

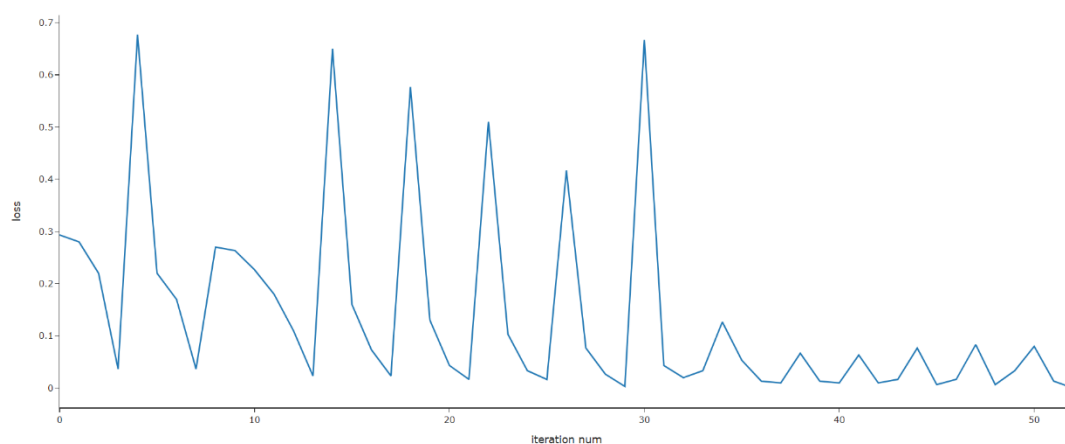
חלק פרקטי

3.1

שאלה 1.

ניתן לראות שאחרי 50 ומשהו איטרציות האלגוריתם מגיע ל $\text{loss} = 0$ ועוצר וזה כי ה data הוא separable

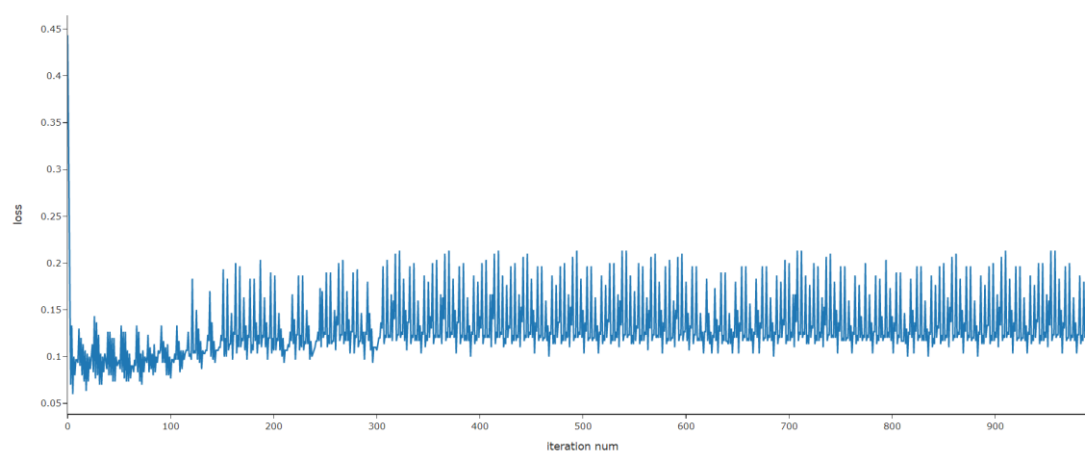
perceptron algorithm's training loss as a function of the training iterations - in Linearly Separable data



שאלה 2.

