Modul 5: Support Vector Machine

03 SVM for Non-linearly Separable Data

IF3270 - Pembelajaran Mesin
(Machine Learning)

Fariska Z. Ruskanda, S.T., M.T.
(fariska@informatika.org)

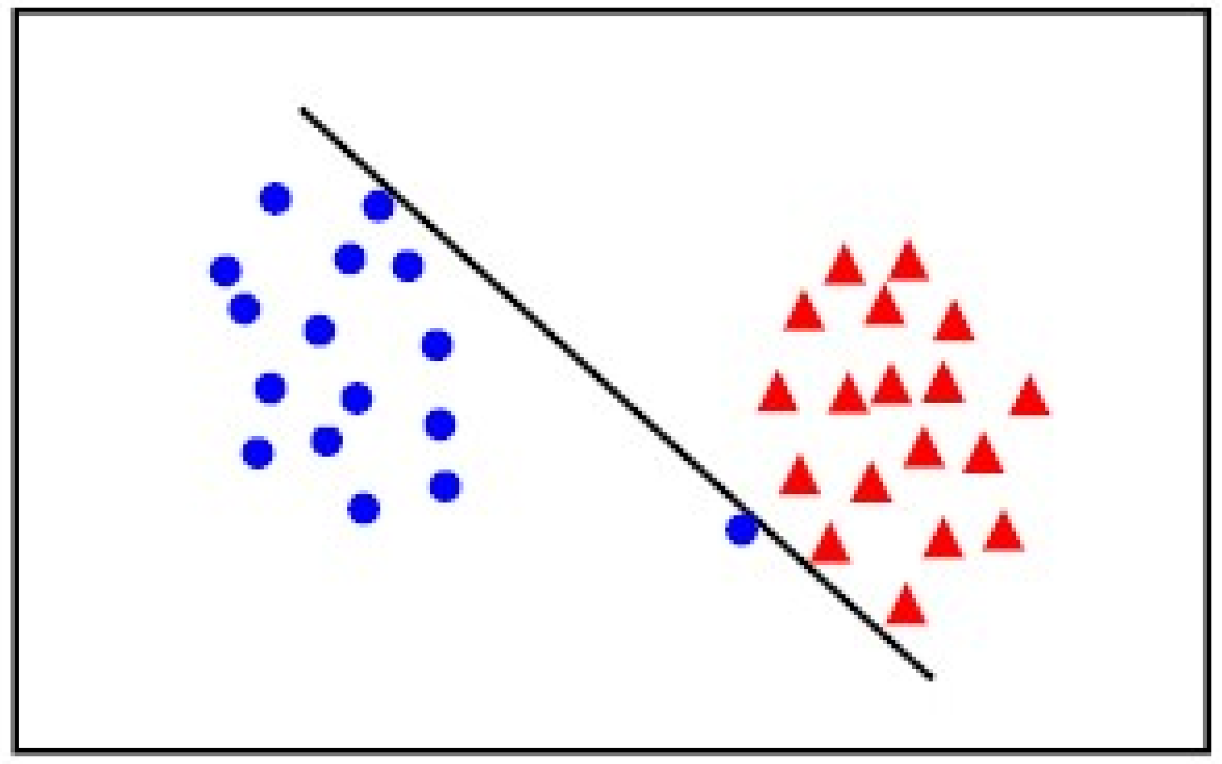
