Support Vector Machine

Lab Seminar - Senseable AI Lab

Byeongjoon Noh

powernoh@sch.ac.kr

SOON CHUN HYANG UNIVERSITY

Most of the slides are available on Senseable AI Lab homepage: https://sailab.space

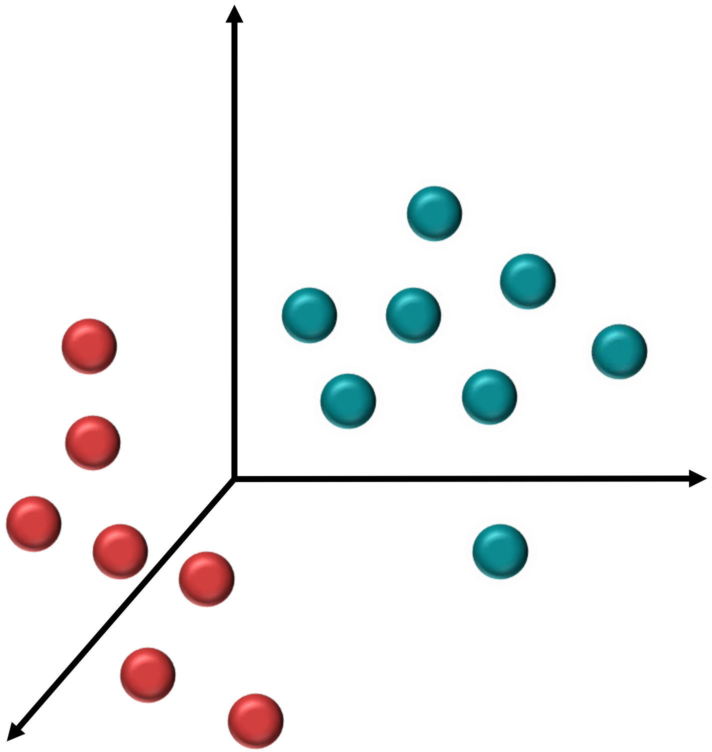
SVM

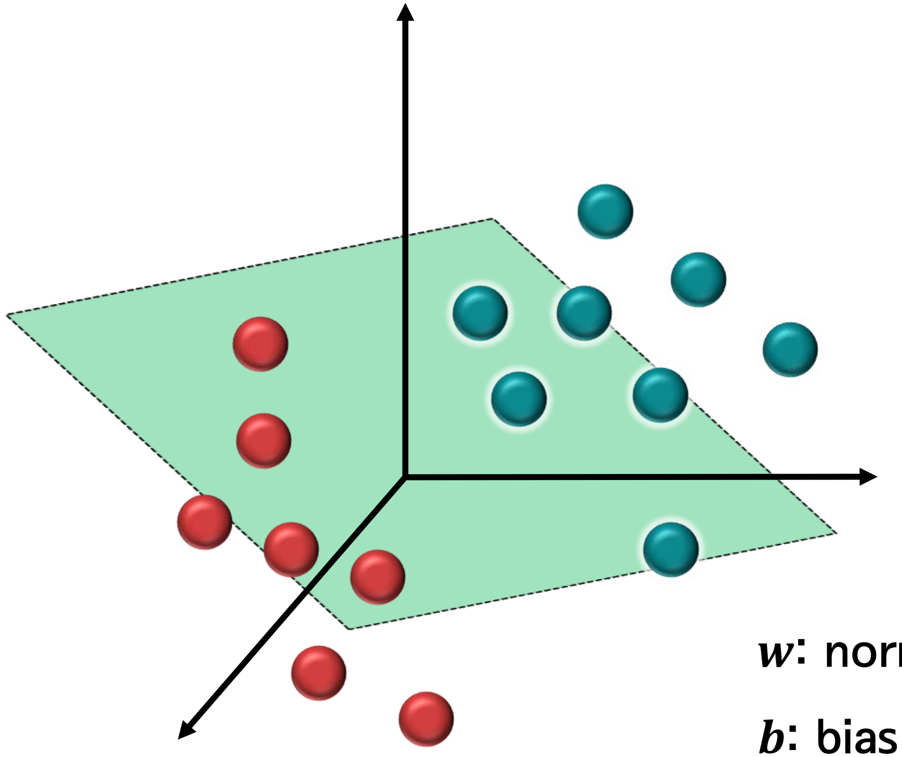
SVM has been shown to be able to achieve good generalization performance for classification of high-dimensional data sets and its training can be framed as solving a quadratic programming problem

