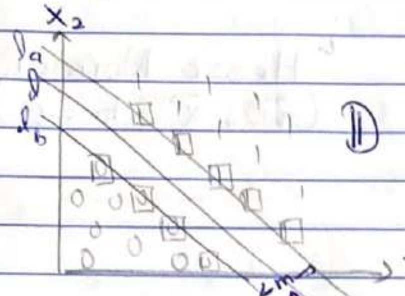
Lecture - 05 02/11/2020.

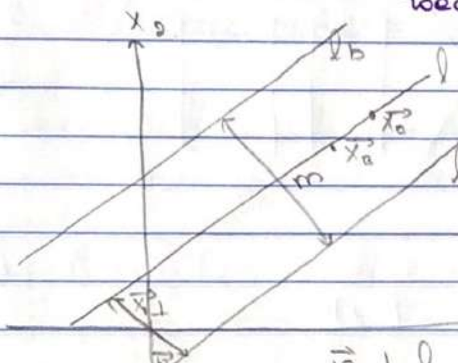
$y = \{0, 1\}$

$H = \left\{ \mathbb{1}_{\vec{w} \cdot \vec{x} + b \geq 0} ; \vec{w} \in \mathbb{R}^p, b \in \mathbb{R} \right\}$



$p = 2$

II $g(\vec{x}) \in y$



$l : x_2 = 2x_1 + 3$

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

$l \; \underbrace{\begin{bmatrix} 2 \\ -1 \end{bmatrix}}_{\vec{w}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\vec{x}} - \underbrace{(-3)}_{b.} = 0$

$l : c(\vec{w} \cdot \vec{x} + b) = 0.$

$\vec{w} \perp l$ normal vector Hesse Normal Form

Let $\|\vec{w}\| = \sqrt{\sum_{i=1}^{p} x_j^2}$ length of $\vec{w}$

$\vec{w_0} = \frac{\vec{w}}{\|\vec{w}\|}$ unit direction vector

$\vec{x_\perp}$; vector $\perp l$ from the origin to l

$\vec{w} \cdot \vec{x_1} - b = 0.$

$\vec{x_\perp} = \underbrace{x}\,\vec{w_0}$ distance from origin to l.

$\vec{W} \cdot \alpha \vec{W_0} - b = 0$.

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

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

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

$\implies w_1 x_1 + w_2 x_2 - b = 0$

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

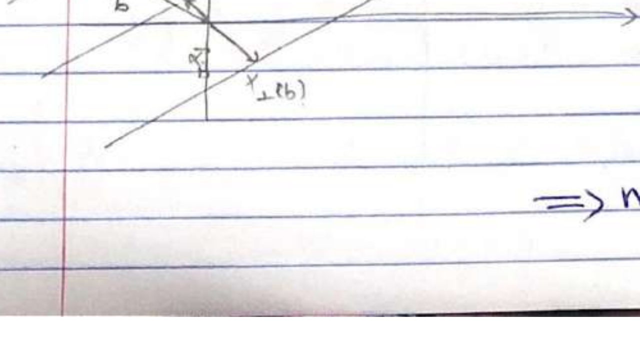
$\implies x_2 = - \frac{w_1}{w_2} x_1 + \frac{b}{w_2} \left(+ \frac{\delta}{w_2} \right), \quad \delta > 0$

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

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

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

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



$m = \alpha_a - \alpha_b$

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

$m = \frac{2\delta}{\|w\|}$

let $\delta = 1$

$\implies m = \frac{2}{\|w\|}$

$\forall y_i = 1 \qquad \vec{w}\cdot\vec{x_i} - (b+1) \geq 0.$

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

$(y_i - \tfrac{1}{2})(\vec{w}\cdot\vec{x_i} - b) \geq (y_i - \tfrac{1}{2}) = \tfrac{1}{2}$

