國立高雄第一科技大学
National Kaohsiung First University of Science and Technology

Support Vector Machine (SVM)

Two Separable Classes

Speaker: Shih-Shinh Huang

Version: v011

Date: February 28, 2021

Outline

- Introduction
- Maximum Margin Formulation
- Dual Representation
- Hyperplane Derivation

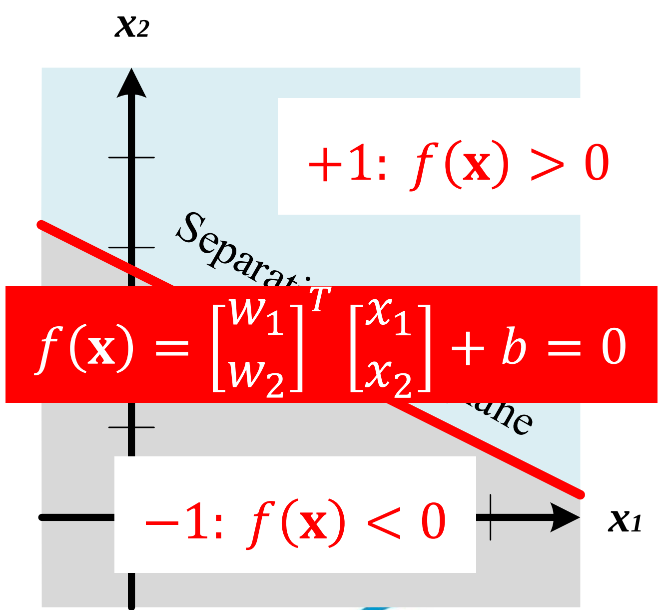
Introduction

- About SVM
  - SVM is original a binary classifier by a decision function $f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b$

$$y = \begin{cases} +1 & if \quad f(\mathrm{x}) > 0 \\ -1 & if \quad f(\mathrm{x}) < 0 \end{cases}$$

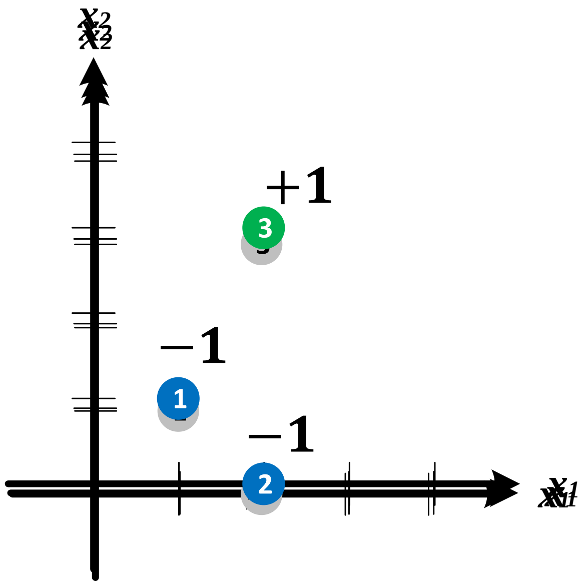
    - $\mathbf{w}$: normal vector of hyperplane
    - b: bias of hyperplane

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$



Introduction

- About SVM
  - SVM is a kind of supervised machine learning.
    - SVM hyperplane is derived by a given training dataset
    - All training samples are well labelled.



