IML 67577
תרגיל 3
יהונתן פרל
312203031

$\mathcal{X}=\mathbb{R}^d \quad \mathcal{Y}=\{\pm 1\}$

נניח התפלגות D על $\mathcal{X}\times\{\pm 1\}$ שהצפיפות שלפיה נתונה

בלתם p.d.f

(1)

$\forall x\in\mathcal{X} \quad h_D(x)\begin{cases}1 & P(y=1|x)\geq\frac{1}{2}\\ -1 & \text{o.w}\end{cases}$

$h_D=\underset{y\in\{\pm 1\}}{\operatorname{argmax}}\, P(x|y)P(y)$

ס.3

לפי נוסחת בייס: $P(y|x)P(x)=P(x|y)P(y)$ אזי

$\underset{y\in\{\pm 1\}}{\operatorname{argmax}}\, P(x|y)P(y) \underset{\text{מבייס}}{=} \underset{y\in\{\pm 1\}}{\operatorname{argmax}}\, P(y|x)P(x) = \underset{y\in\{\pm 1\}}{\operatorname{argmax}}\, P(y|x)P(x)$

כאשר המעבר האחרון נובע מכך ש- $0\leq P(x)$ והוא קבוע עבור כל

בחירת y. נשים לב ש- $(y=0|x)$ ו- $(y=1|x)$ מצביעים

משלימים לכן $P(y=1|x)\geq\frac{1}{2} \Rightarrow P(y=0|x)<\frac{1}{2}$

מקס'

$\underset{y\in\{\pm 1\}}{\operatorname{argmax}}\, P(x|y)P(y) = \begin{cases}y & P(y=1|x)\geq\frac{1}{2}\\ y & P(y=0|x)<\frac{1}{2}\end{cases} = \begin{cases}1 & P(y=1|x)\geq\frac{1}{2}\\ -1 & \text{o.w}\end{cases}$

$=h_D(x)$

מ.ש.ל

(2) נניח $\mathcal{X}=\mathbb{R}^d$ ו- $X|y\sim N(\mu_y,\Sigma)$ עבור וקטור תוחלת $\mu_y\in\mathbb{R}^d$

$f(x|y)=\frac{1}{\sqrt{(2\pi)^d\det(\Sigma)}}e^{-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)}$

נוכיח כי בהינתן μ_{+1},μ_{-1},Σ

אזי ה-classifier האופטימלי של בייס הוא

$h_D(x)=\underset{y\in\{\pm 1\}}{\operatorname{argmax}}\,\delta_y(x)$

כאשר δ_{+1},δ_{-1} הם פונק' $\mathbb{R}^d\to\mathbb{R}$ בלתי תלויות

$\delta_y(x)=x^T\Sigma^{-1}\mu_y-\frac{1}{2}\mu_y^T\Sigma^{-1}\mu_y+\ln Pr(y) \quad y\in\{\pm 1\}$

$$h_0(x) = \underset{y\in\{\pm 1\}}{\operatorname{argmax}} \delta_y(x) \overset{(1)}{=} \underset{y\in\{\pm 1\}}{\operatorname{argmax}} P(x|y)P(y)$$

$\ln$ — מונוטונית עולה, לכן מקסימום נשמר

$$\ln(h_0) = \underset{y\in\{\pm 1\}}{\operatorname{argmax}} \ln P(x|y)P(y) =$$

$$= \underset{y\in\{\pm 1\}}{\operatorname{argmax}} \left(\ln(P(x|y)) + \ln(P(y))\right) =$$

$$= \underset{y\in\{\pm 1\}}{\operatorname{argmax}} \left(\ln\left(\underbrace{\frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}}}_{\text{const}}\right) e^{-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)} + \ln(P(y))\right) =$$

$$= \underset{y\in\{\pm 1\}}{\operatorname{argmax}} \left(-\tfrac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y) + \ln(P(y))\right)$$

$$= \text{-''-} \left(-\tfrac{1}{2}x^T\Sigma^{-1}x + \tfrac{1}{2}\mu_y^T\Sigma^{-1}x + \tfrac{1}{2}x^T\Sigma^{-1}\mu_y - \tfrac{1}{2}\mu_y^T\Sigma^{-1}\mu_y + \ln P(y)\right)$$

$$= \text{-''-} \left(x^T\Sigma^{-1}\mu_y - \tfrac{1}{2}\mu_y^T\Sigma^{-1}\mu_y + \ln P(y)\right) =$$

$$= \underset{y\in\{\pm 1\}}{\operatorname{argmax}} \delta_y(x)$$

לינארי

$S = (x_1, y_1), \ldots, (x_m, y_m)$ (3)