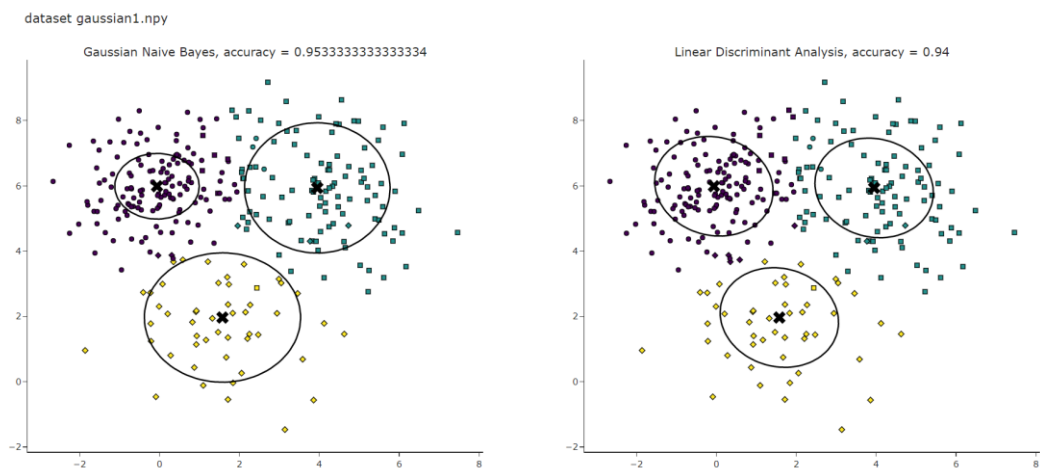
בניגוד לגרף הקודם ניתן לראות שהאלגוריתם לא עצר (עד max iter) – כי לא הגיע ל $\text{loss} = 0$ וזה כי ה data הוא inseparable

perceptron algorithm's training loss as a function of the training iterations - in Linearly Inseparable data



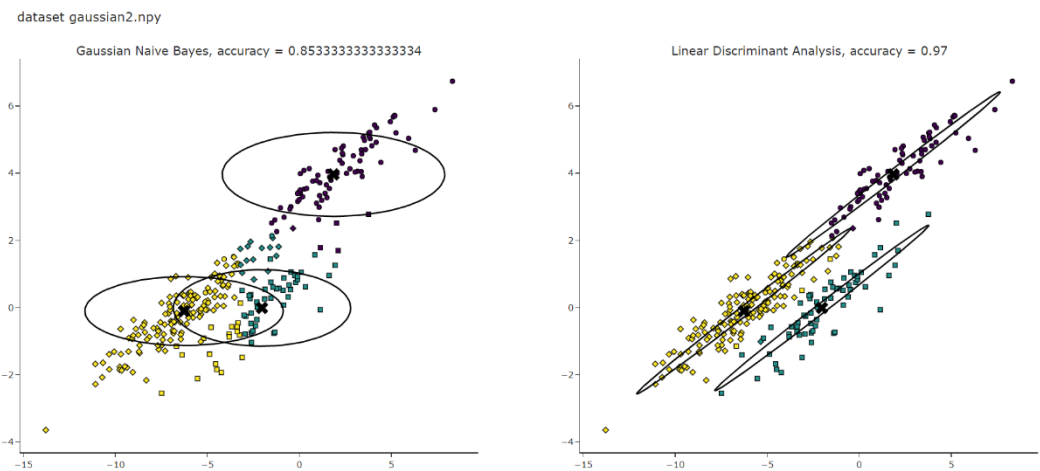
שאלה 1.

מפיזור ה data ניתן לראות שההנחה שהפיצורים הם בלתי תלויים סבירה וכמו כן מצורת האליפסות ניתן לראות שההנחה שמטריצת ה cov זהה נכונה. ואכן המודלים הצליחו טוב.



שאלה 2.

בניגוד למקרה הקודם ניתן לראות שאכן צורת האליפסות דומה ולכן ההנחה של ה LDA לגבי מטריצת ה cov נכונה, לעומת זאת ההנחה של ה GNB לגבי האי תלות לא נכונה וניתן לראות זאת לפי הפיזור של ה data. ולכן ה LDA הצליח לעומת ה GNB שלא הצליח כל כך.



שאלה 1:

ייתנו חזים לגביא למינימום את w, b בלומר את $v = \begin{pmatrix} w \\ b \end{pmatrix}$

$$\argmin_{w, b} \|v\|^2 = \argmin_{w, b} \left\| \begin{pmatrix} w \\ b \end{pmatrix} \right\|^2 = \argmin_{w, b} \begin{pmatrix} w \\ b \end{pmatrix}^T \begin{pmatrix} w \\ b \end{pmatrix} =$$

$$= \argmin_{w, b} \begin{pmatrix} w \\ b \end{pmatrix}^T I \begin{pmatrix} w \\ b \end{pmatrix} = \argmin_{w, b} \frac{1}{2} \begin{pmatrix} w \\ b \end{pmatrix}^T 2I \begin{pmatrix} w \\ b \end{pmatrix} =$$

$$= \argmin_{w, b} \left(\frac{1}{2} v^T Q v + a^T v \right) \quad \begin{cases} a=0 \\ Q=2I \end{cases}$$

וקצת תבין כיצד למצוא את הקטן למינימום.

