IML (67577) | תרגיל 3

שם: יואב שפירא| ת`ז: 312492838

28 באפריל 2022

חלק תאורטי

1. תחילה נשים לב ש

$$argMin_w ||w||^2 = argMin_w \left(\frac{1}{2}||w||^2\right)$$

ולכן אתייחס מעתה לבעיה עם הכפילה בחצי. לבעיות מינימיזציה עם אילוצים ניתן להגדיר את הלגראנז'יאן ולפתור בעזרתו. במגרה זה האילוץ הוא א"ש, ולכן נצטרך להתייחס אליו מאוחר יותר כדי לדאוג שהוא נשמר. נגדיר:

$$L(w,b,\overline{\lambda}) = \frac{1}{2}||w||^2 + \sum_i \lambda_i \left(y_i w_i x_i + y_i b - 1\right)$$

כאשר $\overline{\lambda} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$ כופלי לגראנז'. כעת הבעיה המקורית שקולה לבעיה של חישוב $argMin_{w,b,a}\left(L\left(w,b,\overline{\lambda}\right)\right)$.
נפרש את הביטוי:

$$\begin{aligned} L(w,b,\overline{\lambda}) &= \frac{1}{2}||w||^2 + \sum_i \lambda_i \left(y_i w_i x_i + y_i b - 1\right) \\ &= \frac{1}{2}w^T w + \sum_i \lambda_i y_i w_i x_i + \sum_i \lambda_i y_i b - \sum_i \lambda_i \end{aligned}$$

נבחין כי w האופטימלי לבעיה זו מקיים:

$$argMin_w\left(L\left(w,b,\overline{\lambda}\right)\right) = argMin_w\left(\frac{1}{2}w^T w + \sum_i \lambda_i y_i w_i x_i\right)$$

כלומר החלקים שאינם תלויים בw לא משנים את הפיתרון. לכן נוכל להגדיר:

$$v = w \quad , Q = I, \quad a = \sum \lambda_i y_i x_i \in \mathbb{R}^n$$

ונקבל שהבעיה שאנחנו רוצים לפתור היא:

$$argMin_v \frac{1}{2} v^T Q v + a^T v$$

נשים לב לאילוצים: האילוץ המקורי הוא (בפישוט):

$$\begin{aligned} &\forall_i: \ y_i w_i x_i + y_i b \geq 1 \\ \Longleftrightarrow & -y_i x_i w_i \leq 1 - y_i b \\ \Longleftrightarrow & \left(-X^T \overline{y}\right) w \leq 1 - \overline{y} b \end{aligned}$$

ועל כן נגדיר:

$$A = -X^T \overline{y}, \quad d = 1 - \overline{y} b$$

ונקבל שהאילוץ החדש הוא אכן $Av \leq d$. חשוב להדגיש, שבגלל צורת הפיתרון שנובעת מהלגראנז'יאן, יש לנו גם אילוץ על ה$\overline{\lambda}$: הוא חייב להיות אי שלילי. כלומר לכל i צ"ל $\lambda_i \geq 0$. זוהי עדיין QP: בעיית אופטימיזציה ריבועית תחת אילוצים ליניאריים.

2. בעצם צ"ל ש:

$$\xi_i = \ell^h \left(y_i \langle w, x_i \rangle\right) \iff \xi_i \geq 0 \wedge y_i \langle w, x_i \rangle \geq 1 - \xi_i$$

$\Longrightarrow$: נניח כי $\xi_i \geq 0 \wedge y_i \langle w, x_i \rangle \geq 1 - \xi_i$. אז:

$$\ell^h \left(y_i \langle w, x_i \rangle\right) = max \left\{0, 1 - y_i \langle w, x_i \rangle\right\}$$

נשים לב כי מתקיים מההנחה ש:

$$1 - y_i \langle w, x_i \rangle \leq 1 - (1 - \xi_i) = \xi_i$$

וכמו כן מההנחה ש$\xi_i \geq 0$ ולכן $\ell^h \left(y_i \langle w, x_i \rangle\right) = \xi_i$.
$\Longleftarrow$: נניח כי $\xi_i = \ell^h \left(y_i \langle w, x_i \rangle\right)$. אז זה אומר ש

$$max \left\{0, 1 - y_i \langle w, x_i \rangle\right\} = \xi_i$$

כלומר בכל מקרה $\xi_i \geq 0$ וגם $1 - y_i \langle w, x_i \rangle = \xi_i$, ובפרט מתקיים ש$y_i \langle w, x_i \rangle \geq 1 - \xi_i$.