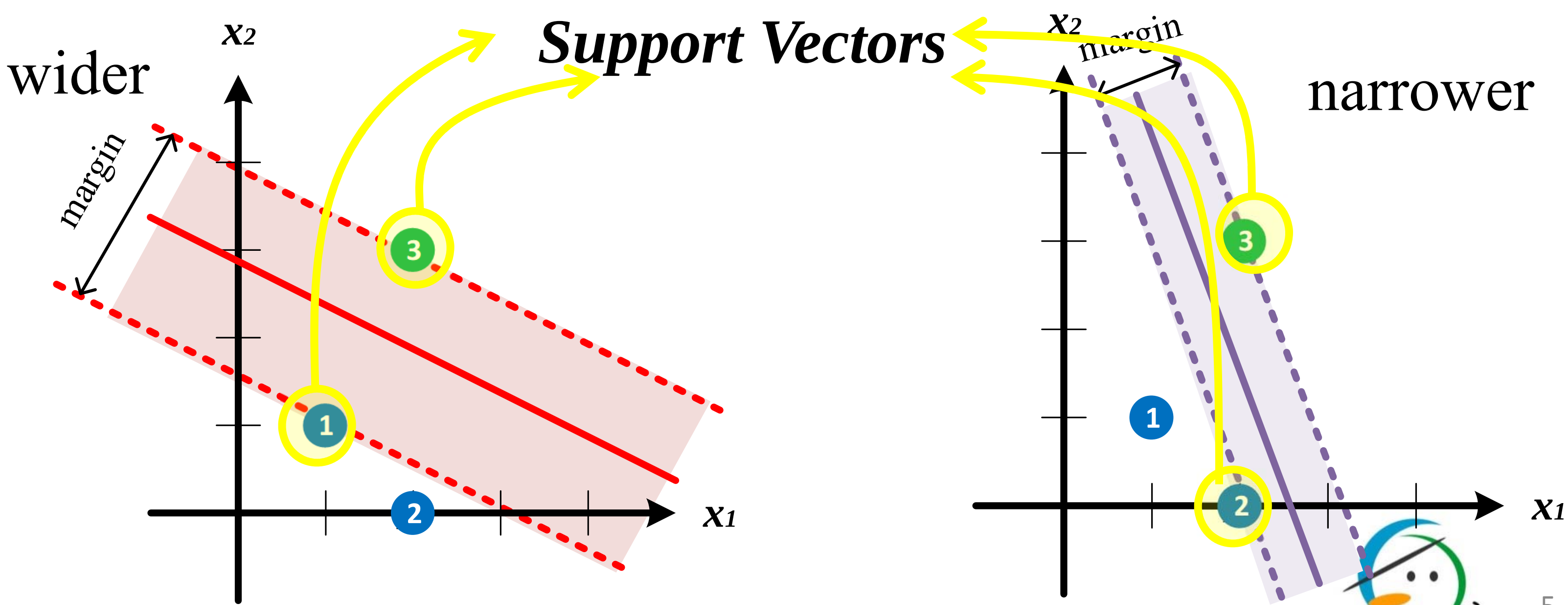
$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad t_1 = -1$$

$$\mathbf{x}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \qquad t_2 = -1$$

$$\mathbf{x}_3 = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \qquad t_3 = +1$$

