香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

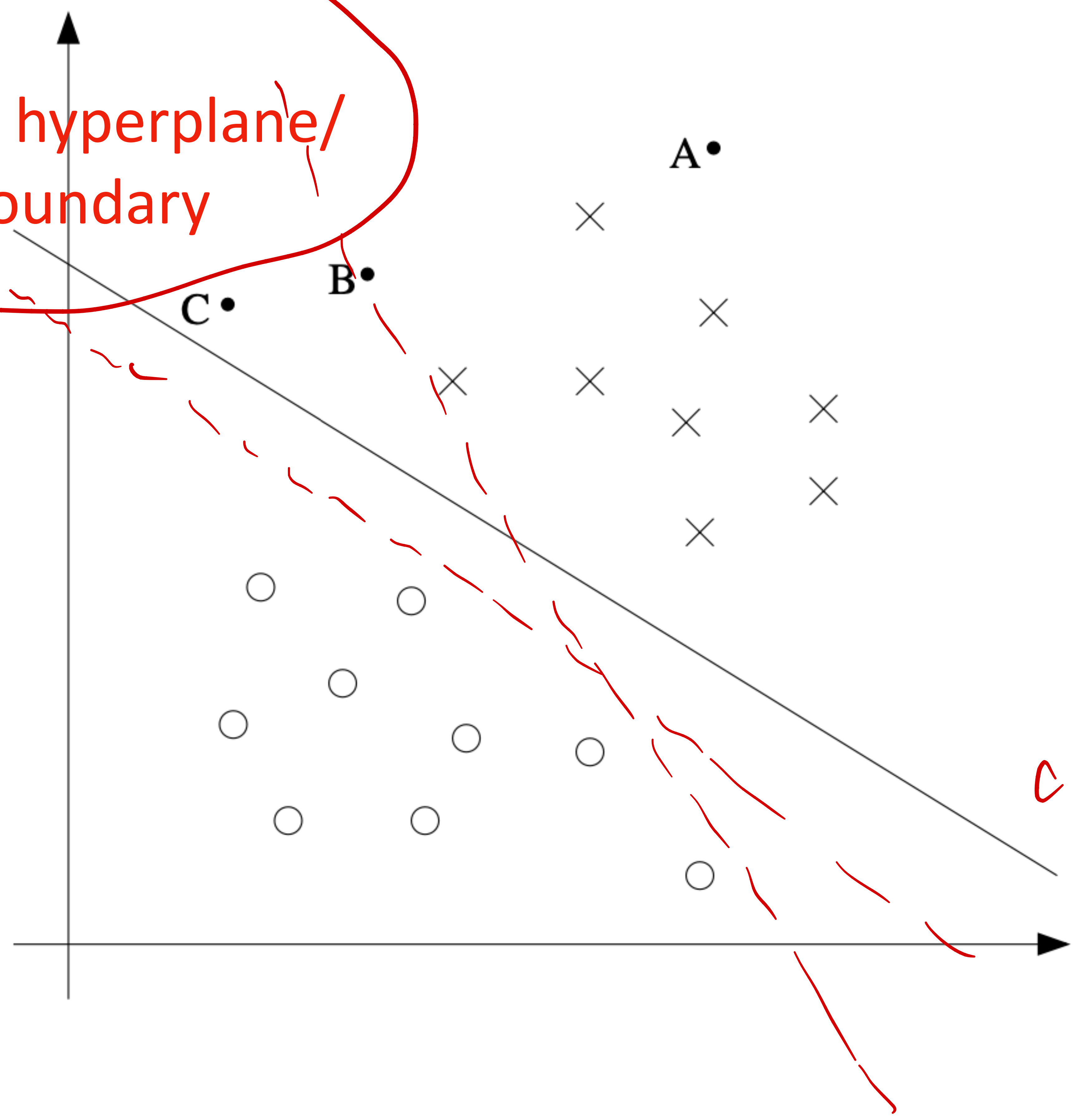
COMP 5212
Machine Learning
Lecture 6

Support Vector Machines

Junxian He
Sep 24, 2024

Recap: Support Vector Machines

Separating hyperplane/
decision boundary

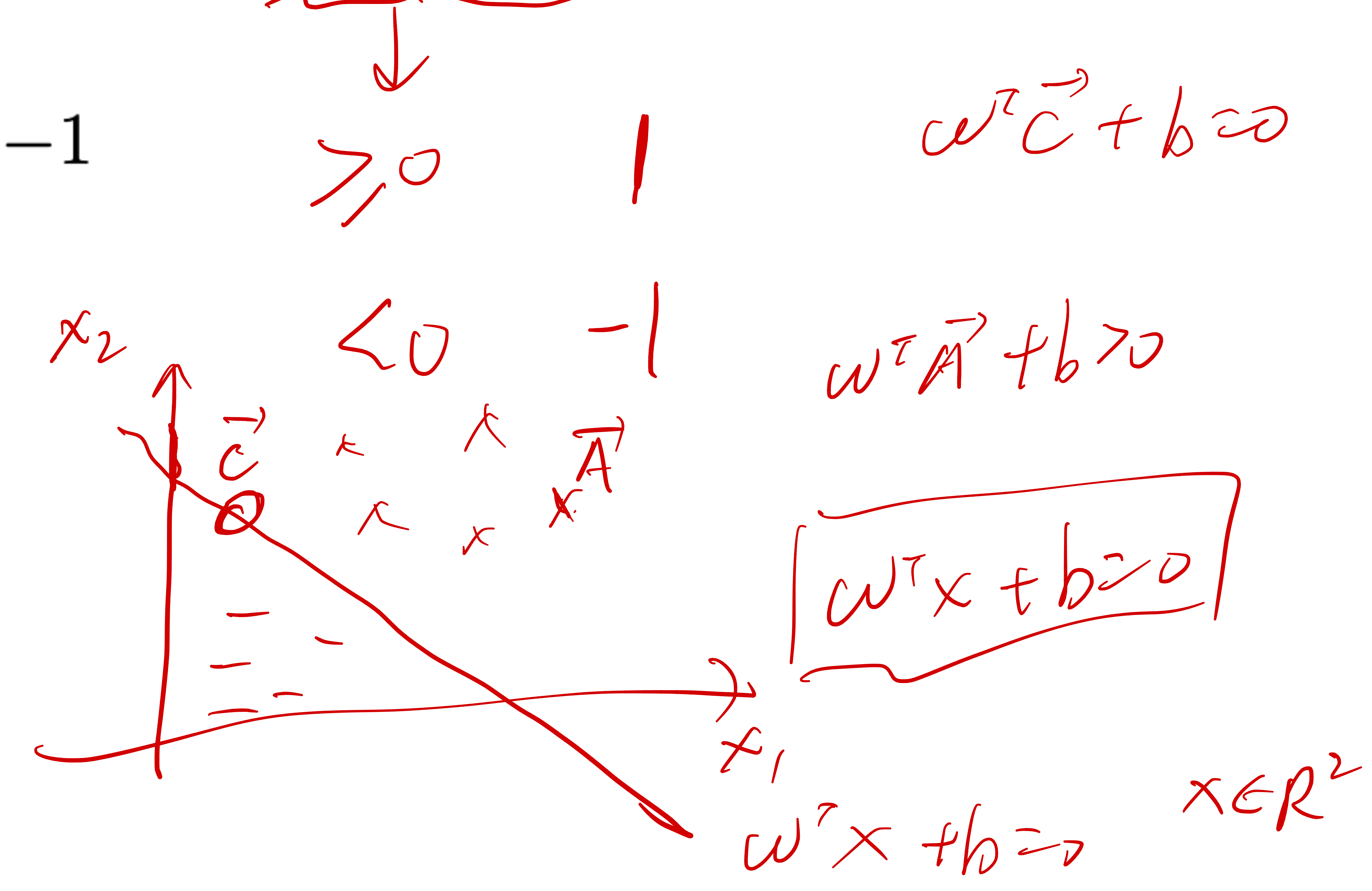


Recap: Notations

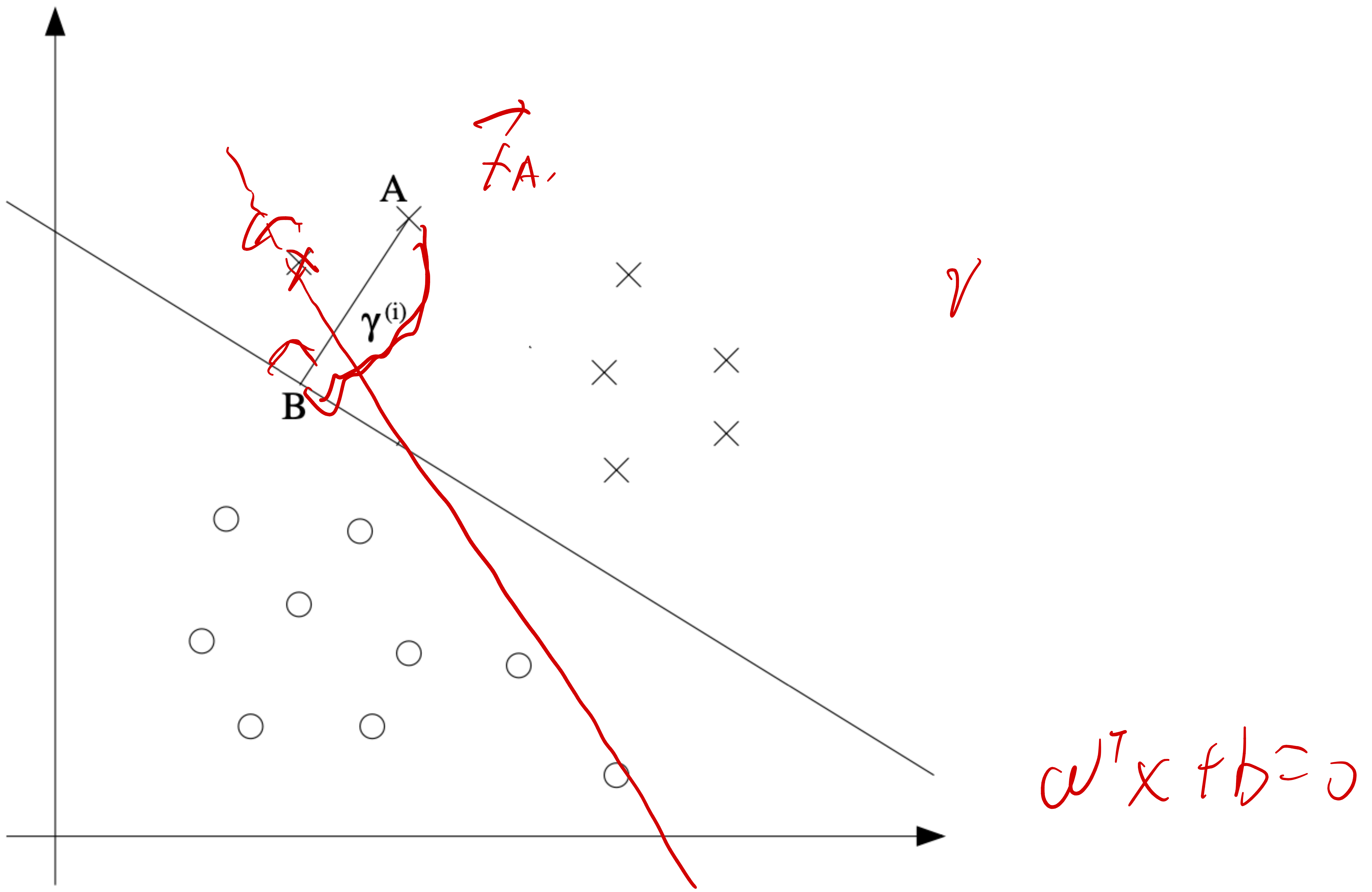
Consider a binary classification problem, with the input feature x and $y \in \{-1, 1\}$ (instead of $\{0, 1\}$), the classifier is:

$$h_{w,b}(x) = g(w^T x + b)$$

$$g(z) = 1 \text{ if } z \geq 0, \text{ and } g(z) = -1$$



Recap: Geometric Margin

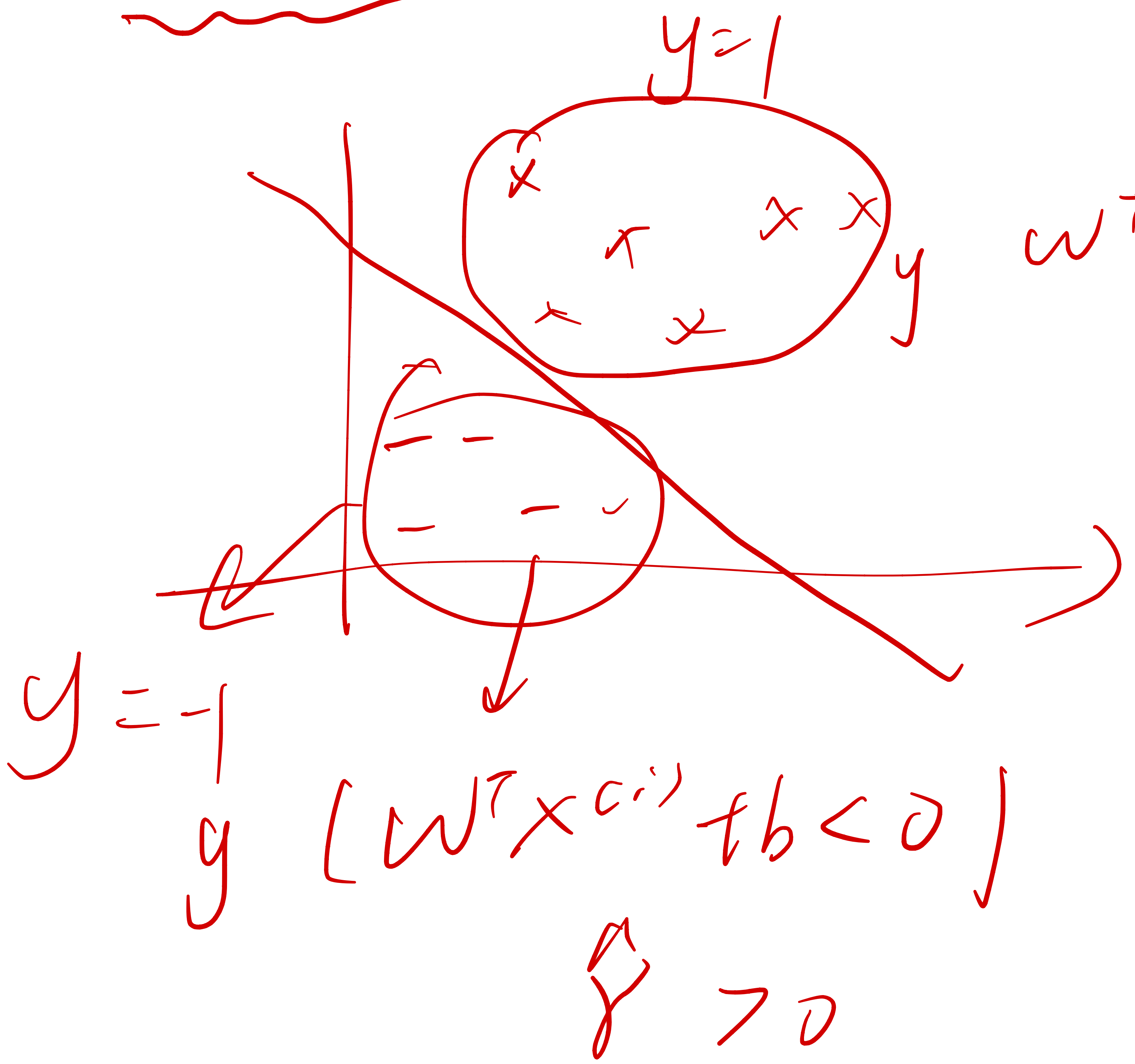


What is the geometric margin?