3. נעשה פה ניתוח של מעריך בייסיאני גאוסיאני בצורה דומה לניתוח שנעשה בתרגול עם הLDA, כשנשים לב לשוני המהותי בין השניים: הLDA הוא ליניארי, ומניח כי כל המחלקות שניתן לסווג אליהן מתפלגות עם אותה שונות רק עם תוחלת שונה. המעריך הגאוסיאני בייסיאני לא מניח שונות שווה, אלא מניח שכל המחלקות מתפלגות נורמלית עם תוחלת ושונות **אחרים**, ובנוסף מניח אי־תלות של הפיצ'רים בדאטא (וכאן הוא שונה מQDA. זה יהיה ההבדל בסעיף ב): כתוצאה מכך יוצא מעריך לא ליניארי, שבנוסף מעריך סטיית תקן **לכל אחת מהמחלקות** . בגלל אי התלות יוצא שלכל מחלקה יש קו־וריאנס אלכסוני (כי כל הפיצ'רים בת"ל).

(א) לפיתוח:

נפתח את הביטוי שעליו צריך למצוא $argMax$:

$$\phi(\Theta|X,Y) = f_{X,Y|\Theta}(\{X_i, y_i\}_{i=1}^m)$$

כלומר למצוא פרמטרים להתפלגויות הנתונות, כך שימקסמו את הביטוי הנ"ל. נסמן בk את המחלקות האפשריות של y_i, כלומר $\mathbb{P}(y_i = k) = \pi_k = \pi_{y_i}$. נפתח:

$$\begin{aligned}
\phi(\Theta|X,Y) &= f_{X,Y|\Theta}(\{X_i, y_i\}_{i=1}^m) \\
{}_{i.i.d} &= \prod_{i=1}^{m}\left(f_{X_i, y_i=k|\Theta}\left(X_i, k\right)\right) \\
&= \prod_{i=1}^{m}\left(f_{X_i|y_i=k}(X_i)\cdot f_{y_i=k|\Theta}(k)\right) \\
&= \prod_{i=1}^{m}\left(\mathcal{N}\left(\mu_{y_i}, \sigma_{y_i}\right)\cdot \pi_{y_i}\right)
\end{aligned}$$

כשהמעבר האחרון מההתפלגויות הנתונות לנו. ניקח לוגריתם:

$$\begin{aligned}
log\left(\phi(\Theta|X,Y)\right) &= log\prod_{i=1}^{m}\left(\mathcal{N}\left(\mu_{y_i}, \sigma_{y_i}\right)\cdot \pi_{y_i}\right) \\
&= \sum_{i=1}^{m} log\left(\mathcal{N}\left(\mu_{y_i}, \sigma_{y_i}\right)\cdot \pi_{y_i}\right) \\
&= \sum_{i=1}^{m} log\left(\frac{1}{\sqrt{2\pi\sigma_{y_i}^2}}\cdot exp\left(\frac{(x_i-\mu_{y_i})^2}{2\sigma_{y_i}^2}\right)\cdot \pi_{y_i}\right) \\
{}_{log\ rules} &= \sum_{i=1}^{m} log\left(\pi_{y_i}\right) - log\left(\sqrt{2\pi\sigma_{y_i}^2}\right) - \frac{(x_i-\mu_{y_i})^2}{2\sigma_{y_i}^2} \\
{}_{more} &= \sum_{i=1}^{m} log\left(\pi_{y_i}\right) - \frac{1}{2}log\left(\sigma_{y_i}^2\right) - \frac{1}{2}log\left(2\pi\right) - \frac{(x_i-\mu_{y_i})^2}{2\sigma_{y_i}^2}
\end{aligned}$$

בצורה דומה לנעשה בתרגול, נסמן בn_k את מספר המופעים של הלייבל הk בדאטא, כלומר $n_k = \sum_{i=1}^m \mathbb{I}_{[y_i=k]}$. ככה ניתן לרשום את הביטוי למעלה כ:

$$log\left(\phi(\Theta|X,Y)\right) = \sum_k n_k log\left(\pi_k\right) - \frac{1}{2}\sum_i \frac{\mathbb{I}_{[y_i=k]}\left(x_i-\mu_k\right)^2}{\sigma_k^2} - \sum \frac{n_k}{2}log\left(\sigma_k^2\right)$$

את הביטוי הזה צריך למקסם לפי הפרמטרים. בנוסף יש אילוץ על π_k : $\sum_k \pi_k = 1$. נרכיב לגראנז'יאן לביטוי:

$$\mathcal{L}\left(\hat{\pi}, \hat{\mu}, \hat{\sigma}, \lambda\right) = log\left(\phi(\Theta|X,Y)\right) - \lambda\left(\sum_k \pi_k - 1\right)$$