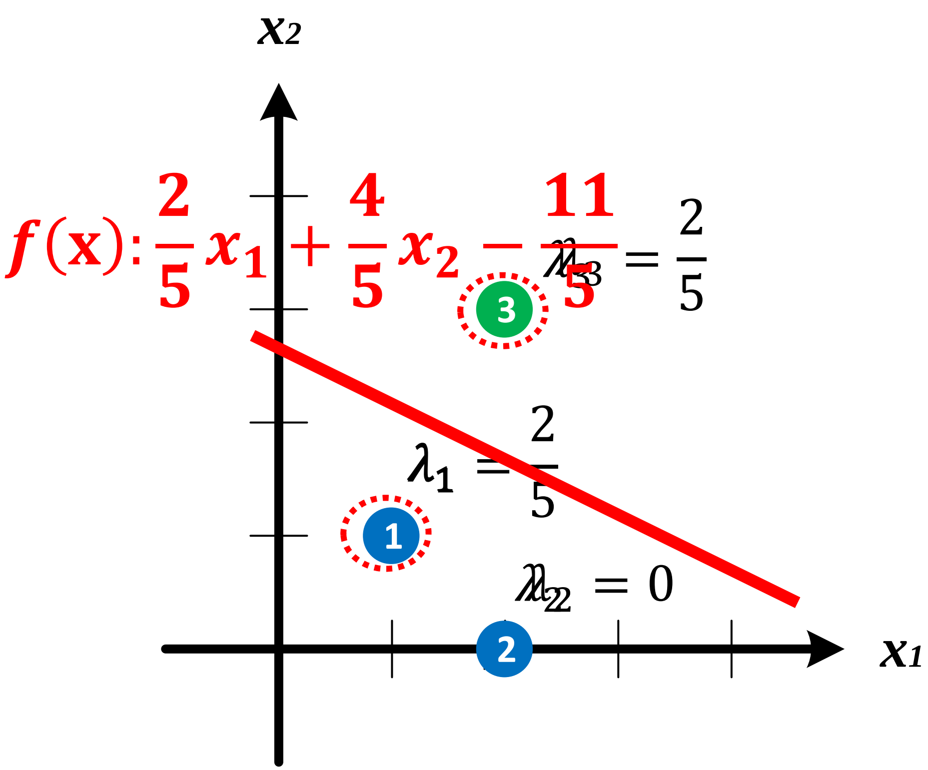
Introduction

- Idea: Maximal Margin
  - SVM hyperplane is the one that maximizes the margin around the separating hyperplane



Introduction

- Hyperplane Determination



Maximal Margin Formulation

Lagrange Multiplier Introduction

Lagrange Multiplier Solving

Normal Vector and Bias Computation

Maximum Margin Formulation

- Formulation Description
  - Given: A training dataset D consists of N training samples

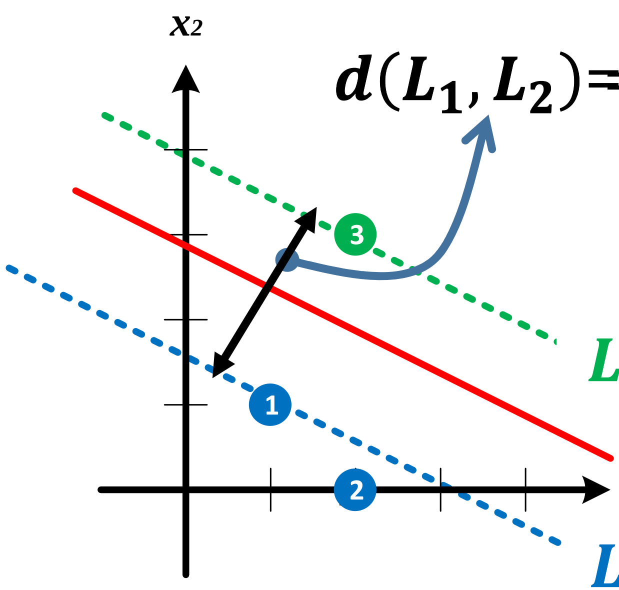
$$\boldsymbol{D} = \{\mathbf{x}_{\boldsymbol{i}}, \boldsymbol{t}_{\boldsymbol{i}}\}_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}$$

    - $\mathbf{x}_i$: feature vector of ith training sample
    - $t_i \in \{+1, -1\}$: label of ith training sample

$$\boldsymbol{D} = \left\{\left(\begin{bmatrix}\mathbf{1}\\\mathbf{1}\end{bmatrix}, -\mathbf{1}\right), \left(\begin{bmatrix}\mathbf{2}\\\mathbf{0}\end{bmatrix}, -\mathbf{1}\right), \left(\begin{bmatrix}\mathbf{2}\\\mathbf{3}\end{bmatrix}, +\mathbf{1}\right)\right\}$$

Maximum Margin Formulation

- Formulation Description
  - Goal: find an optimal hyperplane $\hat{\mathbf{w}}^T\mathbf{x} + \hat{b} = 0$ that has maximal margin



