SVM derive

Tuesday, 27 August BE 2562 13:02

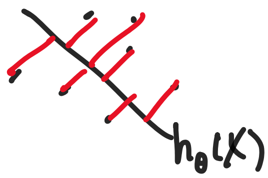
ทบทวน Logistic Regression / Linear Regress

Linear Regression

model : $\hat{y} = h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \dots + \theta_n x_n = \theta^T X$

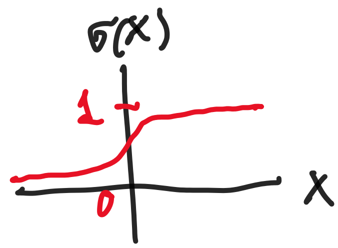
$[\theta_0\ \theta_1 \dots \theta_n] \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$

cost function : $J(\theta) = MSE(\theta) = \sum (y - \hat{y})^2$



$\theta_0 + \nabla_\theta J(\theta)\gamma$

Logistic Regression



model : $\hat{p} = h_\theta(x) = \sigma(\theta^T x)$

decision rule : $\hat{p} > 0.5 \rightarrow$ ทำนายเป็น +

$\hat{p} \leq 0.5 \rightarrow$ ทำนายเป็น −

train cost function $J(\theta) =$ ______

ทำไม $W^T X$

~~vector~~ $\overline{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ \vdots \end{bmatrix}$

$\hat{y} = w_1 x_1 + w_2 x_2 + \dots + w_n x_n$

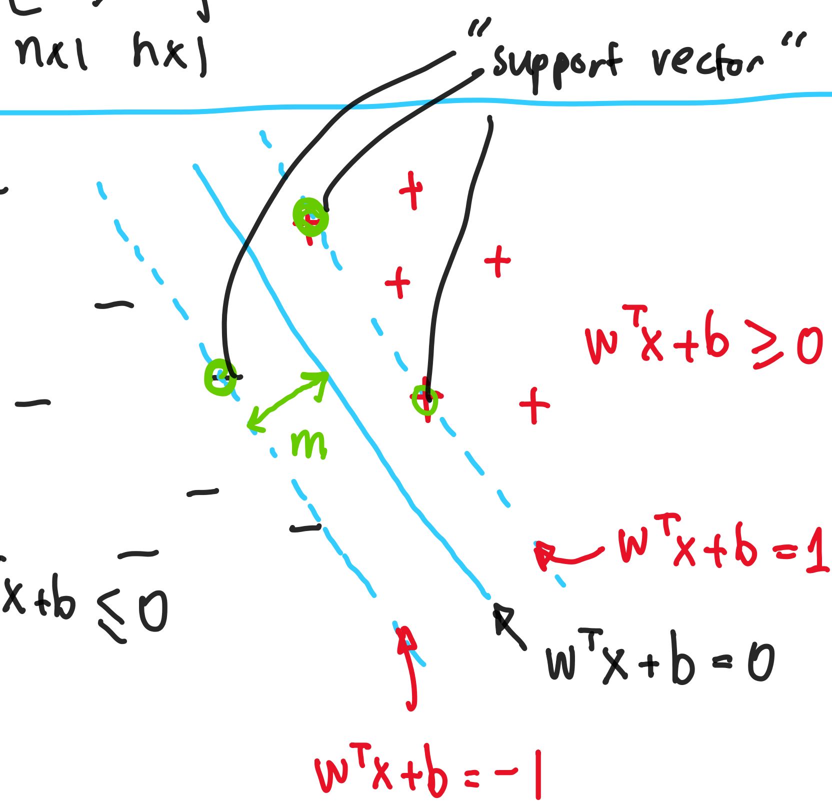
$\begin{bmatrix} w_n \end{bmatrix}$

vector $\bar{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ $\quad = \underbrace{\begin{bmatrix} w_1 & w_2 & \dots & w_n \end{bmatrix}}_{W^T} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} \ \end{bmatrix}}_{1 \times 1}$

$n \times 1$

$WX = ?$ $\underbrace{\begin{bmatrix} \ \\ \ \end{bmatrix}}_{n \times 1} \underbrace{\begin{bmatrix} \ \\ \ \end{bmatrix}}_{n \times 1}$

SVM Math



① How to compute Margin m?

