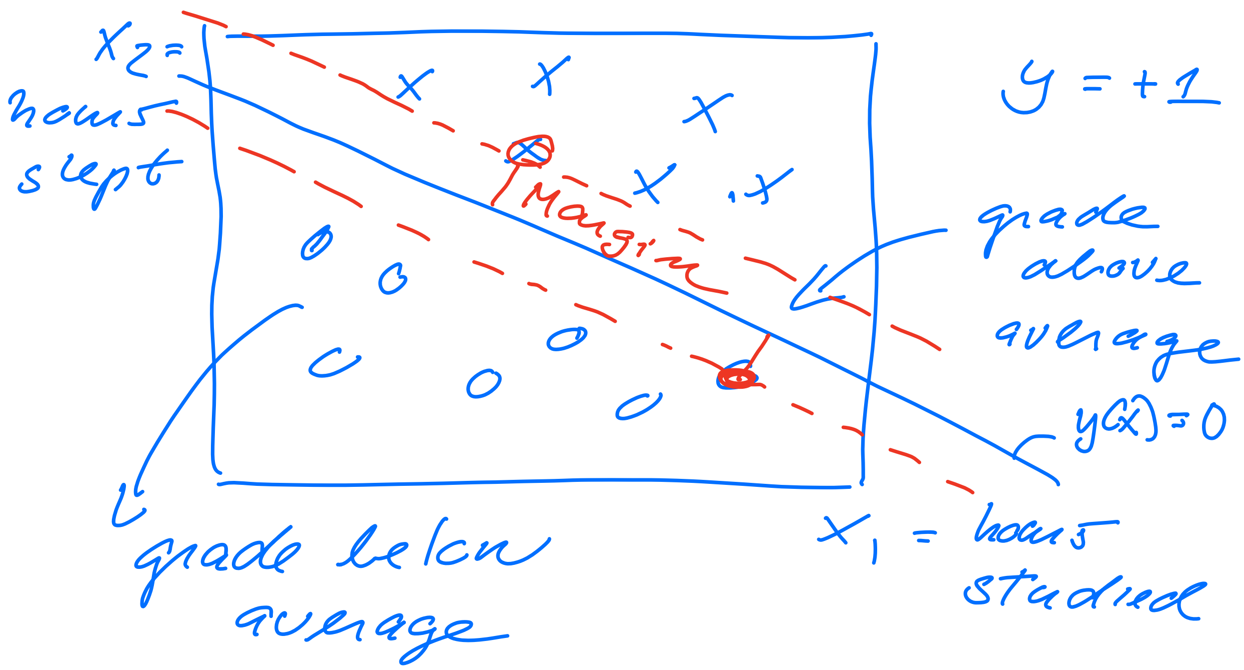
Lecture December 9

Support vector machines



$$x^T = [x_1 \; x_2]$$

outputs $y(x)$

Domain

$$D = \{ (x_0, y_0), (x_1, y_1) \dots \}$$

Model

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

$$= w_1 x_1 + w_2 x_2 + b$$

bias

weights

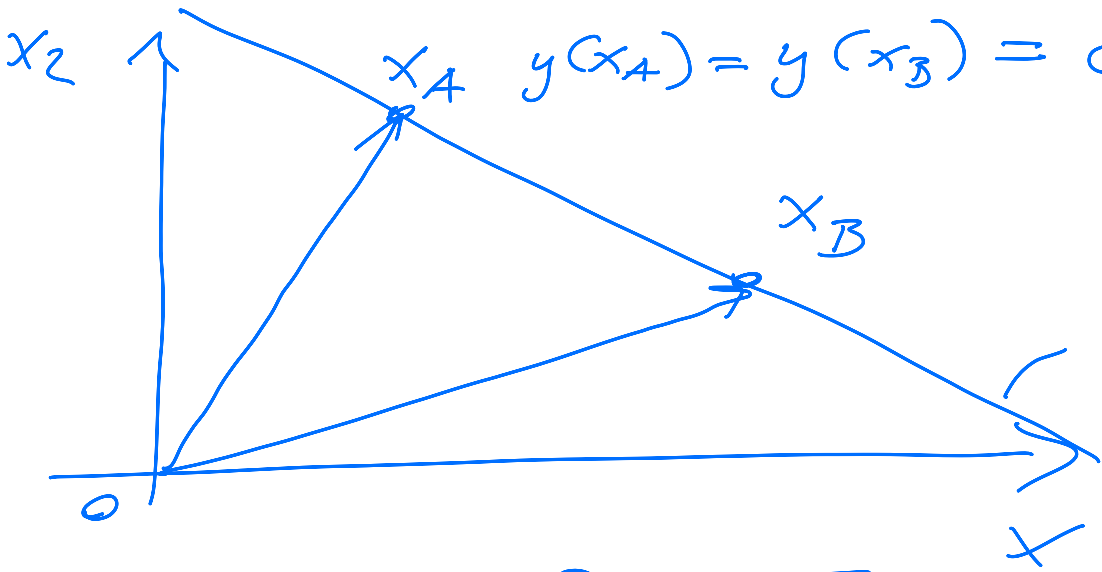
an input vector x is assigned to class C_i

