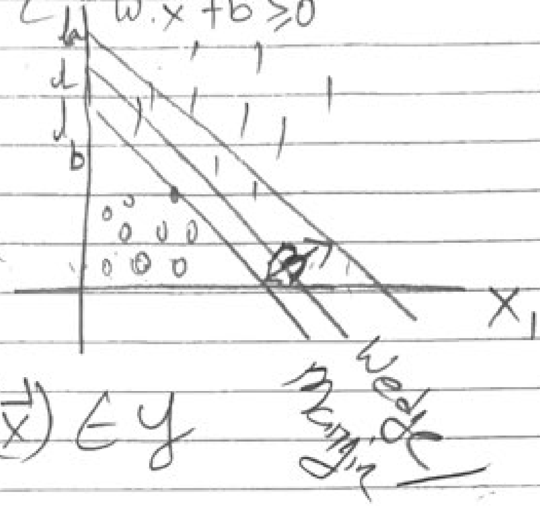
Lecture : 5

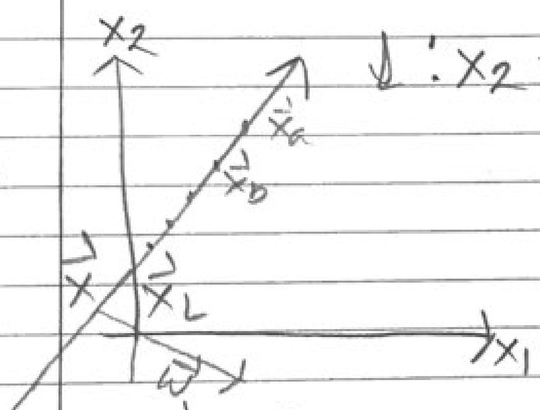
$$\mathcal{H} = \{ 1\!\!1_{\vec{w}\cdot\vec{x}+b \geq 0} : \vec{w} \in \mathbb{R}^p, b \in \mathbb{R} \}$$

$\mathbb{D}$

$p=2$



$y(\vec{x}) \in \mathcal{Y}$



$$\ell : x_2 = 2x_1 + 3 \Rightarrow \ell : c(2x_1 - x_2 + 3) = 0$$
$$c \in \mathbb{R}$$

$$\Rightarrow \ell \ \underbrace{\begin{bmatrix} 2 \\ -1 \end{bmatrix}}_{\vec{w}} \cdot \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\vec{x}} \underbrace{- 3}_{b} = 0$$

$$\ell \cdot c(\vec{w}\cdot\vec{x} + b) = 0$$

Hesse Normal Form

$\vec{w} \perp \ell$

normal vector

$\vec{x}_\perp$: vector $\perp \ell$
from the origin to ℓ.