אילוץ:

$$y_i (\langle w, x_i \rangle + b) \geq 1 \iff i \text{ בס}$$

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

$$-y (\langle w, x \rangle + b) \leq \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix} \iff$$

$$-y X y \begin{pmatrix} w \\ b \end{pmatrix} \leq \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix}$$

אם כמובן הילוכים מתקבלים כאשר $A = -y X y$ $d = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix}$

שאלה 2:

נתון שגמר האינרס $\forall_i y_i \langle w, x_i \rangle \geq 1 - \xi_i$ $0 \leq \xi_i \leq 1$

$$\frac{1}{m} \sum_{i=1}^m \xi_i \quad \text{הוא} \quad \frac{1}{m} \sum_{i=1}^m \ell^{\text{hinge}}(y_i \langle w, x_i \rangle)$$

- האנליזה מתקדמת

$$1 - y_i \langle w, x_i \rangle \geq 1 - (1 - \xi_i) = \xi_i \geq 0$$

וקבל

$$\ell^{\text{hinge}}(y_i \langle w, x_i \rangle) = \max\{0, 1 - y_i \langle w, x_i \rangle\} =$$

$$= 1 - y_i \langle w, x_i \rangle$$

$$\text{כי אם } y_i \langle w, x_i \rangle \geq 1 \quad \text{אז} \quad \ell^{\text{hinge}}(y_i \langle w, x_i \rangle) = 0$$

ξ_i המינימום

$$\left(\begin{array}{l} \text{וקבל} \quad \text{שגמר} \quad \xi_i \leq \text{המינימום של הריבוי} \\ \frac{1}{m} \sum_{i=1}^m \ell^{\text{hinge}}(y_i \langle w, x_i \rangle) \quad \text{הוא} \quad \text{שגמר} \quad \xi_i \end{array} \right)$$

שאלה 3:

$$y \sim \text{Mult}(\pi) \quad \text{הוא}$$

$$\forall_{j \in [d]} x_j | y = \kappa \sim N(\mu_{\kappa j}, \sigma_{\kappa j}^2)$$

$$X | y = \kappa \sim N(\mu_{\kappa}, \sigma_{\kappa}^2) \quad \leftarrow \quad x \in \mathbb{R} \quad (a)$$

parameters σ_k, μ_k, π_k are unknown

$$\arg\max_{k \in [K]} f_{X|Y=k}(x) f_Y(k) = \arg\max_{k \in [K]} \prod_{i=1}^m \frac{\pi_k}{\sigma_k \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x_i - \mu_k}{\sigma_k}\right)^2\right)$$

$$L(\theta | X, Y) = \prod_{i=1}^m f_{X|Y=y_i}(x_i) f_{Y|\theta}(y_i)$$

$$\log L(\theta) = \sum_{i=1}^m \log (N(x_i | \mu_{y_i}, \sigma_{y_i}^2) \text{Mult}(y_i | \pi))$$

$$= \sum_{i=1}^m \log \frac{1}{\sigma_k \sqrt{2\pi}} - \frac{1}{2} \left(\frac{x_i - \mu_k}{\sigma_k}\right)^2 + \log(\pi_{y_i})$$

using

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

$$= \sum_{k \in [K]} [n_k \log(\pi_k) - \sum_{y_i=k} [\log(\sqrt{2\pi}) + \frac{1}{2} \left(\frac{x_i - \mu_k}{\sigma_k}\right)^2]] + m \cdot \log(\sqrt{2\pi})$$

$$J = L(\theta | X, Y) - \lambda g(\pi) \quad \text{where} \quad g(\pi) = \sum_{k \in [K]} \pi_k^{-1} \quad \text{prior on } \pi_k$$