C_1 $y(x) \geq 0$ (above average)

C_2 $y(x) < 0$ (below average)

$y \in \{-1, +1\}$

$y(x_A) = y(x_B) = 0$

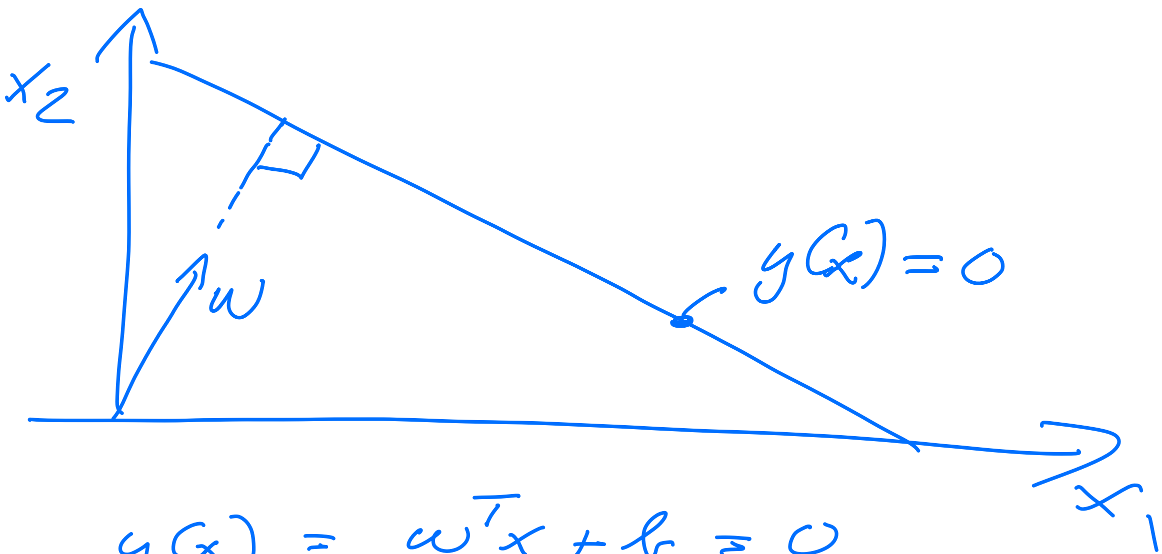


$$y(x_A) - y(x_B) = w^T(x_A - x_B) = 0$$

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

if x is on the surface

$y(x) = 0$



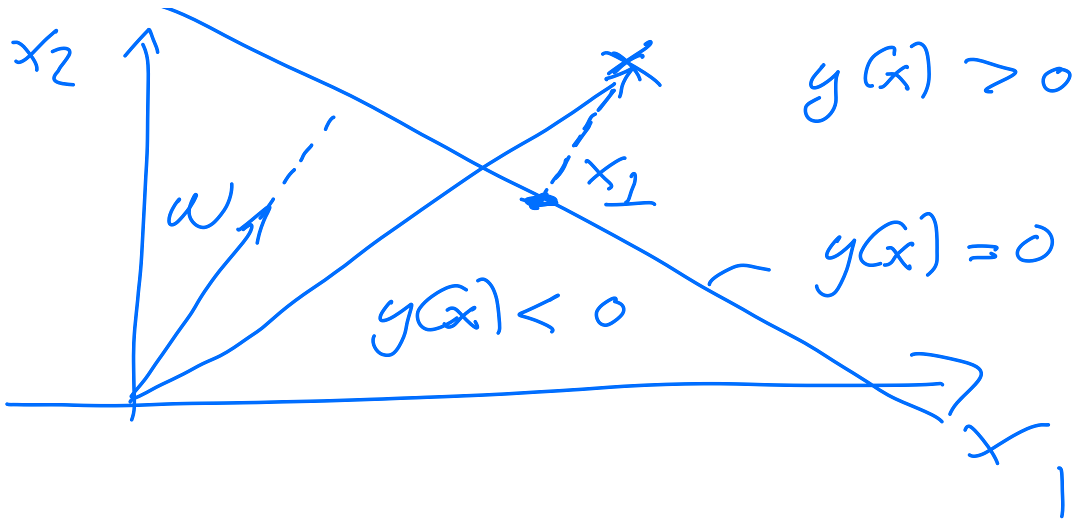
$$y(x) = w^T x + b = 0$$

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

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

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.



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

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

add b and multiply with w^T

$$w^T x + b = \underbrace{w^T x_\perp + b}_{=0} + \delta \underbrace{\frac{w^T w}{\|w\|}}_{\|w\|}$$