Recap: Functional Margin

Given a training example $(x^{(i)}, y^{(i)})$

$$\hat{\gamma}^{(i)} = y^{(i)}(w^T x^{(i)} + b).$$



Recap: Functional Margin

Given a training example $(x^{(i)}, y^{(i)})$

$$\hat{\gamma}^{(i)} = y^{(i)}(w^T x^{(i)} + b).$$

Given a training set $S = \{(x^{(i)}, y^{(i)}); i = 1,\ldots,n\}$

$$\hat{\gamma} = \min_{i=1,\ldots,n} \hat{\gamma}^{(i)}$$

Recap: Functional Margin

Given a training example $(x^{(i)}, y^{(i)})$

$$\hat{\gamma}^{(i)} = y^{(i)}(w^T x^{(i)} + b).$$

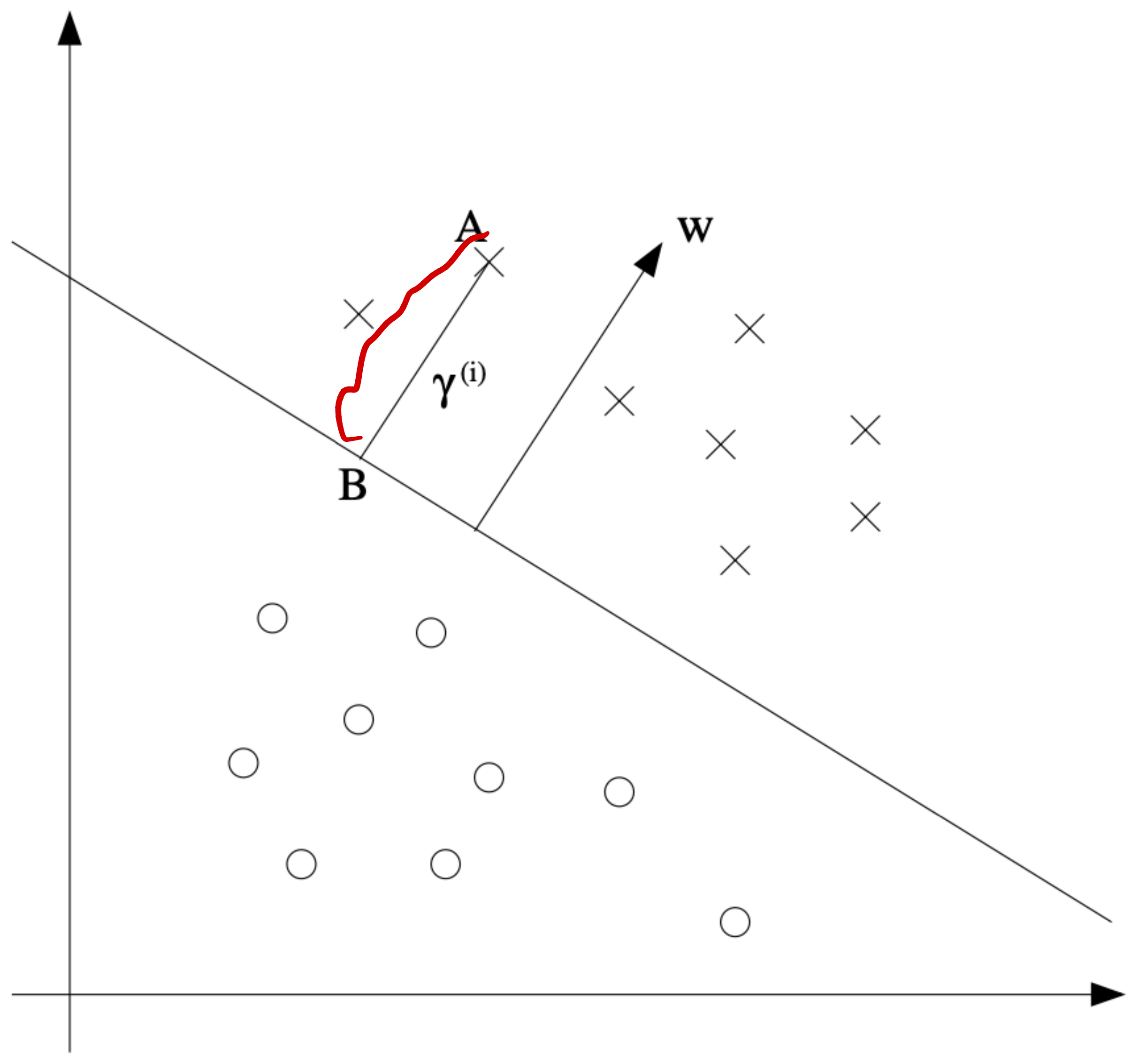
Given a training set $S = \{(x^{(i)}, y^{(i)}); i = 1,\ldots,n\}$

$$\hat{\gamma} = \min_{i=1,\ldots,n} \hat{\gamma}^{(i)}$$

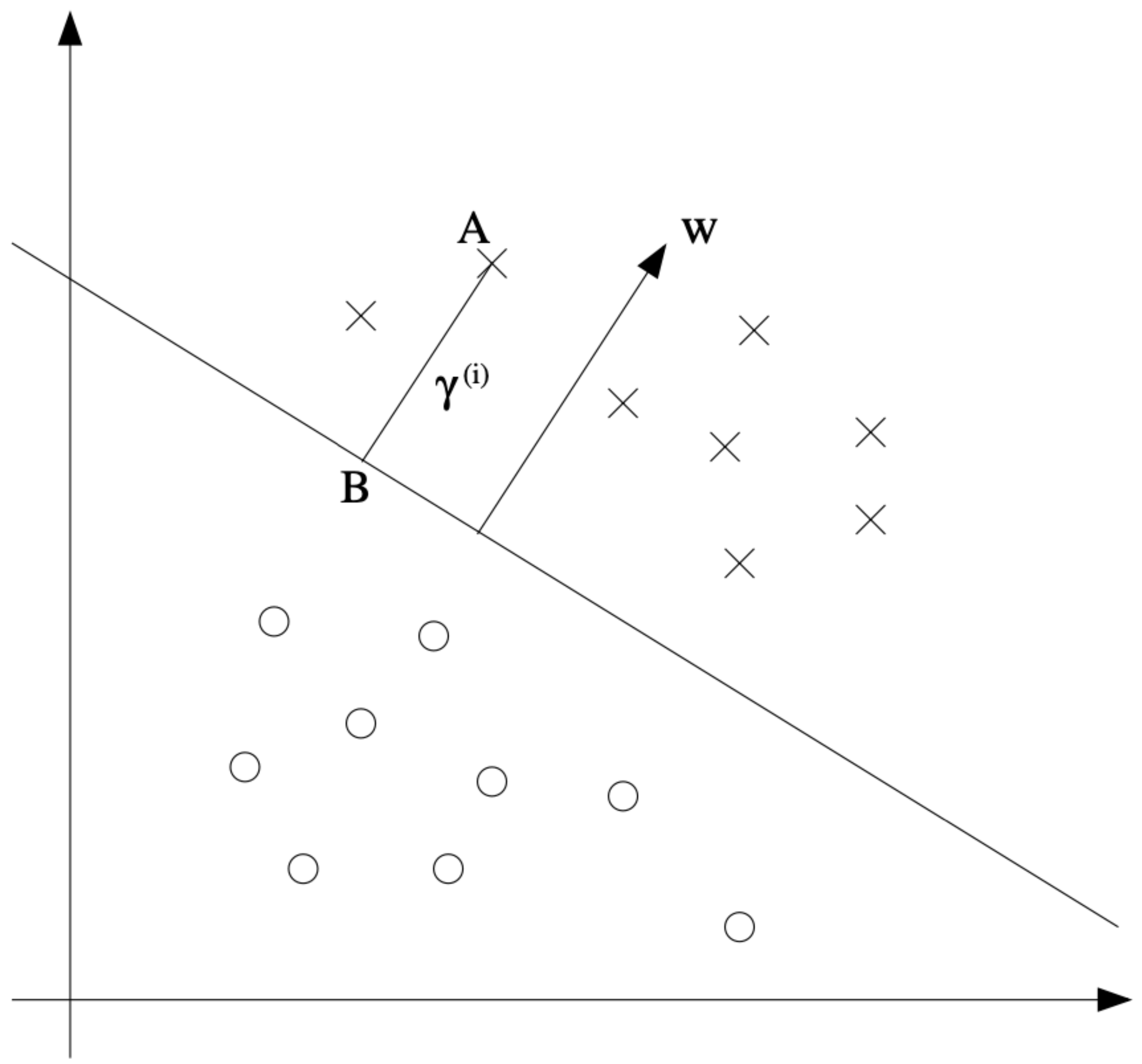
Functional margin changes rescaling parameters, making it a bad objective, e.g. when w->2w, b->2b, the functional margin changes while the separating plane does not really change

Recap: Geometric Margin

Recap: Geometric Margin

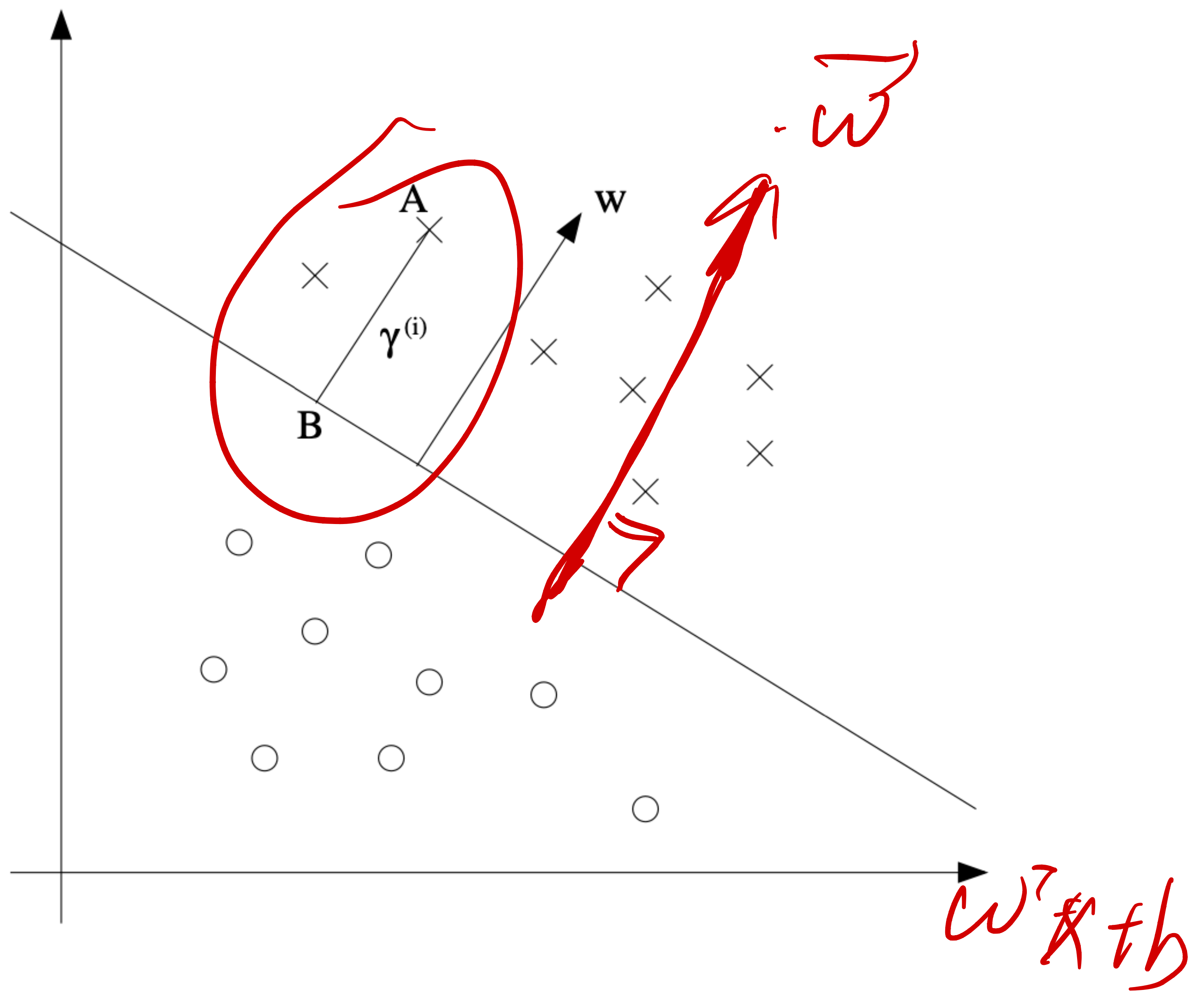


Recap: Geometric Margin



$$w^T \left(x^{(i)} - \gamma^{(i)} \frac{w}{||w||} \right) + b = 0.$$

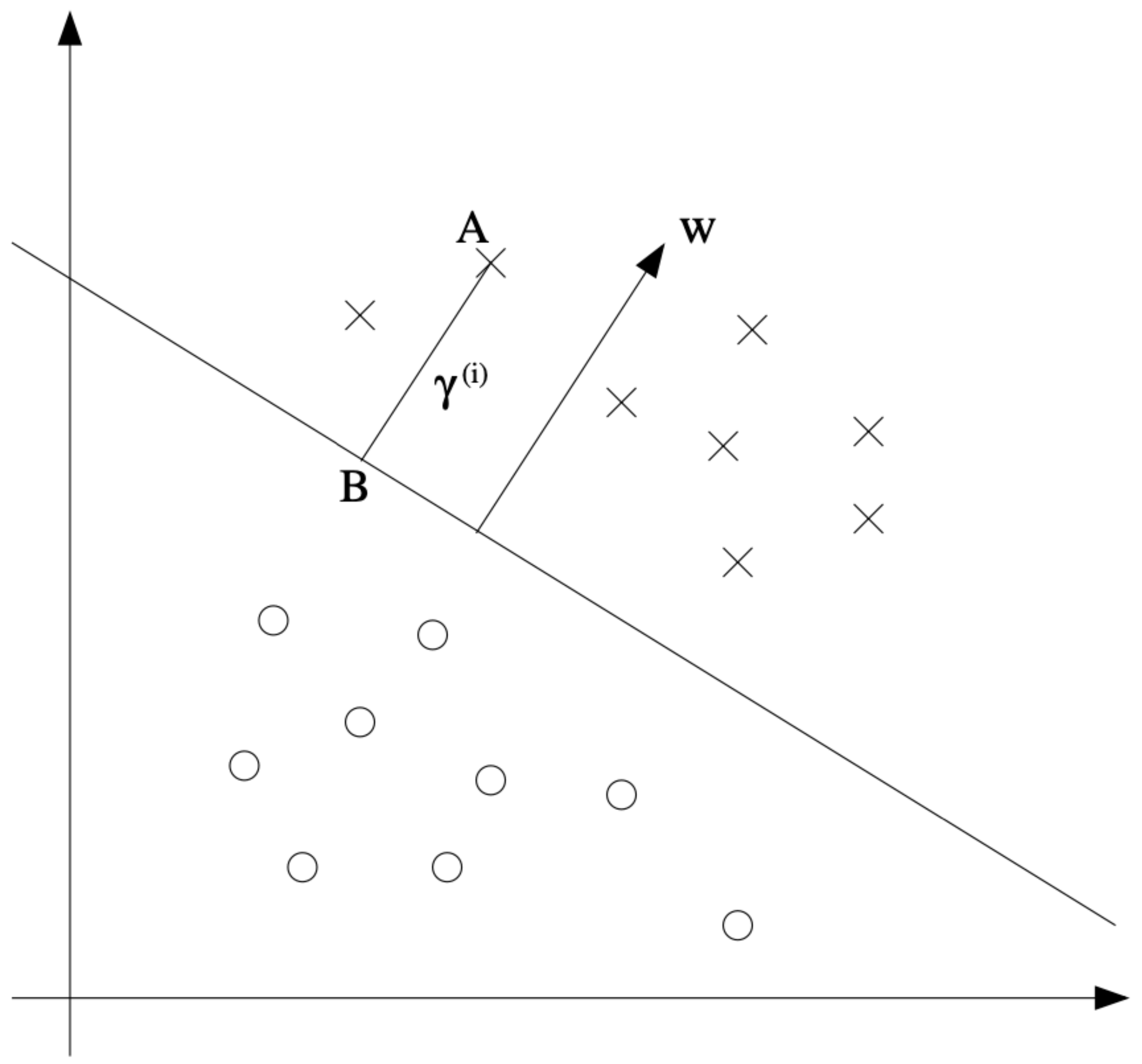
Recap: Geometric Margin



$$w^T\left(x^{(i)} - \gamma^{(i)}\frac{w}{||w||}\right) + b = 0.$$

$$\gamma^{(i)} = \frac{w^T x^{(i)} + b}{||w||} = \left(\frac{w}{||w||}\right)^T x^{(i)} + \frac{b}{||w||}$$

