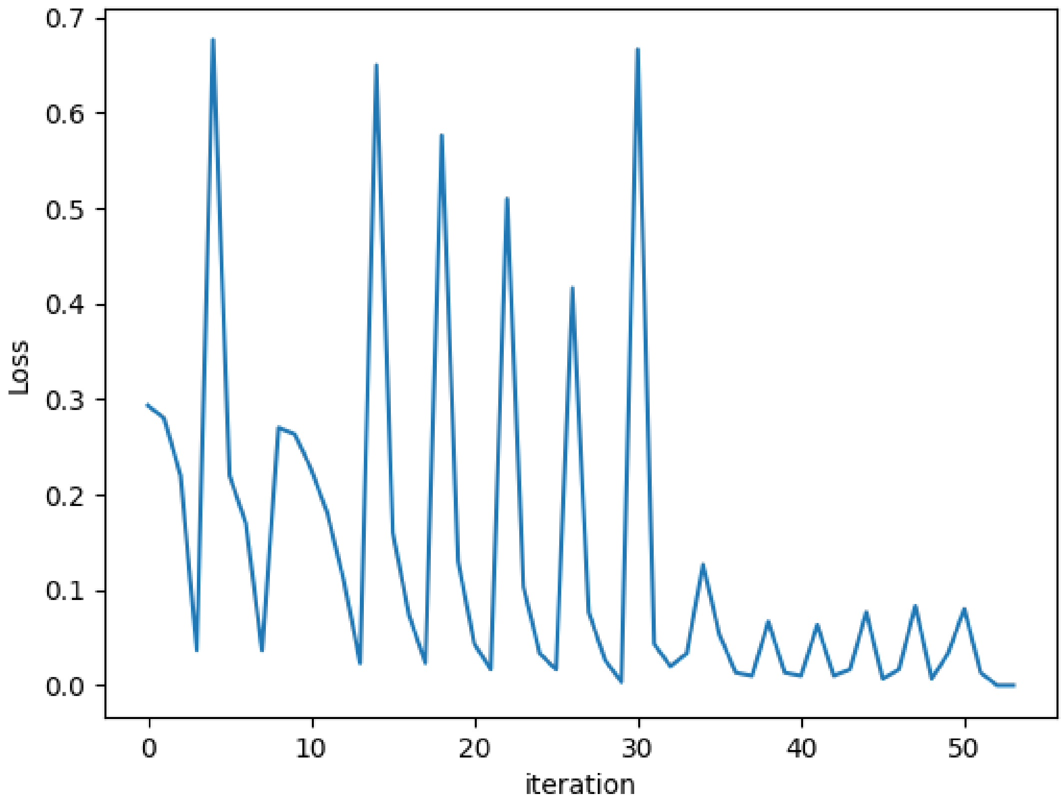
חלק II

:$Practical$

1 :$Perceptron\ Classifier$

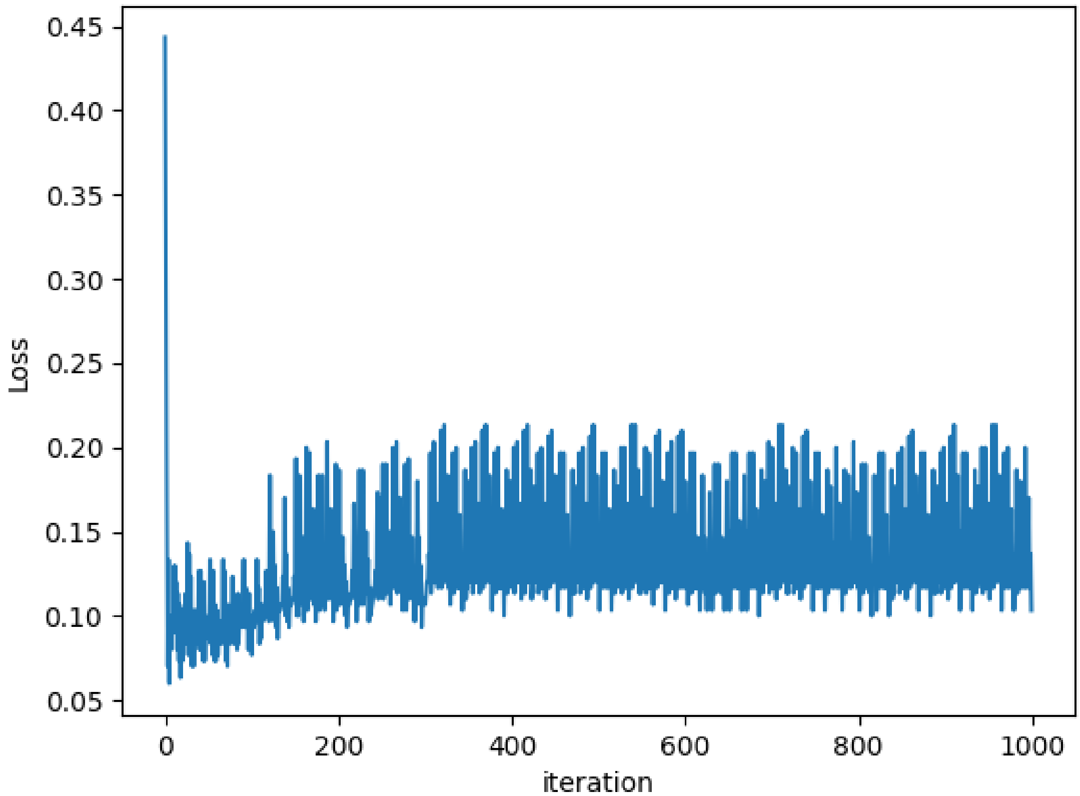
1.1 שאלה 1:

הגרף של ה $Loss$ הוא:



1.2 שאלה 2:

ההבדלים בין הגרף הזה לגרף הקודם הוא שהדאטה לא מופרד למחלקות בצורה ברורה ולכן לא ניתן לקבוע w שיפריד בצורה טובה את הדאטה לשתי מחלקות ולכן ה $loss$ לא יורד בכל איטרציה (הנחת הריליזביליות לא מתקיימת).

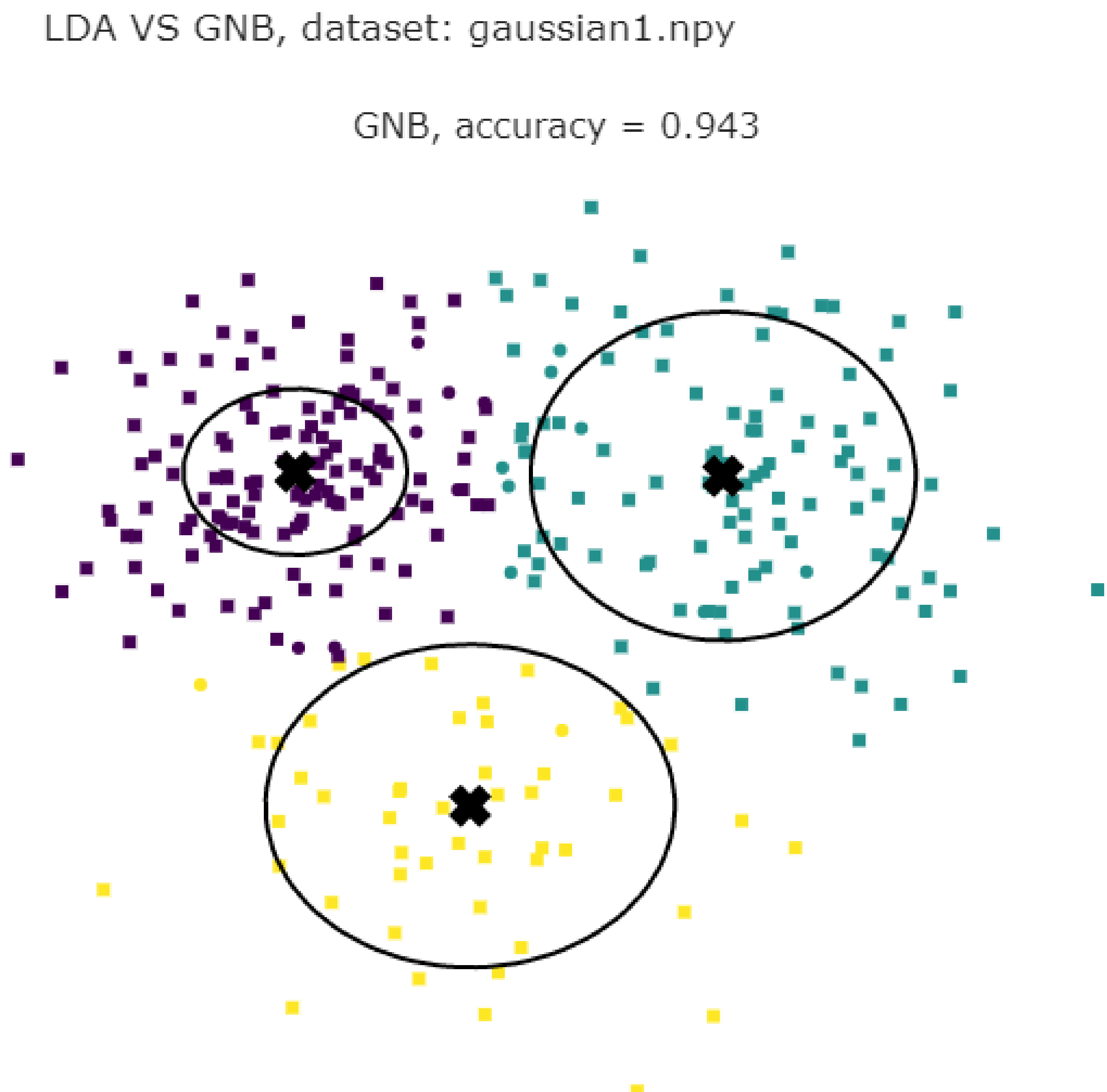


Loss
iteration
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
200
400
600
800
1000