Let $\|\vec{w}\| := \sqrt{\sum_{j=1}^{p} x_j^2}$ Length of $\vec{w}$

$$\vec{w}_0 = \frac{\vec{w}}{\|\vec{w}\|}$$

unit direct vector

$\vec{w} \cdot \vec{x}_\perp - b = 0$

$\vec{x}_\perp = \alpha \vec{w}_0$

$\vec{w} \cdot \alpha \vec{w}_0 - b = 0$

$\alpha \left(\frac{\vec{w} \cdot \vec{w}}{\|w\|} \right) - b = 0$

$\Rightarrow \alpha \frac{\|w\|^2}{\|w\|} - b = 0$

$\Rightarrow \alpha \|w\| = b$

$\Rightarrow \alpha = \frac{b}{\|w\|}$

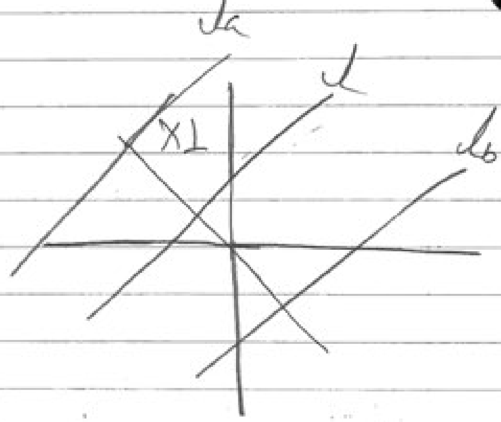
$l_b : \vec{w} \cdot \vec{x} - b = 0$

$\Rightarrow w_1 x_1 + w_2 x_2 - b = 0 \quad \delta > 0$

$\Rightarrow w_2 x_2 = -w_1 x_1 + b$

$\Rightarrow x_2 = \frac{-w_1}{w_2} x_1 + \frac{b}{w_2} \left(+ \frac{\delta}{w_2} \right)$

$l_a : x_2 = \frac{-w_1}{w_2} x_2 + \frac{b}{w_2} \left(- \frac{\delta}{w_2} \right)$



$l : \vec{w} \cdot \vec{x} - b = 0$

$l_a : \vec{w} \cdot \vec{x} - (b + \delta) = 0$

$l_b : \vec{w} \cdot \vec{x} - (b - \delta) = 0$

$\alpha_a = \frac{b + \delta}{\|w\|} = \frac{b + 1}{\|w\|}$

$\alpha_b = \frac{b - \delta}{\|w\|} = \frac{b - 1}{\|w\|}$

$m = \alpha_a - \alpha_b = \frac{b + \delta}{\|w\|} - \frac{b - \delta}{\|w\|}$

$m = \frac{2\delta}{\|w\|}$ $\boxed{\text{Let } \delta = 1}$ $\Rightarrow m = \frac{2}{\|w\|}$

(4) $\forall y_i = 1 \quad \vec{w}\cdot\vec{x}_i - (b+1) \geq 0$

$\Rightarrow \vec{w}\cdot\vec{x}_i - b \geq +1$

$(y_i - 1/2)(\vec{w}\cdot\vec{x}_i - b) \geq (y_i - \frac{1}{2}) = \frac{1}{2}$

$\forall y_i = 0 \quad \vec{w}\cdot\vec{x}_i - (b-1) \leq 0$

$\Rightarrow \vec{w}\cdot\vec{x}_i - b \leq -1$

$-(y_i - \frac{1}{2})(\vec{w}\cdot\vec{x}_i - b) \leq y_i - \frac{1}{2} = -\frac{1}{2}$

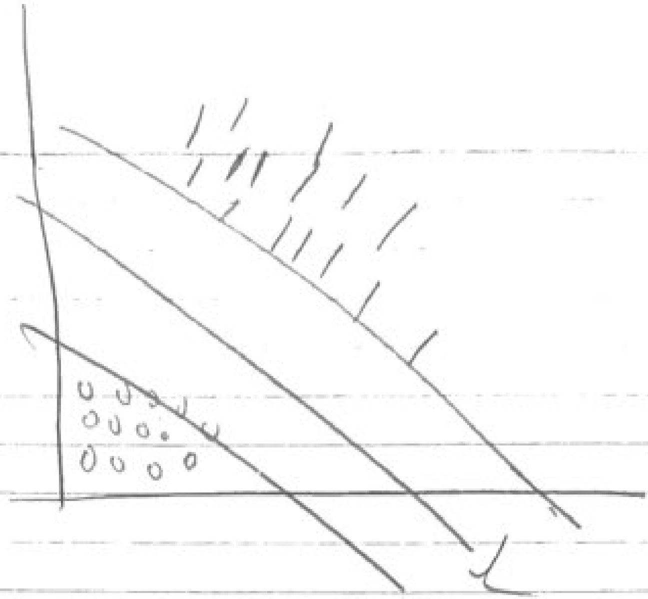
$(y_i - \frac{1}{2})(\vec{w}\cdot\vec{x}_i - b) \geq \frac{1}{2} \; ; \; \forall i$

$$D = \left\langle \underbrace{\begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_n \end{bmatrix}}_{X}, \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\rangle$$

<u>Algorithm</u>: maximize $\frac{2}{\|w\|}$ s.t.