- usually we try to maximize classification performance for the training data
- but, if the classifier is too fit for the training data, the classification ability for unknown data (i.e., the generalization ability) is degraded
- there is a trade-off between the generalization ability and fitting to the training data
- SVM is trained so that the direct decision function maximizes the generalization ability
- SVM is based on statistical learning theory

Separating Hyperplane

Classification problem



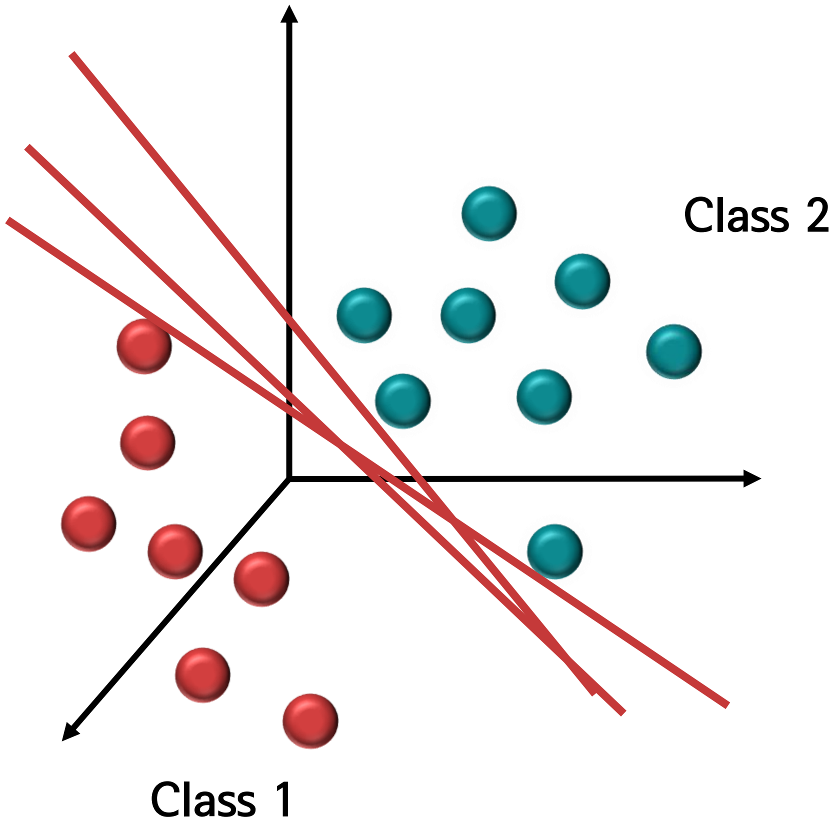


$$\boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{0}$$

$\boldsymbol{w}$: normal vector of the hyperplane

$\boldsymbol{b}$: bias

Separating Hyperplane



- Two class classification problem
- 두 class를 나누는 hyperplane은 무한히 많음
- 어떤 hyperplane이 가장 “좋은“ hyperplane인가?
- “좋다”의 기준은?

Separating Hyperplane

Maximizing margin over the training set

= minimizing generalization error

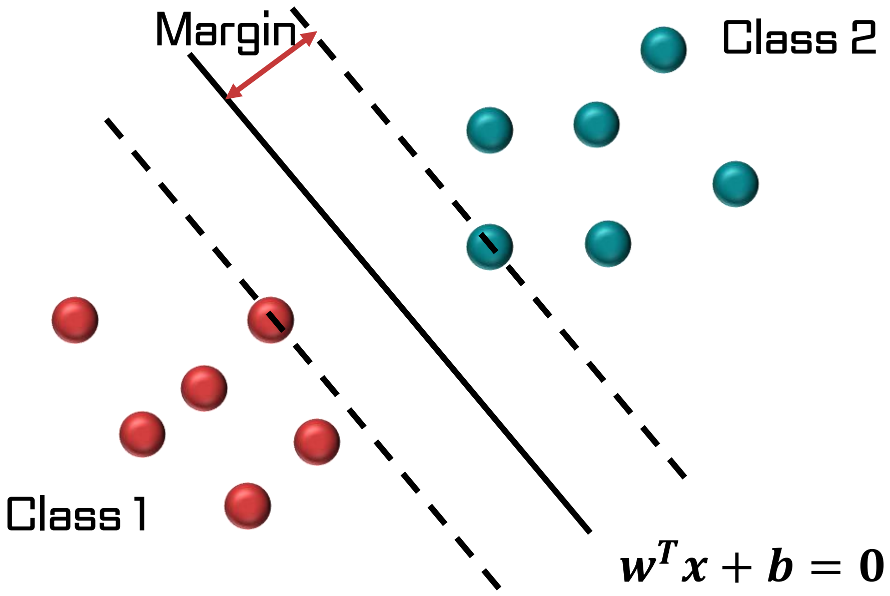
= good prediction performance

So, what is the margin?