$$d(L_1, L_2) = \frac{|+1-(-1)|}{||\mathbf{w}||} = \frac{2}{||\mathbf{w}||}$$

$L_1: \mathbf{w}^T\mathbf{x} + b = +1$

$L: \mathbf{w}^T\mathbf{x} + b = 0$

$L_2: \mathbf{w}^T\mathbf{x} + b = -1$

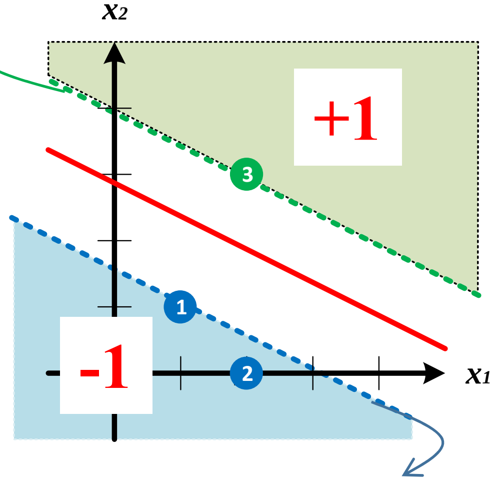
$$(\hat{\mathbf{w}}, \hat{b}) = \text{argmax}_{\mathbf{w},b} \frac{2}{||\mathbf{w}||}$$

$$(\hat{\mathbf{w}}, \hat{b}) = \text{argmin}_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2$$

Maximum Margin Formulation

- Formulation Description
  - Condition: all training samples have to be correctly classified.

$L_1: \mathbf{w}^T\mathbf{x} + b = +1$



$L_2: \mathbf{w}^T\mathbf{x} + b = -1$

$t_i = +1 \Rightarrow f(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i + b \geq +1$

$t_i = -1 \Rightarrow f(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i + b \leq -1$

$$\Rightarrow t_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0 \quad \forall\, i = 1, 2..N$$

Maximum Margin Formulation

- Formulation Description
  - Given: A training dataset $\boldsymbol{D} = \{\mathbf{x}_i, t_i\}_{\boldsymbol{i=1}}^{\boldsymbol{N}}$ consists of N training samples
  - Goal: find an optimal hyperplane $\hat{\mathbf{w}}^T\mathbf{x} + \hat{b} = 0$

$$(\hat{\mathbf{w}}, \hat{b}) = \text{argmin}_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2$$

  - Condition: all training samples have to be correctly classified.

$$t_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0 \quad \forall i = 1, 2, \ldots, N$$

Maximum Margin Formulation

- Formulation: Example
  - Given: $\boldsymbol{D} = \left\{ \left(\begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix}, -\mathbf{1} \right), \left(\begin{bmatrix} \mathbf{2} \\ \mathbf{0} \end{bmatrix}, -\mathbf{1} \right), \left(\begin{bmatrix} \mathbf{2} \\ \mathbf{3} \end{bmatrix}, +\mathbf{1} \right) \right\}$
  - Goal: $(\hat{\mathbf{w}}, \hat{b}) = \text{argmin}_{\mathbf{w},b} \frac{1}{2} ||\mathbf{w}||^2$
  - Three Conditions:

1. $(-\mathbf{1}) \times (\boldsymbol{w_1} + \boldsymbol{w_2} + \boldsymbol{b}) - \mathbf{1} \geq \mathbf{0}$

2. $(-\mathbf{1}) \times (\mathbf{2}\boldsymbol{w_1} + \boldsymbol{b}) - \mathbf{1} \geq \mathbf{0} \quad \geq \mathbf{0}$

3. $(+\mathbf{1}) \times (\mathbf{2}\boldsymbol{w_1} + \mathbf{3}\boldsymbol{w_2} + \boldsymbol{b}) - \mathbf{1} \geq \mathbf{0}$

Maximum Margin Formulation

- Karush-Kuhn-Tucker (KKT) Condition
  - minimize $f(\mathbf{x})$ subject to inequality conditions $g(\mathbf{x}) \geq 0$ is to optimize Lagrange function

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

Lagrange Multiplier

  - KKT Conditions:

$$\frac{\partial L(\mathbf{x}, \lambda)}{\partial \mathbf{x}} = 0 \qquad \lambda \geq 0 \qquad \lambda g(\mathbf{x}) = 0$$

Maximum Margin Formulation

- Karush-Kuhn-Tucker (KKT) Condition
  - Goal: $(\hat{\mathbf{w}}, \hat{b}) = \text{argmin}_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2$
  - Condition: $\underbrace{t_i(\mathbf{w}^T\mathbf{x}_i + b) - 1}_{g_i(\mathbf{x})} \geq 0 \quad \forall i = 1, 2, \ldots N$