Recap: Geometric Margin

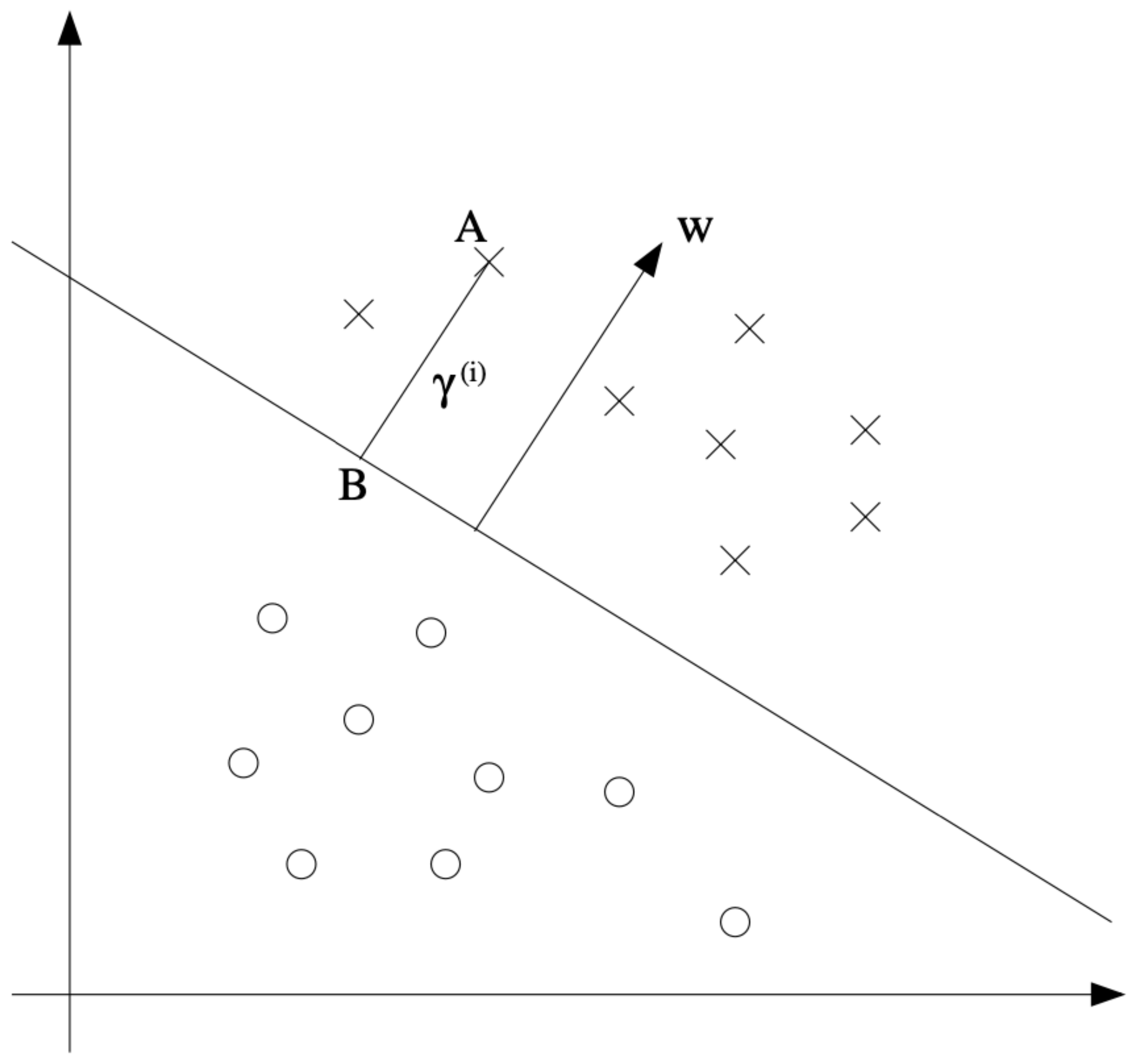


$$w^T \left(x^{(i)} - \gamma^{(i)} \frac{w}{||w||} \right) + b = 0.$$

$$\gamma^{(i)} = \frac{w^T x^{(i)} + b}{||w||} = \left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||}$$

Generally

Recap: Geometric Margin



$$w^T \left(x^{(i)} - \gamma^{(i)} \frac{w}{||w||} \right) + b = 0.$$

$$\gamma^{(i)} = \frac{w^T x^{(i)} + b}{||w||} = \left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||}$$

Generally

$$\gamma^{(i)} = y^{(i)} \left(\left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||} \right)$$

$$= \frac{\hat{\gamma}^{(i)}}{||w||}$$

Recap: Geometric Margin

Given a training set $S = \{(x^{(i)}, y^{(i)}); i = 1,\dots,n\}$

$$\gamma = \min_{i=1,\dots,n} \gamma^{(i)}$$

$$\max_{w,b} \gamma = \max_{w,b} \min_{i=1,\dots,n} \gamma^{(i)}$$

Recap: The Optimization Problem

Recap: The Optimization Problem

Infinite solutions, as $\hat{\gamma}$ can be at any scale without changing the classifier

Recap: The Optimization Problem

Infinite solutions, as $\hat{\gamma}$ can be at any scale without changing the classifier

||w|| is not easy to deal with, non-convex objective

Recap: The Optimization Problem

$$\max_{w,b} \min_{i=1,\ldots,n} \gamma^{(i)}$$

Infinite solutions, as $\hat{\gamma}$ can be at any scale without changing the classifier

||w|| is not easy to deal with, non-convex objective

Recap: The Optimization Problem

$\max_{w,b} \min_{i=1,\ldots,n} \gamma^{(i)}$

Rewrite

$\max_{\gamma,w,b} \ \gamma$

s.t. $y^{(i)}\left(\left(\frac{w}{\|w\|}\right)^T x^{(i)} + \frac{b}{\|w\|}\right) \geq \gamma, \ i = 1, \cdots, n$

nonlinear

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

Infinite solutions, as $\hat{\gamma}$ can be at any scale without changing the classifier

||w|| is not easy to deal with, non-convex objective

Recap: The Optimization Problem

$$\max_{w,b} \min_{i=1,\dots,n} \gamma^{(i)}$$

Rewrite

$$\max_{\gamma,w,b} \ \gamma$$

$$\text{s.t.} \quad y^{(i)}\left(\left(\frac{w}{\|w\|}\right)^T x^{(i)} + \frac{b}{\|w\|}\right) \geq \gamma, \ i = 1, \cdots, n$$

replace γ with $\hat{\gamma}$ $\quad \hat{\gamma} = \|w\| \cdot \gamma$

Linear constraint

Infinite solutions, as $\hat{\gamma}$ can be at any scale without changing the classifier

||w|| is not easy to deal with, non-convex objective

Recap: The Optimization Problem

$\max_{w,b} \min_{i=1,\ldots,n} \gamma^{(i)}$

Rewrite

$\max_{\gamma,w,b} \ \gamma$

s.t. $y^{(i)}\left(\left(\frac{w}{\|w\|}\right)^T x^{(i)} + \frac{b}{\|w\|}\right) \geq \gamma, \ i = 1, \cdots, n$

Linear constraint

$\max_{\hat{\gamma},w,b} \ \frac{\hat{\gamma}}{\|w\|}$

s.t. $y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, \quad i = 1, \ldots, n$

$\hat{\gamma}^*, w^*, b^*$

$2\hat{\gamma}^*, 2w^*, 2b^*$

$k \cdot \hat{\gamma}^*, kw^*, kb^*$

Infinite solutions, as $\hat{\gamma}$ can be at any scale without changing the classifier

$||w||$ is not easy to deal with, non-convex objective

Recap: The Optimization Problem

Recap: The Optimization Problem

$$\max_{\hat{\gamma}, w, b} \quad \frac{\hat{\gamma}}{||w||}$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, \quad i = 1, \ldots, n$$

$w^T x + b = 0$

$w, b \rightarrow kw, kb$

$\hat{\gamma}, w, b$ under-constrained

w, b

$||w|| = 1$

$||w|| = 2$

Recap: The Optimization Problem

$$\max_{\hat{\gamma},w,b} \quad \frac{\hat{\gamma}}{||w||}$$
$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, \quad i = 1, \ldots, n$$

Add constraint $\hat{\gamma} = 1$

$$\max_{w,b} \frac{1}{||w||}, \quad y^{(i)}(w^T x^{(i)} + b) \geq 1$$

$$\min ||w||^2 = w^T w$$

Recap: The Optimization Problem

$$\max_{\hat{\gamma}, w, b} \quad \frac{\hat{\gamma}}{||w||}$$

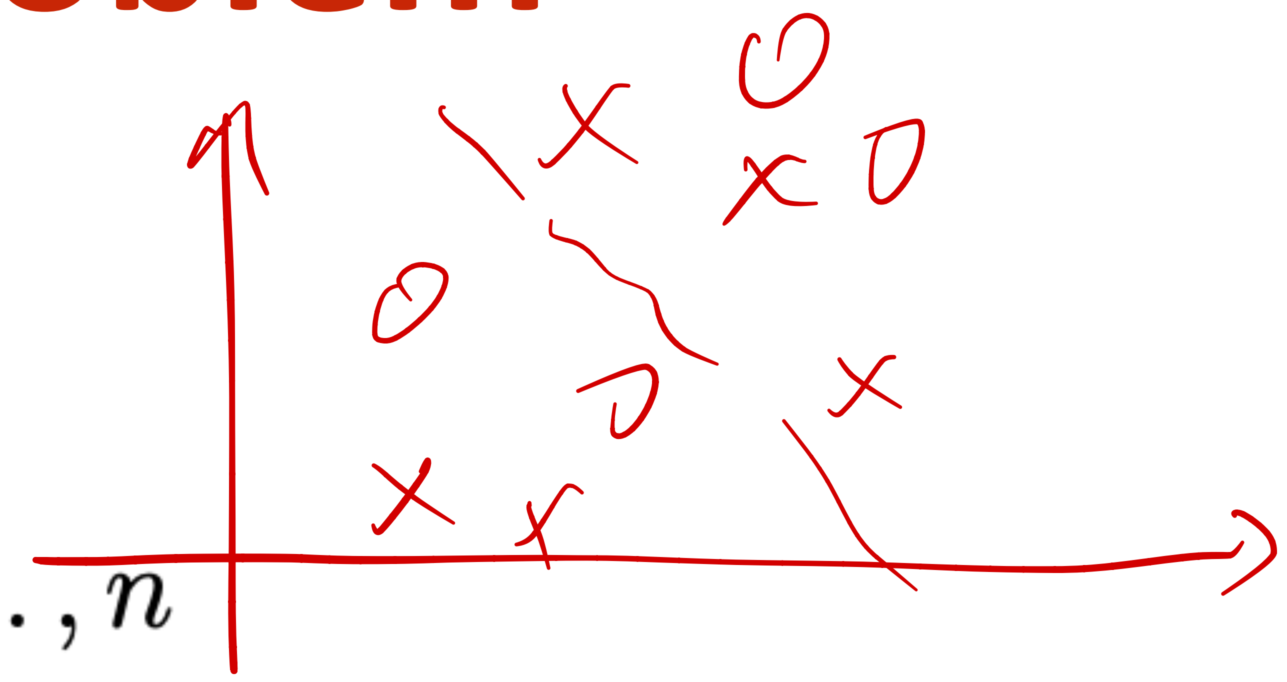
$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, \quad i = 1, \ldots, n$$

Add constraint $\hat{\gamma} = 1$

$$\min_{w,b} \quad \frac{1}{2}||w||^2$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, n$$

Recap: The Optimization Problem



$$\max_{\hat{\gamma}, w, b} \quad \frac{\hat{\gamma}}{||w||}$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, \quad i = 1, \ldots, n$$

assumption: dataset is linearly separable

Add constraint $\hat{\gamma} = 1$

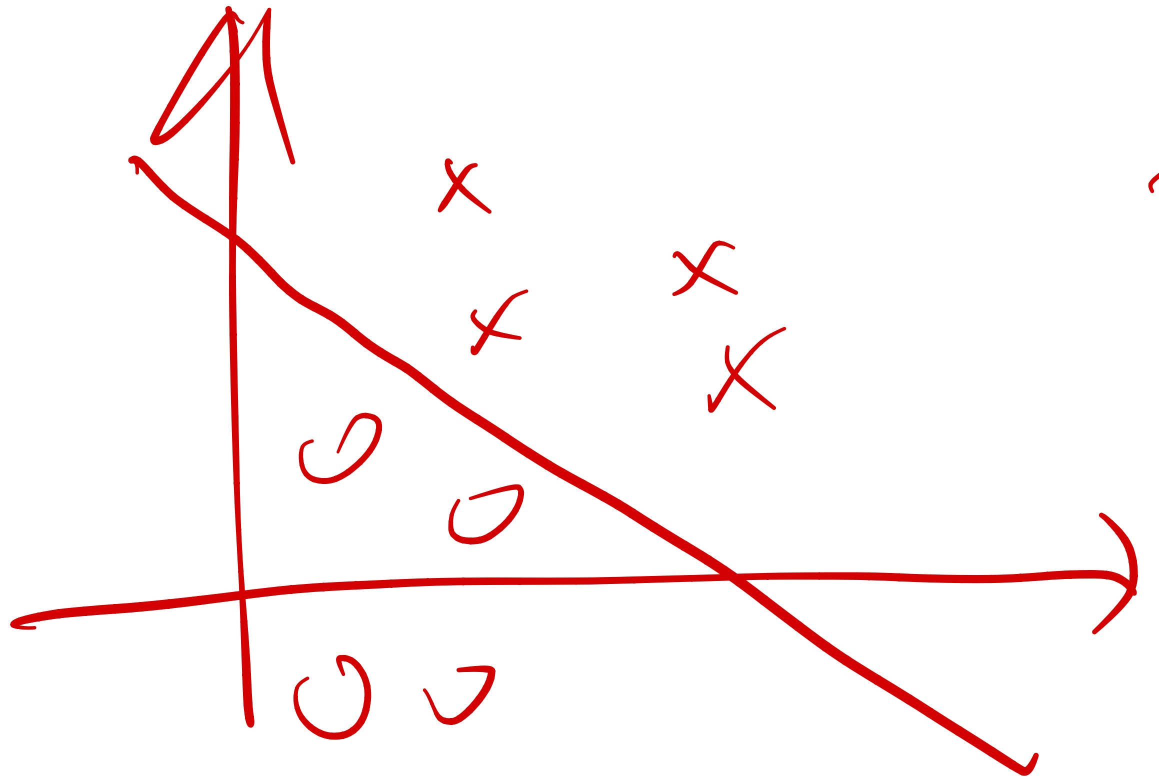
$$\min_{w,b} \quad \frac{1}{2}||w||^2$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, n$$

linear

This is a standard quadratic program problem that can be directly solved with quadratic problem solvers

$y(w^T x + b) > 0$ → correct



1. Objective

2. learning methods

Support vector machines

Variational AE
GANs

same model

model + learning algorithm

Recap: The Optimization Problem

$$\max_{\hat{\gamma},w,b} \quad \frac{\hat{\gamma}}{||w||}$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, \quad i = 1, \ldots, n$$

Add constraint $\hat{\gamma} = 1$

$$\min_{w,b} \quad \frac{1}{2}||w||^2$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, n$$

This is a standard quadratic problem that can be directly solved with quadratic problem solvers

Assumption: the training dataset is linearly separable

$\langle x^{(i)}, x^{(j)} \rangle$

$K(x^{(i)}, x^{(j)})$

Kernel

$\beta \to \theta$

not easy to work with high-dim feature map

$\phi(x^{(i)})$, replace $x^{(i)}$

The Dual Problem in Optimization

original

In optimization, sometimes the primal optimization is hard to solve, then we may find a related alternative optimization problem that can be solved more easily, to solve the orignal problem in an indirect way

Quadratic Program

Quadratic Program

$$\min_{w,b} \quad \frac{1}{2}||w||^2$$
$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, n$$

Quadratic Program

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}||w||^2 \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1, \;\; i = 1, \ldots, n \end{aligned}$$

This is already a standard convex opt problem that is ready to be solved, why are we doing all the rest of things?