KK IF -Teknik Informatika - STEI ITB

Outline

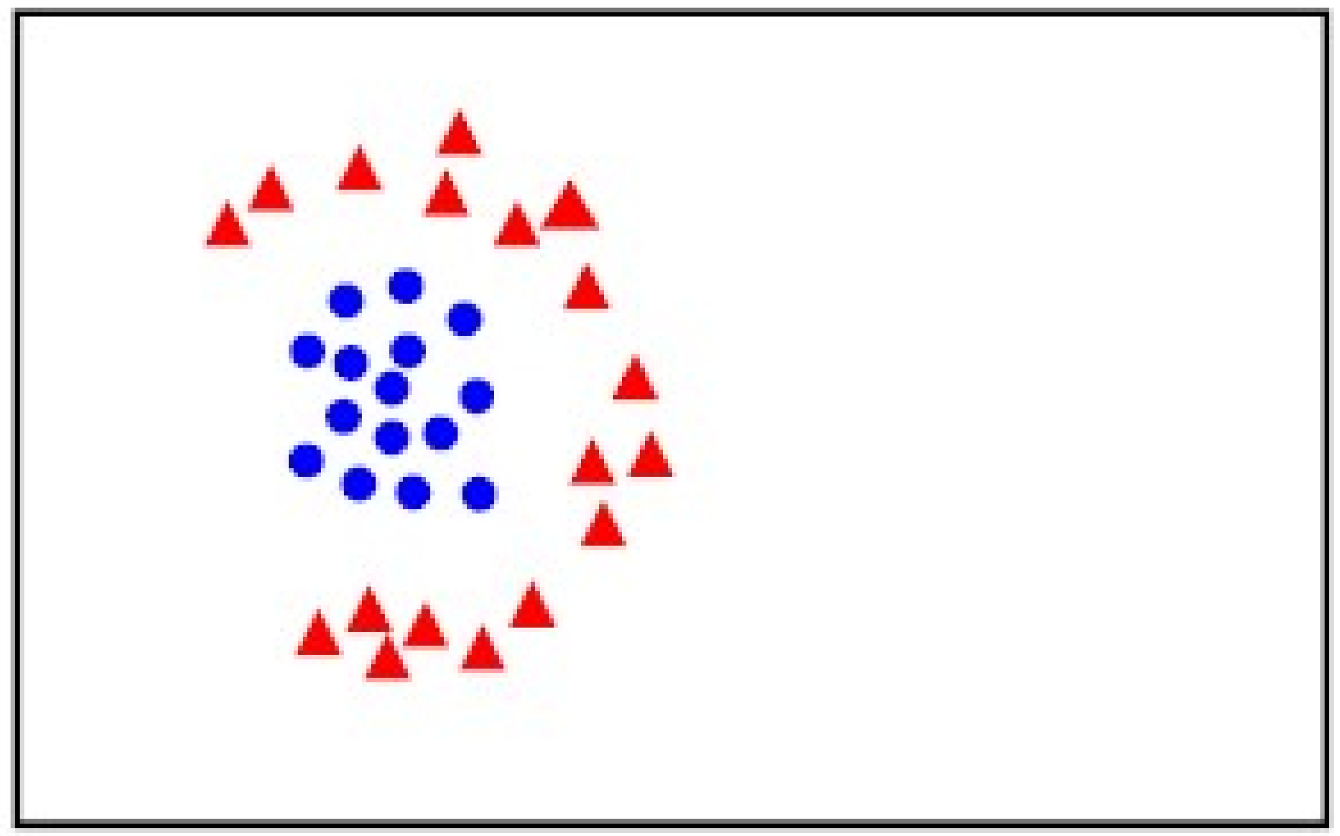
- Slack Variable
- Non-linearly Separable
- Optimization Problem
- SVM for Non-linearly Separable Data
- Non-linear Boundary Transformation