אומדים:

$$\hat{P}(y) = \frac{1}{m}\sum_{i=1}^{m} \delta(y_i = y)$$

$$\hat{\Sigma} = \frac{1}{m}\sum_{y\in\{\pm 1\}}\sum_{i=1}^{m} (x_i - \mu_y)(x_i - \mu_y)^T$$

$$\hat{\mu}_y = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_y)(x_i - \mu_y)^T$$

(4) שתי השגיאות שמופיעות אצלנו הן:

1) סיווג הודעה אמיתית כספאם

2) סיווג הודעה ספאם כאמיתית.

סיווג הודעה אמיתית כספאם היא שגיאה חמורה לא נרצה שתקרה ולכן נרצה שהשגיאה תהיה ה false positive, כאשר נסווג את ההודעות לפי (spam-positive $y=1$), (real-negative $y=-1$)

(5)

QP: $\underset{v\in\mathbb{R}^n}{\arg\min}\left(\frac{1}{2}v^TQv+a^Tv\right)$ s.t $Av\leq d$,

$Q\in\mathbb{R}^{n\times n}, A\in\mathbb{R}^{m\times m}, d\in\mathbb{R}^m$ - fixed

נכתוב את הבעיה לפי צורת Hard-SVM:

$$\underset{(w,b)}{\arg\min}\|w\|^2 \text{ s.t } \forall i, y_i(\langle w,x_i\rangle+b)\geq 1$$

נציג:

$$\arg\min\begin{pmatrix}w\\b\end{pmatrix}^T I\begin{pmatrix}w\\b\end{pmatrix} \text{ s.t } \begin{pmatrix}y_1x_1 & 1\\ \vdots & \vdots\\ y_mx_m & 1\end{pmatrix}\begin{pmatrix}w\\b\end{pmatrix}\geq\underbrace{\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}}_{\text{size } m}$$

$$=\arg\min\frac{1}{2}\begin{pmatrix}w\\b\end{pmatrix}^T[2I]\begin{pmatrix}w\\b\end{pmatrix} \text{ s.t } -\begin{bmatrix}y_1x_1 & 1\\ \vdots & \\ y_mx_m & 1\end{bmatrix}\begin{pmatrix}w\\b\end{pmatrix}\leq-\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}$$

קיבלנו משוואה מהצורה שרצינו עם:

$$A=-\begin{bmatrix}y_1x_1\\ \vdots\\ y_mx_m\end{bmatrix} \quad v=\begin{pmatrix}w\\b\end{pmatrix} \quad a=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix} \quad Q=2I$$

Soft-SVM Problem:

$$\underset{w,\{\xi_i\}}{\operatorname{argmin}} \frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_{i=1}^{m} \xi_i \quad \text{s.t} \quad \forall i, y_i\langle w, x_i\rangle \geq 1 - \xi_i \text{ and } \xi_i \geq 0$$

נראה כי שקולה לבעיה:

$$\underset{w}{\operatorname{argmin}} \frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_{i=1}^{m} \ell^{hinge}(y_i\langle w, x_i\rangle),$$

$$\ell^{hinge}(a) = \max\{0, 1-a\}$$

מכך ש ... קטן ומתקיים מינימום $\xi_i \geq 1 - y_i\langle w, x_i\rangle$:

$$\xi_i \geq \ell^{hinge}(y_i\langle w, x_i\rangle)$$

נניח בשלילה $\xi_i \neq \ell^{hinge}(y_i\langle w, x_i\rangle)$ $\Leftarrow$ $\xi_i > \ell^{hinge}(y_i\langle w, x_i\rangle) \geq 0$

מונוטונית ℓ^{hinge}

$$\ell^{hinge}(\xi_i) \geq \ell^{hinge}(1 - y_i\langle x_i, w\rangle)$$

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

$\Downarrow$

סתירה למינימליות הפונק'

$$\Downarrow \quad \xi_i = \ell^{hinge}(y_i\langle w, x_i\rangle)$$

ומכאן נקבל שקילות הבעיות. מ.ש.ל.