Lagrange Duality — Lagrange Multiplier

Lagrange Duality — Lagrange Multiplier

$$\begin{aligned} \min_w \quad & f(w) \\ \text{s.t.} \quad & h_i(w) = 0, \quad i = 1, \ldots, l. \end{aligned}$$

equality

Lagrange Duality — Lagrange Multiplier

$$\begin{aligned} \min_w \quad & f(w) \\ \text{s.t.} \quad & h_i(w) = 0, \quad i = 1, \ldots, l. \end{aligned}$$

$\beta_i h_i(w)$

$$\mathcal{L}(w, \beta) = f(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

$w, \beta,$

dual

Lagrange Duality — Lagrange Multiplier

$$\min_w \quad f(w)$$
$$\text{s.t.} \quad h_i(w) = 0, \quad i = 1, \ldots, l.$$

$$\mathcal{L}(w, \beta) = f(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

Solve w, β

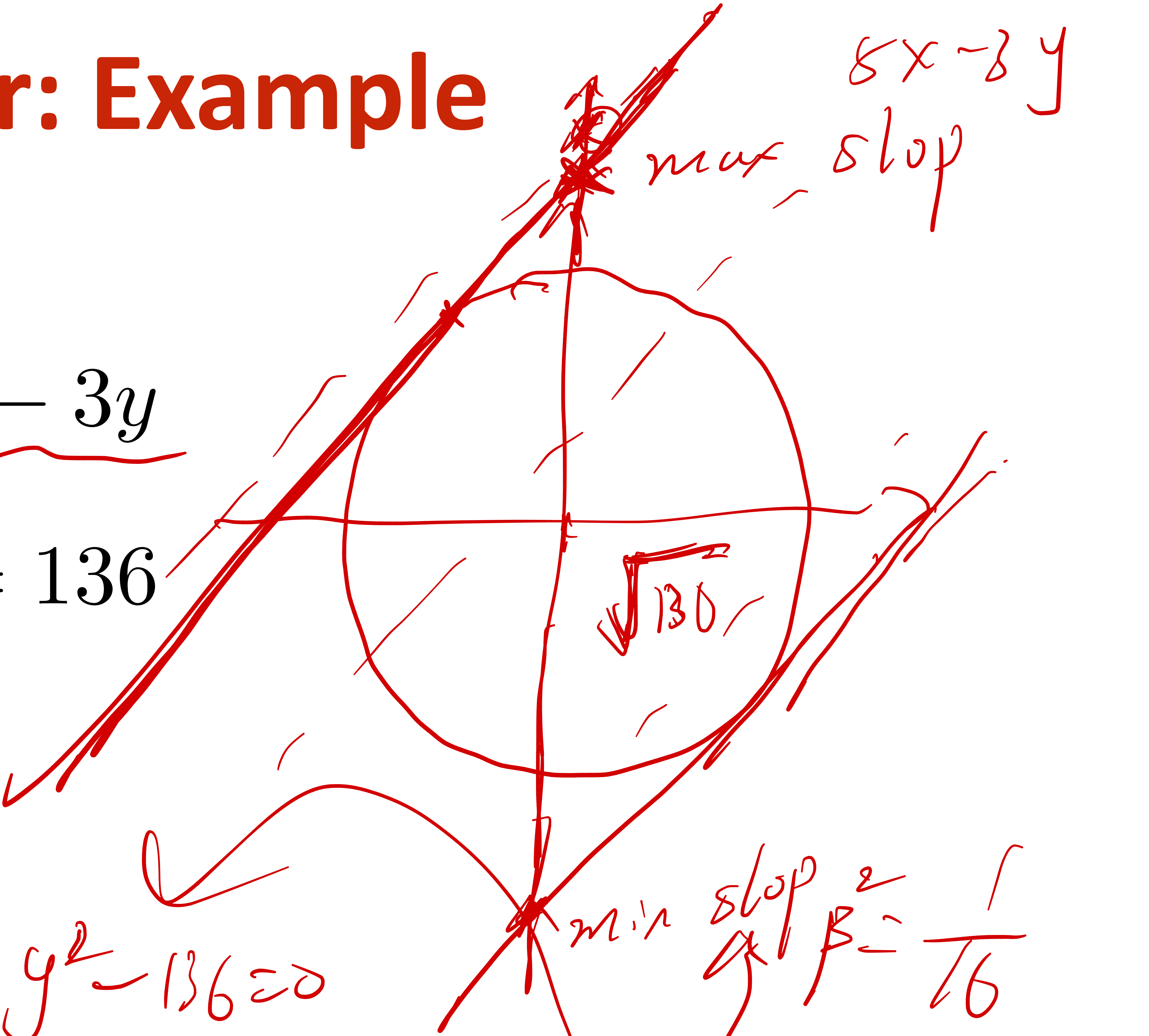
$$\frac{\partial \mathcal{L}}{\partial w_i} = 0; \quad \frac{\partial \mathcal{L}}{\partial \beta_i} = 0,$$

$h_i(w) = 0$

Lagrange Multiplier: Example

$$\min_{x,y} 5x - 3y$$

$$\text{s.t.} \quad x^2 + y^2 = 136$$



$$L(x, y, \beta) = 5x - 3y + \beta(x^2 + y^2 - 136)$$

$$\frac{\partial L}{\partial x} = 5 + 2x\beta = 0$$

$$\frac{\partial L}{\partial y} = -3 + 2\beta y = 0$$

$$\frac{\partial L}{\partial \beta} = x^2 + y^2 - 136 = 0$$

$$\Rightarrow x = -\frac{5}{2\beta}, \quad y = \frac{3}{2\beta}$$

$$\beta^2 = \frac{1}{16}$$

$$\beta = \pm\frac{1}{4}$$

Generalized Lagrangian

Generalized Lagrangian

Primal optimization problem

$$
\begin{aligned}
\min_w \quad & f(w) \\
\text{s.t.} \quad & g_i(w) \leq 0, \quad i = 1, \ldots, k \\
& h_i(w) = 0, \quad i = 1, \ldots, l.
\end{aligned}
$$

svm

$y(w^T x + b) \geq 1$

Generalized Lagrangian

Primal optimization problem

$$\begin{aligned} \min_w \quad & f(w) \\ \text{s.t.} \quad & g_i(w) \leq 0, \quad i = 1, \ldots, k \\ & h_i(w) = 0, \quad i = 1, \ldots, l. \end{aligned}$$

Generalized Lagrangian

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

Generalized Lagrangian

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

Generalized Lagrangian

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

$\beta \in R$

$$\theta_{\mathcal{P}}(w) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

$\alpha_i \geq 0$

Generalized Lagrangian

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

$$\theta_{\mathcal{P}}(w) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

$$\theta_{\mathcal{P}}(w) = \begin{cases} f(w) & \text{if } w \text{ satisfies primal constraints} \\ \infty & \text{otherwise.} \end{cases}$$

$h_i(w) \neq 0$

$+\infty$ $\beta_i h_i(w) \to \infty$

≤ 0

$\alpha_i \geq 0$

$g_i(w) \leq 0$

$h_i(w) = 0$

$f(w)$

$\alpha_i = 0$ $\theta_p(w) = f(w)$. if w satisfies

Generalized Lagrangian

Consider this optimization problem

$$\min_{w} \theta_{\mathcal{P}}(w) = \min_{w} \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \mathcal{L}(w, \alpha, \beta)$$

Generalized Lagrangian

Consider this optimization problem

$$\min_{w} \theta_{\mathcal{P}}(w) = \min_{w} \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \mathcal{L}(w, \alpha, \beta)$$

It has exactly the same solution as our original problem

Generalized Lagrangian

Consider this optimization problem

$$\min_w \theta_{\mathcal{P}}(w) = \min_w \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \mathcal{L}(w, \alpha, \beta)$$

It has exactly the same solution as our original problem

$$p^* = \min_w \theta_{\mathcal{P}}(w)$$

$\min_w \theta_p(w) \iff$ Optimize our original problem

$f(w)$

s.t. $g_i(w) \leq 0$

$h_i(w) = 0$

$\iff \min_w \theta_p(w)$ primal

w

The Dual Problem in Optimization

In optimization, sometimes the primal optimization is hard to solve, then we may find a related alternative optimization problem that can be solved more easily, to solve the orignal problem in an indirect way

The Dual Problem

parameters $\alpha, \beta,$

$$\theta_{\mathcal{D}}(\alpha, \beta) = \min_{w} \mathcal{L}(w, \alpha, \beta)$$

$\alpha, \beta,$

$\theta_D \rightarrow$ dual

The Dual Problem

$$\theta_{\mathcal{D}}(\alpha, \beta) = \min_{w} \mathcal{L}(w, \alpha, \beta)$$

The dual optimization problem

$$\max_{\alpha,\beta\,:\,\alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha, \beta) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_{w} \mathcal{L}(w, \alpha, \beta)$$

The Dual Problem

$$\theta_{\mathcal{D}}(\alpha,\beta) = \min_{w} \mathcal{L}(w,\alpha,\beta)$$

The dual optimization problem

$$\max_{\alpha,\beta\,:\,\alpha_i\geq 0} \theta_{\mathcal{D}}(\alpha,\beta) = \max_{\alpha,\beta\,:\,\alpha_i\geq 0} \min_{w} \mathcal{L}(w,\alpha,\beta)$$

The primal optimization problem

$$\min_{w} \theta_{\mathcal{P}}(w) = \min_{w} \max_{\alpha,\beta\,:\,\alpha_i\geq 0} \mathcal{L}(w,\alpha,\beta)$$

The Dual Problem

$$\theta_{\mathcal{D}}(\alpha, \beta) = \min_{w} \mathcal{L}(w, \alpha, \beta)$$

The dual optimization problem

$$\max_{\alpha,\beta\,:\,\alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha, \beta) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_{w} \mathcal{L}(w, \alpha, \beta)$$

The primal optimization problem

$$\min_{w} \theta_{\mathcal{P}}(w) = \min_{w} \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \mathcal{L}(w, \alpha, \beta)$$

min f(x)
s.t. ...

What is the relation of the two problems?

The Dual Problem

The Dual Problem

$f(w) = \frac{1}{2}||w||^2$

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

$\max_{\alpha, \beta} \min_{w}$

The Dual Problem

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

The dual optimization problem

$$\max_{\alpha,\beta\,:\,\alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha, \beta) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_w \mathcal{L}(w, \alpha, \beta)$$

The Dual Problem

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

The dual optimization problem

$$\max_{\alpha, \beta \,:\, \alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha, \beta) = \max_{\alpha, \beta \,:\, \alpha_i \geq 0} \min_{w} \mathcal{L}(w, \alpha, \beta)$$

$$\nabla_w \mathcal{L}(w, b, \alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0$$

The Dual Problem

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

The dual optimization problem

$$\max_{\alpha,\beta\,:\,\alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha,\beta) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_{w} \mathcal{L}(w,\alpha,\beta)$$

$$\nabla_w \mathcal{L}(w,b,\alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0 \qquad w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

The Dual Problem

min L(w, α, β)
w

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

The dual optimization problem

$$\max_{\alpha,\beta\,:\,\alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha,\beta) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_{w} \mathcal{L}(w,\alpha,\beta)$$

$$\nabla_w \mathcal{L}(w,b,\alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0$$

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

$$\frac{\partial}{\partial b}\mathcal{L}(w,b,\alpha) = \sum_{i=1}^{n} \alpha_i y^{(i)} = 0$$

The Dual Problem

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

The dual optimization problem

$$\max_{\alpha,\beta\,:\,\alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha,\beta) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_{w} \mathcal{L}(w,\alpha,\beta)$$

$$\nabla_w \mathcal{L}(w,b,\alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0 \qquad w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} \qquad \frac{\partial}{\partial b}\mathcal{L}(w,b,\alpha) = \sum_{i=1}^{n} \alpha_i y^{(i)} = 0$$

$$\theta(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j (x^{(i)})^T x^{(j)}$$

The Dual Problem

max

$\alpha_i \geq 0$

$$\theta(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j (x^{(i)})^T x^{(j)}$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \ldots, n$$

The Dual Problem

$$\theta(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j (x^{(i)})^T x^{(j)}$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \ldots, n$$

$$\nabla_w \mathcal{L}(w, b, \alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0$$

The Dual Problem

$$\theta(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j (x^{(i)})^T x^{(j)}$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \ldots, n$$

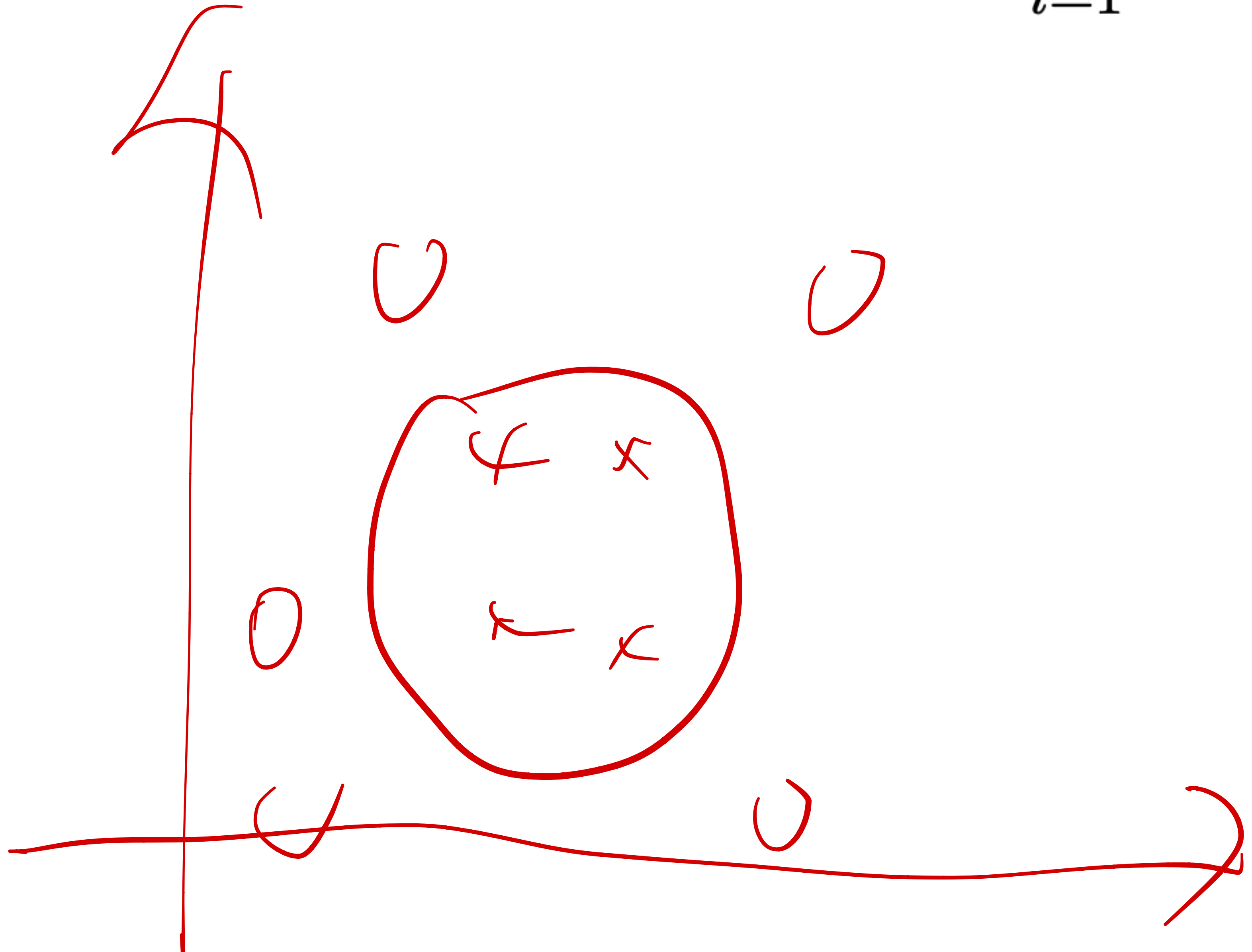
$$\nabla_w \mathcal{L}(w, b, \alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0 \qquad w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