- Lagrange Function:

$$L(\mathbf{w}, b, \lambda) = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \lambda_i \times (t_i(\mathbf{w}^T\mathbf{x}_i + b) - 1)$$

$$\left\{ \begin{array}{l} \lambda_1 (\underbrace{t_1(\mathbf{w}^T\mathbf{x}_1 + b) - 1}_{g_1(\mathbf{x})}) + \\ \lambda_2 (\underbrace{t_2(\mathbf{w}^T\mathbf{x}_2 + b) - 1}_{g_2(\mathbf{x})}) + \\ \quad \cdots . \\ \lambda_N (\underbrace{t_N(\mathbf{w}^T\mathbf{x}_N + b) - 1}_{g_N(\mathbf{x})}) \end{array} \right\}$$

Maximum Margin Formulation

- Karush-Kuhn-Tucker (KKT) Condition
  - Goal: $(\hat{\mathbf{w}}, \hat{b}) = \text{argmin}_{\mathbf{w},b} \frac{1}{2} ||\mathbf{w}||^2$
  - Condition: $t_i(\mathbf{w}^T\mathbf{x}_i + b) - \mathbf{1} \geq \mathbf{0} \quad \forall \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots \boldsymbol{N}$

- KKT Conditions:

1. $\frac{\partial L(\mathbf{w}, b, \lambda)}{\partial \mathbf{w}} = 0$
2. $\frac{\partial L(\mathbf{w}, b, \lambda)}{\partial b} = 0$
3. $\lambda_i \geq \mathbf{0} \quad \forall \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots \boldsymbol{N}$
4. $\lambda_i(t_i(\mathbf{w}^T\mathbf{x}_i + b) - \mathbf{1}) = \mathbf{0} \quad \forall \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots \boldsymbol{N}$

Maximum Margin Formulation

- KKT Condition: Example
  - Goal: $(\hat{\mathbf{w}}, \hat{b}) = \text{argmin}_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2$
  - Conditions: $(-1) \times (w_1 + w_2 + b) - 1 \geq 0$

    $(-1) \times (2w_1 + b) - 1 \geq 0$

    $(+1) \times (2w_1 + 3w_2 + b) - 1 \geq 0$

- Lagrange Function

$$L(\mathbf{w}, b, \lambda) = \frac{1}{2}||\mathbf{w}||^2 - \sum \begin{pmatrix} \lambda_1\big((-1) \times (w_1 + w_2 + b) - 1\big) + \backslash \\ \lambda_2\big((-1) \times (2w_1 + b) - 1\big), \\ \lambda_3\big((+1) \times (2w_1 + 3w_2 + b) - 1\big) \end{pmatrix}$$

$g_1(\mathbf{x})$, $g_2(\mathbf{x})$, $g_3(\mathbf{x})$

Maximum Margin Formulation

- KKT Condition: Example
  - KKT Conditions:

$$\frac{\partial L(\mathbf{w}, b, \lambda)}{\partial \mathbf{w}} = 0$$

$$\frac{\partial L(\mathbf{w}, b, \lambda)}{\partial b} = 0$$

$$\lambda_1 \geq \mathbf{0}$$

$$\lambda_2 \geq \mathbf{0}$$

$$\lambda_3 \geq \mathbf{0}$$

$$\lambda_1\big((-\mathbf{1}) \times (\boldsymbol{w}_1 + \boldsymbol{w}_2 + \boldsymbol{b}) - \mathbf{1}\big) = 0$$

$$\lambda_2\big((-\mathbf{1}) \times (\mathbf{2}\boldsymbol{w}_1 + \boldsymbol{b}) - \mathbf{1}\big) = 0$$

$$\lambda_3\big((+\mathbf{1}) \times (\mathbf{2}\boldsymbol{w}_1 + \mathbf{3}\boldsymbol{w}_2 + \boldsymbol{b}) - \mathbf{1}\big) = 0$$