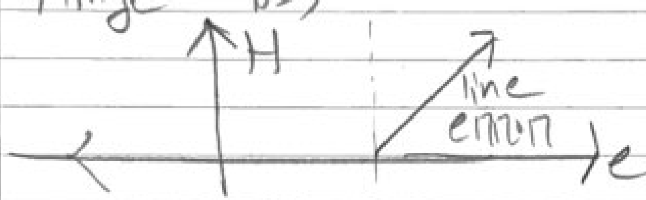
$\forall i \; (y_i - \frac{1}{2})(\vec{w}\cdot\vec{x} - b) \geq \frac{1}{2}$

$\vec{w}^*, b^* = \underset{\vec{w}, \vec{b}}{\arg\max} \left\{ \frac{2}{\|w\|} \text{ s.t. } \angle \right\}$



$$H_i := \max\left\{0, \frac{1}{2} - \left(y_i - \frac{1}{2}\right)\left(\vec{w}\cdot\vec{x}_i - b\right)\right\}$$

Hinge loss



IF I are above the line by d

$$\left(y_i - \frac{1}{2}\right)\left(\vec{w}\cdot\vec{x}_i - b\right) = \frac{1}{2} + d$$

$$H_i = \max\left\{0, \frac{1}{2} - \left(\frac{1}{2} - d\right)\right\} - \max\{0, +d\} = d$$

$$SHE = \sum_{i=1}^{n} \max\left\{0, \frac{1}{2} - \left(y_i - \frac{1}{2}\right)\left(\vec{w}\cdot\vec{x}_i - b\right)\right\}$$

sum of Hinge error

$$A: \vec{w}^*, b^* = \underset{\vec{w}, b}{\arg\min}\{SHE\}$$

Vapnik (1963) !

$$A : \underset{\vec{w} \in \mathbb{R}^p, b \in \mathbb{R}}{\text{argmin}} \left\{ \underbrace{\frac{1}{n} \text{SHE}}_{\text{average hinge error}} + \underbrace{\lambda \|\vec{w}\|^2}_{\text{"maximize the}} \right\}$$

$$g(\vec{w}^*, b^*)$$

"minimize error distance"

What is λ?
- Hyperparameter (you specify)

$$g = A(\mathbb{D}, \mathcal{H}_i, \lambda)$$

λ controls the tradeoff between the two considerations:

λ High: margin more important

λ Low: error more important

Null Model where $\vec{x}$ infactor. $x = \text{NULL}$

$$g_0(\vec{x}) = \text{Mode}[\vec{y}]$$

$$\mathbb{D} = \langle \vec{y} \rangle$$

$\mathcal{Y} = \{A, B, C \ldots\}$

L levels

$\mathbb{D} = \langle x, \vec{y} \rangle$

Want g.

Null: $g(\vec{x}) = \text{Mode}[\vec{y}]$

$$\mathbb{D} = \left\{ \begin{array}{l} \vec{x}_1, \text{ Red} \\ \vec{x}_2, \text{ Blue} \\ \vec{x}_3, \text{ Blue} \\ \vdots \\ x_n, \text{ Green} \end{array} \right\}$$

$g = \mathcal{A}(\mathbb{D}, \mathbb{H})$

g: function that finds closest $\vec{x}_i$ and returns y_i.

Hyperparameter

$$g(\vec{x}^*) = y_i \text{ s.t. } i = \underset{i \in \{1,2,\ldots,n\}}{\text{argmin}} \{ d(\vec{x}_i, \vec{x}^*) \}$$

Nearest Neighbour.

usually $d = \|\vec{x}_i - \vec{x}^*\|^2 = \sum_{j=1}^{p} (x_{ij} - x_j^*)^2$

Classification Model:

K is a hyperparameter.