Non-Linearly Separable

- Existence of the noise



- Nature of data : Non linear boundary



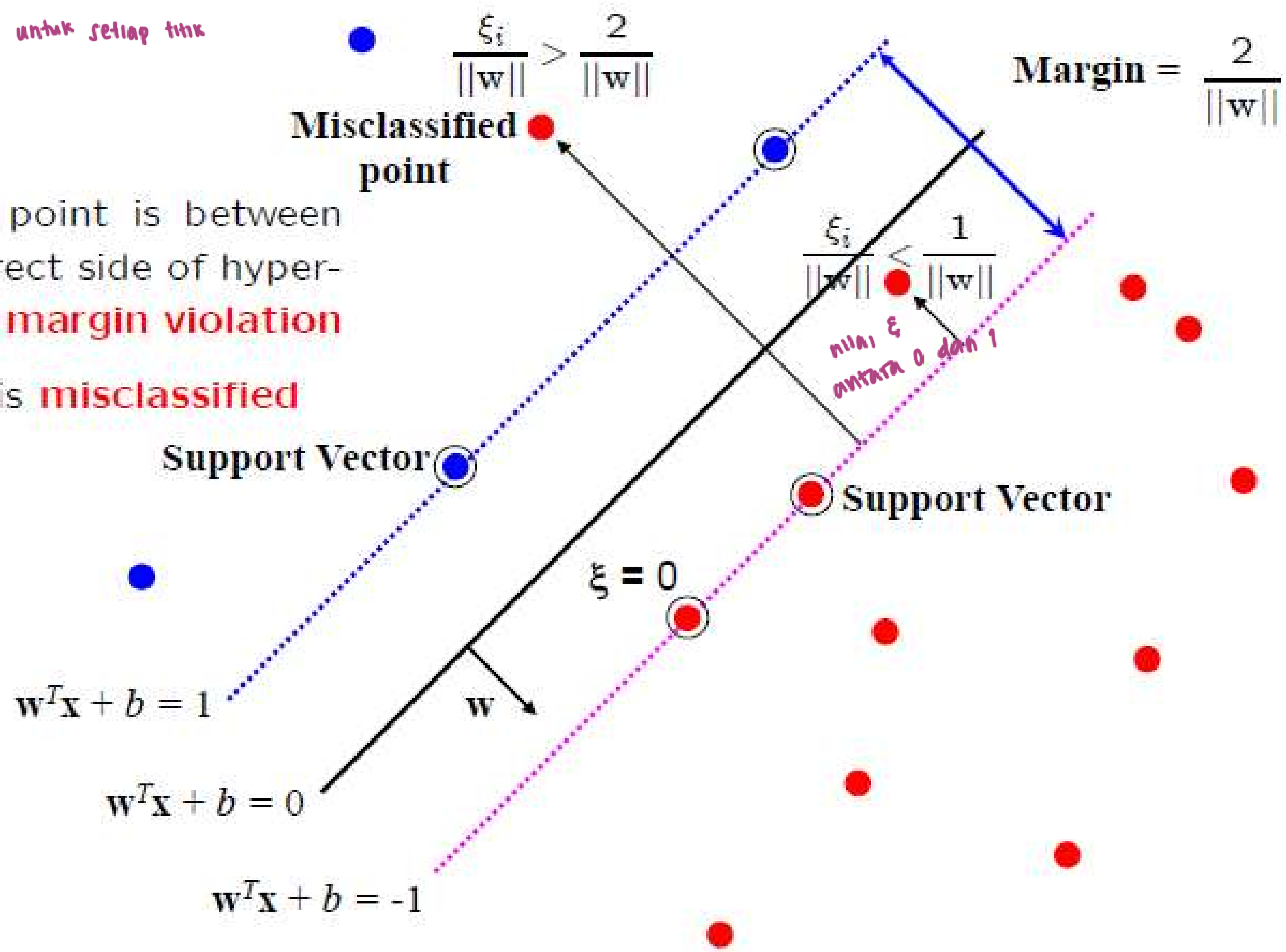
Slack Variable

nilai error

$$\xi_i \geq 0$$

untuk setiap titik

- for $0 < \xi \leq 1$ point is between margin and correct side of hyper-plane. This is a **margin violation**
- for $\xi > 1$ point is **misclassified**



https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf

Noise in Data