The Dual Problem

$$\theta(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j (x^{(i)})^T x^{(j)}$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \ldots, n$$

$$\nabla_w \mathcal{L}(w, b, \alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0 \qquad w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} \qquad \frac{\partial}{\partial b} \mathcal{L}(w, b, \alpha) = \sum_{i=1}^{n} \alpha_i y^{(i)} = 0$$



The Dual Problem

$$\theta(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j (x^{(i)})^T x^{(j)}$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \ldots, n$$

$$\nabla_w \mathcal{L}(w, b, \alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0 \qquad w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} \qquad \frac{\partial}{\partial b} \mathcal{L}(w, b, \alpha) = \sum_{i=1}^{n} \alpha_i y^{(i)} = 0$$

What is the relation between solving this dual problem and solving the original problem

The Dual Problem

$$d^* = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_w \mathcal{L}(w, \alpha, \beta) \leq \min_w \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \mathcal{L}(w, \alpha, \beta) = p^*$$

The Dual Problem

$$d^* = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_w \mathcal{L}(w,\alpha,\beta) \leq \min_w \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \mathcal{L}(w,\alpha,\beta) = p^*$$

$$\max_x \min_y f(x,y) \leq \min_y \max_x f(x,y)$$

The Dual Problem

$$d^* = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_w \mathcal{L}(w,\alpha,\beta) \leq \min_w \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \mathcal{L}(w,\alpha,\beta) = p^*$$

$$\max_x \min_y f(x,y) \leq \min_y \max_x f(x,y)$$

Under certain conditions: $d^* = p^*$

The Dual Problem

$$d^* = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_w \mathcal{L}(w,\alpha,\beta) \leq \min_w \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \mathcal{L}(w,\alpha,\beta) = p^*$$

$$\max_x \min_y f(x,y) \leq \min_y \max_x f(x,y)$$

Under certain conditions: $d^* = p^*$ Zero-duality Gap

The Dual Problem

$$d^* = \max_{\alpha,\beta:\alpha_i\geq 0} \min_w \mathcal{L}(w,\alpha,\beta) \leq \min_w \max_{\alpha,\beta:\alpha_i\geq 0} \mathcal{L}(w,\alpha,\beta) = p^*$$

$$\max_x \min_y f(x,y) \leq \min_y \max_x f(x,y)$$

Under certain conditions: $d^* = p^*$ Zero-duality Gap

What are the conditions?

Slater's Condition

$$\begin{aligned} \min_w \quad & f(w) \\ \text{s.t.} \quad & g_i(w) \leq 0, \quad i = 1, \ldots, k \\ & h_i(w) = 0, \quad i = 1, \ldots, l. \end{aligned}$$

Slater's Condition

$$\begin{aligned} \min_w \quad & f(w) \\ \text{s.t.} \quad & g_i(w) \leq 0, \quad i = 1, \ldots, k \\ & h_i(w) = 0, \quad i = 1, \ldots, l. \end{aligned}$$

- f(w) and g(w) are convex
- $h_i(w)$ is affine (i.e. linear)
- $g_i(w)$ are strictly feasible for all i, which means there exists some w so that $g_i(w) < 0$ for all i

Slater's Condition

$$\begin{aligned} \min_w \quad & f(w) \\ \text{s.t.} \quad & g_i(w) \le 0, \quad i = 1, \ldots, k \\ & h_i(w) = 0, \quad i = 1, \ldots, l. \end{aligned}$$

- f(w) and g(w) are convex
- $h_i(w)$ is affine (i.e. linear)
- $g_i(w)$ are strictly feasible for all i, which means there exists some w so that $g_i(w) < 0$ for all i

If slater's condition holds, then $d* = p*$

Slater's Condition

$$\begin{aligned} \min_w \quad & f(w) \\ \text{s.t.} \quad & g_i(w) \leq 0, \quad i = 1, \ldots, k \\ & h_i(w) = 0, \quad i = 1, \ldots, l. \end{aligned}$$

- f(w) and g(w) are convex
- $h_i(w)$ is affine (i.e. linear)
- $g_i(w)$ are strictly feasible for all i, which means there exists some w so that $g_i(w) < 0$ for all i

If slater's condition holds, then $d* = p*$

The primal optimization problem of SVM satisfies the slater's condition

KKT Conditions

KKT Conditions

Denote the solution to the primal problem as $w*$, the solution to the dual problem as $\alpha*, \beta*$, then zero duality gap is sufficient and necessary (i.e. equivalent) to satisfy KKT Conditions:

KKT Conditions

Denote the solution to the primal problem as $w*$, the solution to the dual problem as $\alpha*, \beta*$, then zero duality gap is sufficient and necessary (i.e. equivalent) to satisfy KKT Conditions:

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

KKT Conditions

Denote the solution to the primal problem as w^*, the solution to the dual problem as α^*, β^*, then zero duality gap is sufficient and necessary (i.e. equivalent) to satisfy KKT Conditions:

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

$$\begin{aligned}
\frac{\partial}{\partial w_i} \mathcal{L}(w^*, \alpha^*, \beta^*) &= 0, \quad i = 1, \ldots, d \\
\frac{\partial}{\partial \beta_i} \mathcal{L}(w^*, \alpha^*, \beta^*) &= 0, \quad i = 1, \ldots, l \\
\alpha_i^* g_i(w^*) &= 0, \quad i = 1, \ldots, k \\
g_i(w^*) &\leq 0, \quad i = 1, \ldots, k \\
\alpha^* &\geq 0, \quad i = 1, \ldots, k
\end{aligned}$$

KKT Conditions

Denote the solution to the primal problem as w^*, the solution to the dual problem as α^*, β^*, then zero duality gap is sufficient and necessary (i.e. equivalent) to satisfy KKT Conditions:

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

$$\begin{aligned}
\frac{\partial}{\partial w_i} \mathcal{L}(w^*, \alpha^*, \beta^*) &= 0, \quad i = 1, \ldots, d \\
\frac{\partial}{\partial \beta_i} \mathcal{L}(w^*, \alpha^*, \beta^*) &= 0, \quad i = 1, \ldots, l \\
\alpha_i^* g_i(w^*) &= 0, \quad i = 1, \ldots, k \\
g_i(w^*) &\leq 0, \quad i = 1, \ldots, k \\
\alpha^* &\geq 0, \quad i = 1, \ldots, k
\end{aligned}$$

Normal Lagrange multiplier equations

KKT Conditions

Denote the solution to the primal problem as w^*, the solution to the dual problem as α^*, β^*, then zero duality gap is sufficient and necessary (i.e. equivalent) to satisfy KKT Conditions:

$$\mathcal{L}(w,\alpha,\beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

$$\begin{aligned}
\frac{\partial}{\partial w_i}\mathcal{L}(w^*,\alpha^*,\beta^*) &= 0, \quad i = 1,\ldots,d \\
\frac{\partial}{\partial \beta_i}\mathcal{L}(w^*,\alpha^*,\beta^*) &= 0, \quad i = 1,\ldots,l \\
\alpha_i^* g_i(w^*) &= 0, \quad i = 1,\ldots,k \\
g_i(w^*) &\leq 0, \quad i = 1,\ldots,k \\
\alpha^* &\geq 0, \quad i = 1,\ldots,k
\end{aligned}$$

Normal Lagrange multiplier equations

The original constraints

KKT Conditions

Denote the solution to the primal problem as w^*, the solution to the dual problem as α^*, β^*, then zero duality gap is sufficient and necessary (i.e. equivalent) to satisfy KKT Conditions:

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

$$\begin{aligned}
\frac{\partial}{\partial w_i}\mathcal{L}(w^*, \alpha^*, \beta^*) &= 0, \quad i = 1, \ldots, d \\
\frac{\partial}{\partial \beta_i}\mathcal{L}(w^*, \alpha^*, \beta^*) &= 0, \quad i = 1, \ldots, l \\
\alpha_i^* g_i(w^*) &= 0, \quad i = 1, \ldots, k \\
g_i(w^*) &\leq 0, \quad i = 1, \ldots, k \\
\alpha^* &\geq 0, \quad i = 1, \ldots, k
\end{aligned}$$

KKT Conditions

Denote the solution to the primal problem as $w*$, the solution to the dual problem as $\alpha*, \beta*$, then zero duality gap is sufficient and necessary (i.e. equivalent) to satisfy KKT Conditions:

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

$$\begin{aligned}
\frac{\partial}{\partial w_i} \mathcal{L}(w^*, \alpha^*, \beta^*) &= 0, \quad i = 1, \ldots, d \\
\frac{\partial}{\partial \beta_i} \mathcal{L}(w^*, \alpha^*, \beta^*) &= 0, \quad i = 1, \ldots, l \\
\alpha_i^* g_i(w^*) &= 0, \quad i = 1, \ldots, k \\
g_i(w^*) &\leq 0, \quad i = 1, \ldots, k \\
\alpha^* &\geq 0, \quad i = 1, \ldots, k
\end{aligned}$$