$$= y(x) \quad \Rightarrow$$

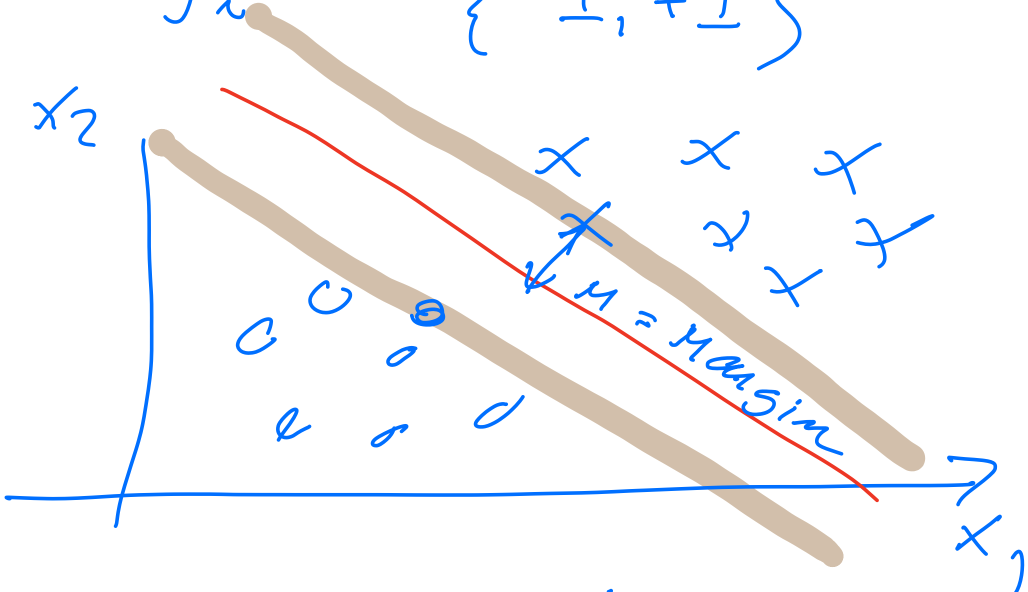
$$\delta = \frac{y(x)}{\|w\|}$$

How do we find w?

Define a model

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

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



signed distance

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

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

we want (classification)

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

Simple approach:

cost functions which contains all misclassified results

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

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

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

leads to many different lines that separate C_1 and C_2.

