$\forall x\in X \quad h_D(x)=\begin{cases} 1 & Pr(y=1|x)\geq\frac{1}{2} \\ -1 & \text{otherwise}\end{cases}$ Bayes Optimal and LDA (1

צריך להראות כי $h_D=\arg\max_{y\in\{-1,1\}} Pr(x|y)p(y)$.

(i) $Pr(x|y)p(y)=Pr(y|x)p(x)$ של שוויון זה כי זה כלל בייס.

נשים לב שאם $Pr(y=1|x)\geq\frac{1}{2}$ אז $Pr(y=1|x)=\max\{Pr(y|x)\}$ מאחר

ו- $\sum_y P(y|x)=1$ ו- $y\in\{-1,1\}$. באותו אופן ל- $P(y=-1,x)$.

$$\Rightarrow h_D(x)=\arg\max_{y\in\{-1,1\}} Pr(y|x)\overset{(*)}{=}\arg\max_{y\in\{-1,1\}} Pr(y|x)p(x)=$$

$$\overset{(i)}{=}\arg\max_{y\in\{-1,1\}} Pr(x|y)pr(y)$$

(*) x הוא קבוע ל- h_D ולכן הכפלה ב- $Pr(x)$ ב- $\{p(y|x)\}_y$ משמרת את

ה האיברים באותו וקטור.

(2 מונוטוניות של הפונק' $\ln$ אנו יודעים כי

$$\arg\max_y g(y)=\arg\max_y \ln(y)$$

$$\Rightarrow h_D(x)=\arg\max_y \ln(pr(x|y)p(y))=\arg\max_y\{\ln(pr(x|y))+\ln(pr(y))\}$$

$$=\arg\max_y\Big\{\ln\Big(\underbrace{\frac{1}{\sqrt{(2\pi)^d\det(\Sigma)}}}_{A}e^{-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)}\Big)+\ln(pr(y))\Big\}$$

$$=\arg\max_y\{\ln(A)-\frac{1}{2}(x-\mu_y)^T\Sigma^{-1}(x-\mu_y)+\ln(pr(y))\}=$$

$$=\arg\max_y\{\ln(A)-\frac{1}{2}[x^Tx-\mu_y^T\Sigma^{-1}x+x^T\Sigma^{-1}\mu_y+\mu_y^T\mu_y]+\ln(pr(y))\}$$

(שים לב כי $\ln(A)$ ו- $-\frac{1}{2}x^Tx$ אינם תלויים ב- y.
ולכן

$$= \arg\max_y \left\{ -\frac{1}{2}\left[-\mu_y^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu_y + \mu_y^T\Sigma^{-1}\mu_y\right] + \ln(\Pr(y))\right\}$$

קודם מאחר ו- Σ^{-1} הוא הרמטי ~~?~~ אז - ~~$\mu_y^T\Sigma^{-1}x = \Sigma^{-1}\mu_y^Tx = \Sigma^{-1}x^T\mu_y$~~ $^{(*)}$

$$\Rightarrow h_D(x) = \arg\max_y \left\{ x^T\Sigma^{-1}\mu_y - \frac{1}{2}\mu_y^T\Sigma^{-1}\mu_y + \ln(\Pr(y))\right\} =$$

$$= \arg\max_y \delta_y(x)$$

$$\mu_y^T\Sigma^{-1}x = \langle\mu_y^T, \Sigma^{-1}x\rangle \underset{\uparrow}{=} \langle\Sigma^{-1}\mu_y^T, x\rangle = \langle\Sigma^{-1}x^T, \mu_y\rangle \quad (*)$$

Σ^{-1} סימטרית
(מטריצת cov).

~~$\tilde{\Pr}(y) =$~~

3) את $\Pr(y)$ נעריך לפי הממוצע האמפירי:

$$\underset{i\in\{1,-1\}}{\tilde{\Pr}(y=i)} = \frac{1}{m}\sum_{(x_j,y_j)} 1(\{y_j = i\})$$

הקורדינטה ה-i של μ_1.

$$\tilde{\mu}_1^i = \underset{(x_j,y_j)}{\overset{\text{ממוצע}}{\mathrm{Mean}}}\left\{x_j^i : 1\{y_j = 1\}\right\}$$

הווקטור μ_1 - ממוצע כל קורדינטה
כל הדגימות עבורם $y=1$.

באופן דומה נחשב גם ל- $\tilde{\mu}_{-1}^i$.

את מטריצת ה-cov נעריך כרגיל: (לפי הגדרה קודמת).

~~cov~~

spam 4

נראה שההבדל על סמך הגודל (נרמלי) —— יהיה Type-2 error

ולכן

{not-spam, true} - False positive

{spam, false} - False negative

SVM - Formulation 5

ניבחר את $Q = 2I$ ואת $\underline{a} = \underline{0}$ ונראה כי:

$$\forall_j \quad y_j \cancel{\text{sign}} \cdot (\langle w, x_j \rangle + b) \geq 1$$

$$\Longleftrightarrow \quad y_j \langle w, x_j \rangle \geq 1 - b y_j$$

$$\Rightarrow \begin{bmatrix} - y_1 - \\ - y_2 - \\ \vdots \\ - y_m - \end{bmatrix} \begin{bmatrix} - w - \\ - w - \\ \vdots \\ - w - \end{bmatrix} \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_m \\ | & | & & | \end{bmatrix} \geq \begin{bmatrix} 1 - b y_1 \\ \vdots \\ 1 - b y_m \end{bmatrix}$$

$$\Rightarrow \underbrace{- \begin{bmatrix} - y_1 - \\ \vdots \\ - y_j - \end{bmatrix} \begin{bmatrix} - w - \\ \vdots \\ - w - \end{bmatrix}}_{A = -y^T \mathbb{1}_{m \times m} \, w} \underbrace{\begin{bmatrix} | & & | \\ x_1 & \cdots & x_m \\ | & & | \end{bmatrix}}_{v} \leq \underbrace{\begin{bmatrix} 1 - b y_1 \\ \vdots \\ 1 - b y_m \end{bmatrix}}_{d}$$

מטריצה

של 1-ים

Soft-SVM

צריך להוכיח שקולות סעיף

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

$$\arg\min_{w} \frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_{i=1}^{m} \ell^{hinge}(y_i\langle w,x_i\rangle) \qquad -1$$

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

(בלי נוש ההגבלה).

$$\min_{w,\xi_i} \frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_{i=1}^{m}\xi_i = \min_{w,\xi_i}\left\{\frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum_{i=1}^{m}\min \xi_i\right\} =$$

$$= \min_{w}\left\{\frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum^{m}\begin{cases}1-y_i\langle w,x_i\rangle & 1-y_i\langle w,x_i\rangle \geq 0 \text{ אם} \\ 0 & \text{אחרת}\end{cases}\right\}$$

$$= \min_{w}\left\{\frac{\lambda}{2}\|w\|^2 + \frac{1}{m}\sum^{m}\max\{0, 1-y_i\langle w,x_i\rangle\}\right\}$$

Scanned with CamScanner