If $\alpha_i^* > 0$, then $g_i(w^*) = 0$, the inequality is actually equality

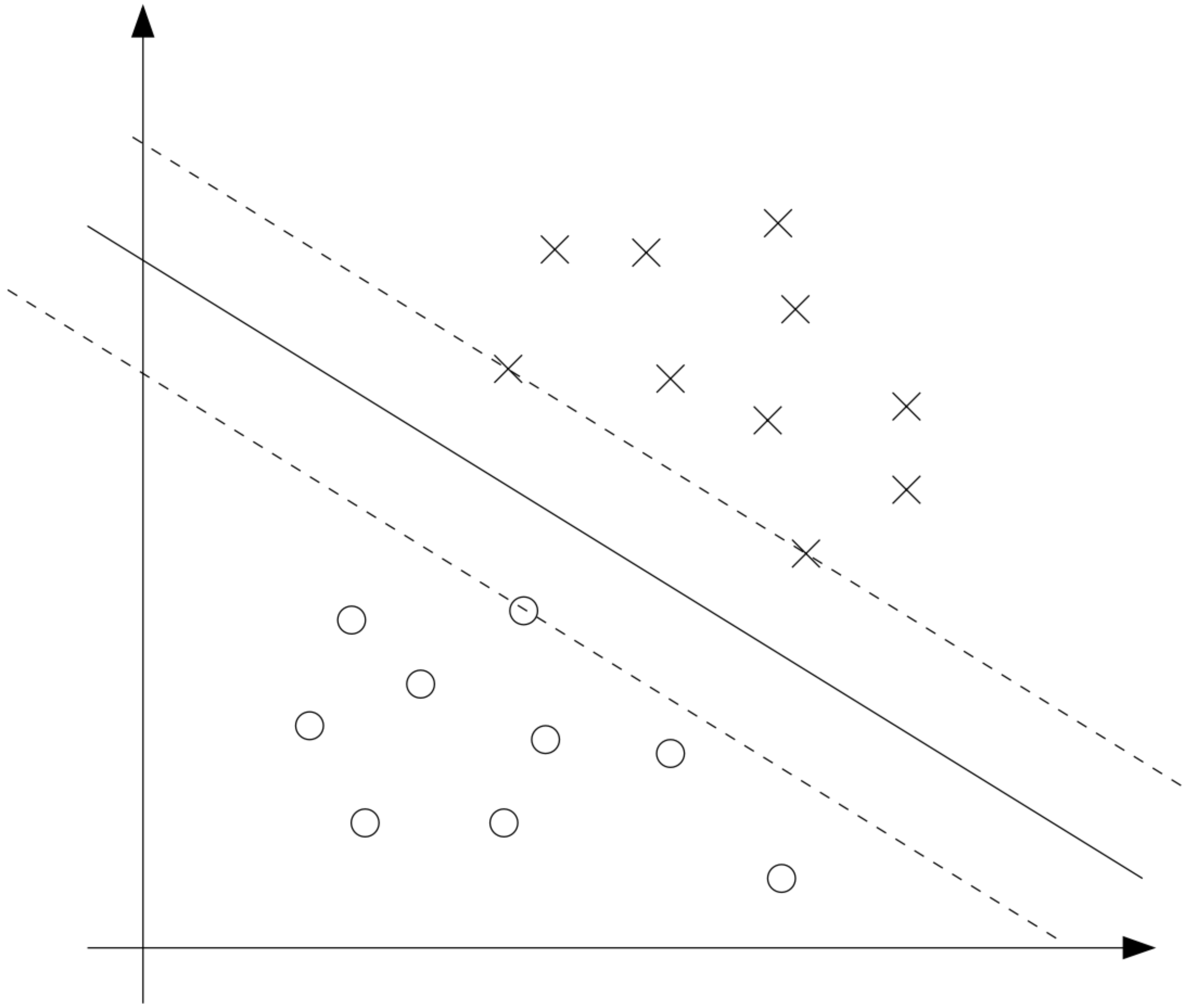
Supporting Vectors

Supporting Vectors

$$\alpha_i^* g_i(w^*) \quad = \quad 0, \quad i = 1, \ldots, k$$

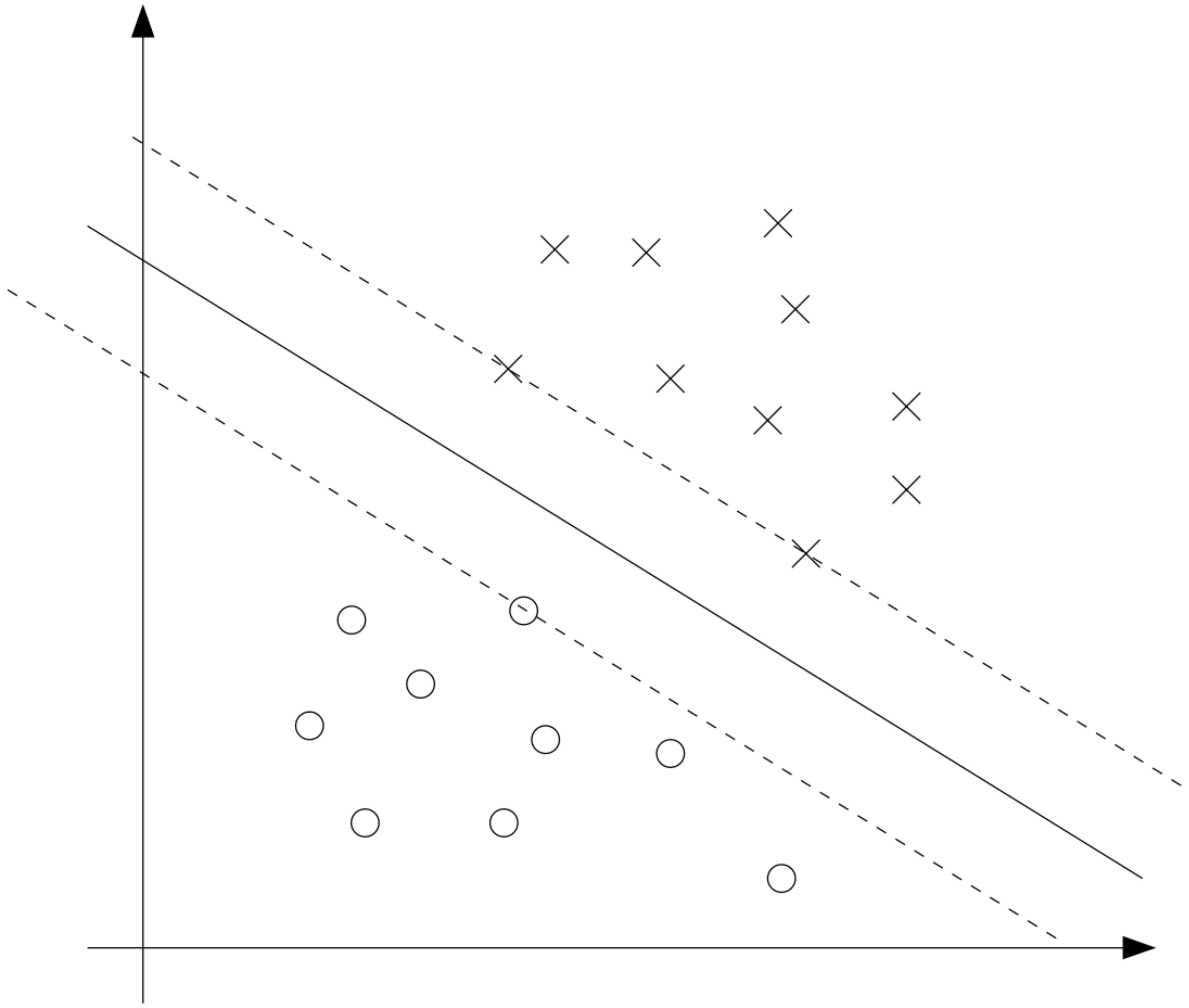
Supporting Vectors

$$\alpha_i^* g_i(w^*) \quad = \quad 0, \;\; i = 1, \ldots, k$$



Supporting Vectors

$$\alpha_i^* g_i(w^*) \quad = \quad 0, \quad i = 1, \ldots, k$$



Only the 3 points have non-zero α_i, and they are called supporting vectors

Lagrangian for SVM

Lagrangian for SVM

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

Lagrangian for SVM

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

The dual optimization problem

$$\max_{\alpha,\beta\,:\,\alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha, \beta) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_{w} \mathcal{L}(w, \alpha, \beta)$$

Lagrangian for SVM

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

The dual optimization problem

$$\max_{\alpha, \beta \,:\, \alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha, \beta) = \max_{\alpha, \beta \,:\, \alpha_i \geq 0} \min_{w} \mathcal{L}(w, \alpha, \beta)$$

$$\nabla_w \mathcal{L}(w, b, \alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0$$

Lagrangian for SVM

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

The dual optimization problem

$$\max_{\alpha,\beta\,:\,\alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha,\beta) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_{w} \mathcal{L}(w,\alpha,\beta)$$

$$\nabla_w \mathcal{L}(w,b,\alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0 \qquad w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

Lagrangian for SVM

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

The dual optimization problem

$$\max_{\alpha,\beta\,:\,\alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha,\beta) = \max_{\alpha,\beta\,:\,\alpha_i \geq 0} \min_{w} \mathcal{L}(w,\alpha,\beta)$$

$$\nabla_w \mathcal{L}(w,b,\alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0$$

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

$$\frac{\partial}{\partial b}\mathcal{L}(w,b,\alpha) = \sum_{i=1}^{n} \alpha_i y^{(i)} = 0$$

Lagrangian for SVM

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \left[y^{(i)}(w^T x^{(i)} + b) - 1 \right]$$

The dual optimization problem

$$\max_{\alpha, \beta : \alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha, \beta) = \max_{\alpha, \beta : \alpha_i \geq 0} \min_{w} \mathcal{L}(w, \alpha, \beta)$$

$$\nabla_w \mathcal{L}(w, b, \alpha) = w - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0 \qquad w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} \qquad \frac{\partial}{\partial b} \mathcal{L}(w, b, \alpha) = \sum_{i=1}^{n} \alpha_i y^{(i)} = 0$$

$$\theta(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j (x^{(i)})^T x^{(j)}$$

The Dual Problem of SVM

The Dual Problem of SVM

$$
\begin{aligned}
\max_\alpha \quad & W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle \\
\text{s.t.} \quad & \alpha_i \geq 0, \quad i = 1, \ldots, n \\
& \sum_{i=1}^{n} \alpha_i y^{(i)} = 0,
\end{aligned}
$$

The Dual Problem of SVM

$$
\begin{aligned}
\max_\alpha \quad & W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle \\
\text{s.t.} \quad & \alpha_i \geq 0, \quad i = 1, \ldots, n \\
& \sum_{i=1}^{n} \alpha_i y^{(i)} = 0,
\end{aligned}
$$

Kernel is all we need!

The Dual Problem of SVM

$$\begin{aligned}\max_{\alpha} \quad & W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle \\ \text{s.t.} \quad & \alpha_i \geq 0, \quad i = 1, \ldots, n \\ & \sum_{i=1}^{n} \alpha_i y^{(i)} = 0,\end{aligned}$$

Kernel is all we need!

After solving α (we'll talk about how later)

The Dual Problem of SVM

$$\max_\alpha \quad W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \dots, n$$

$$\sum_{i=1}^{n} \alpha_i y^{(i)} = 0,$$

Kernel is all we need!

After solving α (we'll talk about how later)

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

The Dual Problem of SVM

$$\max_\alpha \quad W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \ldots, n$$

$$\sum_{i=1}^{n} \alpha_i y^{(i)} = 0,$$

Kernel is all we need!

After solving α (we'll talk about how later)

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

From KKT Conditions

The Dual Problem of SVM

$$\max_\alpha \quad W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \ldots, n$$

$$\sum_{i=1}^{n} \alpha_i y^{(i)} = 0,$$

Kernel is all we need!

After solving α (we'll talk about how later)

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

$$b^* = -\frac{\max_{i:y^{(i)}=-1} {w^*}^T x^{(i)} + \min_{i:y^{(i)}=1} {w^*}^T x^{(i)}}{2}$$

From KKT Conditions

The Dual Problem of SVM

$$\max_\alpha \quad W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle$$

$$\text{s.t.} \quad \alpha_i \geq 0, \quad i = 1, \ldots, n$$

$$\sum_{i=1}^{n} \alpha_i y^{(i)} = 0,$$

Kernel is all we need!

After solving α (we'll talk about how later)

$$w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

From KKT Conditions

$$b^* = -\frac{\max_{i:y^{(i)}=-1} w^{*T} x^{(i)} + \min_{i:y^{(i)}=1} w^{*T} x^{(i)}}{2}$$

From the original constraints

Inference

Inference

$$\begin{aligned} w^T x + b &= \left(\sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} \right)^T x + b \\ &= \sum_{i=1}^{n} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle + b. \end{aligned}$$

Inference

$$
\begin{aligned}
w^T x + b &= \left(\sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}\right)^T x + b \\
&= \sum_{i=1}^{n} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle + b.
\end{aligned}
$$

We never need to really compute w

Inference

$$
\begin{aligned}
w^T x + b &= \left(\sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} \right)^T x + b \\
&= \sum_{i=1}^{n} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle + b.
\end{aligned}
$$

We never need to really compute w

$$
\alpha_i^* g_i(w^*) = 0, \quad i = 1, \ldots, k
$$

Inference

$$
\begin{aligned}
w^T x + b &= \left(\sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}\right)^T x + b \\
&= \sum_{i=1}^{n} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle + b.
\end{aligned}
$$

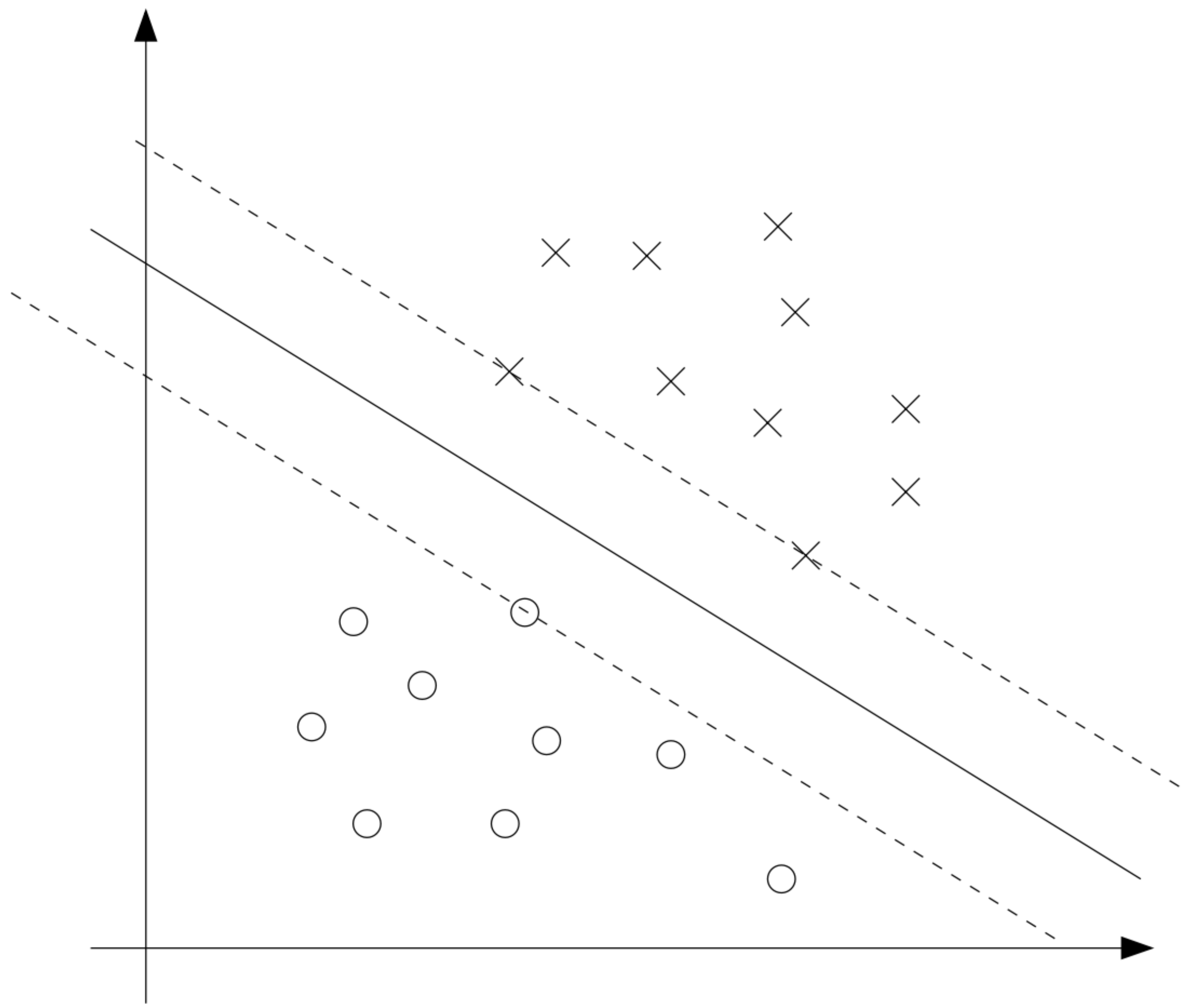
We never need to really compute w

$$\alpha_i^* g_i(w^*) = 0, \quad i = 1, \ldots, k$$

Most α_i are 0, only the supporting examples will influence the final prediction

Inference

$$w^T x + b = \left(\sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}\right)^T x + b$$
$$= \sum_{i=1}^{n} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle + b.$$

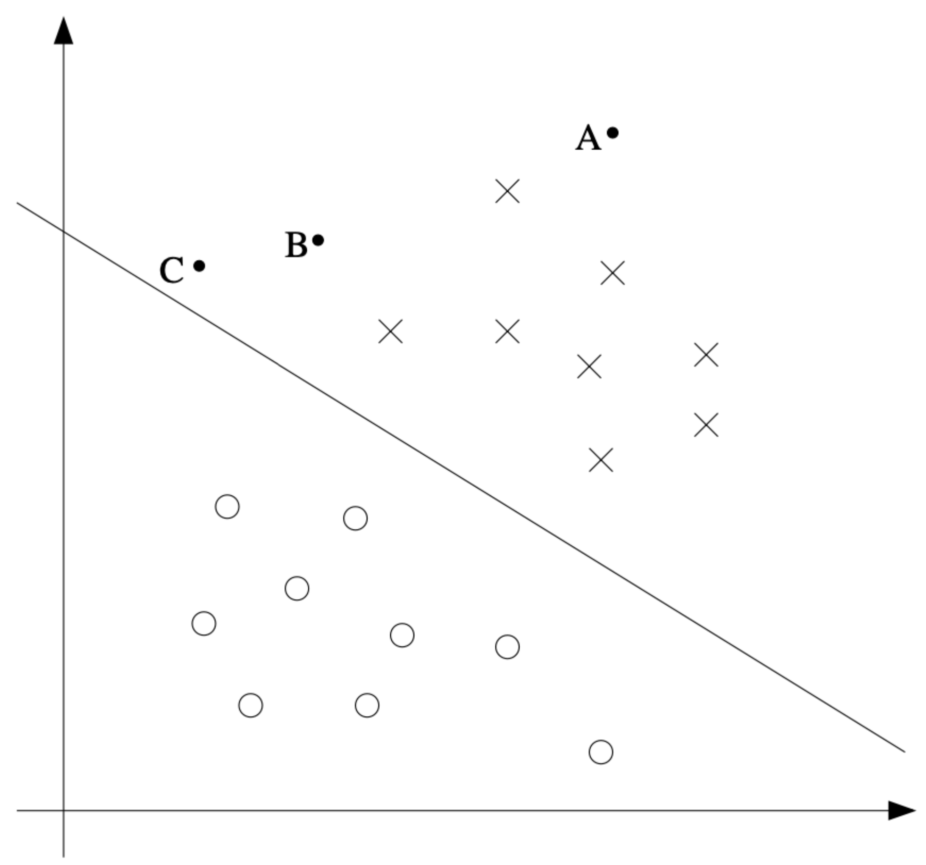


We never need to really compute w

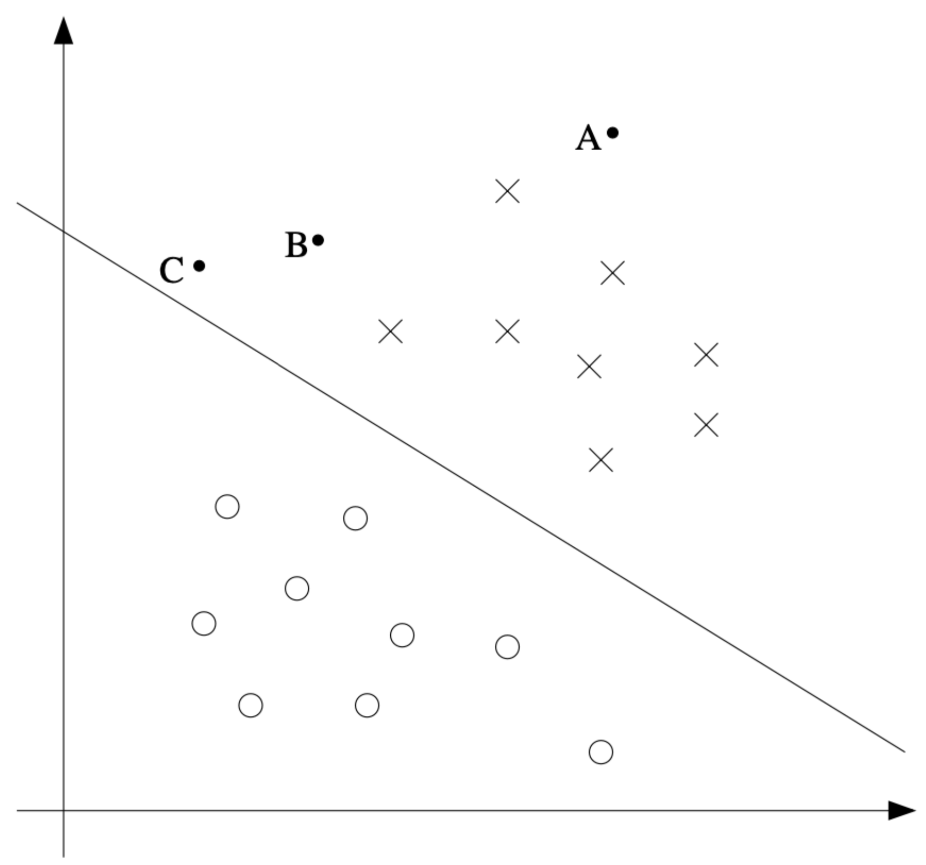
$$\alpha_i^* g_i(w^*) = 0, \quad i = 1, \ldots, k$$

