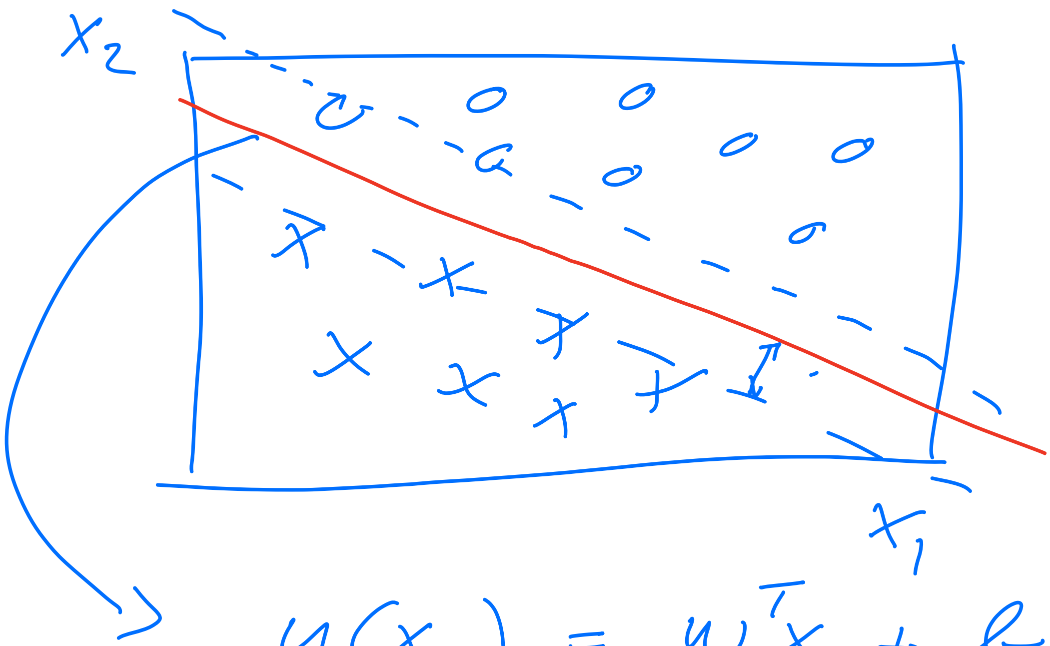
Lecture November 25

SVM



$$y(x) = w^T x + b \quad (= f(x))$$

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

$$D = \{ (x_0\, y_0), (x_1\, y_1) \ldots (x_{n-1}\, y_{n-1}) \}$$

Binary classification

Two classes C_1 and C_2

$$C_1: \quad y(x) \geq 0$$

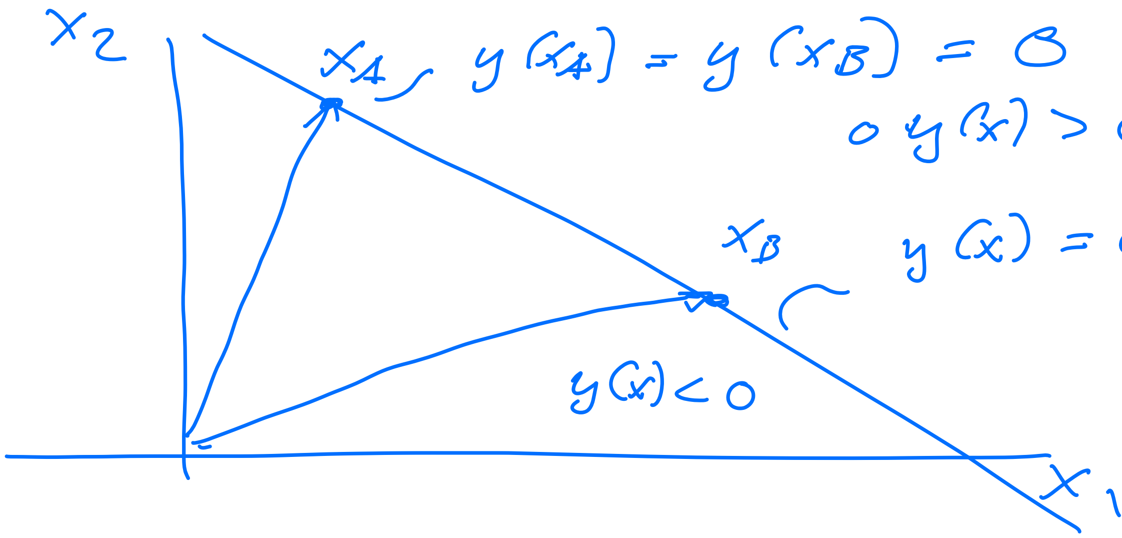
