203907092 ניר מנסור

תרגיל בית 2

Theoretical part

1) לכל $i \in [m]$

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

נגדיר $d = \begin{pmatrix} b - \frac{1}{y_1} \\ \vdots \\ b - \frac{1}{y_m} \end{pmatrix}$ $A = -1 \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$ וכך קיבלנו אילוץ כנדרש.

כמו כן נגדיר $a = 0$ $Q = 2I_n$ ולכן $\frac{1}{2} v^T Q v + a^T v = \frac{1}{2} v^T 2 I_n v = \|v\|^2$

ולכן הבעיה שקולה לבעיה

2.a) נחשב את פונ' הנראות -

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

$$= \prod_{i=1}^{m} N(x_i | \mu_{y_i}, \sigma^2_{y_i}) \, mult(y_i | \pi)$$

ונחשב את הלוג -

$$\ell(\Phi | X, y) = \sum_i \log(mult(y_i | \pi)) + \log(N(x_i | \mu_{y_i}, \sigma^2_{y_i})) =$$

$$= \sum_i \log(\pi_{y_i}) + \log\left(\frac{1}{\sqrt{2\pi\sigma^2_{y_i}}} e^{-\frac{(x_i - \mu_{y_i})^2}{2\sigma^2_{y_i}}}\right) =$$

$$= \sum_i \log(\pi_{y_i}) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2_{y_i}) - \frac{(x_i - \mu_{y_i})^2}{2\sigma^2_{y_i}} =$$

$$= -\frac{m}{2}\log(2\pi) + \sum_k n_k \log(\pi_k) - \frac{1}{2} n_k \log(\sigma^2_k) - \frac{1}{2\sigma^2_k} \sum_{\{i | y_i = k\}} (x_i - \mu_k)^2$$

נגזור לפי μ_k לכל $k\in[K]$

$$\frac{dl}{d\mu_k} = -\frac{1}{\sigma^2_k}\sum_{\{i|y_i=k\}} x_i-\mu_k = 0 \Rightarrow \mu_k^{mle} = \bar{x}_k$$

נציב את ה-mle בגזירה

$$\frac{dl}{d\sigma^2_k} = -\frac{n_k}{2\sigma^2_k} - \frac{1}{2(\sigma^2_k)^2}\sum_{\{i|y_i=k\}}(x_i-\mu_k)^2 \Rightarrow (\sigma^2)_k^{mle} = \frac{1}{n_k}\sum_{\{i|y_i=k\}}(x_i-\mu_k^{mle})^2$$

כאשר $n_k=\sum_i 1_{y_i=k}$ ו- $\bar{x}=\frac{1}{n_k}\sum_{i|y_i=k} x_i$

נכתוב את הלגרנז' לפי π_k ונגזור את הלגרנזיאן

$$L = l(\Phi|X,y) - \lambda\left(\sum_k \pi_k - 1\right)$$

נגזור לפי π_k ונשווה ל-0

$$\frac{dL}{d\pi_k} = \frac{dl(\Phi|X,y)}{d\pi_k} - \lambda = \frac{n_k}{\pi_k} - \lambda = 0 \Rightarrow \pi_k = \frac{n_k}{\lambda}$$

נמצא את λ

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

ולכן ה-mle של π_k הוא

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

2.b נכתוב את ה לייקליהוד-

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

$$= \prod_{j=1}^{d}\prod_{i=1}^{m} N(x_{ij}|\mu_{y_i,j},\sigma^2_{y_i,j})\, mult(y_i|\pi)$$

נפעיל את ה-log

$$l(\Phi|X,y) = \sum_i\left(\log\left(mult(y_i|\pi)\right) + \sum_j \log\left(N(x_{ij}|\mu_{y_i,j},\sigma^2_{y_i,j})\right)\right) =$$

$$= \sum_k n_k \log(\pi_k) + \sum_k\sum_j 1_{y_i=k}\log\left(N(x_{ij}|\mu_{y_i,j},\sigma^2_{y_i,j})\right)$$