Concept of Margin

Margin: 각 클래스에서 가장 가까운 관측치 사이의 거리

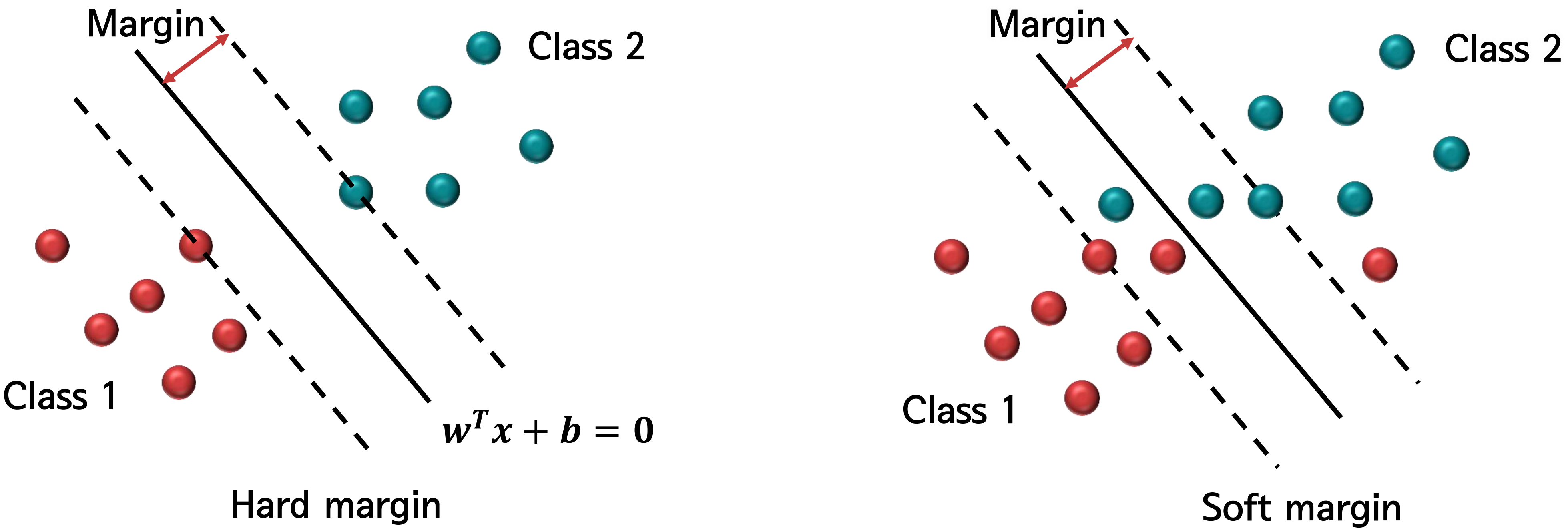
- Margin은 w(기울기)로 표현가능



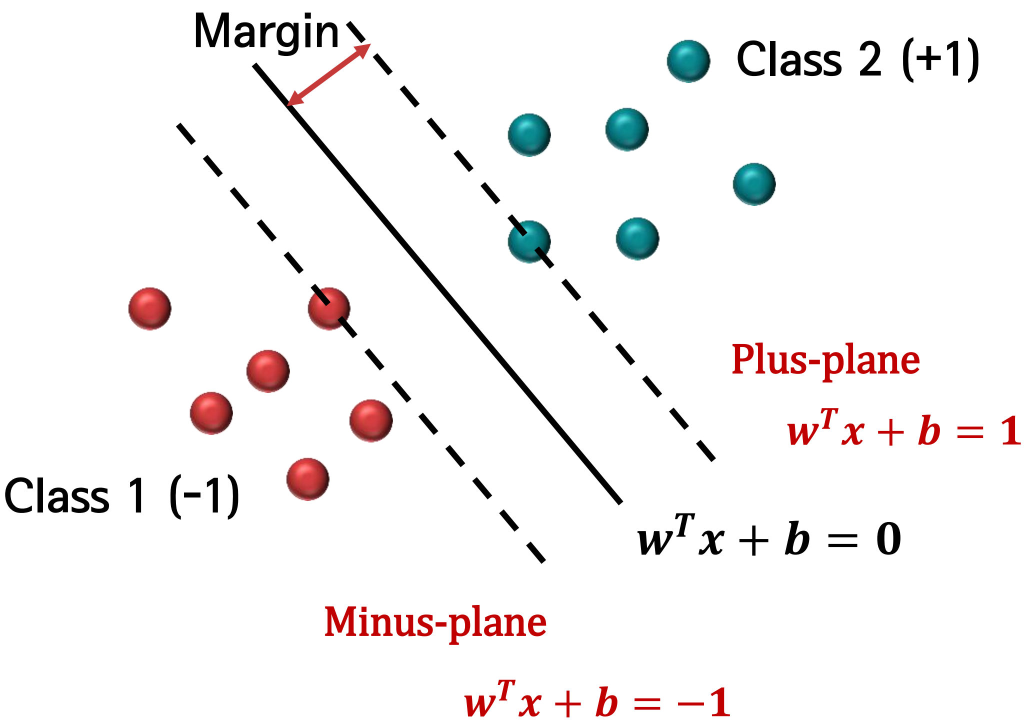
