Naive Bayes

*0과 1을 예측하는 생성 알고리즘
생성 model $p(x|y)$ $p(y)$

$p(x|y) = \prod_{j=1} p(x_j|y)$, 파라미터

$p(y=1) = \phi_y$
$p(x_j=1|y=0) = \phi_{j|y=0}$
$p(x_j=1|y=1) = \phi_{j|y=1}$

Prediction time

$$* \ p(y=1|x) = \frac{p(x|y=1)\,p(y=1)}{p(x|y=1)\,p(y=1) + p(x|y=0)\,p(y=0)}$$

$\prod_{i=1} p(x_i|y)$

문제점이 1가 처음 등장하는 경우 $p(y=1|x)=0$으로 측정

문제 해결하기 위해서

Laplace Smoothing

ex) 4연패일 경우 다음 경기 승리 확률 : $\frac{0}{0+4} = 0$

Laplace smoothing은 각 경우에 '+1' 해줌 $\frac{1}{1+5} = \frac{1}{6}$

↳ 이 방법이 실제로도 최적임!

더 일반적으로, $X \in \{1, \dots k\}$

Estimate $P(x=j) = \frac{\sum_{j=1}^{m} 1\{x^{(i)}=j\} +1}{M +1}$

$$\phi_{j|y=0} = \frac{\sum_{i=1}^{m} 1\{x_j^{(i)}=1,\ y^{(i)}=0\} +1}{\sum_{i=1}^{m} 1\{y^{(i)}=0\} +2}$$

특성이 다항식 값인 경우, 변수를 이산화해주면 된다. 보통 10개.

ex) $x = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ a, aardvark, buy ← 800, drug ← 1600, now ← 6200 $\quad x_i \in \{0,1\}$

$\xrightarrow{\text{New Representation}}$ $x \in \begin{bmatrix} 1600 \\ 800 \\ 1600 \\ 6200 \end{bmatrix}$ 사전 등장 index $\in \mathbb{R}^n$

$x_j \in \{1, 2, \cdots, 10000\}$

n_i = len of email

'Drugs! Buy Drugs now'

* Drug가 2번 등장한다는 정보 손실.

Multivariate Bernoulli event model ⟶ Multinomial event model

→ $P(x, y) = P(x|y) \cdot P(y)$

가정 $= \prod_{j=1}^{n} p(x_j|y) \cdot p(y)$

파라미터 $\begin{cases} \phi_y = P(y=1) \\ \phi_{k|y=0} = p(x_j = k | y=0) \end{cases}$

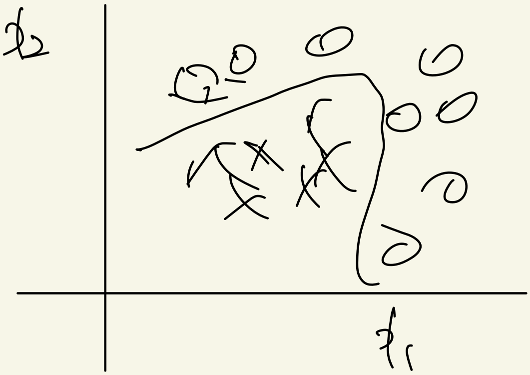
$$\phi_{k|y=0} = \frac{\sum_{i=1}^{m} 1\{y^{(i)}=0\} \sum_{j=1}^{n_i} 1\{x_j^{(i)}=k\} + 1}{\sum_{i=1}^{m} 1\{y^{(i)}=0\}\, n_i + 10{,}000}$$

(ㄴ) 스팸이 아닌 e-mail 중 'k' 단어가 포함된 e-mail은 몇분의 몇?

* 단어가 사전에 없으면 { (1) 그냥 버린다 (2) UNK 토큰으로 대체함

* 언제 사용? { (1) 계산 Easy (2) 구현이 쉬움 (3) GD 반복 X

SVM : support Vector Machine



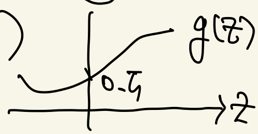
'커널'

※ 비선형적인 결정경계를 찾는 알고리즘 / 상대적으로 소수의 매개변수
수가 많지 않다.

→ $\boxed{\text{Optimal margin classifier}}$ - 분리가능한 경우만 다룸.

정의) Functional Margin : 얼마나 잘 분리하는지.