[illegible] $j\in[d]$ [illegible]

$$\pi_k^{mle} = \frac{n_k}{m},\quad \mu_{k,j}^{mle} = \frac{1}{n_k}\sum_{\{i|y_i=k\}} x_{ij},\quad (\sigma^2)_{kj}^{mle} = \frac{1}{n_k}\sum_{\{i|y_i=k\}}(x_{ij}-\mu_{kj}^{mle})^2$$

3.a נכתוב את הלייקליהוד-

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

$$= \prod_{i=1}^{m} poi(x_i|\lambda_{y_i}) \, mult(y_i|\pi)$$

נחשב גם ל-log

$$l(\Phi|X,y) = \sum_i \log(poi(x_i|\lambda_{y_i})) + \log(mult(y_i|\pi)) =$$

$$= \sum_i \log\left(\frac{\lambda_{y_i}^{x_i} e^{-\lambda_{y_i}}}{x_i!}\right) + \log(\pi_{y_i}) = \sum_i x_i \log(\lambda_{y_i}) - \lambda_{y_i} - \log(x_i!) + \log(\pi_{y_i}) =$$

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

נגזור לפי את ונשווה ל0

$$\frac{dl}{d\lambda_k} = \frac{1}{\lambda_k} \sum_{\{i|y_i=k\}} x_i - n_k = 0 \Rightarrow \lambda_k^{mle} = \frac{1}{n_k} \sum_i 1_{(y_i=k)} x_i$$

הערך של π_k ניתן בשאלה קודמת ולכן. $\pi_k^{mle} = \frac{n_k}{m}$

3.b נכתוב את הלייקליהוד-

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

$$= \prod_{i=1}^{m} \left(\prod_{j=1}^{d} poi(x_{ij}|\lambda_{y_i j}) \right) mult(y_i|\pi)$$

ונחשב את ה log

$$l(\Phi|X,y) = \sum_i \sum_j \log(poi(x_{ij}|\lambda_{y_i,j})) + \sum_i \log(mult(y_i|\pi)) =$$

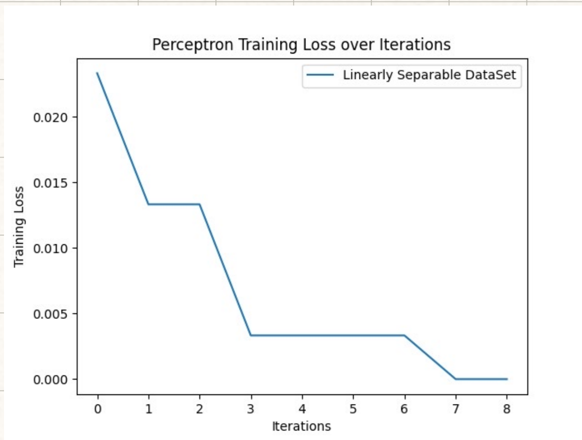
$$= \sum_k \sum_j \left(\log(\lambda_{kj}) \sum_{\{i|y_i=k\}} x_{ij} - n_k \lambda_{kj} + n_k \log \pi_k \right) - \sum_{i,j} \log(x_{ij}!)$$

ובדומה נקבל-

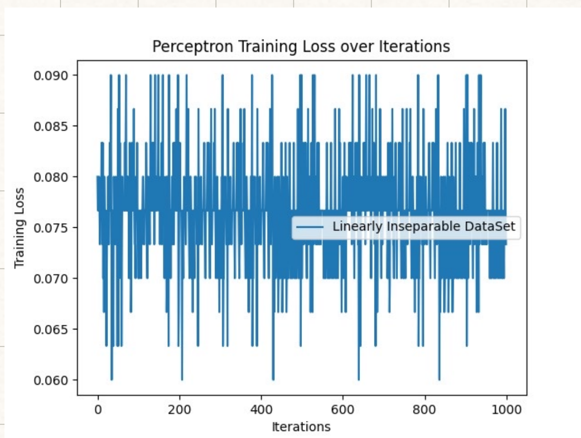
$$\lambda_{kj}^{mle} = \frac{1}{n_k} \sum_i 1_{(y_i=k)} x_{ij} \, , \; \pi_k^{mle} = \frac{n_k}{m}$$

practical part

(3.1)