$$C_2: \quad y(x) < 0$$

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

Boundary given by (line)

$$y(x) = 0$$

D-dimensional case, hyperplane of dim $D-1$



$$y(x_A) - y(x_B) = w^T(x_A - x_B) = 0$$

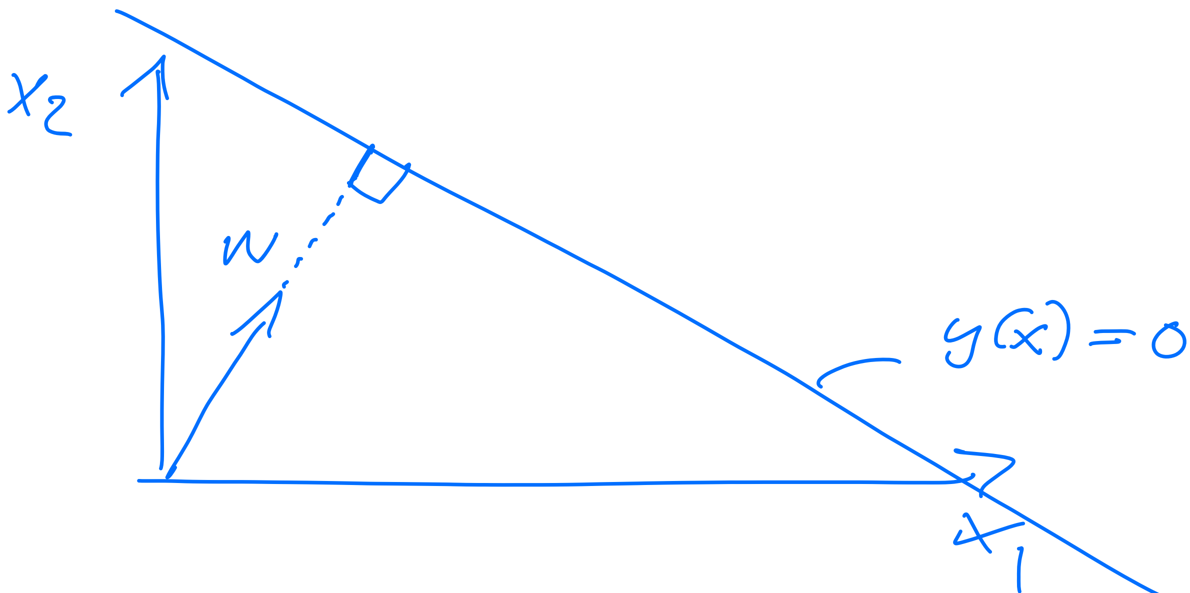
w is orthogonal to every vector lying within the decision (boundary) surface

