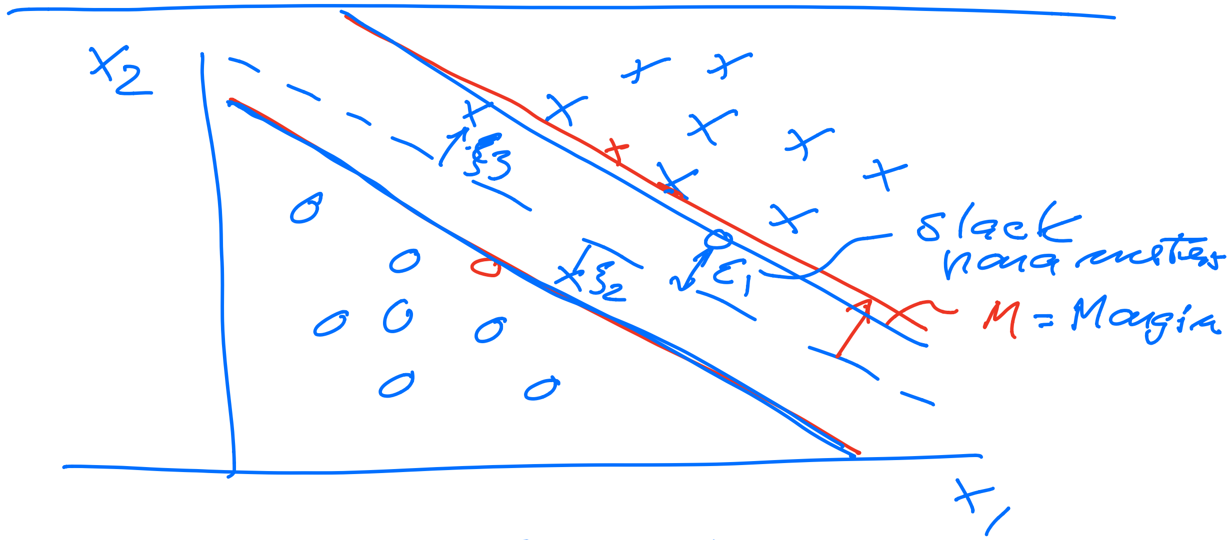
Lecture November 26



$\mathbf{X} = [x_1, x_2]$ $\qquad$ x_m support vector.

based on $x_i = [x_{i1}, x_{i2}]$
have an observation

$$y_i = \{-1, 1\}$$

our model

$$f(x_i) = \mathbf{x}_i^T w + b$$

$$w^T = [w_1, w_2]$$

$$f(x_i) = x_{i1} w_1 + x_{i2} w_2 + b$$

$$f(x_i) = \{-1, 1\}$$

studied problem $\quad y_i f(x_i)$

$= \begin{cases} +1 & \text{correct classific} \\ -1 & \text{wrong} - \text{ı} - \end{cases}$

$y_i(x_i^T w + b) \geq 1$

Want to minimize $\|w\|^2$

where $M = \dfrac{1}{\|w\|^2}$

$\underline{M} = \dfrac{1}{w^T w}$

Hard Margin

minimize $\frac{1}{2} w^T w$

s.t. $y_i(x_i^T w + b) \geq 1$

support vectors are defined by $y_m(x_m^T w + b) = 1$

Lagrangian :

$\mathcal{L}(w, b, \lambda, x) =$

$$\frac{1}{2} \underline{w^T w} - \sum_{i=0}^{n-1} \underline{\lambda_i} \left(y_i (x_i^T w + b) - 1 \right)$$

Lagrangian multiplier

$\dfrac{\partial \mathcal{L}}{\quad} \quad \dfrac{\partial \mathcal{L}}{\quad} \ldots$

$\overline{\partial w}$, $\overline{\partial b}$

$$w = \boxed{\sum_i \lambda_i y_i x_i} \qquad \lambda_i \geq 0$$

$$\sum \lambda_i y_i = 0$$

s.t. $\underline{\lambda_i \geq 0}$

and

$$\lambda_i \left[y_i (x_i^T w + b) - 1 \right] = 0$$

$\lambda_i \neq 0 \Rightarrow y_i (x_i^T w + b) - 1 = 0$

which defines the support vectors.

$$\mathcal{L}(\lambda) = \frac{1}{2} \sum_{ij} \lambda_i \lambda_j y_i y_j x_i^T x_j$$

$$- \sum_{ij} \lambda_i \lambda_j y_i y_j x_i^T x_j$$

$$- \sum_i \lambda_i \underbrace{y_i b}_{=0} + \sum_i \lambda_i \Rightarrow$$

$$\boxed{\mathcal{L}(\lambda) = \sum_{i=0}^{n-1} \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i \lambda_j y_i y_j x_i^T x_j}$$

s.t. $\lambda_i \geq 0$

$\wedge \quad \sum^{n-1} \lambda_i y_i = 0$

$\lambda > 0$ —

and

$$\Rightarrow \boxed{y_i[(x_i^T w + b) - 1] = 0}$$

$$\Rightarrow \underline{w} = \sum \underline{\lambda_i} y_i x_i$$

$$b = \frac{1}{y_i} - x_i^T w$$

Def N_S = # support vectors

optimal b

$$\hat{b} = \frac{1}{N_S} \sum_{\substack{i \in \\ \text{support} \\ \text{vectors}}} \left(\frac{1}{y_i} - \sum_{j=0}^{n-1} \lambda_j y_j x_j^T x_i \right)$$