$$\text{Minimize} \ : \ \Phi(\mathbf{w}) = \tfrac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$\text{Subject to} \ : \ d_i(\mathbf{w}^T\mathbf{x_i} + b) \geq 1 \quad \forall i$$

$$\text{Introduce slack variables} \quad \xi_i \geq 0$$

$$\text{Minimize} \ : \ \Phi(\mathbf{w}) = \tfrac{1}{2}\mathbf{w}^T\mathbf{w}$$

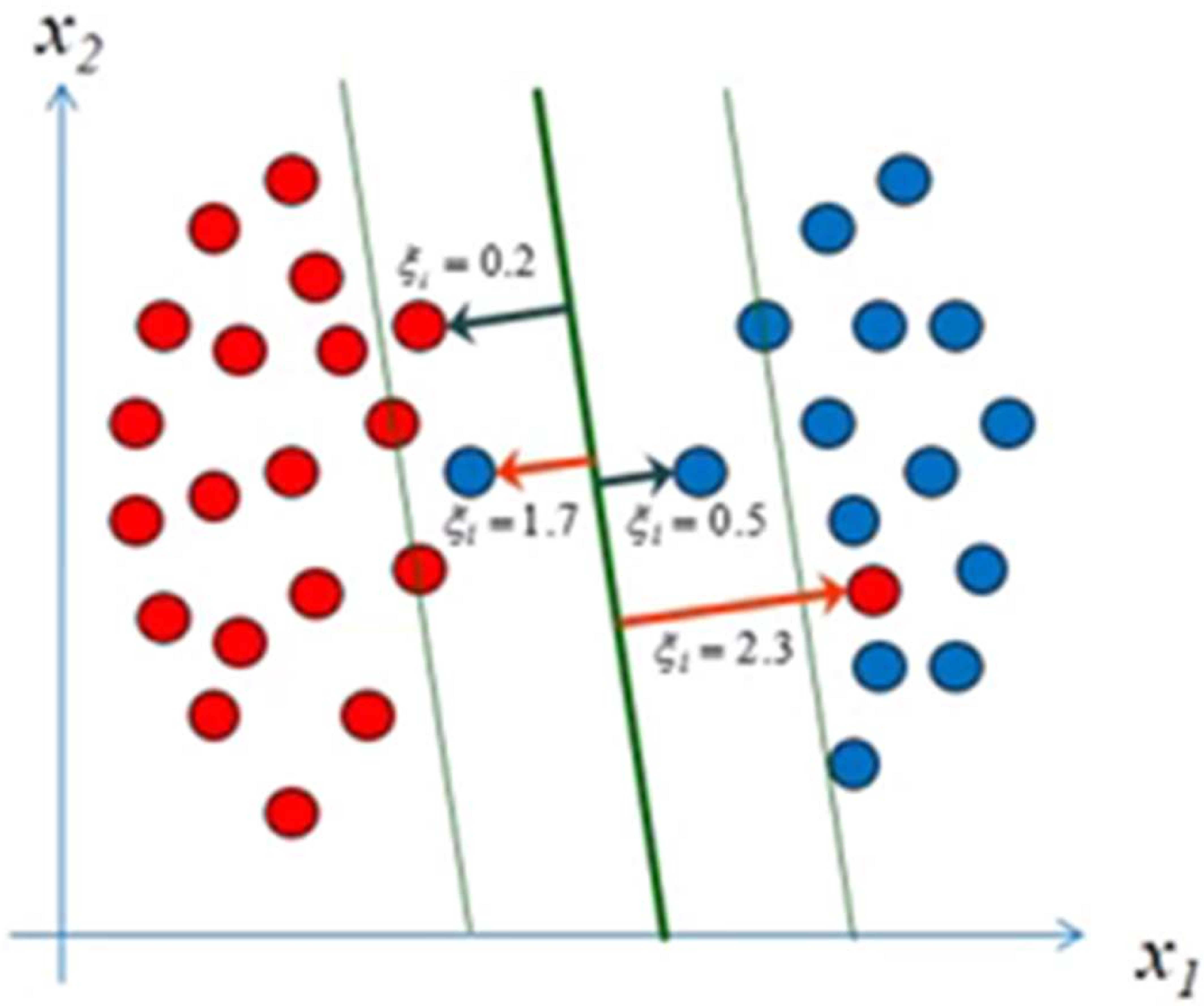
$$\text{Subject to} \ : \ d_i(\mathbf{w}^T\mathbf{x_i} + b) \geq 1 - \xi_i \quad \forall i$$