Most α_i are 0, only the supporting examples will influence the final prediction

Review of the High-Level Logic

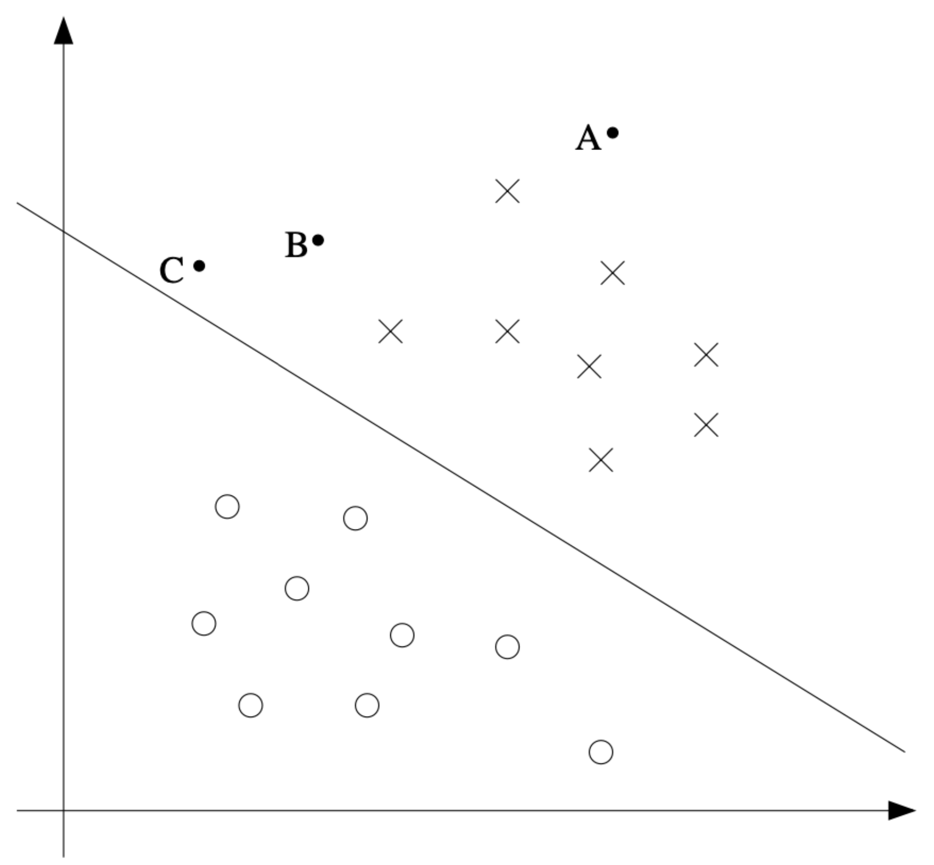


Review of the High-Level Logic



Review of the High-Level Logic

$$h_{w,b}(x) = g(w^T x + b).$$

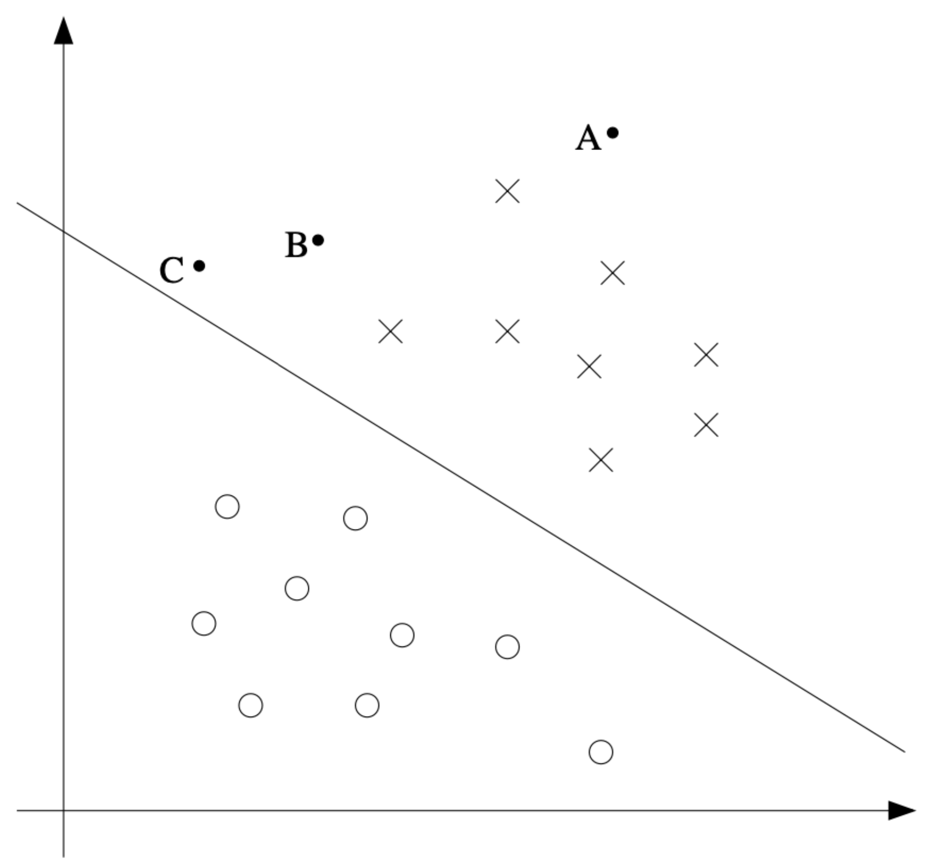


Review of the High-Level Logic

$$h_{w,b}(x) = g(w^T x + b).$$

Maximize geometric margin

$$\gamma^{(i)} = y^{(i)} \left(\left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||} \right)$$



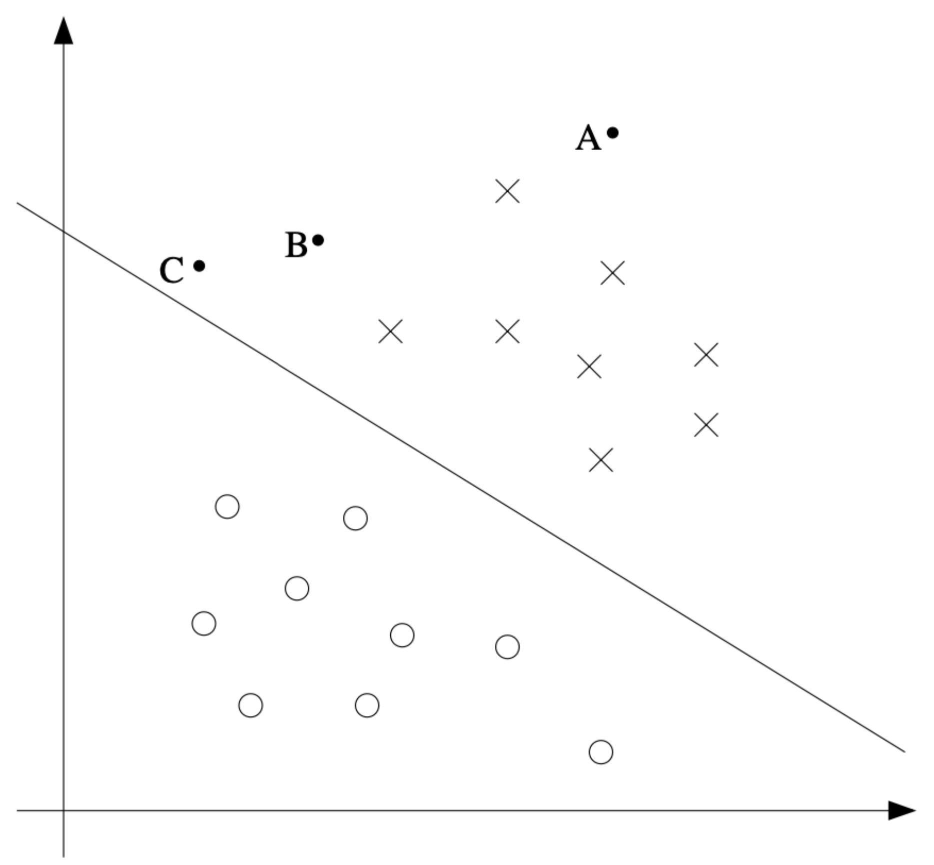
Review of the High-Level Logic

$$h_{w,b}(x) = g(w^T x + b).$$

Maximize geometric margin

Problem rewriting →

$$\gamma^{(i)} = y^{(i)} \left(\left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||} \right)$$



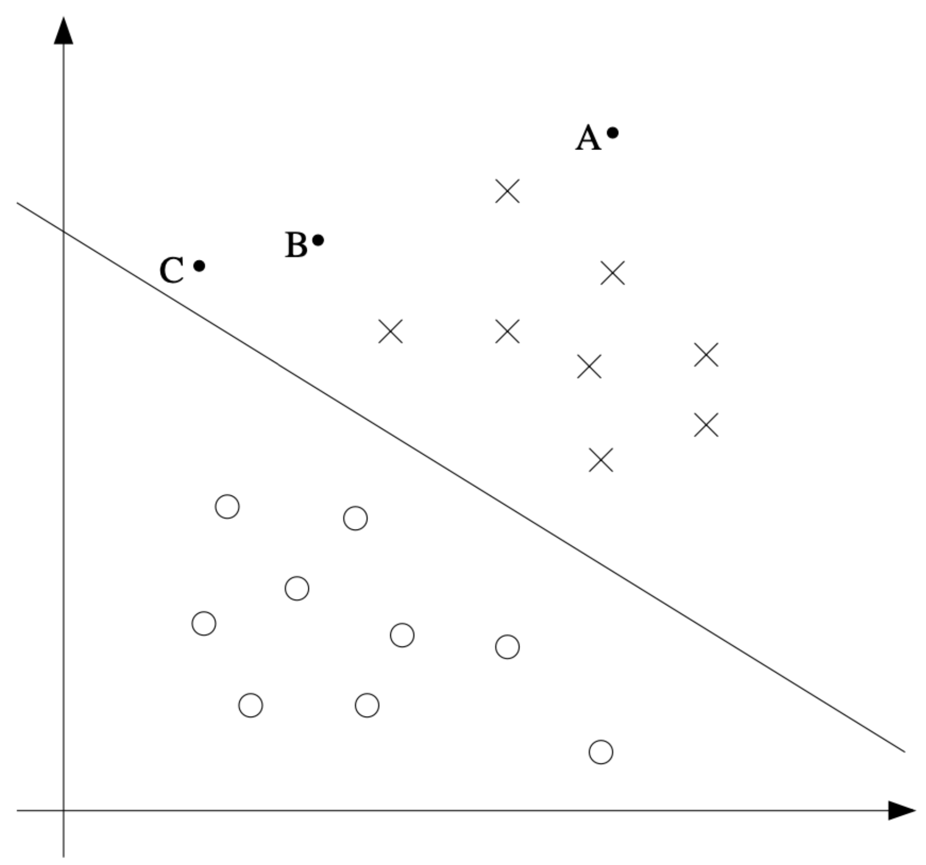
Review of the High-Level Logic

$$h_{w,b}(x) = g(w^T x + b)$$

Maximize geometric margin → (Problem rewriting) → Quadratic Optimization Problem

$$\gamma^{(i)} = y^{(i)} \left(\left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||} \right)$$

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}||w||^2 \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, n \end{aligned}$$



Review of the High-Level Logic

$$h_{w,b}(x) = g(w^T x + b)$$

Maximize geometric margin

Problem rewriting →

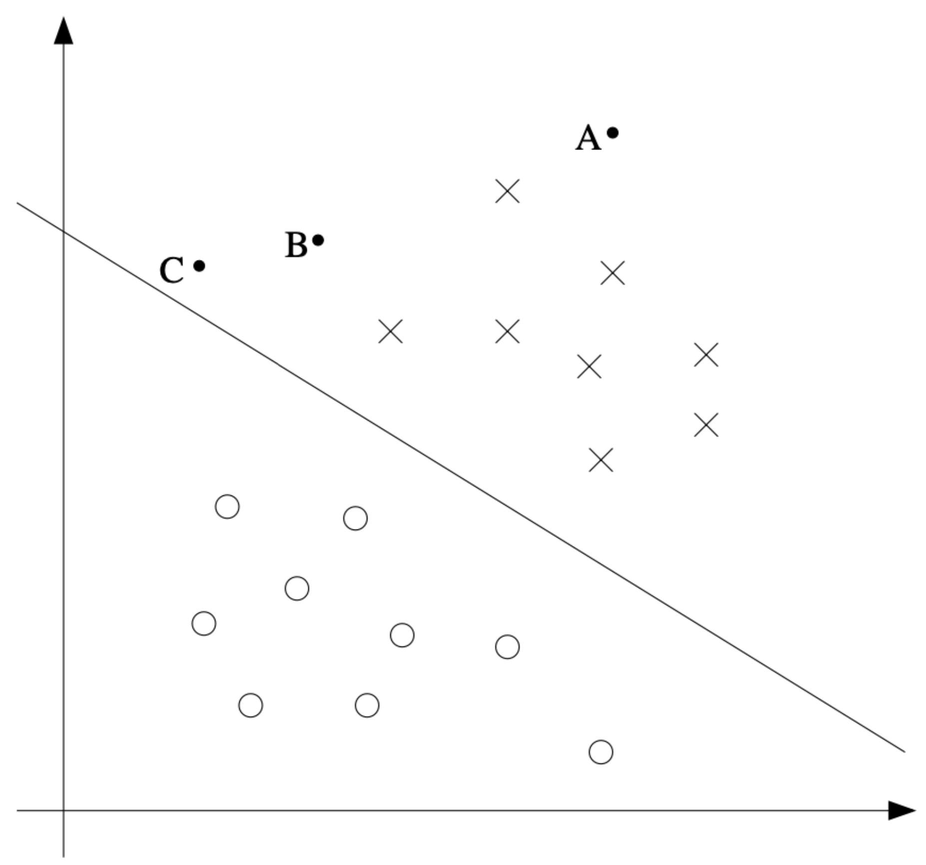
Quadratic Optimization Problem

$$\gamma^{(i)} = y^{(i)} \left(\left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||} \right)$$