$$f(x_i) = x_i^T \hat{w} + \hat{b}$$

slack parameter

ξ_i = slack variable

$$y_i(x_i^T w + b) \geq 1 - \xi_i$$

New optimization problem:

$$\frac{1}{2} w^T w + C \sum_{i=0}^{n-1} \xi_i$$

s.t. $\quad y_i(x_i^T w + b) \geq 1 - \xi_i \ \forall i$

s.t. $y_i(x_i^T w + b) \geq 1 - \xi_i$

$$\sum \xi_i \leq \text{constant}$$

$$\mathcal{L}(w, b, \lambda, \xi, \gamma) =$$

$$\frac{1}{2} w^T w + C \sum_{i=0}^{m-1} \xi_i$$

$$- \sum_{i=0}^{m-1} \lambda_i \left(y_i (x_i^T w + b) - (1 - \xi_i) \right)$$

$$- \sum_i \underline{\gamma_i} \xi_i$$

$$w = \sum \lambda_i y_i x_i$$

$$\sum \lambda_i y_i = 0$$

$$\lambda_i = C - \underline{\gamma_i}$$

$$\mathcal{L} = -\frac{1}{2} \lambda^T P \lambda + \mathbb{1} \lambda$$

$$P = \begin{bmatrix} y_0 y_0 x_0^T x_0 & y_0 y_1 x_0^T x_1 & \cdots & y_0 y_{m-1} x_0^T x_{m-1} \\ \vdots & & & \\ y_{m-1} y_0 x_{m-1}^T x_0 & - & - & - \end{bmatrix}$$

$$\mathcal{L} = \frac{1}{2} \lambda^T P \lambda - \mathbb{1} \lambda$$

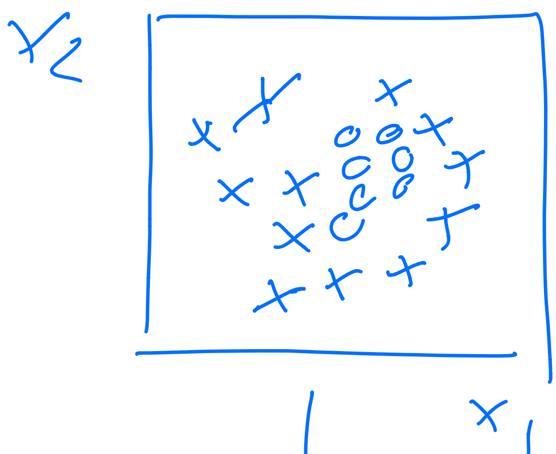
s.t. $\sum \lambda_i y_i = 0$

$\lambda_i \geq 0$ $\boxed{0 \leq \lambda_i \leq C}$

$y_i(\underline{x_i}^T w + b) \geq 1 - \xi_i$

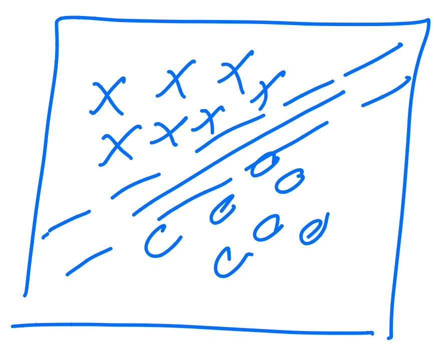
$\xi_i \geq 0$

<u>what about?</u>



linear classifier not meaningful,

kernel transformations



$f(x_i) = x_i^T w + b$

in 2 Dims $\underline{x_i}^T = [x_{i1}, x_{i2}]$

$$y(x_i) = \{-1, 1\} \qquad w^T = [w_1, w_2]$$

kernel $\quad z_i^T = \phi(x_i)^T$

$$= [1, x_{i1}, x_{i2}, x_{i1}^2, x_{i2}^2, x_{i1} x_{i2}]$$

$$\mathcal{L} = \frac{1}{2} \lambda^T P \lambda - 1\lambda$$

$$P = \begin{bmatrix} y_0 y_0 x_0^T x_0 & y_0 y_1 x_0^T x_1 & \cdots \\ \vdots & & \\ y_{n-1} y_0 x_{n-1}^T x_0 & \cdots & y_{n-1} y_{n-1} x_{n-1}^T x_{n-1} \end{bmatrix}$$

$$\tilde{P} = \begin{bmatrix} y_0 y_0 z_0^T z_0 & y_0 y_1 z_0^T z_1 & \cdots & y_0 y_{n-1} z_0^T z_{n-1} \\ \vdots & & & \end{bmatrix}$$

$$K(x, x') = z^T z' = \phi(x)^T \phi(x')$$

kernel,

$$K(x, x') = (1 + x^T x')^2$$

$$x^T = [x_1, x_2]$$

$$x' = [x_1' \; x_2']$$

$$\phi(x) = (\cancel{1}, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

$$\boxed{\begin{aligned} K(x, x') = \; & 1 + 2x_1x_1' + 2x_1x_2' + \\ & 2(x_1x_2)(x_1'x_2') + x_1^2 x_1'^2 + x_2^2 x_2'^2 \end{aligned}}$$

$$\phi(x) = (\sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2)$$