Concept of Margin

Goal of SVM model

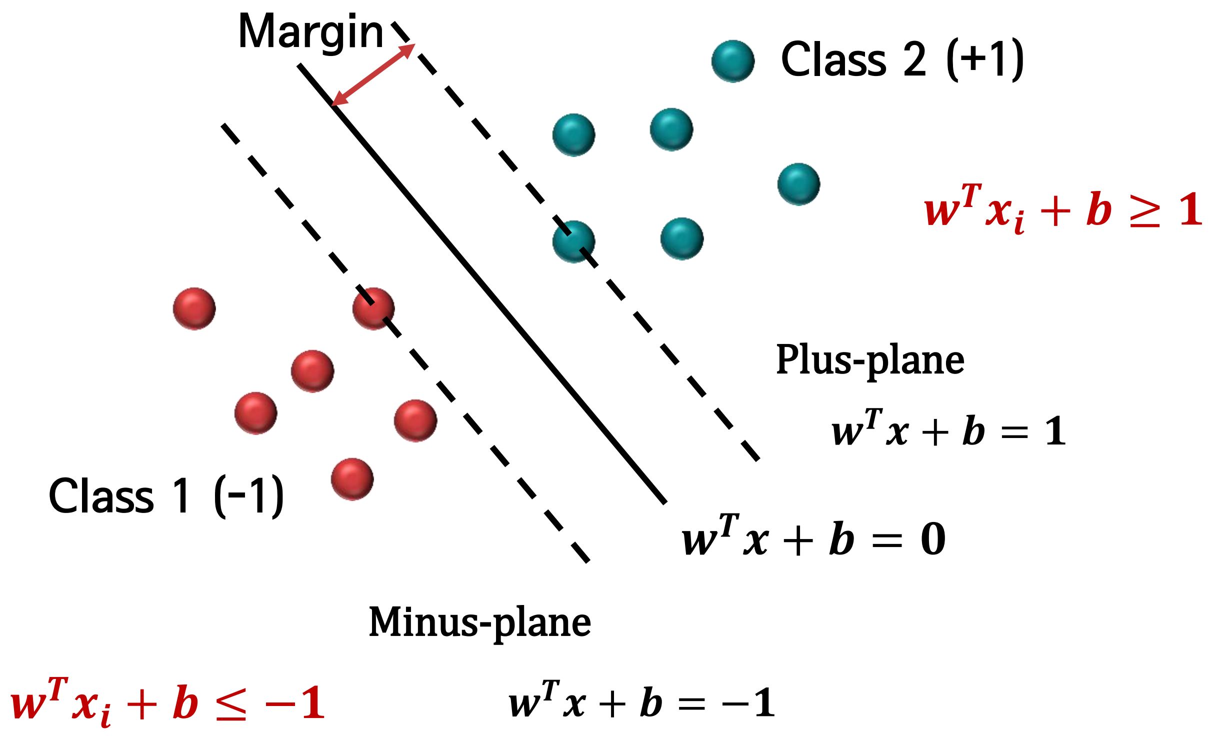
- 주어진 데이터를 통해 결정경계(decision boundary, maximum margin)를 찾는 것