נגזור תחילה לפי π ונקבל:

$$\frac{\partial\mathcal{L}}{\partial\pi_k}=\frac{\partial log\left(\phi(\Theta|X,Y)\right)}{\partial\pi_k}-\lambda\cdot\frac{\partial\mathcal{L}}{\partial\pi_k}=\frac{n_k}{\pi_k}-\lambda$$

נשווה ל-0 ונקבל

$$\hat{\pi}_k=\frac{n_k}{\lambda}$$

מהתנאי ש-$\sum_k\pi_k=1$ נקבל כי $\lambda=m$.
הפרמטר μ מביא את אותו ביטוי שראינו בתרגול (רק הפעם חד מימדי, יותר פשוט) ולכן נקבל

$$\hat{\mu}_k=\frac{1}{n_k}\sum\mathbb{I}_{[y_i=k]}x_i\in\mathbb{R}$$

כלומר התוחלת של כל הדגימות שהלייבל שלהם הוא k. זוהי תוצאה לתוצאה שקיבלנו ב-LDA.
עבור ה-σ אנחנו מקבלים הפעם ביטוי קצת שונה בגלל ההנחות השונות: במקום שונות אחת משותפת לכל הלייבלים כמו ב-LDA, אנחנו רוצים לבטא את השונות של כל לייבל בנפרד. כלומר:

$$\hat{\sigma}_k^2=\frac{1}{n_k}\sum\mathbb{I}_{[y_i=k]}\left(x_i-\hat{\mu}_k\right)^2\in\mathbb{R}$$

(ב) כעת נרחיב ל-d פיצ'רים של הדאטא, ונסמן $\overline{x_i}\in\mathbb{R}^d$ דגימה מהדאטא. נחזור ל-$likelihood$ (לוג) מסעיף קודם ונחיל את הנוסחה של גאוסיאן רב מימדי:

$$\begin{aligned}
log\left(\phi(\Theta|X,Y)\right)&=log\prod_{i=1}^{m}\left(\mathcal{N}\left(\overline{\mu}_{y_i},\Sigma_{y_i}\right)\cdot\pi_{y_i}\right)\\
&=\sum_{i=1}^{m}log\left(\mathcal{N}\left(\overline{\mu}_{y_i},\Sigma_{y_i}\right)\cdot\pi_{y_i}\right)\\
&=\sum_{i=1}^{m}log\left(\frac{1}{\sqrt{(2\pi)^d\left|\Sigma_{y_i}\right|}}\cdot exp\left(-\frac{1}{2}(x_i-\overline{\mu}_{y_i})^T\Sigma_{y_i}^{-1}(x_i-\overline{\mu}_{y_i})\right)\cdot\pi_{y_i}\right)\\
{\scriptstyle log\ rules}&=\sum_{i=1}^{m}log\left(\pi_{y_i}\right)-log\left(\sqrt{(2\pi)^d\left|\Sigma_{y_i}\right|}\right)-\frac{1}{2}(x_i-\overline{\mu}_{y_i})^T\Sigma_{y_i}^{-1}(x_i-\overline{\mu}_{y_i})\\
{\scriptstyle more}&=\sum_{i=1}^{m}log\left(\pi_{y_i}\right)-\frac{d}{2}log\left(2\pi\right)-\frac{1}{2}log\left(\left|\Sigma_{y_i}\right|\right)-\frac{1}{2}(x_i-\overline{\mu}_{y_i})^T\Sigma_{y_i}^{-1}(x_i-\overline{\mu}_{y_i})
\end{aligned}$$

נרצה למקסם את הביטוי, ונבחין כי $-\frac{d}{2}log\left(2\pi\right)$ הוא קבוע ולכן נתעלם ממנו בשלב זה.
באותה צורה כמו בחד מימדי, נשתמש באינדיקטורים כדי לרשום את הנוסחה בצורה קצת שונה:

$$log\left(\phi(\Theta|X,Y)\right)=\sum_k n_k log\left(\pi_k\right)-\frac{1}{2}log\left(\left|\Sigma_k\right|\right)-\frac{1}{2}\sum_i\mathbb{I}_{[y_i=k]}(x_i-\overline{\mu}_k)^T\Sigma_k^{-1}(x_i-\overline{\mu}_k)$$