$$\frac{\partial J}{\partial \pi_k} = \frac{\partial L}{\partial \pi_k} - \lambda \frac{\partial g}{\partial \pi_k} = \frac{n_k}{\pi_k} - \lambda = 0$$

$$\pi_k = \frac{n_k}{\lambda} \rightarrow \pi_k = \frac{n_k}{m} \quad \text{where } \lambda = m$$

$$\frac{\partial J}{\partial \mu_k} = \sum_{y_i=k} \frac{x_i - \mu_k}{\sigma_k^2} = 0, \quad n_k \mu_k = \sum_{y_i=k} x_i$$

$$\rightarrow \mu_k = \frac{1}{n_k} \sum_{y_i=k} x_i$$

$$\frac{\partial L}{\partial \sigma_k^2} = \sum_{y_i=k} \frac{1}{2} \left(\frac{x_i - \mu_k}{\sigma_k^2} \right)^2 = \frac{n_k}{2\sigma_k^2}$$

$$\rightarrow \sum_{y_i=k} (x_i - \mu_k)^2 = n_k \sigma_k^2$$

$$\rightarrow \sigma_k^2 = \sum_{y_i=k} \frac{(x_i - \mu_k)^2}{n_k}$$

(b) מהי הסתברות? מהי תכונה?

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

$$\mu_k = \frac{1}{n_k} \sum_{y_i=k} x_i$$

$$\sigma_k^2 = \sum_{y_i=k} \frac{(x_i - \mu_k)(x_i - \mu_k)^T}{n_k}$$

השאלה 4:

היחס π_k, λ_k בין

$$\arg \max_{k \in [K]} \frac{P_{X|Y=k}(x) P_Y(y)}{P_X(x)} = \arg \max_{k \in [K]} \pi(k) \prod_{i=1}^n \frac{e^{-\lambda_k} (\lambda_k)^{x_i}}{x_i!}$$

MLE ላይ ማሳሰቢያ

$$\begin{aligned} L(\theta | x, y) &= \prod_{i=1}^m P_{X|Y=y_i}(x_i, y_i) \cdot P_{Y|\theta}(y_i) = \\ &= \prod_{i=1}^m \pi_{y_i} \frac{e^{\lambda_{y_i} x_i}}{x_i!} \end{aligned}$$

ሎግ ማድረግ

$$\begin{aligned} L(\theta | x, y) &= \sum_{i=1}^m \log(\pi(y_i)) + \log\left(\frac{e^{\lambda_{y_i} x_i}}{x_i!}\right) - \lambda_{y_i} = \\ &= \sum_{i=1}^m \log(\pi(y_i)) + x_i \log(\lambda_{y_i}) - \log(x_i!) - \lambda_{y_i} = \\ &= \sum_{k \in \{k\}} n_k \log(\pi_k) + x_i \log(\lambda_k) - \log(x_i!) - n_k \lambda_k \end{aligned}$$

፡ ማሳሰቢያ ለሎግ ማድረግ

$$J = L(\theta | x, y) - \lambda g(\pi)$$

$$\frac{\partial L}{\partial \pi_k} = \frac{n_k}{\pi_k} - \lambda \xrightarrow{\text{ማዕከላዊ}} \pi_k = \frac{n_k}{\lambda} \quad , \text{ ለምሳሌ}$$

$$\sum_k \pi_k = 1 \rightarrow \sum_k \frac{n_k}{\lambda} = 1 \rightarrow \frac{n}{\lambda} = 1 \rightarrow n = \lambda$$

$$\pi_k = \frac{n_k}{n} \quad \text{ሆኖ}$$

ማሳሰቢያ ለሎግ ማድረግ

$$\frac{\partial L}{\partial \lambda_k} = \sum_k x_i \log(\lambda_k) - n_k \lambda_k = 0 \Rightarrow$$

$$\Rightarrow \sum_k \frac{x_i \log(\lambda_k)}{\lambda_k} - n_k = 0 \Rightarrow$$

$$\Rightarrow \frac{1}{n_k} \sum_k x_i = \lambda_k$$

: סך כל הנתונים - מספר הנקודות (b)

$$p_k = \frac{n_k}{n} , \quad \lambda_k = \frac{1}{n_k} \sum_k x_i$$