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}||w||^2 \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, n \end{aligned}$$

Not suitable for non-linear cases (high-dim feature map)

Review of the High-Level Logic



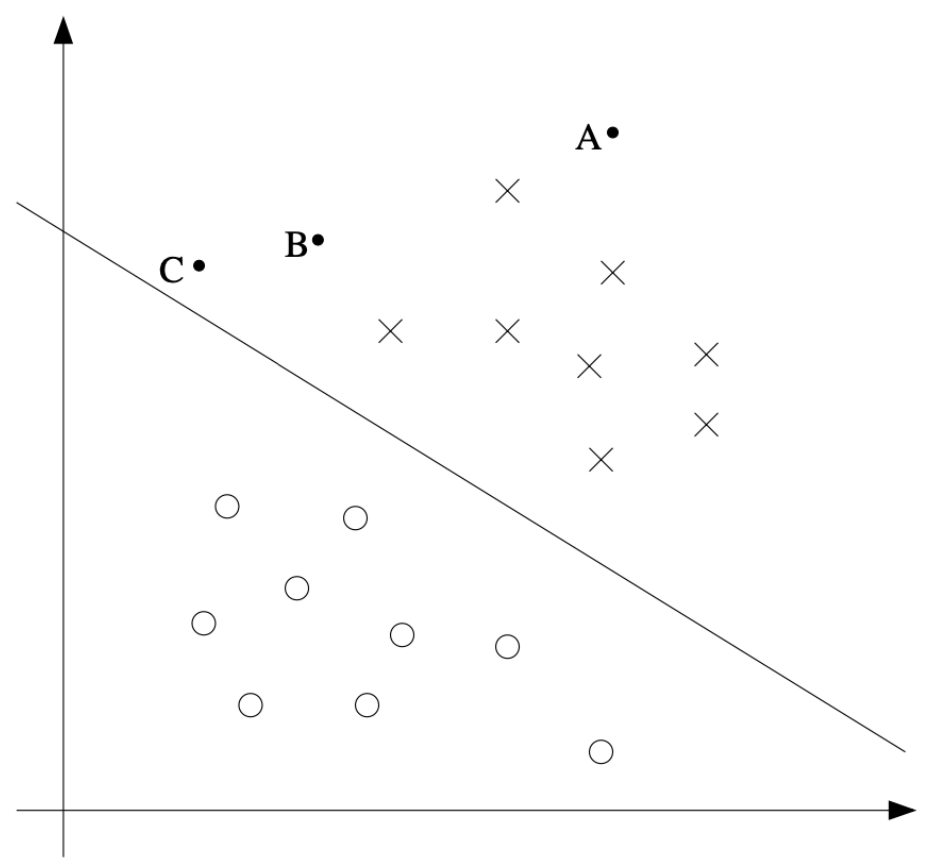
$$h_{w,b}(x) = g(w^T x + b)$$

Maximize geometric margin → *Problem rewriting* → **Quadratic Optimization Problem** → *Finding a related optimization problem that is easier*

$$\gamma^{(i)} = y^{(i)} \left(\left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||} \right)$$

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}||w||^2 \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, n \end{aligned}$$

Not suitable for non-linear cases (high-dim feature map)



Review of the High-Level Logic

$$h_{w,b}(x) = g(w^T x + b).$$

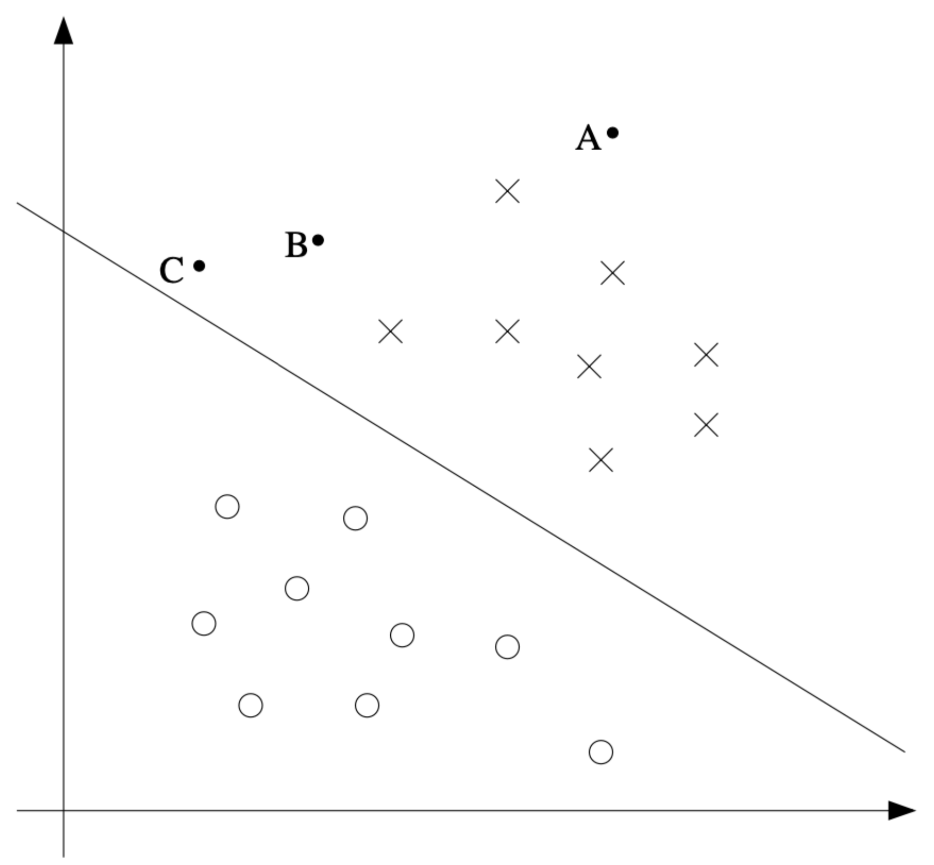
Maximize geometric margin → *Problem rewriting* → **Quadratic Optimization Problem** → *Finding a related optimization problem that is easier* → **Dual optimization problem**

$$\gamma^{(i)} = y^{(i)} \left(\left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||} \right)$$

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}||w||^2 \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, n \end{aligned}$$

$$\begin{aligned} \max_\alpha \quad & W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle \\ \text{s.t.} \quad & \alpha_i \geq 0, \quad i = 1, \ldots, n \\ & \sum_{i=1}^n \alpha_i y^{(i)} = 0, \end{aligned}$$

Not suitable for non-linear cases (high-dim feature map)



Review of the High-Level Logic

$$h_{w,b}(x) = g(w^T x + b).$$

Maximize geometric margin → (Problem rewriting) → **Quadratic Optimization Problem** → (Finding a related optimization problem that is easier) → **Dual optimization problem**

$$\gamma^{(i)} = y^{(i)} \left(\left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||} \right)$$

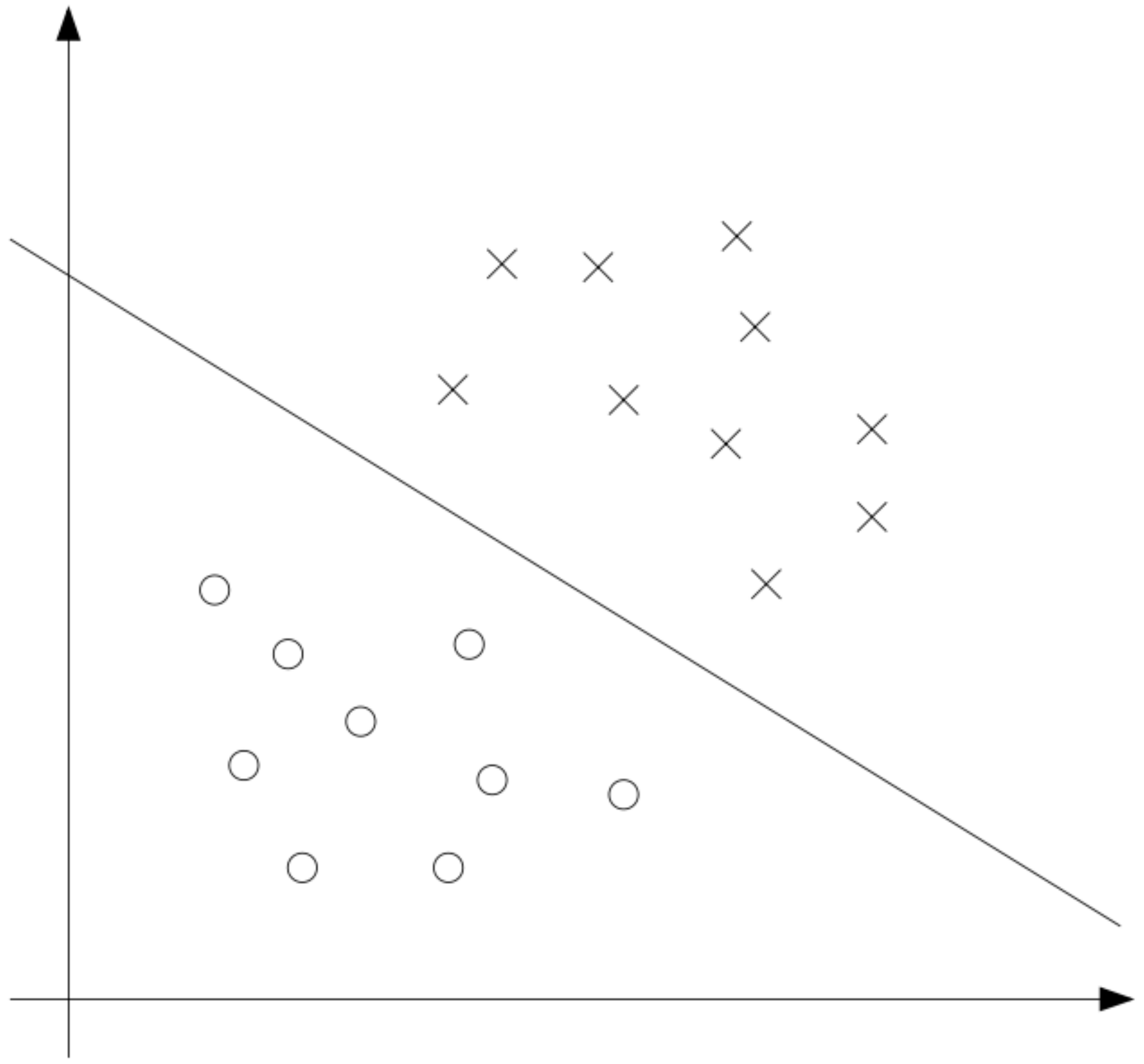
$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}||w||^2 \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i = 1, \ldots, n \end{aligned}$$

Not suitable for non-linear cases (high-dim feature map)

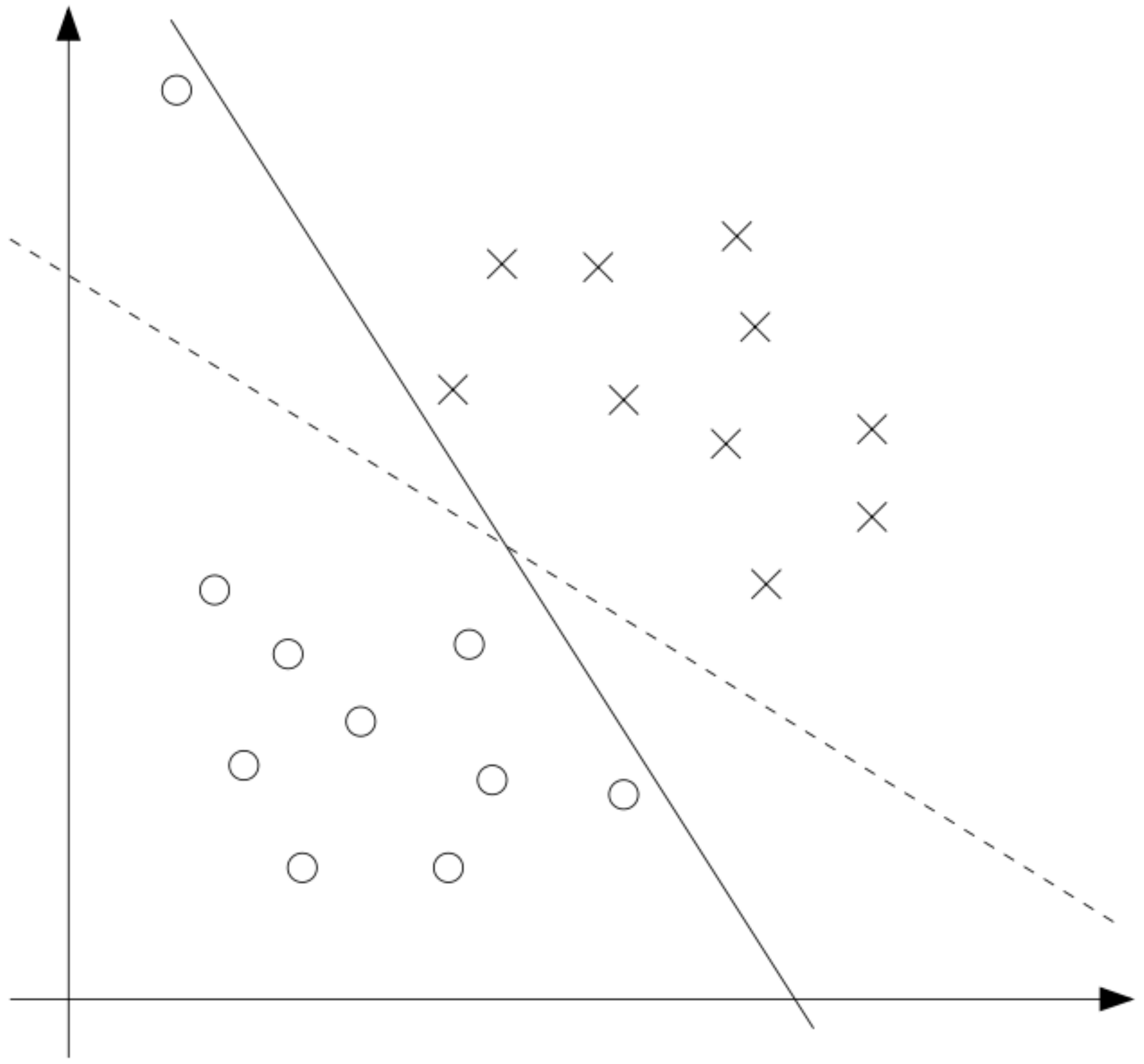
$$\begin{aligned} \max_\alpha \quad & W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle \\ \text{s.t.} \quad & \alpha_i \geq 0, \quad i = 1, \ldots, n \\ & \sum_{i=1}^n \alpha_i y^{(i)} = 0, \end{aligned}$$

Kernel makes it very flexible in non-linear cases!

The Non-Separable Case



Linearly Separable



Linearly Non-Separable

The Non-Separable Case

Primal opt problem:

$$\begin{aligned}\min_{\gamma,w,b} \quad & \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \quad i = 1, \ldots, n \\ & \xi_i \geq 0, \quad i = 1, \ldots, n.\end{aligned}$$

Dual opt problem

$$\begin{aligned}\max_{\alpha} \quad & W(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y^{(i)}y^{(j)}\alpha_i\alpha_j\langle x^{(i)}, x^{(j)}\rangle \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, n \\ & \sum_{i=1}^{n}\alpha_i y^{(i)} = 0,\end{aligned}$$

Thank You!

Q & A