2 $GNB\&LDA$:

2.1 שאלה 1:

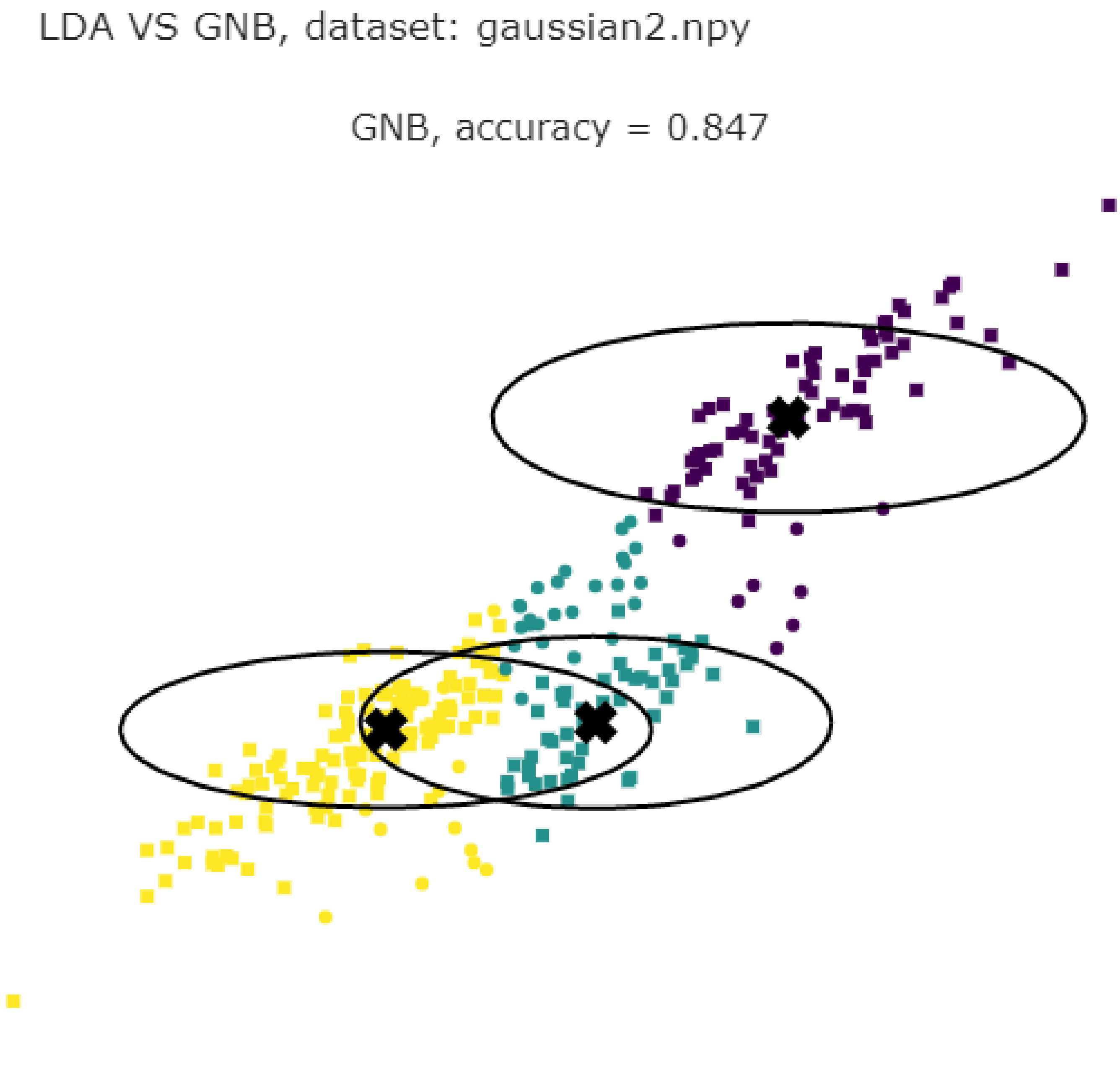
הגרפים שקיבלנו הם:



ניתן לראות כי LDA מסווג את הדאטה טוב יותר, אך בגדול התתפלגויות נראות שוות.

2.2 שאלה 2:

הגרפים שקיבנו הם:



ההבדל בין שתי התרחישים הוא פיזור הדאטה. בתרחיש הראשון הדאטה היה יחסית מופרד, בעוד במקרה זה הדאטה צמוד

יותר.

ניתן לראות כי כל מסווג הניח ההתפלגות שונה של הדאטה.

המסווג הטוב יותר הוא LDA ניתן לראות זאת לפי ערך הפונקציה $accuracy$ שגדול יותר מערך הפונקציה של GNB.