if x is a point on the decision surface

$$y(x) = 0$$

$y(x) = 0$

w determines the orientation of the boundary surface,



$$y(x) = w^T x + b = 0$$

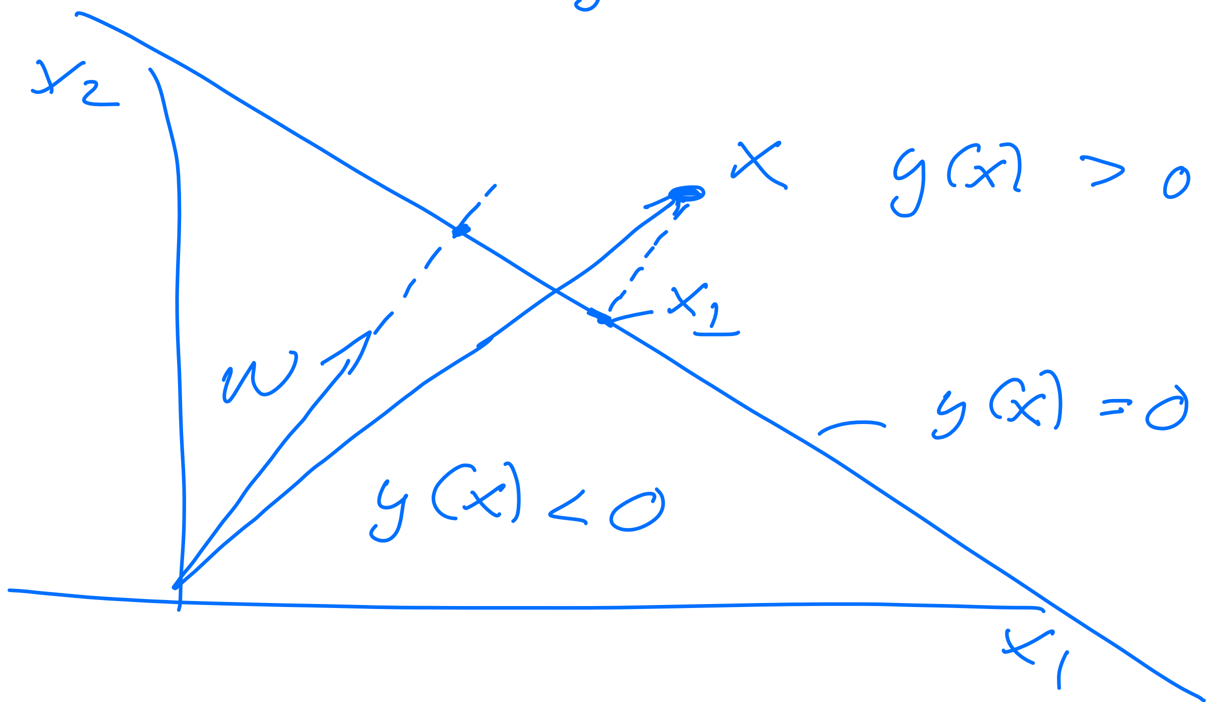
$$\|w\|_2 = \sqrt{w^T w} = \|w\|$$

$$\frac{w^T x}{\|w\|} = -\frac{b}{\|w\|}$$

b determines the location of the surface,

The value of $y(x)$ gives

also a signed measure of the perpendicular distance of a point x from the decision boundary.



$$x = x_\perp + \delta \frac{w}{\|w\|}$$

$$y(x_\perp) = w^T x_\perp + b = 0$$

$w^T x$ + add b

$$b + w^T x = w^T x_\perp + w_0 + \delta \frac{w^T w}{\|w\|}$$

$$= y(x) \Rightarrow$$

$$\delta = \frac{y(x)}{\|w\|}$$

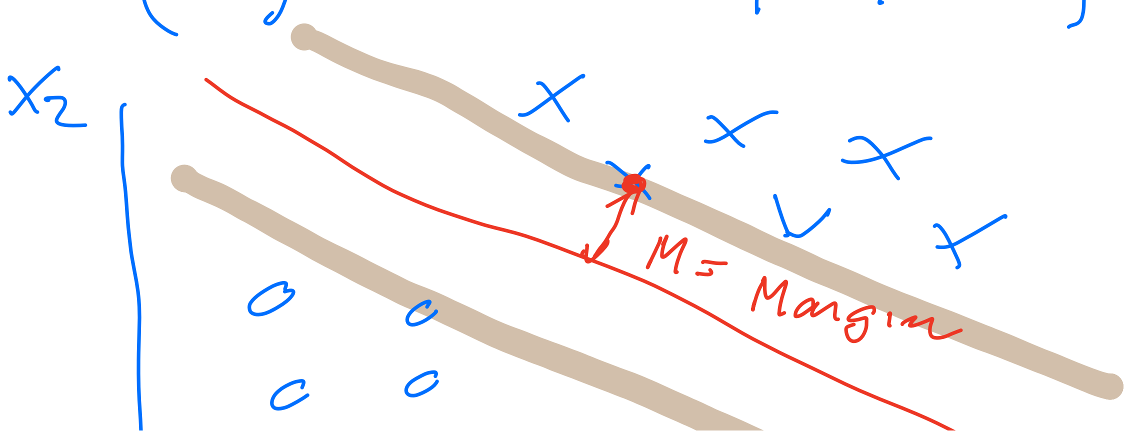
w are the unknown quantities.