נרכיב לגראנז'יאן שיכלול את האילוץ על π_k:

$$\mathcal{L}\left(\hat{\pi},\hat{\overline{\mu}},\hat{\Sigma},\lambda\right)=log\left(\phi(\Theta|X,Y)\right)-\lambda\left(\sum_{k}\pi_k-1\right)$$

ניתן לראות בקלות שגזירת לפי π_kנותנת בדיוק את אותה תוצאה:

$$\hat{\pi}_k=\frac{n_k}{m}$$

הגזירה לפי $\overline{\mu_k}$ זהה לגזירה שראינו בתרגול הפעם. נקבל:

$$\hat{\overline{\mu}}_k=\frac{1}{n_k}\sum\mathbb{I}_{[y_i=k]}\overline{x}_i\in\mathbb{R}^d$$

ולמעשה סיימנו־ השונות מוגדרת ע"י ההנחה של האומד, והיא שונות אחרת לכל k וגם פיצ'רים בת"ל. תחת ההנחה שכל הפיצ'רים הם בת"ל, המטריצת שונות תהיה **אלכסונית**, (בגודל d על d) ־ אם כל הפיצ'רים הם בת"ל, אז אין ביניהם קורלציה, וכל האיברים מחוץ לאלכסון שווים ל0. לכן נוכל פשוט לבטא את השונות של לייבל k ע"י לקיחת כל האיברים על האלכסון והרכבת וקטור עם d כניסות. כלומר נגדיר:

$$\hat{\overline{\sigma}}_k^2=Diag\left(\sum\mathbb{I}_{[y_i=k]}\left(x_i-\hat{\overline{\mu}}_k\right)\left(x_i-\hat{\overline{\mu}}_k\right)^T\right)$$

ובסופו של דבר לצורך המימוש בקוד, נרכיב מטריצת שונויות שתבוטא כך:

$$Vars=\begin{bmatrix}\hat{\overline{\sigma}}_1^2\\ \vdots\\ \hat{\overline{\sigma}}_k^2\end{bmatrix}\in\mathbb{R}^{k\times d}$$

.4

(א) **פואסון חד מימדי:** נתון כי

$$f_y(k)=\pi_k\ ,f_{x_i|y_i}(x_i)=\frac{\lambda_{y_i}^{-x_i}\cdot e^{-\lambda_{y_i}}}{x_i!}$$

והאילוץ הרגיל על $\sum_k\pi_k=1$. צריך למקסם את הביטוי:

$$\begin{aligned}\phi(\Theta|X,Y)&=f_{X,Y|\Theta}(\{X_i,y_i\}_{i=1}^m)\\ {}_{i.i.d}&=\prod_{i=1}^m\left(f_{x_i,y_i|\Theta}\left(x_i,y_i\right)\right)\\ &=\prod_{i=1}^m\left(f_{x_i|y_i}(x_i)\cdot f_{y_i}(y_i)\right)\\ &=\prod_{i=1}^m\left(\frac{\lambda_{y_i}^{-x_i}\cdot e^{-\lambda_{y_i}}}{x_i!}\cdot\pi_{y_i}\right)\end{aligned}$$

צריך למקסם אז אפשר לקחת לוגריתם:

$$\begin{aligned}
log\left(\phi(\Theta|X,Y)\right) &= \sum_{i}^{m} log\left(\frac{\lambda_{y_i}^{x_i}\cdot e^{-\lambda_{y_i}}}{x_i!}\cdot \pi_{y_i}\right) \\
&= \sum_{i}^{m}\left[log\left(\lambda_{y_i}^{x_i}\right)+log\left(e^{-\lambda_{y_i}}\right)+log\left(\pi_{y_i}\right)-log\left(x_i!\right)\right] \\
&= -\sum_{i}^{m} log\left(x_i!\right)+\sum_{i}^{m}\left[x_i log\left(\lambda_{y_i}\right)-\lambda_{y_i}+log\left(\pi_{y_i}\right)\right]
\end{aligned}$$

נבחין כי הביטוי הראשון הוא קבוע ולכן נשמיט אותו מכאן. נשתמש באינדיקטורים כדי לספור מופעים על $y_i = k$, ונכתוב את הביטוי כך:

$$log\left(\phi(\Theta|X,Y)\right) = \sum_{i}^{m}\mathbb{I}_{[y_i=k]}\cdot x_i log\left(\lambda_{y_i}\right) - n_k\lambda_k + n_k log\left(\pi_k\right)$$

נרכיב את הלגראנז'יאן:

$$\mathcal{L}\left(\hat{\pi},\hat{\lambda},\tau\right) = log\left(\phi(\Theta|X,Y)\right) - \tau\left(\sum_{k}\pi_k - 1\right)$$