A: KNN ("K-nearest neighbour")

For $\vec{x}^*$ find $x_{i_{(1)}}, x_{i_{(2)}}, \ldots, x_{i_{(K)}}$

where $d(\vec{x}^*, \vec{x}_{i_{(K)}})$'s are the K smallest.

and let $\hat{y} = \text{Mode}[y_{i_{(1)}}, y_{i_{(2)}}, \ldots, y_{i_{(K)}}]$

* Let $y = \mathbb{R}$, i.e, a continuous response.

These model are called "Regression Models" for historical reason only.

Null $g_0(\vec{x}) = \bar{y}$

Regression Hypothesis Sets for p features

$$\mathcal{H} = \{\underbrace{\vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1}}_{\text{The set of all linear models}}\} \qquad \vec{x} = [1 \ \underset{\substack{\uparrow \\ x_1, \ldots, x_p}}{\vec{x}}]$$

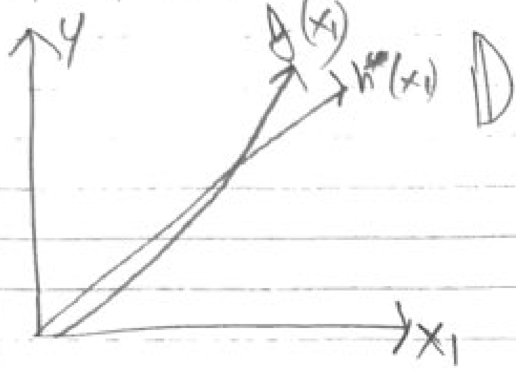
$$\vec{y} = \underset{\substack{\| \\ \hat{y}}}{g} + (h^* - g) + \underbrace{(f - h^*) + (t - f)}_{\varepsilon}$$

$$h^*(\vec{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

$Y = \beta_0 + \beta_1 x + \ldots + \beta_p x_p + \varepsilon$

Beta Possible w_j value

IF $P=1$



$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = SSE$$

Sum of squared Errors.

Least squared Regression

$$A \doteq \underset{\vec{w} \in \mathbb{R}^{p+1}}{\arg\min} \left\{ \sum_{i=1}^{n} \left(y_i - (w_0 + w_1 x_{1i} + \dots + w_p x_{pi})\right)^2 \right\}$$

$$SSE = \Sigma (y_i - w_0 - w_1 x_i)^2$$

$$= \Sigma (y_i^2 + w_0^2 + w_1^2 x_i^2 - 2 y_i w_0 - 2 y_i w_1 x_i + 2 w_0 w_1 x_i)$$

$$= (\Sigma y_i^2) + n w_0^2 + w_1^2 (\Sigma x_i^2) - 2 w_0 n \bar{y} - 2 w_1 \Sigma x_i y_i + 2 w_0 w_1 n \bar{x}$$

$$\frac{\partial}{\partial w_0} [SSE] \overset{set}{=} 0$$

$$2 n w_0 - 2 n \bar{y} + 2 n w_1 \bar{x} \overset{set}{=} 0$$

$$\Rightarrow w_0 - \bar{y} + w_1 \bar{x} = 0$$

$$\Rightarrow \widehat{w_0} = \bar{y} - w_1 \bar{x}$$

$$\frac{\partial}{\partial w_1}[SSE] \stackrel{set}{=} 0$$

$$2w_1 \Sigma x_i^2 - 2\Sigma x_i y_i + 2w_0 n\bar{x} \stackrel{set}{=} 0$$

$$\Rightarrow \Sigma x_i^2 w_1 = \Sigma x_i y_i + n\bar{x} w_0$$

$$\Rightarrow \Sigma x_i^2 w_1 = \Sigma x_i y_i + n\bar{x}(\bar{y} - w_1 \bar{x})$$

$$\Sigma x_i^2 w_1 = \Sigma x_i y_i + n\bar{x}\bar{y} - w_1 n\bar{x}^2$$

$$(\Sigma x_i^2 + n\bar{x}^2) w_1 = \Sigma x_i y_i + n\bar{x}\bar{y}$$

$$\Rightarrow \hat{w}_1 = \frac{\Sigma x_i y_i + n\bar{x}\bar{y}}{\Sigma x_i^2 + n\bar{x}}$$