$\forall y_i = 0 \qquad \vec{w}\cdot\vec{x_i} - (b-1) \leq 0$

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

$-(y_i - \tfrac{1}{2})(\vec{w}\cdot\vec{x_i} - b) \leq y_i - \tfrac{1}{2} = -\tfrac{1}{2}$

same

mult. by (-1)

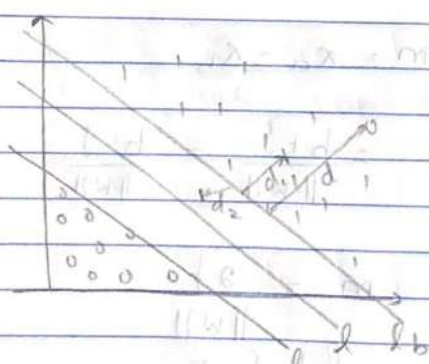
$\Rightarrow (y_i - \tfrac{1}{2})(\vec{w}\cdot\vec{x_i} - b) \geq \tfrac{1}{2}$

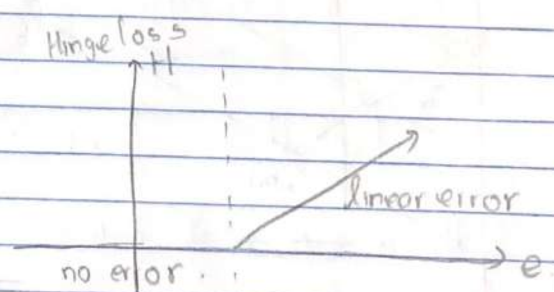
$$ⅱ = \left\langle \underbrace{\begin{bmatrix} \vec{x_1} \\ \vec{x_2} \\ \vdots \\ \vec{x_n} \end{bmatrix}}_{X} \overset{\forall i}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}} \right\rangle$$

A: maximize $\frac{2}{\|w\|}$ s.t. $\forall_i (y_i - \tfrac{1}{2})(\vec{w}\cdot\vec{x_i} - b) \geq \tfrac{1}{2}$

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

H_i : $\max [0, \tfrac{1}{2} - (y_i - \tfrac{1}{2})(\vec{w}\cdot\vec{x_i} - b)]$





If it above the line d.

above d.

$$\left\{ \begin{array}{l} (y_i - \tfrac{1}{2})(\vec{w} \cdot \vec{x_i} - b) = \tfrac{1}{2} + d \\ \\ H_i = \max[0, \tfrac{1}{2} - (\tfrac{1}{2} + d)] = \max[0, +d] = -d \end{array} \right.$$

below d.

$$\left\{ \begin{array}{l} (y_i - \tfrac{1}{2})(\vec{w} \cdot \vec{x_i} - b) = \tfrac{1}{2} - d \\ \\ H_i = \max[0, \tfrac{1}{2} - (\tfrac{1}{2} - d)] = \max[0, +d] = d \quad (\text{loss of } d) \end{array} \right.$$

sum of Hinge

$$SHE = \sum_{i=1}^{n} \max[0, \tfrac{1}{2} - (y_i - \tfrac{1}{2})(\vec{w} \cdot \vec{x_i} - b)]$$

$$A : \vec{w}^*, \vec{b}^* = \underset{\vec{w}, b}{\operatorname{argmin}} \{SHE\}$$

Vapnik (1963) suggested

$$A : \underset{\vec{w} \in \mathbb{R}^p,\, b \in \mathbb{R}}{\operatorname{argmin}} \{ \underbrace{\tfrac{1}{n} SHE}_{\text{average hinge error}} + \underbrace{\lambda \|\vec{w}\|^2}_{\text{says "maxmize the margin"}} \}$$

$g(\vec{w}^*, b^*)$
says "minimize error distance"

What is λ?
Hyper parameter.
(You specify)

$g = A(\mathbb{D}, \mathcal{H}; \lambda) \Rightarrow \lambda$ controls the trade off between the two considerations.