Dual Representation

- **w** and b Elimination

$$L(\mathbf{w}, b, \lambda) = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \lambda_i \times \left(t_i(\mathbf{w}^T \mathbf{x}_i + b) - \mathbf{1}\right)$$

$$\frac{\partial L(w, b, \lambda)}{\partial b} = 0 \quad \Rightarrow \quad \frac{\partial}{\partial b}\left[\frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \lambda_i \times \left(t_i(\mathbf{w}^T \mathbf{x}_i + b) - \mathbf{1}\right)\right] = 0$$

$= 0$; $= t_i$

$$\Rightarrow \sum_{i=1}^{N} \lambda_i t_i = 0$$

$$\frac{\partial L(w, b, \lambda)}{\partial \mathbf{w}} = 0 \quad \Rightarrow \quad \frac{\partial}{\partial \mathbf{w}}\left[\frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \lambda_i \times \left(t_i(\mathbf{w}^T \mathbf{x}_i + b) - \mathbf{1}\right)\right] = 0$$

$= \mathbf{w}$; $= t_i \mathbf{x}_i$

$$\Rightarrow \mathbf{w} - \sum_{i=1}^{N} \lambda_i t_i \mathbf{x}_i = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{N} \lambda_i t_i \mathbf{x}_i$$

Dual Representation

- $\mathbf{w}$ and b Elimination

$$\sum_{i=1}^{N} \lambda_i t_i = 0$$

$$\mathbf{w} = \sum_{i=1}^{N} \lambda_i t_i \mathbf{x}_i$$

$$L(\mathbf{w}, b, \lambda) = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \lambda_i \times \left(t_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) - \mathbf{1}\right)$$

$$= \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \lambda_i t_i \mathbf{w}^T \mathbf{x}_i + \lambda_i t_i b - \lambda_i$$

$$= \frac{1}{2}||\mathbf{w}||^2 - \left[\mathbf{w}^T \underbrace{\sum_{i=1}^{N} \lambda_i t_i \mathbf{x}_i}_{= \mathbf{w}} + b \underbrace{\sum_{i=1}^{N} \lambda_i t_i}_{= 0} - \sum_{i=1}^{N} \lambda_i\right]$$

$$= \frac{1}{2}||\mathbf{w}||^2 - ||\mathbf{w}||^2 + \sum_{i=1}^{N} \lambda_i = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}||\mathbf{w}||^2$$

Dual Representation

- $\mathbf{w}$ and b Elimination

$$\sum_{i=1}^{N} \lambda_i t_i = 0$$

$$\mathbf{w} = \sum_{i=1}^{N} \lambda_i t_i \mathbf{x}_i$$

$$= \mathbf{w}^{\mathrm{T}}\mathbf{w}$$

$$L(\mathbf{w}, b, \lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}||\mathbf{w}||^2 = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$$= \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i t_i \mathbf{x}_i^T\right)\left(\sum_{j=1}^{N} \lambda_j t_j \mathbf{x}_j\right)$$

$$(\lambda_1 t_1 \mathbf{x}_1^T + \cdots + \lambda_N t_N \mathbf{x}_N^T)(\lambda_1 t_1 \mathbf{x}_1 + \cdots + \lambda_N t_N \mathbf{x}_N)$$

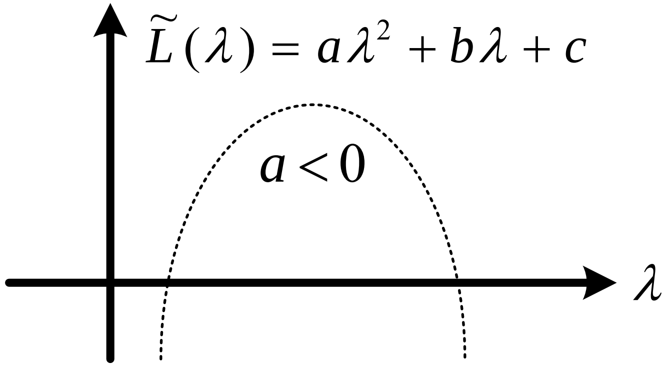
$$= \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j t_i t_j \mathbf{x}_i^T \mathbf{x}_j$$