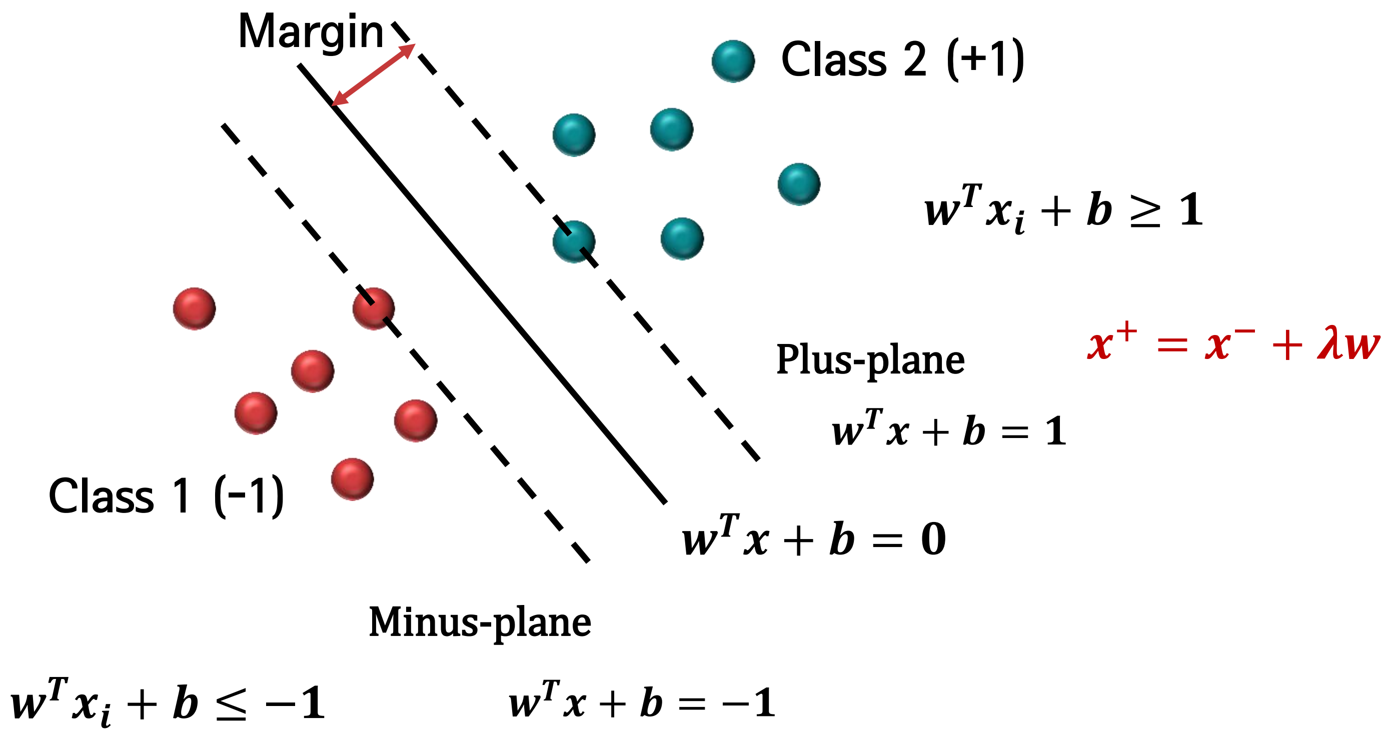
Geometric Margin



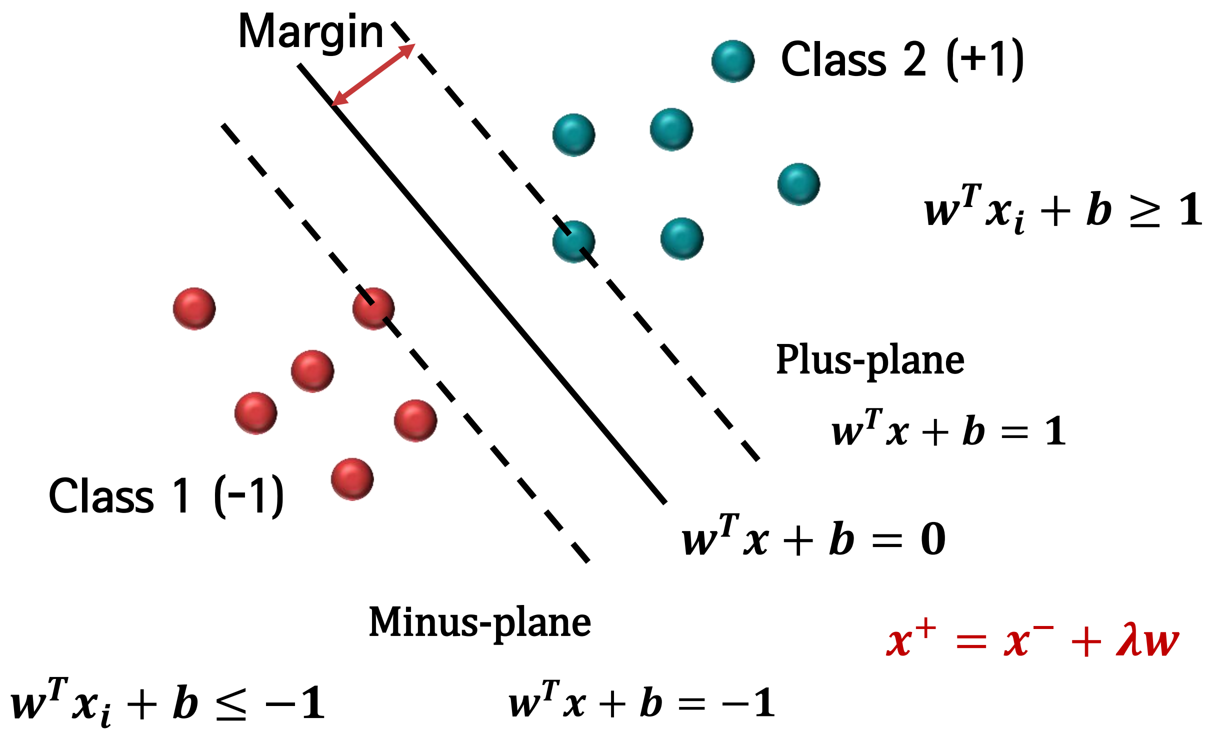
Geometric Margin



Geometric Margin



Geometric Margin



$w^T x^+ + b = 1$ ➔ x^+가 plus−plane 위의 점

$w^T(x^- + \lambda w) + b = 1$ ($\because x^+ = x^- + \lambda w$)

$w^T x^- + b + \lambda w^T w = 1$

$-1 + \lambda w^T w = 1$

x^-는 minus−plane 위의 점

$$\therefore \lambda = \frac{2}{w^T w}$$

Vector norm $\|W\|_p$ ($p = 1, 2, 3, \ldots$)

* Vector norm

$$\|\boldsymbol{W}\|_p = \left(\sum_i |\boldsymbol{w_i}|^p\right)^{1/p}$$

$\boldsymbol{L_2}$ norm

$$\|\boldsymbol{W}\|_2 = \left(\sum_i |\boldsymbol{w_i}|^2\right)^{1/2} = \sqrt{\boldsymbol{w_1^2} + \boldsymbol{w_2^2} + \cdots + \boldsymbol{w_n^2}} = \sqrt{\boldsymbol{W^T W}}$$

$$\boldsymbol{W^T} = (\boldsymbol{w_1}, \boldsymbol{w_2}, \dots, \boldsymbol{w_n})$$

Geometric Margin

$$
\begin{aligned}
\text{Margin} &= \boldsymbol{distance}(\boldsymbol{x}^{+}, \boldsymbol{x}^{-}) \\
&= \|\boldsymbol{x}^{+} - \boldsymbol{x}^{-}\|_{2} \\
&= \|(\boldsymbol{x}^{-} + \boldsymbol{\lambda w}) - \boldsymbol{x}^{-}\|_{2} \\
&= \|\boldsymbol{\lambda w}\|_{2} \\
&= \boldsymbol{\lambda}\sqrt{\boldsymbol{w}^{T}\boldsymbol{w}} \\
&= \frac{2}{\boldsymbol{w}^{T}\boldsymbol{w}}\sqrt{\boldsymbol{w}^{T}\boldsymbol{w}} \\
&= \frac{2}{\boldsymbol{w}^{T}\boldsymbol{w}} \\
&= \frac{2}{\|\boldsymbol{w}\|_{2}}
\end{aligned}
$$