$h_\theta(x) = g(\theta^T x)$

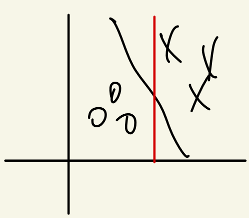


$$\begin{cases} 1 & \text{if } \theta^T x \geq 0 \\ & (h_\theta(x) = g(\theta^T x) \geq 0.5) \\ 0 & \text{otherwise} \end{cases}$$

if $y^{(i)} = 1$, hope that $\theta^T x^{(i)} \gg 0$

$y^{(i)} = 0$, hope that $\theta^T x^{(i)} \ll 0$

정의) Geometric Margin



'검정' 선이 더 좋음! → 기하학적으로 더 견고함.

Notation

labels $y \in \{-1, +1\}$, have h output values in $\{-1, +1\}$

$$g(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

표기법 $h_{w,b}(x) = g(\underbrace{w^T x + b}_{\sum_{i=1}^{n} w_i x_i + b})$, Drop $x_0 = 1$

$$\begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \end{bmatrix} \begin{matrix} \} b \\ \\ \} w \\ \\ \end{matrix}$$

Functional Margin of hyperplane define by (w, b)

$\rightarrow$ $\hat{\gamma}^{(i)}$ → 가 양수든 커져야 함.

$$\hat{\gamma}^{(i)} = y^{(i)} (w^T x^{(i)} + b)$$

$$\begin{cases} \text{if } y^{(i)} = 1, & \text{want } w^T x^{(i)} + b \gg 0 \\ \text{if } y^{(i)} = -1, & \text{want } w^T x^{(i)} + b \ll 0 \end{cases}$$

Functional Margin with training set $\hat{\gamma} = \min_{i} \hat{\gamma}^{(i)}$

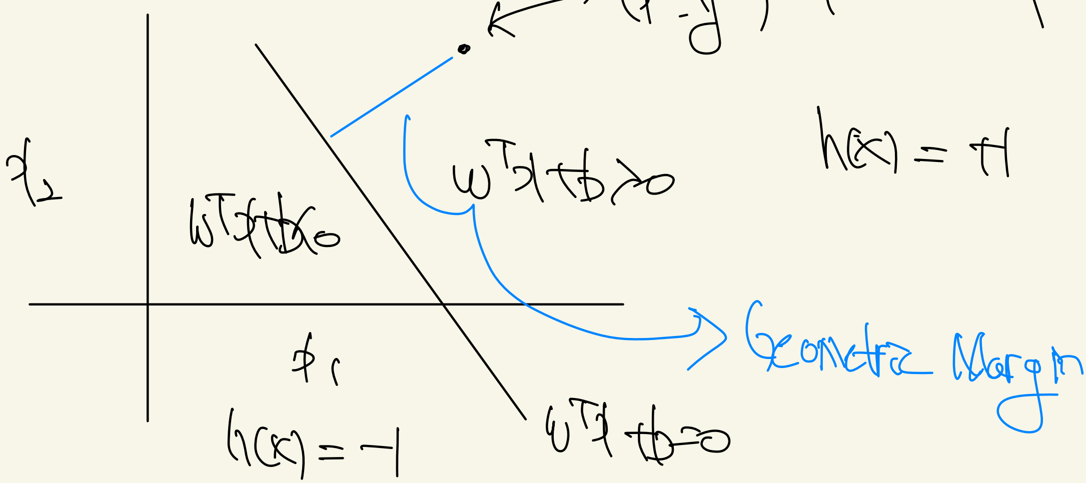
$i = 1, 2, \ldots, n$

∵ 훈련세트의 최악의 예를 얼마나 잘 수행?

↳ 매개변수 발산되기 방지를 위한 정규화

가정: $\|w\| = 1$, $(w, b) \rightarrow \left(\frac{w}{\|w\|}, \frac{b}{\|w\|} \right)$

Geometric Margin



$$\rightarrow \gamma^{(i)} = \frac{y^{(i)}(w^T x^{(i)} + b)}{\|w\|}$$ ('유클리드 공식')

$$\gamma \text{ 기하학적 Margin} = \frac{\text{함수적 Margin}}{\|w\|} \leftarrow \hat{\gamma}$$

In training set, $\gamma = \min_i \gamma^{(i)}$ (분류의 최소거리)

∴ Optimal margin classifier :

Max γ 하는 w, b 를 선택

Max γ

s.t. $\dfrac{y^{(i)}(w^T x^{(i)} + b)}{\|w\|} \geq \gamma$ $\quad i = 1, 2, \dots, m$

합업

$$\left\{ \begin{array}{l} = \min_{w,b} \|w\|^2 \\ \text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1 \end{array} \right.$$