למנוע בלבול: כאן τ הם כופלי לגראנז. נגזור לפי π_k:

$$\frac{\partial\mathcal{L}}{\partial\pi_k} = \frac{n_k}{\pi_k} - \tau$$

נקבל את אותה תוצאה כמו בשאלה 3:

$$\hat{\pi}_k = \frac{n_k}{m}$$

נגזור לפי λ_k:

$$\frac{\partial\mathcal{L}}{\partial\lambda_k} = -n_k + \sum_{i}^{m}\mathbb{I}_{[y_i=k]}\cdot x_i log\left(\lambda_{y_i}\right) \stackrel{!}{=} 0$$

נקבל ש:

$$\lambda_k = \frac{\sum_{i}^{m}\mathbb{I}_{[y_i=k]}\cdot x_i}{n_k}$$

כלומר־ פשוט התוחלת של כל הסאמפלים שהלייבל שלהם הוא k.

(ב) פואסון רב מימדי: כעת $f_{x_i|y_i}(x_i) = \prod_{j=1}^{d} \frac{\lambda_{y_{i,j}}^{x_{ij}} \cdot e^{-\lambda_{y_{i,j}}}}{x_{ij}!}$ כלומר לכל $\overline{x_i}$ יש d פיצ'רים שכולם מתפלגים פואסון עם פרמטר $\lambda_{y_i,j}$ באופן בת"ל. נפתח את הלוג-לייקליהוד:

$$\begin{aligned}
\phi(\Theta|X,Y) &= f_{X,Y|\Theta}(\{X_i, y_i\}_{i=1}^{m}) \\
{}_{i.i.d} &= \prod_{i=1}^{m} \left(f_{x_i,y_i|\Theta}\left(x_i, y_i\right)\right) \\
&= \prod_{i=1}^{m} \left(f_{x_i|y_i}(x_i) \cdot f_{y_i}(y_i)\right) \\
&= \prod_{i=1}^{m} \left(\prod_{j=1}^{d} \frac{\lambda_{y_{i,j}}^{x_{ij}} \cdot e^{-\lambda_{y_{i,j}}}}{x_{ij}!} \cdot \pi_{y_i}\right)
\end{aligned}$$

ניקח לוג:

$$\begin{aligned}
log\left(\phi(\Theta|X,Y)\right) &= \sum_{i}^{m} log\left(\prod_{j=1}^{d} \frac{\lambda_{y_{i,j}}^{x_{ij}} \cdot e^{-\lambda_{y_{i,j}}}}{x_{ij}!} \cdot \pi_{y_i}\right) \\
&= \sum_{i}^{m}\sum_{j}^{d}\left[log\left(\lambda_{y_{i,j}}^{x_{ij}}\right) + log\left(e^{-\lambda_{y_{i,j}}}\right) + log\left(\pi_{y_i}\right) - log\left(x_{ij}!\right)\right] \\
&= -\sum_{i}^{m}\sum_{j}^{d} log\left(x_i!\right) + \sum_{i}^{m}\sum_{j}^{d}\left[x_i log\left(\lambda_{y_i}\right) - \lambda_{y_i} + log\left(\pi_{y_i}\right)\right]
\end{aligned}$$

הראשון קבוע ויושמט, ועכשיו נשתמש בטריק האינדיקטור:

$$= \sum_{k}\sum_{j}^{d}\left[\sum_{i} \mathbb{I}_{[y_i=k]} x_i log\left(\lambda_{y_i}\right) + n_k\left(log\left(\pi_k\right) - \lambda_{k,j}\right)\right]$$

נרכיב לגראנז'יאן כמו מקודם ($\mathcal{L}$ הביטוי ו-τ הכופלים) ונגזור לפי π_k ונקבל:

$$\frac{\partial \mathcal{L}}{\partial \pi_k} = \frac{dn_k}{\pi_k} - \tau \overset{!}{=} 0 \Longrightarrow \pi_k = \frac{dn_k}{\tau}$$

מכך ש $\sum_k \pi_k = 1$ נקבל ש

$$\sum_{k} \frac{dn_k}{\tau} = 1 \Longrightarrow \tau = md$$

ומכאן נקבל בסהכ כי $\hat{\pi}_k = \frac{n_k}{m}$ באופן מאוד לא מפתיע.
נגזור לפי λ_{kj}:

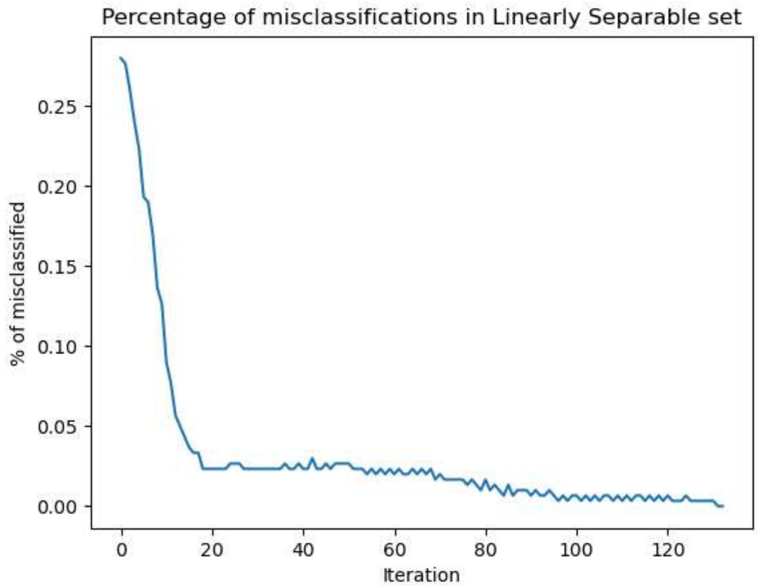
$$\frac{\partial \mathcal{L}}{\partial \lambda_{kj}} = -n_k + \sum_{i} \mathbb{I}_{[y_i=k]} x_i log\left(\lambda_{y_i}\right)$$

בדיוק כמו בסעיף החד מימדי. נקבל ש $\hat{\lambda}_{k,j} = \frac{\sum_i \mathbb{I}_{[y_i=k]} x_{i,j}}{n_k}$, התוחלת של הפיצ'ר ה$j$ בתוך כל הסאמפלים שהלייבל שלהם הוא k.

חלק פרקטי

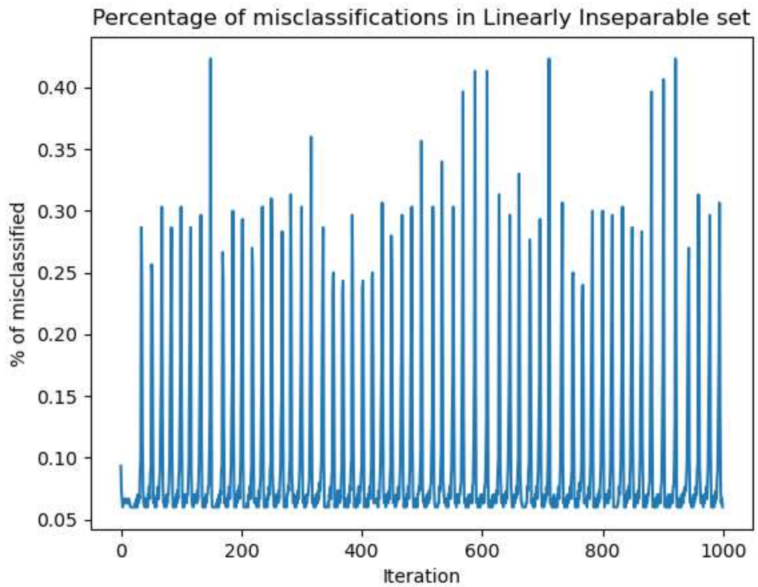
Perceptron classifier

1. הגרף הבא מציג את אחוז השגיאות, כלומר $Misclassification$, עבור הדאטא שניתן להפרדה ליניארית כפונקצייה של האיטרציות בפונקציית הFit של הפרספטרון:



ניתן לראות שלא הגענו למקסימום איטרציות, כלומר הגענו למצב שבו **כל** הקלסיפיקציות נכונות - מצאנו מישור שמפריד ליניארית את הדאטא.

2. הגרף הבא מציג את אותה הפונקצייה אך עבור הדאטא שאינו ניתן להפרדה:

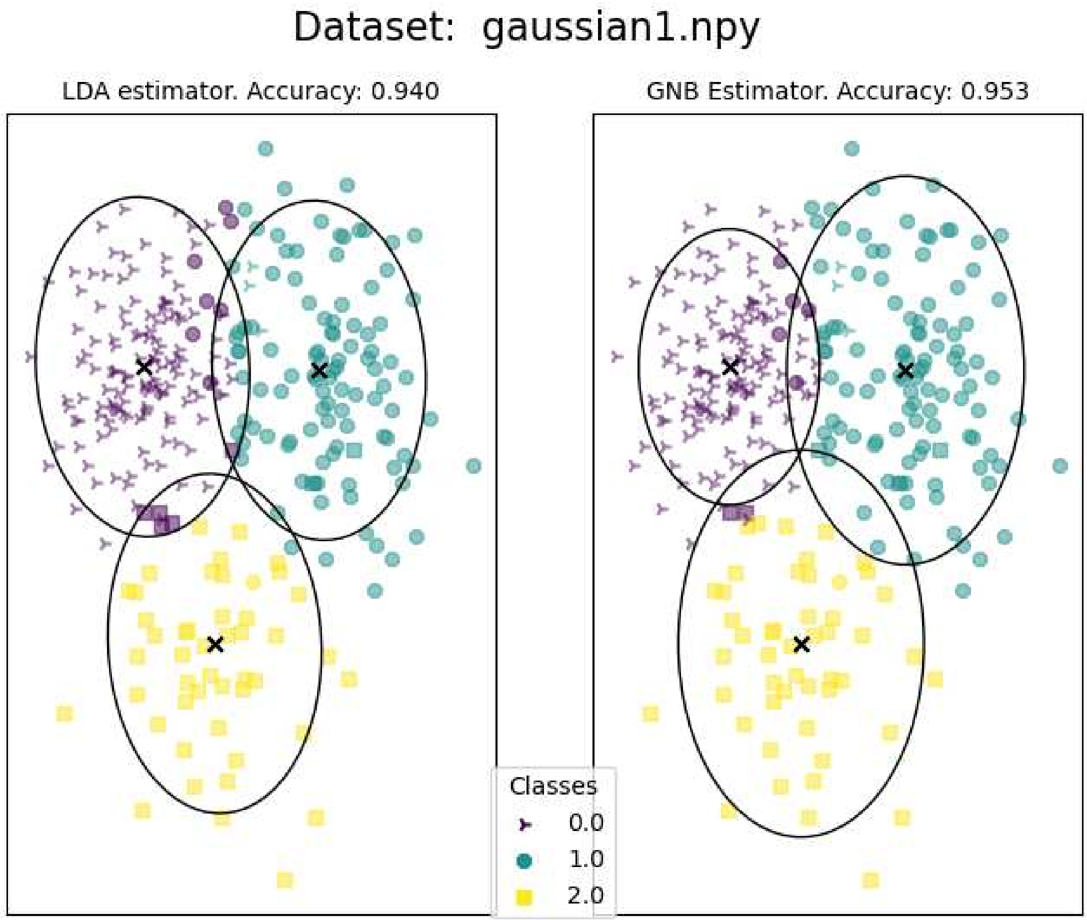


הגרף מתנהג בצורה משוגעת כי אנחנו מחפשים מישור שיהיה 'צודק' לגבי כל הנקודות בדאטא, ומכיוון שזה דאטא שאי אפשר להפריד ליניארית זה מצב לא אפשרי. זו גם הסיבה שהגענו למקסימום איטרציות שאפשרנו. אם היינו 'מתפשרים' על $Soft-SVM$ כלומר על מישור שנותן מרווח טעות, אך מינימלית - היה אפשר להוציא תוצאה די טובה, בערך 0.06% טעות, המינימום שמופיע בגרף. רוב הדאטא בעולם הוא לא ניתן להפרדה ליניארית או לפחות לא ידוע לנו עליו שכן, והתעקשות על $Hard-SVM$ הרבה פעמים תביא תוצאות לא כל כך טובות כמו זו.