Geometric Margin

$$\max Margin = \max \frac{2}{\|w\|_2} \Leftrightarrow \min \frac{1}{2}\|w\|_2$$

$$\|w\|_2 = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}$$

- w의 L_2 norm이 제곱근을 포함하고 있기 때문에 계산이 어려움
- ➔ 계산상의 편의를 위해 목적함수 변형

$$min \frac{1}{2}\|w\|_2 \Leftrightarrow min \frac{1}{2}\|w\|_2^2$$

Hard Margin

Convex optimization problem

- Objective function (목적식) $\boldsymbol{minimize\ \frac{1}{2}\|w\|_2^2}$
  - objective function은 separating hyperplane으로부터 정의된 margin의 역수

- Constraint (제약식) $\boldsymbol{subject\ to\ y_i\left(w^T x_i + b\right) \geq 1, i = 1, 2, \dots, n}$
  - decision variable: $\boldsymbol{w, b}$
  - constraint는 training data를 완벽하게 separating하는 조건

- Objective function is quadratic and constraint is linear
  - quadratic programming ➔ convex optimization

    ➔ globally optimal solution exists (전역최적해가 존재)
  - training data가 linearly separable한 경우에만 해가 존재함

Lagrangian Formulation

Original problem

$$minimize\ \frac{1}{2}\|w\|_2^2$$

$$subject\ to\ y_i\left(w^T x_i + b\right) \geq 1, i = 1, 2, \dots, n$$

Lagrangian multiplier를 이용하여 문제 변환

$$\max_{\alpha} \min_{w,b} \mathcal{L}(w, b, \alpha) = \max_{\alpha} \min_{w,b} \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{n} \alpha_i\left(y_i\left(w^T x_i + b\right) - 1\right)$$

$$subject\ to\ \alpha_i \geq 0, i = 1, 2, \dots, n$$

Lagrangian Formulation

Original problem

$$minimize\ \frac{1}{2}\|w\|_2^2$$

$$subject\ to\ y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, n$$

Lagrangian multiplier를 이용하여 문제 변환

$$\max_{\alpha} \min_{w,b} \mathcal{L}(w, b, \alpha) = \max_{\alpha} \min_{w,b} \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{n} \alpha_i (y_i(w^T x_i + b) - 1)$$

$$subject\ to\ \alpha_i \geq 0, i = 1, 2, \dots, n$$

Lagrangian Primal problem

Lagrangian Formulation

$$\min_{w,b} \mathcal{L}(w,b,\alpha) = \min_{w,b} \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(w^T x_i + b \right) - 1 \right)$$

$$subject\ to\ \alpha_i \geq 0, i = 1, 2, \dots, n$$

Solving Lagrangian Primal problem

- convex, continuous 이기 때문에 미분값=0인 지점에서 최솟값

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

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

Lagrangian Formulation

Recall: $\frac{\partial \mathcal{L}(w,b,\alpha)}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i y_i x_i$

Solving Lagrangian Primal problem

$$\frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(w^T x_i + b\right) - 1\right)$$

$$\frac{1}{2}\|w\|_2^2 = \frac{1}{2} w^T w = \frac{1}{2} w^T \sum_{j=1}^{n} \alpha_j y_j x_j$$

$$= \frac{1}{2} \sum_{j=1}^{n} \alpha_j y_j \left(w^T x_j\right)$$

$$= \frac{1}{2} \sum_{j=1}^{n} \alpha_j y_j \left(\sum_{i=1}^{n} \alpha_i y_i x_i^T x_j\right)$$

$$= \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Lagrangian Formulation

Recall

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

Solving Lagrangian Primal problem

$$\frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{n} \alpha_i \left(y_i\left(w^T x_i + b\right) - 1\right)$$

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

$$= -\sum_{i=1}^{n} \alpha_i y_i w^T x_i - b\sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{n} \alpha_i$$

Lagrangian Formulation

Solving Lagrangian Primal problem

$$\min_{w,b} \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{n} \alpha_i\left(y_i\left(w^T x_i + b\right) - 1\right)$$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j x_i^T x_j \quad where \sum_{i=1}^{n} \alpha_i y_i = 0$$