<u>สัญลักษณ์</u> $\left.\begin{matrix} y_i \\ \bar{x}_i \end{matrix}\right\}$ ข้อมูลชุด train โดยที่

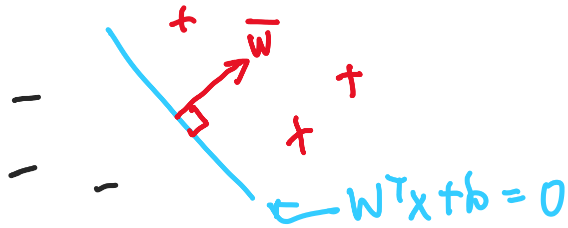
$\bar{x}_i = \begin{bmatrix} \vdots \\ \vdots \end{bmatrix}$ (ราคาซื้อ, จำนวนห้องนอน) เป็น feature vector ของ e.g. บ้านหลังที่ i

$y_i \in \{-1, +1\}$ เป็น label e.g. มะเร็ง/ไม่มะเร็ง

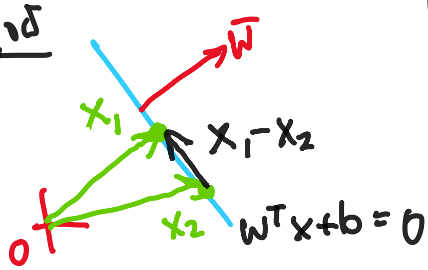
$\hat{y}_i = \begin{cases} +1 \ ; \ w^Tx_i + b \geq 0 \\ -1 \ ; \ w^Tx_i + b < 0 \end{cases}$

1.1) ต้องเข้าใจก่อนว่า

" $\bar{w}$ จะตั้งฉาก (orthogonal) กับ $w^Tx+b=0$ เสมอ "



พิสูจน์



จะได้

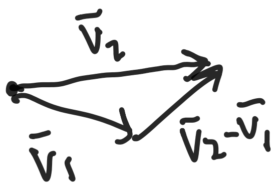
$$w^Tx_1+b=0 \quad \cdots(1)$$
$$w^Tx_2+b=0 \quad \cdots(2)$$

(1)-(2): $w^T(\bar{x}_1-\bar{x}_2)=0$

$$\bar{w}\cdot(\bar{x}_1-\bar{x}_2)=0$$

dot product

ทบทวน ลบ vector :



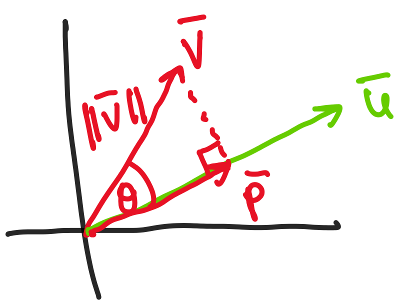
ถ้า dot product $\bar{u}\cdot\bar{v}=0$
แสดงว่า $\bar{u}$ กับ $\bar{v}$ จะตั้งฉากกัน

เพราะ: $\bar{u}\cdot\bar{v} = \|\bar{u}\|\|\bar{v}\|\cos\theta = 0$

$$\therefore \theta = \frac{\pi}{2} \ (90^\circ)$$

1.2) ทวน projection

จะหา projection ของ $\bar{v}$ ฉายลงบน $\bar{u}$



$$\bar{p} = \underbrace{\|\bar{v}\|\cos\theta}_{\text{ขนาด } \bar{p}} \cdot \underbrace{\frac{\bar{u}}{\|\bar{u}\|}}_{\text{ทิศ } \bar{p} \text{ (unit vector)}}$$

$$\bar{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\therefore \|\bar{p}\| = \|\bar{v}\|\cos\theta$$
$$= \|\bar{u}\|\|\bar{v}\|\cos\theta$$

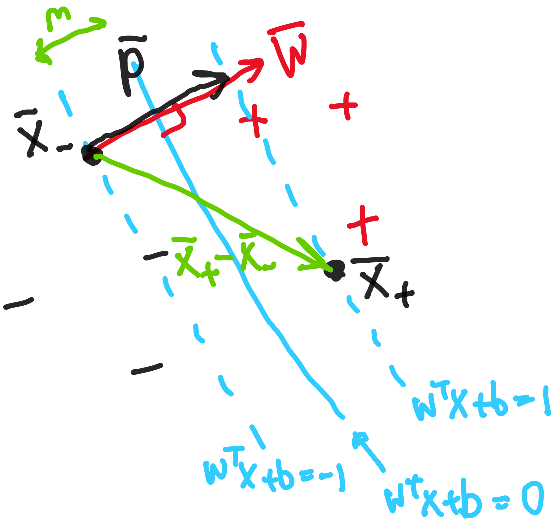
$$\|\bar{u}\|$$

$$\|\bar{v}\| = \sqrt{\bar{v}\cdot\bar{v}}$$

$$= \sqrt{[1\,0\,1]\begin{bmatrix}1\\0\\1\end{bmatrix}}$$

$$= \frac{\bar{u}\cdot\bar{v}}{\|\bar{u}\|}$$

1.3) ทำการ project $\bar{X}_+ - \bar{X}_-$ ลงบน $\bar{w}$ เพื่อหา Margin



$$\|\bar{p}\| = 2m$$

$$\therefore \quad 2m = \frac{\bar{w}\cdot(\bar{x}_+ - \bar{x}_-)}{\|\bar{w}\|}$$

$$m = \frac{1}{2}\frac{\bar{w}\cdot(\bar{x}_+ - \bar{x}_-)}{\|\bar{w}\|} \quad \ldots\ldots (4)$$