λ high : margin more
λ low : error more

Null model without $\vec{x}$ information $X = \text{NULL}$

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

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

$\mathcal{Y} = \{A, B, C, \ldots\}$ $\quad \mathbb{D} = \left\langle \begin{array}{l} \vec{x}_1, \text{Red} \\ \vec{x}_2, \text{Blue} \\ \vec{x}_3, \text{Blue} \\ \vdots \\ \vec{x}_n, \text{Green} \end{array} \right\rangle$

L levels $\quad \mathbb{D} = \langle X, \vec{y} \rangle$

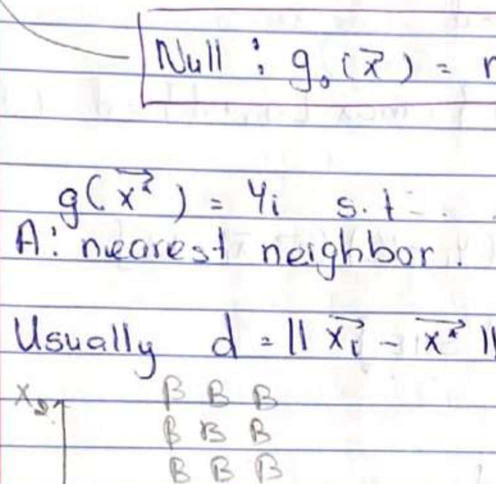
want $g \Rightarrow g(\vec{x}_*) \Rightarrow g$: function that finds closest $\vec{x}_i$ and return y_i

$\boxed{\text{Null}: g_0(\vec{x}) = \text{mode}[\vec{y}]}$

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

A: nearest neighbor.

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



Classification model

A: KNN ("K - nearest neighbors") (K is hyperparameter)
For $\vec{x}^*$ find $x_{i_{(1)}}, x_{i_{(2)}}, \ldots, x_{i_{(k)}}$
whose $d(\vec{x}^*, \vec{x}_{i_{(k)}})$'s are the k smallest and,

let $\hat{y} = \text{mode}[y_{i(1)}, y_{i(2)}, \ldots, y_{i(k)}]$

Let $\mathcal{Y} = \mathbb{R}$. i.e. a continuous response. These models are called "regression" models for historical reasons only.

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

Regression Hypothesis for p features

$$H = \{ \underbrace{\overset{w_0 + w_1 x_1 + \dots + w_p x_p}{\overset{\parallel}{\vec{w}\cdot\vec{x}}}}_{\text{The set of all linear models}} \; ; \vec{w} \in \mathbb{R}^{p+1} \}$$

$\vec{x} = [1, \vec{x}]$

$\uparrow x_1, \dots, x_p$

$$y = g + (h^* - g) + \underbrace{(f - h^*) + (t - f)}_{\varepsilon}$$

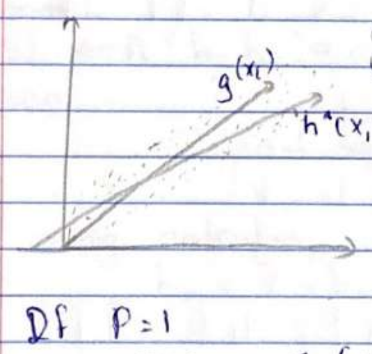
g

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

Best possible w_j values

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon$$

Df $p = 1$



I Least square regression

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

sum of square error

$\parallel$

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

Df $p = 1$

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

$$= \sum \left(y_i^2 + w_0^2 + w_1^2 x_i^2 - 2 y_i w_0 - 2 y_i w_1 x_i + 2 w_0 w_1 x_i \right)$$

$\frac{\sum y_i}{n} = \bar{y}$

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

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

$$2nw_0 - 2n\bar{y} + 2nw_1\bar{x} \overset{set}{=} 0$$

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

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

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

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

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

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

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

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