חלק תיאורטי

שאלה 1:

אנחנו רוצים להביא למינימום את w,b כלומר את $v=\binom{w}{b}$

$$\operatorname*{argmin}_{w,b} \|v\|^2 = \operatorname*{argmin}_{w,b} \left\|\binom{w}{b}\right\|^2 = \operatorname*{argmin}_{w,b} \binom{w}{b}^T \binom{w}{b} =$$

$$= \operatorname*{argmin}_{w,b} \binom{w}{b}^T I \binom{w}{b} = \operatorname*{argmin}_{w,b} \frac{1}{2}\binom{w}{b}^T 2I \binom{w}{b} =$$

$$= \operatorname*{argmin}_{w,b} \left(\frac{1}{2} v^T Q I + a^T v\right)$$

עבור $\begin{cases} a=0 \\ Q=2I \end{cases}$ ו-

ובעזרת תכנון ריבועי שניתן להביא את הביטוי למינימום.

אילוצים:

לכל i $\iff$ $y_i(\langle w,x_i\rangle + b) \geq 1$

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

$$-y(w^T X + b) \leq \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix} \iff$$

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

וזה בצורת האילוצים המקובלת כאשר $A = -y X y$, $d = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix}$

שאלה 2:

נראה שתחת האילוצים $\forall i\ y_i\langle w,x_i\rangle \ge 1-\xi_i$ ו $\xi_i \ge 0$

$\frac{1}{m}\sum_{i=1}^{m} \ell^{hinge}(y_i\langle w,x_i\rangle)$ שקול ל- $\frac{1}{m}\sum_{i=1}^{m} \xi_i$

- אם האילוצים מתקיימים אזי

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

ונקבל

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

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

כי אם לא כן - $\ell^{hinge}(y_i\langle w,x_i\rangle)=0$ ולא נסכום את

ξ_i המתאימים -

(ולכן נקבל שבסכום $< \xi_i$ ולכן המינימום של הביטוי $\frac{1}{m}\sum_{i=1}^{m}\xi_i$ קטן או שווה משל הביטוי $\frac{1}{m}\sum_{i=1}^{m}\ell^{hinge}(y_i\langle w,x_i\rangle)$)

שאלה 3:

נתון: $y\sim Mult(\pi)$

$$\forall j\in[d]\ \ x_j \mid y=k \sim N(\mu_{kj},\sigma_{kj}^2)$$

(a) $x\in\mathbb{R}$. $\leftarrow$ $x \mid y=k \sim N(\mu_k,\sigma_k^2)$

נרצה למצוא את π_k, μ_k, σ_k שממקסמים

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

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

log [illegible] $= \sum_{i=1}^{m} \log\left(N(x_i|\mu_{y_i}, \sigma_{y_i}^2)\, Mult(y_i|\pi)\right)$

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

מחוקי לוג

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

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

אם נסמן $g(\pi) = \sum_{k\in[K]} \pi_k - 1$ ולגרנז': $\mathcal{L} = \ell(\theta|X,Y) - \lambda g(\pi)$

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

ונקבל: $\pi_k = \frac{n_k}{\lambda} \xrightarrow[\lambda=m \text{ כי}]{} \pi_k = \frac{n_k}{m}$

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

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

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

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

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

b) פשוט נציב וקטור במקום סקלר ונקבל

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

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

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

שאלה 4:

נרצה למצוא את λ_k , π_k שמקיימים

$$\underset{k\in[K]}{\operatorname{argmax}} \frac{P_{X|Y=k}(x)\,P_Y(y)}{P_X(x)} = \underset{k\in[K]}{\operatorname{argmax}}\ \pi(k)\prod_{i=1}^{m}\frac{e^{-\lambda_k}(\lambda_k)^{x_i}}{x_i!}$$

ולפי עיקרון ה MLE

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

$$= \prod_{i=1}^{m} \pi_{y_i} \frac{e^{-\lambda_k}(\lambda_k)^{x_i}}{x_i!}$$

נפעיל log

$$\mathcal{L}(\theta|X,Y) = \sum_{i=1}^{m} \log(\pi(y_i)) + \log\left(\frac{(\lambda_{y_i})^{x_i}}{x_i!}\right) - \lambda_{y_i} =$$

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

$$= \sum_{k\in\{k\}} n_k \log(\pi_k) + x_i \log(\lambda_k) - \log(x_i!) - n_k\lambda_k$$

נגזור לפי כל אחד מהמשתנים:

$$\mathcal{L} = l(\theta|X,Y) - \lambda g(\pi)$$

נקבל,

$$\frac{\partial L}{\partial \pi_k} = \frac{n_k}{\pi_k} - \lambda \xrightarrow{\text{כשנשווה ל-0}} \pi_k = \frac{n_k}{\lambda}$$

כעת

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

לכן

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

נגזור לפי λ ונשווה ל-0

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

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

$$\Longrightarrow \quad \frac{1}{n_k} \sum_{k} x_i = \lambda_k$$

(b) נחליף את הסקלר - בוקטור ונקבל:

$$\pi_k = \frac{n_k}{m} \quad , \quad \lambda_k = \frac{1}{n_k} \sum_{k} x_i$$