1.4) ใช้คุณสมบัติของ X_+, X_-

$$w^T x_+ + b = 1 \Rightarrow w^T x_+ = 1 - b$$

$$w^T x_- + b = -1 \Rightarrow w^T x_- = -1 - b$$

$$(4): \quad m = \frac{1}{2}(\bar{x}_+ - \bar{x}_-)\cdot\frac{\bar{w}}{\|\bar{w}\|}$$

$$= \frac{1}{2\|\bar{w}\|}\left(w^T x_+ - w^T x_-\right)$$

$$= \frac{1}{2\|\bar{w}\|}\left((1-\cancel{b}) - (-1-\cancel{b})\right)$$

$$= \frac{1}{\|\bar{w}\|}$$

1.5) ได้ optimization goal : max margin

$$\max \frac{1}{\|\bar{w}\|}$$

ซึ่ง มีค่าเท่ากับ

goal
$$\min_{w,b} \frac{1}{2}\|\bar{w}\|^2$$

subject to : $\underbrace{y_i}_{\text{label}}\underbrace{(w^T x_i + b)}_{\text{ทำนาย}} \underbrace{\geq 1}_{\text{ต้องเครื่องหมายเดียวกัน + } x_i \text{ ต้องอยู่นอกถนน}}$; สำหรับทุก i

---

② แก้โจทย์ optimize โดยใช้ KKT condition (คล้ายๆ Lagrange Multiplier)

แปลงโจทย์ให้เป็น dual form

$$L(\bar{w},b) = \underbrace{\frac{1}{2}\|\bar{w}\|^2}_{\text{goal}} - \underbrace{\sum \overset{?}{\alpha_i}\left[y_i(w^T x_i + b) - 1\right]}_{\text{subject to ...}}$$

partial derivation :

$$\frac{\partial L}{\partial \bar{w}} = \bar{w} - \sum \alpha_i y_i x_i = 0 \Rightarrow \boxed{\bar{w} = \sum \alpha_i y_i x_i}$$ solution

$$\frac{\partial L}{\partial b} = -\sum \alpha_i y_i = 0 \Rightarrow \boxed{\sum \alpha_i y_i = 0}$$

ย้าย

เอาไปแทนใน $L(w,b)$: $L(\bar{\alpha}) = \frac{1}{2}\left(\sum_i \alpha_i y_i x_i\right)\left(\sum_j \alpha_j y_j x_j\right) - \sum\left[\alpha_i \ldots\right]$

$$\therefore\ L(\bar{\alpha}) = \sum \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i \alpha_j y_i y_j \underbrace{\bar{x}_i \cdot \bar{x}_j}$$

dot product ของข้อมูลทุกคู่

dual problem : $\max_{\alpha} L(\bar{\alpha})$

subject to $\sum \alpha_i y_i = 0$, $\alpha_i \geq 0$

แล้วก็ใช้ Quadratic Programming Solver ได้

---

③ เนื่องจากเราได้ว่า $\bar{w} = \sum \alpha_i y_i x_i$

จะได้ว่า decision rule

$$\hat{y} = \begin{cases} +1 ; & \overbrace{\sum \alpha_i y_i \bar{x}_i}^{w^T} \cdot \bar{x} + b \geq 0 \\ -1 ; & \sum \alpha_i y_i \underbrace{\bar{x}_i \cdot \bar{x}} + b \leq 0 \end{cases}$$

dot product ของข้อมูลกับ $\bar{x}$

---

④ ถ้าเราอยากได้ SVM ที่ non-linear

$$\hat{y} = \begin{cases} +1 ; & \ldots\ldots \\ -1 ; & \sum \alpha_i y_i \underbrace{\phi(\bar{x}_i)\cdot\phi(\bar{x})}_{\text{expensive}} + b \leq 0 \end{cases}$$

$$\sum \alpha_i y_i \underbrace{\phi(\bar{x}_i)\phi(\bar{x})}_{\text{expensive transform}} \xrightarrow[\text{trick}]{\text{kernel}} \sum \alpha_i y_i \underbrace{K(\bar{x}_i, \bar{x})}_{\text{cheap}}$$

เลือก Kernel function

$$\boxed{K(a,b) = \phi(a)\phi(b)}$$