Dual Representation

- Dual Form
  - Goal: maximize Lagrange function $\tilde{L}(\lambda)$ that is quadratic function of λ

$$\tilde{L}(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j t_i t_j \mathbf{x}_i^T \mathbf{x}_j$$

  - Conditions:
    - $\lambda_i \geq \mathbf{0} \qquad \forall \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots \boldsymbol{N}$
    - $\sum_{i=1}^{N} \lambda_i t_i = 0$



Dual Representation

- Dual Form: Example
  - Given: $\boldsymbol{D} = \left\{ \left(\begin{bmatrix}1\\1\end{bmatrix}, -1 \right), \left(\begin{bmatrix}2\\0\end{bmatrix}, -1 \right), \left(\begin{bmatrix}2\\3\end{bmatrix}, +1 \right) \right\}$

$$\tilde{L}(\boldsymbol{\lambda}) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j t_i t_j \mathbf{x}_i^T \mathbf{x}_j$$

$$\tilde{L}(\boldsymbol{\lambda}) = \lambda_1 + \lambda_2 + \lambda_3 - \frac{1}{2} \begin{pmatrix} 2\lambda_1^2 + 2\lambda_1\lambda_2 - 5\lambda_1\lambda_3 + \\ 2\lambda_1\lambda_2 + 4\lambda_2^2 - 4\lambda_2\lambda_3 + \\ -5\lambda_1\lambda_3 - 4\lambda_2\lambda_3 + 13\lambda_3^2 \end{pmatrix}$$

$$\sum_{i=1}^{N} \lambda_i$$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j t_i t_j \mathbf{x}_i^T \mathbf{x}_j$$

$$-\frac{1}{2} \begin{pmatrix} \lambda_1\lambda_1(-1)(-1) \begin{bmatrix}1\\1\end{bmatrix}^T \begin{bmatrix}1\\1\end{bmatrix} + \lambda_1\lambda_2(-1)(-1) \begin{bmatrix}1\\1\end{bmatrix}^T \begin{bmatrix}2\\0\end{bmatrix} + \lambda_1\lambda_3(-1)(+1) \begin{bmatrix}1\\1\end{bmatrix}^T \begin{bmatrix}2\\3\end{bmatrix} + \\ \lambda_2\lambda_1(-1)(-1) \begin{bmatrix}2\\0\end{bmatrix}^T \begin{bmatrix}1\\1\end{bmatrix} + \lambda_2\lambda_2(-1)(-1) \begin{bmatrix}2\\0\end{bmatrix}^T \begin{bmatrix}2\\0\end{bmatrix} + \lambda_2\lambda_3(-1)(+1) \begin{bmatrix}2\\0\end{bmatrix}^T \begin{bmatrix}2\\3\end{bmatrix} + \\ \lambda_3\lambda_1(+1)(-1) \begin{bmatrix}2\\3\end{bmatrix}^T \begin{bmatrix}1\\1\end{bmatrix} + \lambda_3\lambda_2(+1)(-1) \begin{bmatrix}2\\3\end{bmatrix}^T \begin{bmatrix}2\\0\end{bmatrix} + \lambda_3\lambda_3(+1)(+1) \begin{bmatrix}2\\3\end{bmatrix}^T \begin{bmatrix}2\\3\end{bmatrix} \end{pmatrix}$$

Dual Representation

- Dual Form: Example
  - Given: $\boldsymbol{D} = \left\{ \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, -1 \right), \left(\begin{bmatrix} 2 \\ 0 \end{bmatrix}, -1 \right), \left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}, +1 \right) \right\}$
  - Goal: maximize Lagrange function $\tilde{L}(\lambda)$

$$\tilde{L}(\lambda) = \lambda_1 + \lambda_2 + \lambda_3 - \frac{1}{2} \begin{pmatrix} 2\lambda_1^2 + 2\lambda_1\lambda_2 - 5\lambda_1\lambda_3 + \\ 2\lambda_1\lambda_2 + 4\lambda_2^2 - 4\lambda_2\lambda_3 + \\ -5\lambda_1\lambda_3 - 4\lambda_2\lambda_3 + 13\lambda_3^2 \end{pmatrix}$$