Define a model

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

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

$$\left(f(x) = w^T \phi(x) + b \right)$$



signed distance

$$\delta = \frac{y(x)}{\|w\|} = \frac{w^T x + b}{\|w\|} = \frac{f(x)}{\|w\|}$$

targets

we want

$$y \cdot f > 0 \qquad y_i \in \{-1, 1\}$$

Simple approach:
cost function which contains all misclassified results

$$C(w,b) = - \sum_{i \in MisC} y_i \overbrace{(w^T x_i + b)}^{-1}$$

$$\frac{\partial C}{\partial b} = - \sum_i y_i = 0$$

$$\frac{\partial c}{\partial w} = -\sum_i y_i x_i$$

$$\frac{1}{\|w\|} y_i (w^T x_i + b) \geq M$$

(Margin)

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

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

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

want to optimize

$$M \rightarrow w^T w$$

Lagrangian formalism

Example

$$f(x_1 x_2) = x_1 + 3x_2$$

Subject to (s.t.)

$$x_1^2 + x_2^2 = 10$$

$$g(x_1 x_2) = x_1^2 + x_2^2 - 10$$

Def $\mathcal{L}(x, \lambda)$

$$= f(x_1 x_2) - \lambda(x_1^2 + x_2^2 - 10)$$

lagrangian multiplier

Minimize/maximize wrt n-variables x_i and m-lagrangian multiplier λ_j

$$\frac{\partial \mathcal{L}}{\partial x_1} = 1 - 2\lambda x_1 = 0$$

$$x_1 = \frac{1}{2\lambda}$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 0 \Rightarrow x_2 = 3/2\lambda$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = x_1^2 + x_2^2 - 10 = 0$$

Max $\longrightarrow \lambda = +1/2 \quad x_1 = 1 \quad x_2 = 3$

Min $\longrightarrow \lambda = -1/2 \quad x_1 = -1 \quad x_2 = -3$

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

$$b \sum_{i=0}^{n-1} \lambda_i y_i$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 = -\sum_i y_i \lambda_i$$

$$\frac{\partial \mathcal{L}}{\partial w} = 0 = w - \sum_i \lambda_i y_i \cancel{x}_i$$

$$w = \sum_i \lambda_i y_i X_i$$

$$\mathcal{L} = \sum_i \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i \lambda_j y_i y_j$$

$$X^T X$$

$\times\, x_i\, x_j$

Subject to

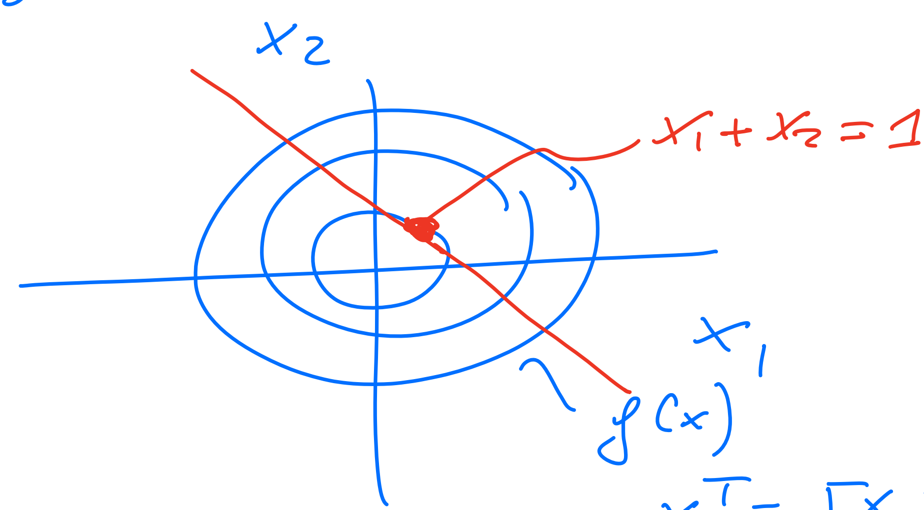
$$\begin{cases} \lambda_i \geq 0 \quad \text{and} \\ \lambda_i \left[y_i \left(w^T x_i + b \right) - 1 \right] = 0 \end{cases}$$