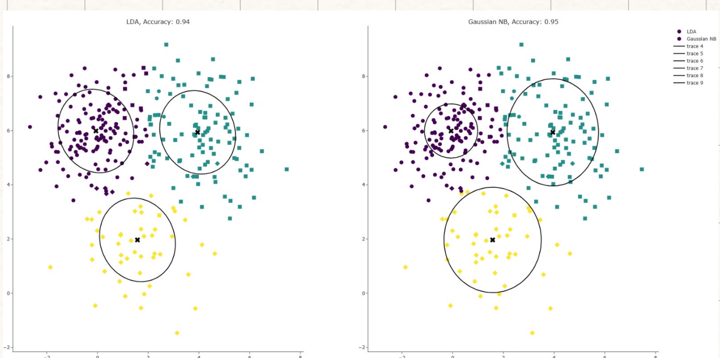


(1) אנחנו רואים לאחר שמונה איטרציות.
ה-loss יורד ויורד ומגיע לאפס לבסוף
שהאלגוריתם מצא ~~את~~ מפריד את hyperplane
המפריד את הנקודות בצורה מושלמת
של hyperplane ה מסווג בצורה מושלמת.
מה שמוביל ל- training loss=0 ניתן לראות
שה data הוא linearly seperatable, כי נמצא
hyperplane בלי loss

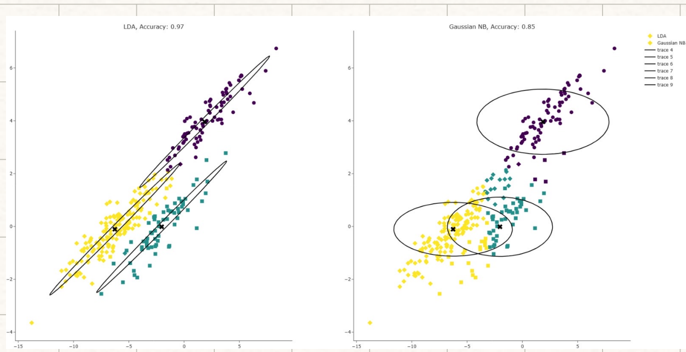


(2) האלגוריתם לא מתכנס כי ה-data, ה-seperable אין
hyperplane שמפריד את הdata לשני קבוצות
כיוון כי ה- inseperatable לפי הגדרה בצורה מושלמת.
בכל שלב ניתן לראות כי יש בו נקודות ה
hyperplane מסווג לא נכון אם שני חלקים מכל אחד
ה-loss לא יורד לאפס לכן נשאר בין ערכים,
בערך כמו כסוג מקרי.

(3.2)



(a) בגלל שמקבלים את ההנחה אחד
כשאני גוזל' הגאוסיאנית מספק לנו
תמונה כמות כמו רואים בגאוסיאן)
ובגלל שאנחנו הגאוס מפרד כמה האנונים
גרסאה מספק לנו תמונה ברורה
באמצעות ה-LDA.



② בגלל שהדאטה לא מפוזר כעיגול
ולכן נראה שהפיצ'רים תלויים אחד בשני
והקלסיפיירים הם לא שונים, ה- LDA
יכול ללמוד מלאה יותר כיוונים (לציפצים)
בגלל שהוא מתאר קווריאנס משותף
לכל הקלאסים. ואז לגבי ה GB -
שהקווריאנס אלכסונית אצלו מוגבל יותר.
בגלל כך גם לא חופף דומה הוא מגוון
כיוון שהם נוגע בכיוון שלוש ופנות
אליפטיות. גם LDA מצליח יותר
מאפיינת שגיאה למולה.