Also minimize training error $\sum_{i=1}^{N} \mathrm{I}(\xi_i \geq 1)$ or $\sum_{i=1}^{N} \xi_i$

$$\text{Minimize} \ : \ \Phi(\mathbf{w}) = \tfrac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N} \xi_i$$

$$\text{Subject to} \ : \ d_i(\mathbf{w}^T\mathbf{x_i} + b) \geq 1 - \xi_i \ ; \quad \xi_i \geq 0 \ , \quad \forall i$$

parameter yg menunjukkan
penalti thdp error



Dual Form dengan Slack

- Forming the Lagrangian and converting to dual, we get:

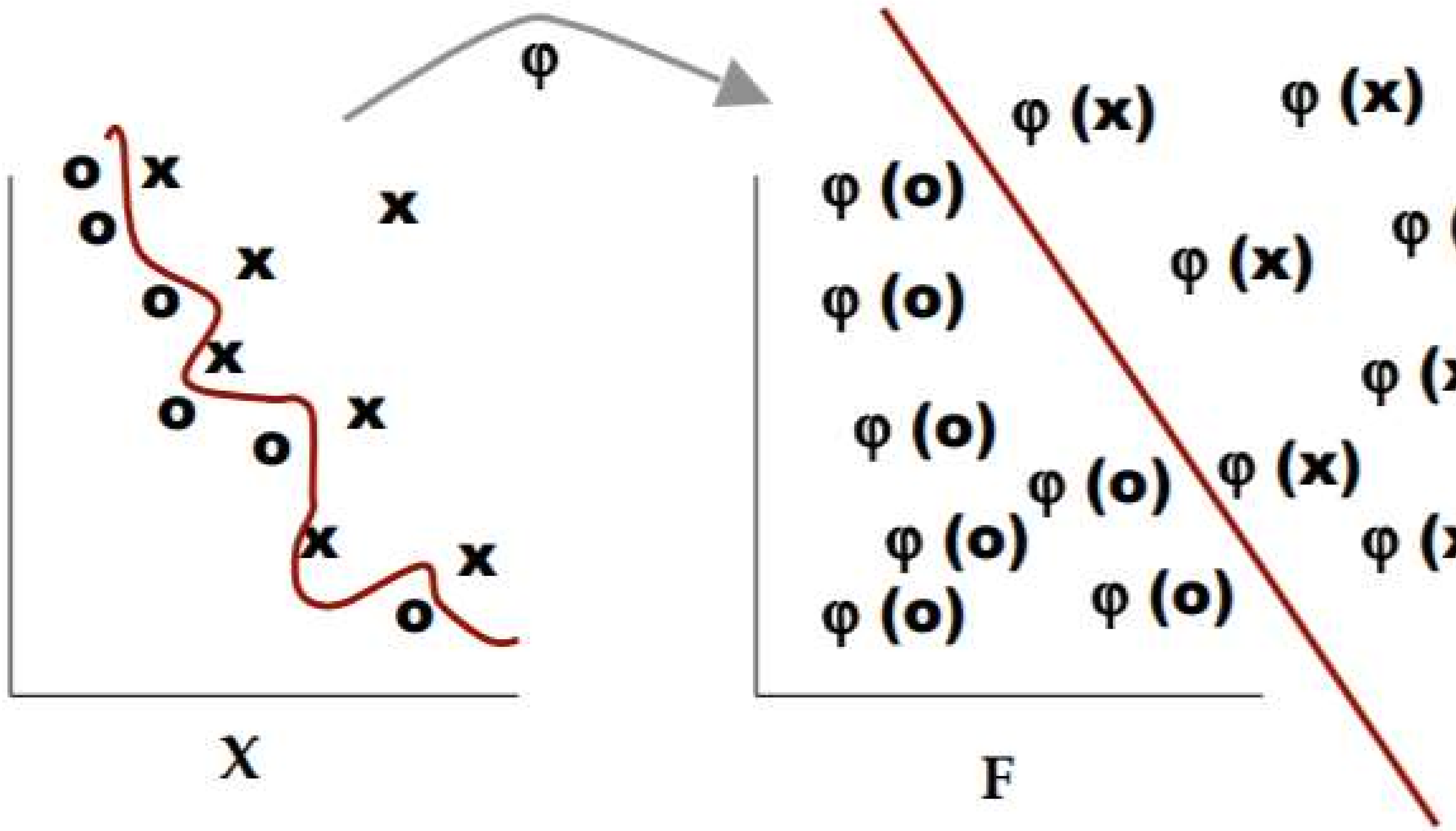
$$Q(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i - \tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \mathbf{x_i}^T \mathbf{x}_j$$