Karush-Kuhn-Tucker,

$$\mathcal{L}(x, \lambda) = f(x) + \lambda g(x)$$

$$f(x) = 1 - x_1^2 - x_2^2$$

$$g(x) = x_1 + x_2 - 1$$



$x^T = [x_1\ x_2]$

when $y(x) > 0$, the constraint $g(x)$, does not play any role.

Stationary points

$\nabla f(x) = 0$ with $\lambda = 0$
$g(x) > 0$

when on the boundary

$g(x) = 0$ and $\lambda \neq 0$

$\lambda g(x) = 0$ for either case

$\Rightarrow$

$$g(x) \geq 0$$

$$\lambda \geq 0$$

$$\lambda g(x) = 0$$

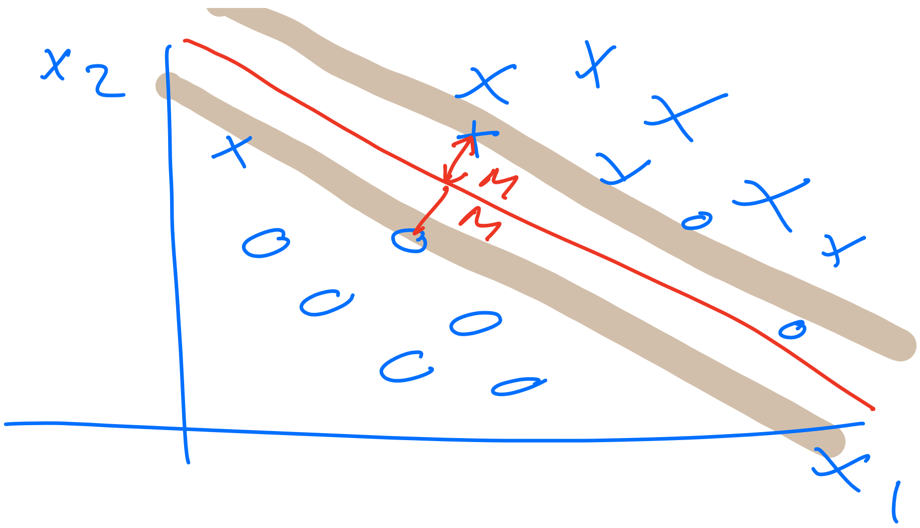
To minimize

$$\mathcal{L}(x, \lambda) = f(x) - \lambda g(x)$$

$\lambda_i > 0 \Rightarrow$

$$y_i (w^T x_i + b) = \underline{1}$$

M = hard Margin



Introduce a slack parameter

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

$$\xi_i \geq 0$$

Total violation $\sum_{i=0}^{n-1} \xi_i < \infty$

New optimization

$$\mathcal{L} = \frac{1}{2} w^T w$$

$$- \sum \lambda_i \left(y_i (w^T x_i + b) - (1 - \xi_i) \right)$$

$$+ C \sum_i \xi_i - \sum_i \gamma_i \xi_i$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 = -\sum \lambda_i y_i$$

$$\frac{\partial \mathcal{L}}{\partial w} = 0 = w - \sum_i \lambda_i y_i x_i$$

$$\lambda_i = C - \gamma_i \quad \forall i$$

$$\mathcal{L} = \sum_i \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j x_i^T x_j$$

subject to

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

$$\gamma_i \xi_i = 0$$

$$y_i (w^T x_i + b) - (1 - \xi_i) \geq 0$$