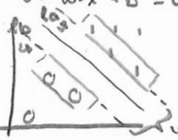
Math 390.4
lecture 5 2/11

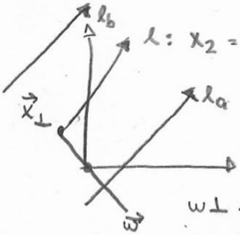
$y = \{0, 1\}$

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



$p = 2$ — wedge

$g(x) \in y$



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

$c \in \mathbb{R}$

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

$w \perp l$ (normal vector)

$x_\perp$: vector $\perp l$ from the origin to l

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

$\vec{w}_0 = \frac{\vec{w}}{\|\vec{w}\|}$ unit direction vector

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

$w_1 x_1 + w_2 x_2 - b = 0$

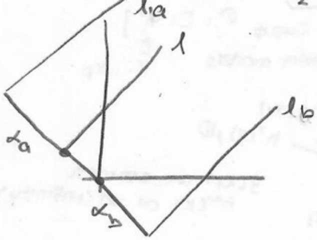
$w_2 x_2 = -w_1 x_1 + b$

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

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

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

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



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

$\alpha_a = \frac{b+d}{\|w\|}$

$\alpha_b = \frac{b-d}{\|w\|}$

$m = \alpha_a - \alpha_b$

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

let $d = 1$

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

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

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

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

$w \cdot x - b - 1 \geq 0$

$w \cdot x - b \geq 1$

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

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

$w \cdot x - b + 1 \leq 0$

$w \cdot x - b \leq -1$

$-(y_i - \frac{1}{2})(w \cdot x - b) \leq -\frac{1}{2}$ } mult -1 to both sides

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

same $\forall_i$

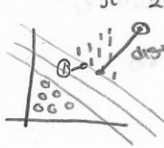
$\mathcal{A}$: maximize $\frac{2}{\|w\|}$

s.t. $\forall_i \; (y_i - \frac{1}{2})(\vec{w} \cdot \vec{x} - b) \geq \frac{1}{2}$

$\vec{w}, b = \text{argmax} \left\{ \frac{2}{\|w\|} \text{ st } \uparrow \right\}$

or

$\text{argmin} \{ \|w\| \text{ st } \uparrow \}$

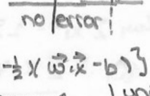


distance (d)

l_m

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

Hinge loss



no error | linear error

If above the line by d

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

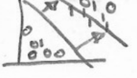
distance above : (1 above l_m) (correct placement)

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

under l_m : $= \max \{ 0, d \} = d$

SHE (sum of hinge error) $= \sum_{i=1}^{n} \max \{ 0, \frac{1}{2} - (y_i - \frac{1}{2})(\vec{w} \cdot \vec{x} - b) \}$

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



$\leftarrow$ no linearly separability

Vapnik (1963)

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

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

what is λ?

Hyperparameter (you specify)

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

λ controls the trade off between the two considerations

λ high: margin more important

λ low: error more important

Null model (without $\vec{x}$ information)

$g_0(\vec{x}) = \text{mode}(\vec{y})$, $\mathbb{D} = \langle \vec{y} \rangle$

$y = \{A, B, C, \dots\}$

L levels; $\mathbb{D} = \langle X, \vec{y} \rangle$; want g

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

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

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

nearest neighbor   d: hyperfunction

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

A: KNN ("K nearest neighbors") (K is a hyper parameter)

For x^* find $x_{i(1)}, x_{i(2)}, \dots, x_{i(K)}$

where $d(\vec{x}^*, x_{i(K)})$'s are the K smallest

and let $\hat{y} = \text{mode}[y_{i(1)}, y_{i(2)}, \dots, y_{i(K)}]$

let $y = \mathbb{R}$ ie. a continuous response

Those models are called the regression model (historical reasons only)

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

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

$h^*(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ ← best possible w_j values

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

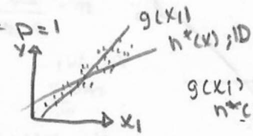
Regression Hypothesis set for p features

$\mathcal{H} = \{ \vec{w} \cdot \vec{x} \;:\; \vec{w} \in \mathbb{R}^{p+1} \}$

$w_0 + w_1 x_1 + \dots w_p x_p$ $\vec{x} : [1 \;\; \vec{x}]$

set of all linear models $x_1 \dots x_p$

IF $p = 1$



g(x1) will approach h*(x) at ∞ (infinity)

$$\mathcal{A}: \underset{\vec{w} \in \mathbb{R}^p}{\operatorname{argmin}} \Big\{ \sum_{i=1}^{n} (y_i - (w_0 + w_1 x_1 \dots + w_p x_p))^2$$

least squared regression

SSE (sum of squared error)

IF $p = 1$

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

$\sum y_i^2 + w_0^2 + w_1^2 x_i^2 - 2 y_i w_0 - 2 y_i w_1 x_i + 2 w_0 w_1 x_i$

$\frac{d}{dw_0}[SSE] = 0$

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

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

$\hat{w}_0 = \bar{y} - w_1 \bar{x}$

$\frac{d}{dw_1}[SSE] = 0$

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

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