  - Conditions:

$$\lambda_1 \geq \mathbf{0}$$

$$\lambda_2 \geq \mathbf{0}$$

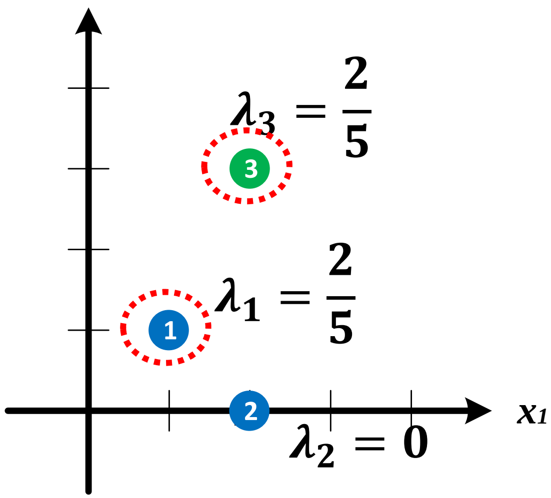
$$\lambda_3 \geq \mathbf{0}$$

$$\sum_{i=1}^{3} \lambda_i t_i = \lambda_1 t_1 + \lambda_2 t_2 + \lambda_3 t_3 = 0$$

$$\Rightarrow (-1)\lambda_1 + (-1)\lambda_2 + (+1)\lambda_3 = 0$$

Hyperplane Derivation

- **w** Computation
  - solve the quadratic programming problem to find the Lagrange multipliers $\{\lambda_i | \boldsymbol{i} = \mathbf{0}, \mathbf{1}, \dots, \boldsymbol{N}\}$
  - compute $\mathbf{w} = \sum_{i=1}^{N} \lambda_i t_i \mathbf{x}_i$



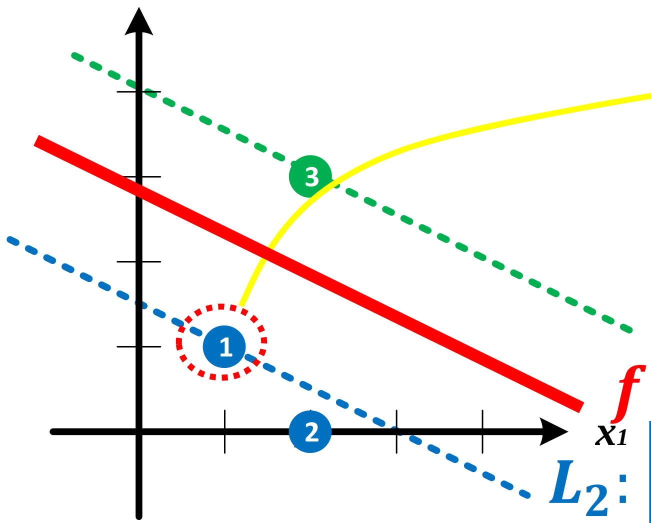
$$\mathbf{w} = \frac{2}{5} \times (-1) \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0 \times (-1) \times \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \frac{2}{5} \times (+1) \times \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} & \frac{4}{5} \end{bmatrix}^T$$

Hyperplane Derivation

- b Computation
  - compute b using support vectors ($\lambda_i \neq 0$)

$$t_i = -1 \Rightarrow (\mathbf{w}^T\mathbf{x}_i + b) = -1$$

or $$t_i = +1 \Rightarrow (\mathbf{w}^T\mathbf{x}_i + b) = +1$$



$$\begin{bmatrix} \frac{2}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + b = -1$$

$$\Rightarrow \frac{2}{5} + \frac{4}{5} + b = -1$$

$$b = -\frac{11}{5}$$

$$f(\mathbf{x}): \frac{2}{5}x_1 + \frac{4}{5}x_2 - \frac{11}{5} = 0$$

$$L_2: \begin{bmatrix} 2/5 \\ 4/5 \end{bmatrix}^T \mathbf{x} + b = -1$$

國立高雄第一科技大學
National Kaohsiung First University of Science and Technology

Nkfustcc