Bayes classifiers

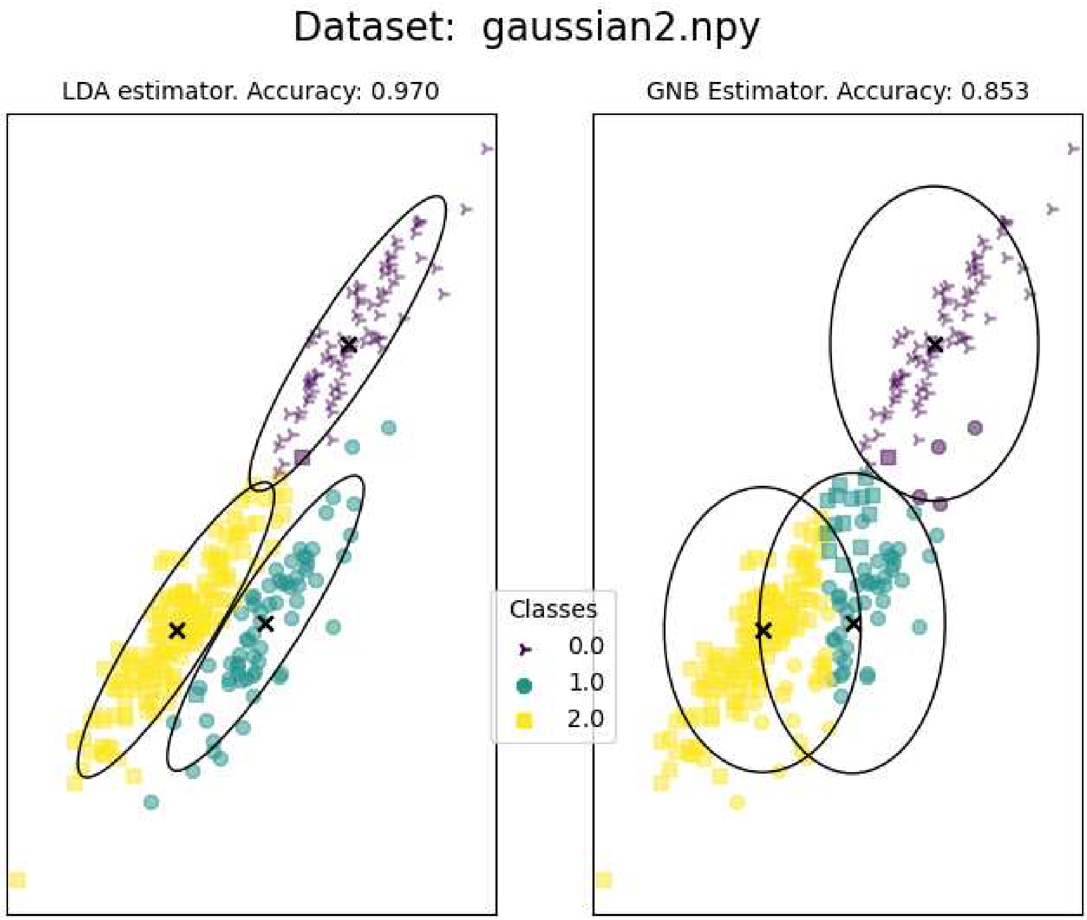
בגרפים הבאים לכל מחלקה ניתן צבע וצורה המופיעים במקרא. נקודה על הגרף שאיננה תואמת למקרא, משמעותה טעות בסיווג.

1. הגרף הבא מציג ריצה של LDA ושל GNB על דאטא ללא קורלציה, או קורלציה חלשה, בין הפיצ'רים:



ניתן לראות שעבור שהGNB הצליח יותר בחיזוי שלו, ולכן ניתן להניח שהדאטא מקיים (או כמעט מקיים) את ההנחות שלו: מכיוון שהוא מניח דאטא שמגיע מהתפלגויות שונות (גאוסיאנים, אבל פרמטרים שונים), ניתן לומר שההתפלגות שממנה נדגם הדאטא באמת מתנהגת כך: לכל לייבל יש התפלגות שונה. כמו כן הGNB מניח אי תלות בין הפיצ׳רים של הדאטא, שאת זה ניתן לראות בעיניים שמתקיים כאן, ולכן הוא זכה להצלחה גדולה. אף על פי כן, מכך שהLDA **הלינארי** הצליח גם לא רע יחסית, אפשר ללמוד שבין כל התפלגות של שני לייבלים, אפשר למצוא איזשהו מישור מפריד שיהיה יחסית מוצלח (ואכן אפשר לראות בעיניים שזה המצב).

2. הגרף הבא מציג ריצה של LDA ושל GNB על דאטא עם קורלציה חיובית חזקה בין הפיצ׳רים:



ה-LDA מוצלח יותר מה-GNB.

מכיוון שההצלחה של LDA גבוהה, ניתן ללמוד שהדאטא מקיים (או כמעט מקיים) את ההנחות שלו: כל הלייבלים מתפלגים בצורה זהה (כלומר זהים בשונות) אך עם תוחלת אחרת. ההנחה של ה-GNB שהדאטא מקיים אי־תלות בין הפיצ׳רים לא נכונה כאן, ורואים שאחוזי ההצלחה שלו נמוכים בהרבה, לעומת ה-LDA שפשוט לא מניח כלום לגבי תלות בין הפיצ׳רים. כמו כן LDA הוא ליניארי, ואכן אפשר לראות שניתן למתוח מישורים מפרידים בין כל שני לייבלים שיצליחו להפריד בין שניהם בצורה טובה.