$$\text{Subject to} \quad 0 \le \alpha_i \le C \quad \forall_i \quad \text{and} \quad \sum_{i=1}^{N} \alpha_i d_i = 0$$

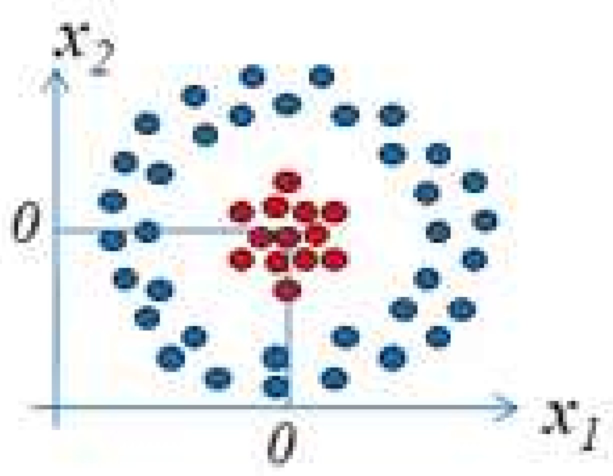
$d = y$

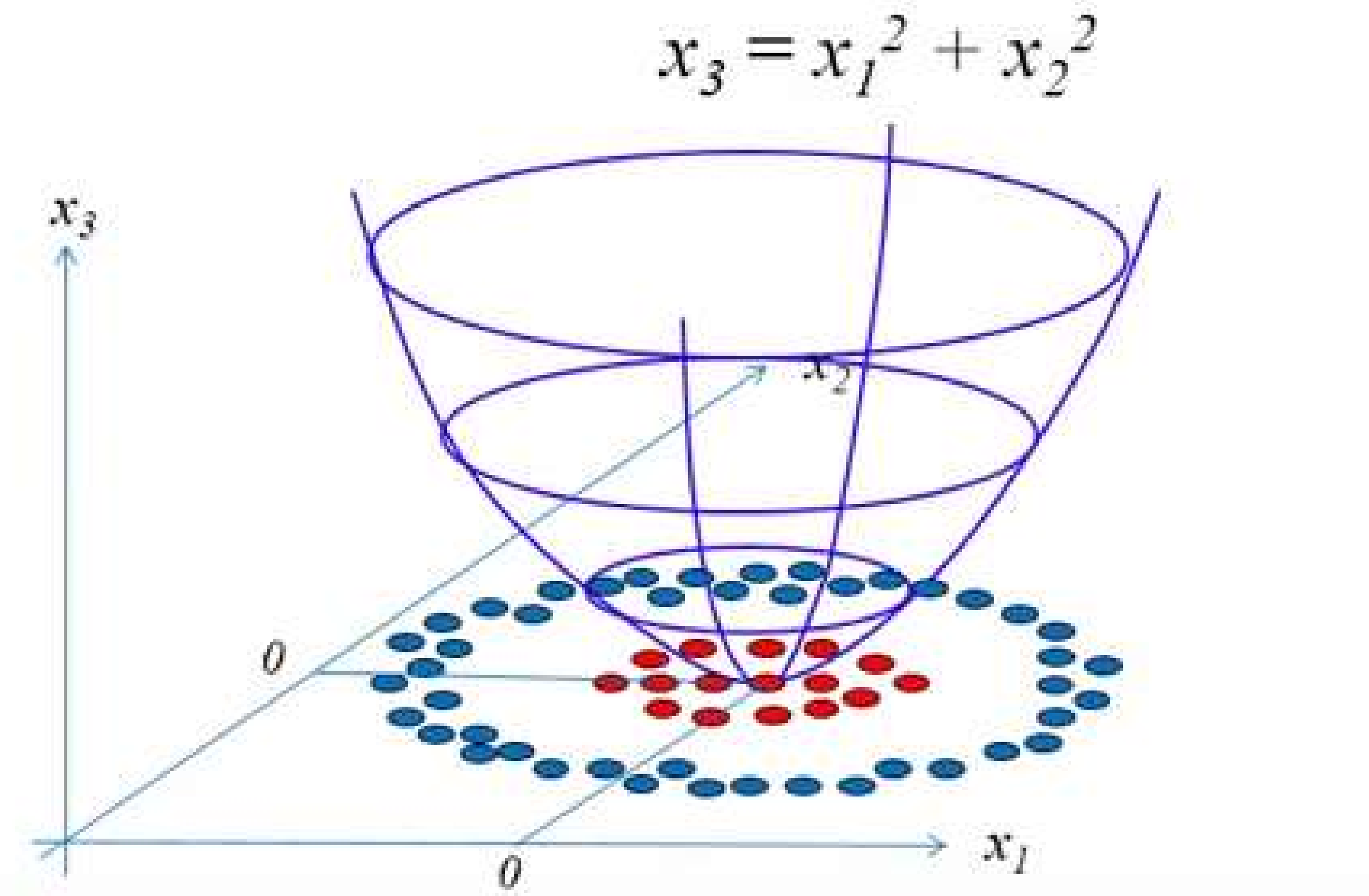
- Note that neither the slack variables, nor their Lagrange multipliers appear in the dual.
- The only change is the additional constraint on α_i
- The parameter C controls the relative weight between training error and the VC dimension.