A better approach is to define a Margin M

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

for all $i = 0, 1, 2, \ldots, n-1$

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

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

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

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

we want to find the norm $w^T w$ subject to the condition that

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

$$\mathcal{L}(w, b, x) = \frac{1}{2} w^T w$$

Lagrangian multiplier

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

$\lambda_i \geq 0$, $\mathcal{L}$ = Lagrangian

Example : Lagrangian formalism

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

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

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

Def : $\mathcal{L}(x, \lambda) = f(x_1, x_2)$
$- \lambda g(x_1, x_2)$

n - variables x_i
m - lagrangian multipliers
λ_j

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

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

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

Max : $\lambda = +1/2$, $x_1 = 1$ $x_2 = 3$

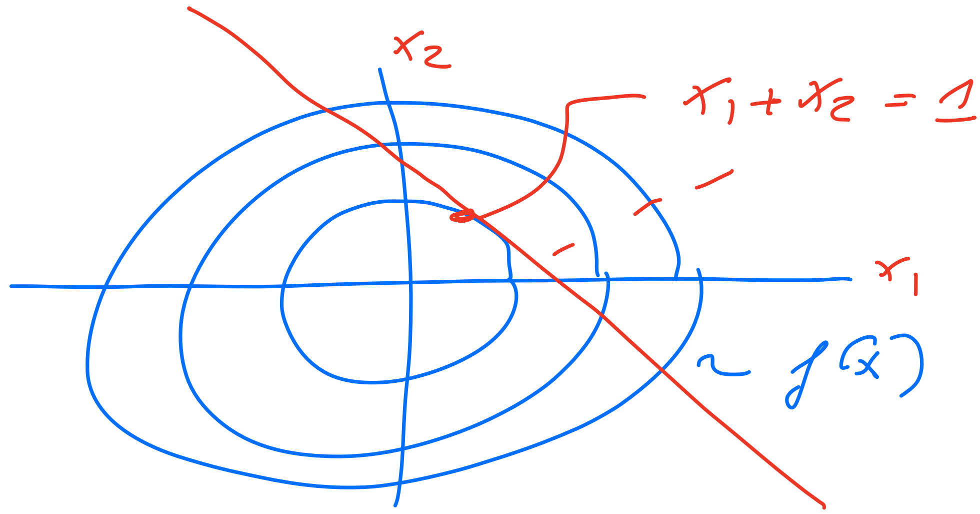
Min : $\lambda = -1/2$ $x_1 = -1$ $x_2 = -3$

Example 2

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

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

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



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

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

- on the boundary
  $g(x) = 0$ and $\lambda \neq 0$

$\Rightarrow$

$$g(x) \geq 0$$

$$\lambda \geq 0$$

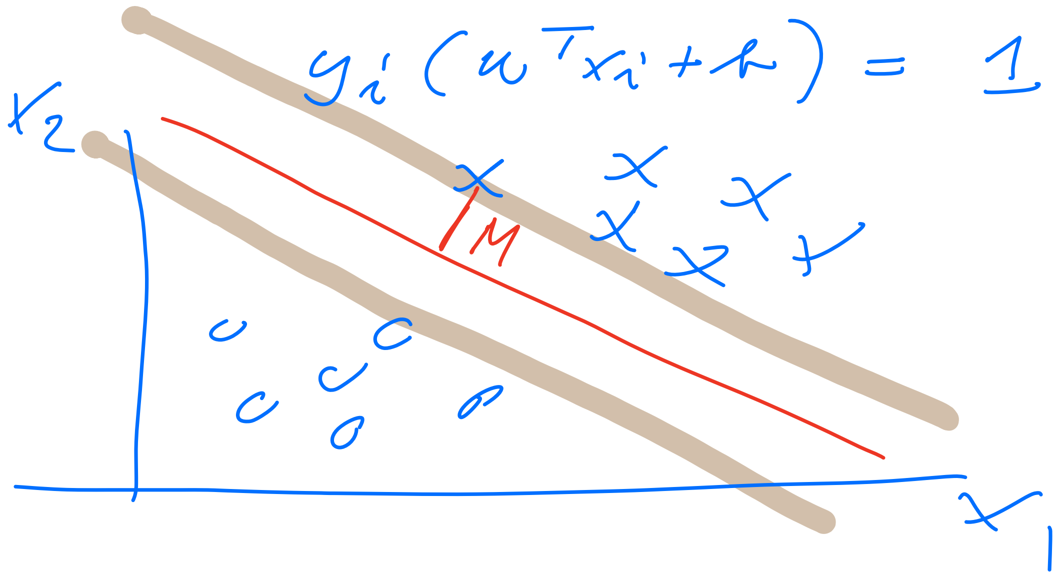
$$\lambda g(x) = 0$$

if we minimize

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

$(\lambda > 0)$

Hard Margin



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

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

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

$$w = \sum_i \lambda_i y_i x_i$$

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

subject to

$\lambda_i \geq 0$ and

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

with λ_i we can find

$$w = \sum_i \lambda_i y_i x_i$$

$y_i (w^T x_i + b) = 1$ on the margin

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

$$b = \frac{1}{N_S} \sum_{j \in N_S} \left(y_j - \sum_{i=0}^{n-1} \lambda_i y_i x_i^T x_j \right]$$

$$y_i = \text{sign}(w' x_i + b)$$

$$K = \begin{bmatrix} y_1 y_1 x_1^T x_1 & y_1 y_2 x_1^T x_2 & \cdots & y_1 y_n x_1^T x_n \\ y_2 y_1 x_2^T x_1 & & & \\ \vdots & & & \\ y_n y_1 x_n^T x_1 & - & - & - \end{bmatrix}$$

$$\mathcal{L} = \lambda + \frac{\lambda^T K \lambda}{2}$$

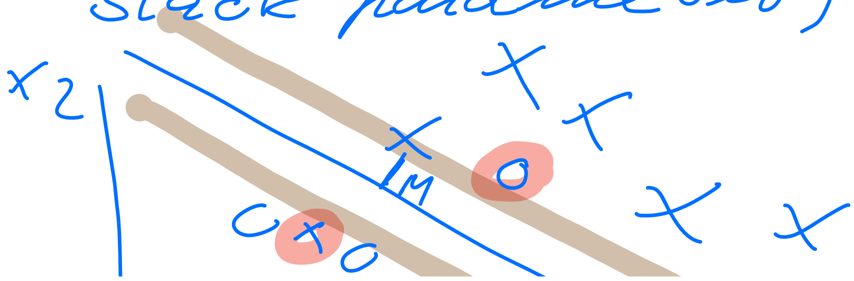
subject to $y^T \lambda = 0$

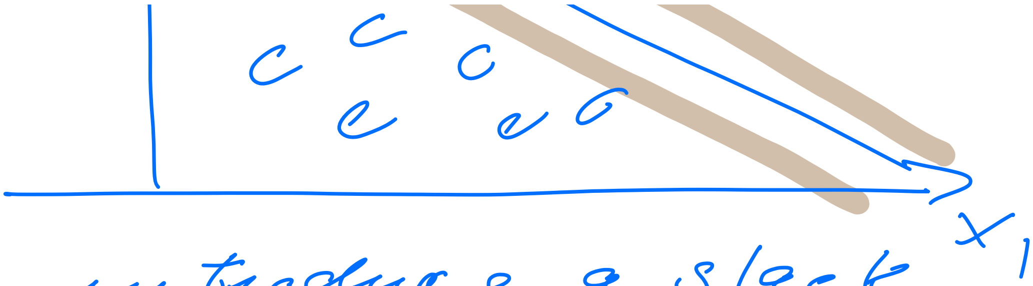
$$\lambda^T = [\lambda_1, \lambda_2, \ldots \lambda_n]$$

$$y^T = [y_1 \; y_2, \ldots y_n]$$

Hard margin problem

soft margin problem (slack parameter)





introduce a slack parameter

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

$$\xi_i \geq 0$$

Total misclassification

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

New optimization

$$\mathcal{L} = \frac{1}{2} w^T w - \sum_i \lambda_i \left[y_i (w^T x_i + b) - (1 - \xi_i) \right] + C \sum_{i=0}^{n-1} \xi_i + \sum_i \gamma_i \xi_i$$

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

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

$$\frac{}{\partial w} = 0 - w \qquad \sum_i$$

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