Non-Linear Boundary Transformation



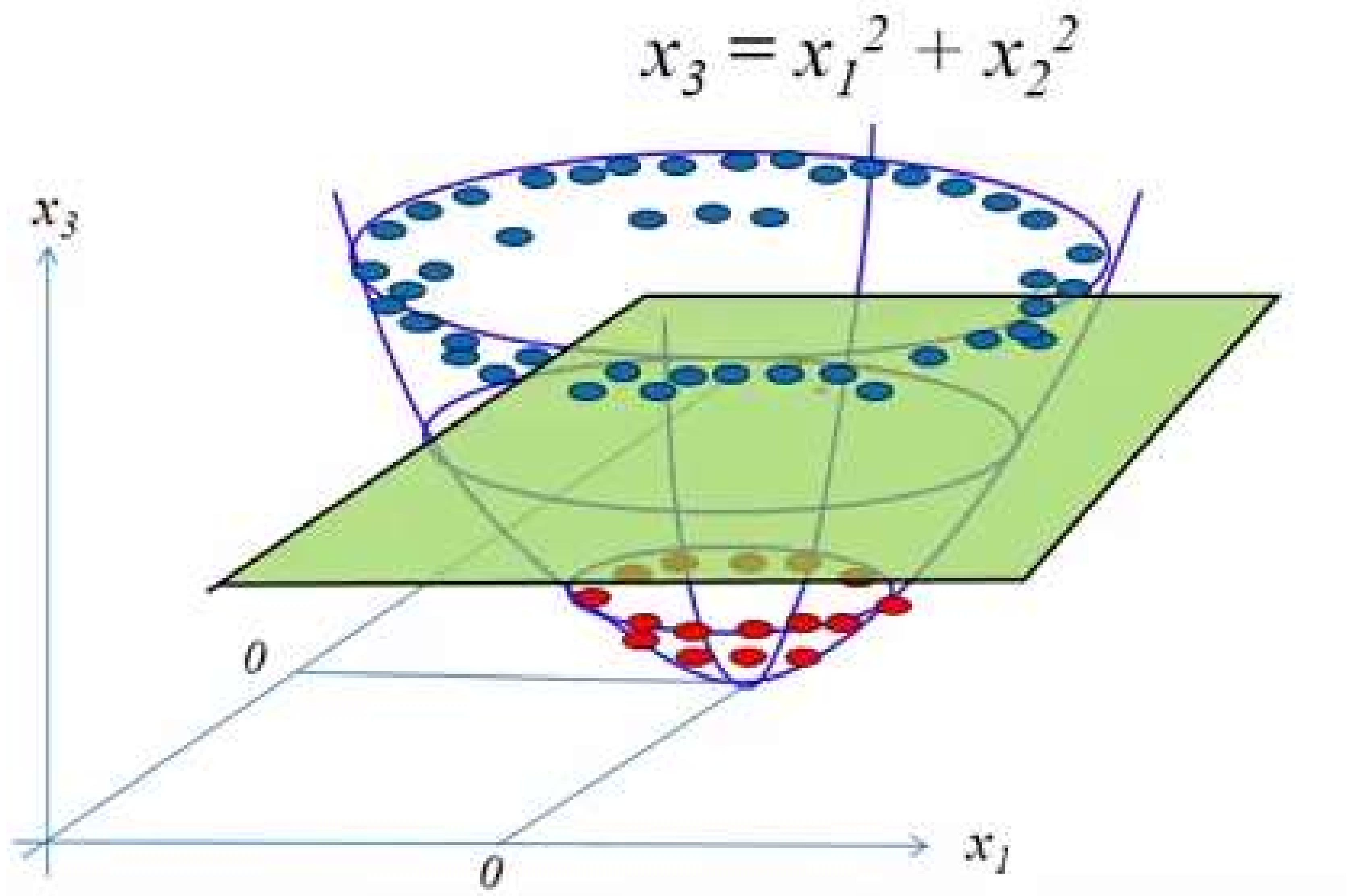
Non-Linear Boundary Transformation





Lower Dimension: Non-linearly Separable Data

Higher Dimension: Linearly Separable Data

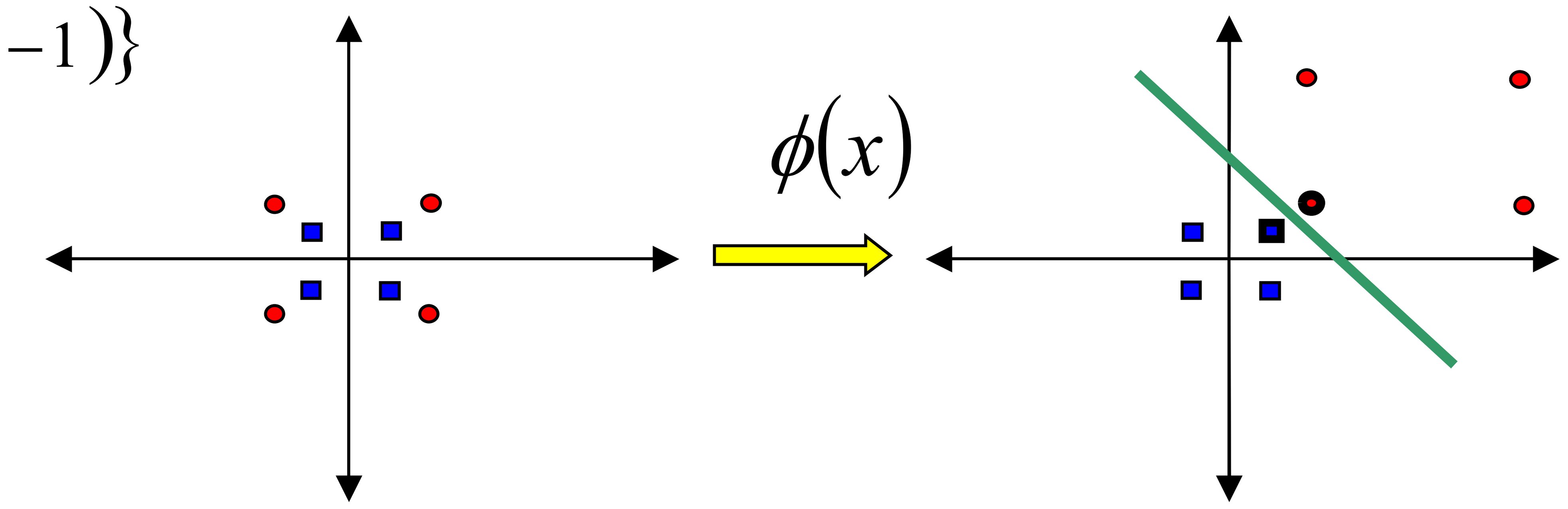


SVM pada *non-linearly separable data (2)*

- Contoh:

Misalkan dataset

  - Data kelas positif $\{(2,2),(2,-2),(-2,2),(-2,-2)\}$
  - Data kelas negatif $\{(1,1),(1,-1),(-1,1),(-1,-1)\}$



$$\phi(x_1,x_2) = \begin{cases} \sqrt{{x_1}^2 + {x_2}^2} > 2 \rightarrow (4 - x_2 + |x_1| - |x_2|, 4 - x_1 + |x_1| - |x_2|) \\ \sqrt{{x_1}^2 + {x_2}^2} \leq 2 \rightarrow (x_1, x_2) \end{cases}$$

- Dengan transformasi diperoleh
  - Data kelas positif $\{(2,2),(6,2),(6,6),(2,6)\}$
  - Data kelas negatif $\{(1,1),(1,-1),(-1,1),(-1,-1)\}$

SVM pada *non-linearly separable data (3)*

Klasifikasi: $f(x) = \sum_{i=1}^{ns} \alpha_i y_i x_i . x + b$ ➡ $\boxed{f(x) = \sum_{i=1}^{ns} \alpha_i y_i \phi(x_i) \phi(x) + b}$

- Sulit untuk mengetahui $\phi(x)$ dan feature space biasanya memiliki dimensi yang lebih besar
- Solusinya "*kernel trick*", yang perlu diketahui adalah $K(x_i, x) = \phi(x_i)\phi(x)$
- Dengan fungsi K (fungsi Kernel), maka fungsi $\phi(x)$ tidak perlu diketahui

Klasifikasi: $\boxed{f(x) = \sum_{i=1}^{ns} \alpha_i y_i K(x_i, x) + b}$

Fungsi Kernel yang umum digunakan:

Linear Kernel → $K(x_i, x_j) = x_i^T x_j$

Polynomial kernel → $K(x_i, x_j) = (\gamma . x_i^T x_j + r)^p, \gamma > 0$ bisa 2 dst

RBF kernel → $K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2), \gamma > 0$

Sigmoid kernel → $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

Example

- Suppose we have 5 (1D) data points
  - x_1=1, x_2=2, x_3=4, x_4=5, x_5=6,
  - with 1, 2, 5 as class 1 and 3, 4 as class 2 $\Rightarrow$ y_1=1, y_2=1, y_3=-1, y_4=-1, y_5=1
- We use the polynomial kernel of degree 2
  - $K(x,y) = (xy+1)^2$
  - C is set to 100
- We first find a_i (i=1, …, 5) by

kernelnya

$$\text{max.} \quad \sum_{i=1}^{5} \alpha_i - \frac{1}{2} \sum_{i=1}^{5} \sum_{j=1}^{5} \alpha_i \alpha_j y_i y_j (x_i x_j + 1)^2$$

$$\text{subject to } 100 \geq \alpha_i \geq 0, \sum_{i=1}^{5} \alpha_i y_i = 0$$

A Simple Introduction to Support Vector Machines; Martin Law; Michigan State Univ.

$-7.9992 + b = 1$

$-9.9975 + b = -1$

$-7.9968 + b = 1$

+

$-17.9943 + 2b = 0$

$2b = 17.9943$

$b \approx 9.$

kalo $\alpha = 0$ gabakal dipake jd support vector

Example

- $x_1=1$, $x_2=2$, $x_3=4$, $x_4=5$, $x_5=6$,
- $y_1=1$, $y_2=1$, $y_3=-1$, $y_4=-1$, $y_5=1$

- By using a QP solver, we get
  - $\alpha_1=0$, $\alpha_2=2.5$, $\alpha_3=0$, $\alpha_4=7.333$, $\alpha_5=4.833$
  - Note that the constraints are indeed satisfied
  - The support vectors are $\{x_2=2, x_4=5, x_5=6\}$
- The discriminant function is

$$
\begin{aligned}
f(z) &= 2.5(1)(2z+1)^2 + 7.333(-1)(5z+1)^2 + \underbrace{4.833}_{\alpha_5}\,\underbrace{(1)}_{y_5}\,\underbrace{(6z+1)^2}_{K(z,x_5)} + b \\
&= 0.6667z^2 - 5.333z + b
\end{aligned}
$$

- b is recovered by solving f(2)=1 or by f(5)=-1 or by f(6)=1, as x_2 and x_5 lie on the line $\phi(\mathrm{w})^T\phi(\mathrm{x}) + b = 1$ and x_4 lies on the line $\phi(\mathrm{w})^T\phi(\mathrm{x}) + b = -1$
- All three give b=9

$$\Longrightarrow \quad f(z) = 0.6667z^2 - 5.333z + 9$$

04 SVM for Multi-class Data

IF3270 Pembelajaran Mesin