Lagrangian Formulation

$$\max_{\alpha}\min_{w,b}\mathcal{L}(w,b,\alpha)=\max_{\alpha}\min_{w,b}\frac{1}{2}\|w\|_2^2-\sum_{i=1}^{n}\alpha_i\big(y_i\big(w^Tx_i+b\big)-1\big)$$

$$subject\ to\ \alpha_i\geq 0, i=1,2,\dots,n$$

$$\min_{w,b}\frac{1}{2}\|w\|_2^2-\sum_{i=1}^{n}\alpha_i(y_i(w^Tx_i+b)-1)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_jy_iy_jx_i^Tx_j\quad where\sum_{i=1}^{n}\alpha_iy_i=0$$

Lagrangian Dual

$$\max_{\alpha}\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_jy_iy_jx_i^Tx_j$$

$$subject\ to\ \textstyle\sum_{i=1}^{n}\alpha_iy_i=0, \alpha_i\geq 0 i=1,2,\dots,n$$

Lagrangian Formulation

Solving Lagrangian Dual problem

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$subject\ to \sum_{i=1}^{n} \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, n$$

- decision variable: α
- original problem formulation (primal formulation)보다 풀기 쉬운 형태
- objective function is quadratic (α_i, α_j) and constraint is linear ($\sum_{i=1}^{n} \alpha_i y_i = 0$)
  - ➔ quadratic programming ➔ convex optimization ➔ globally optimal solution exists

KKT (Karush-Kuhn-Tucker) condition

w, b, α가 Lagrangian dual problem의 최적해가 되기 위한 조건

- 1) stationarity

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

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

- 2) primal feasibility $y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, n$

- 3) dual feasibility $\alpha_i \geq 0, i = 1, 2, \dots, n$

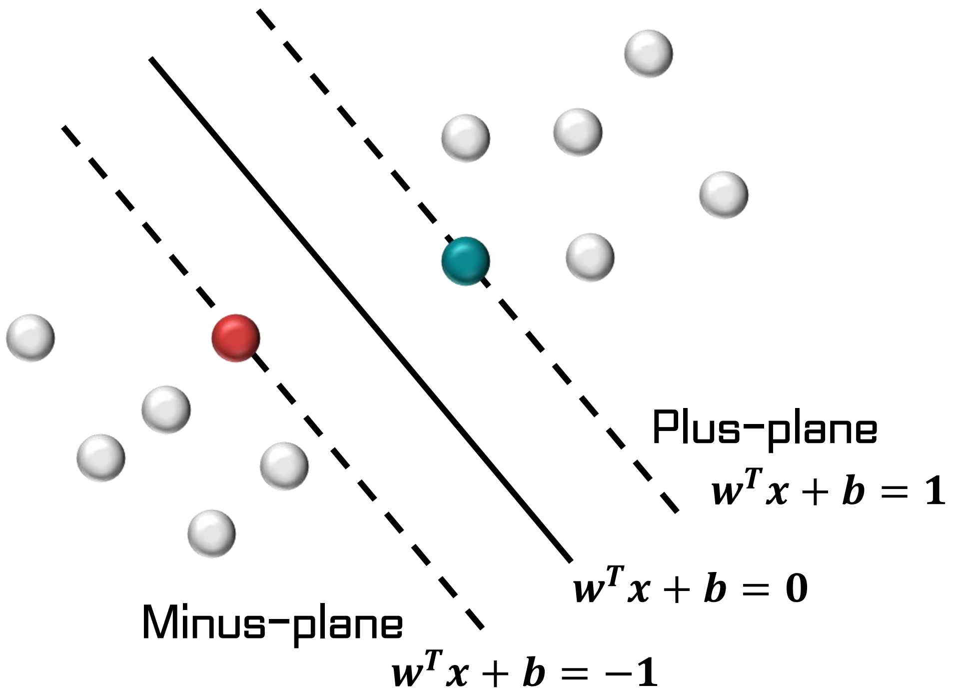
- 4) complementary slackness $\alpha_i(y_i(w^T x_i + b) - 1) = 0, i = 1, 2, \dots, n$

Characteristics of the Solutions

$$\alpha_i\left(y_i\left(w^T x_i + b\right) - 1\right) = 0, i = 1, 2, \dots, n$$

1) $\alpha_i > 0$ and $y_i\left(w^T x_i + b\right) - 1 = 0$

- x_i가 plus-plane 또는 minus-plane (margin) 위에 있음
- margin 위에 있는 x_i를 support vector라고 함

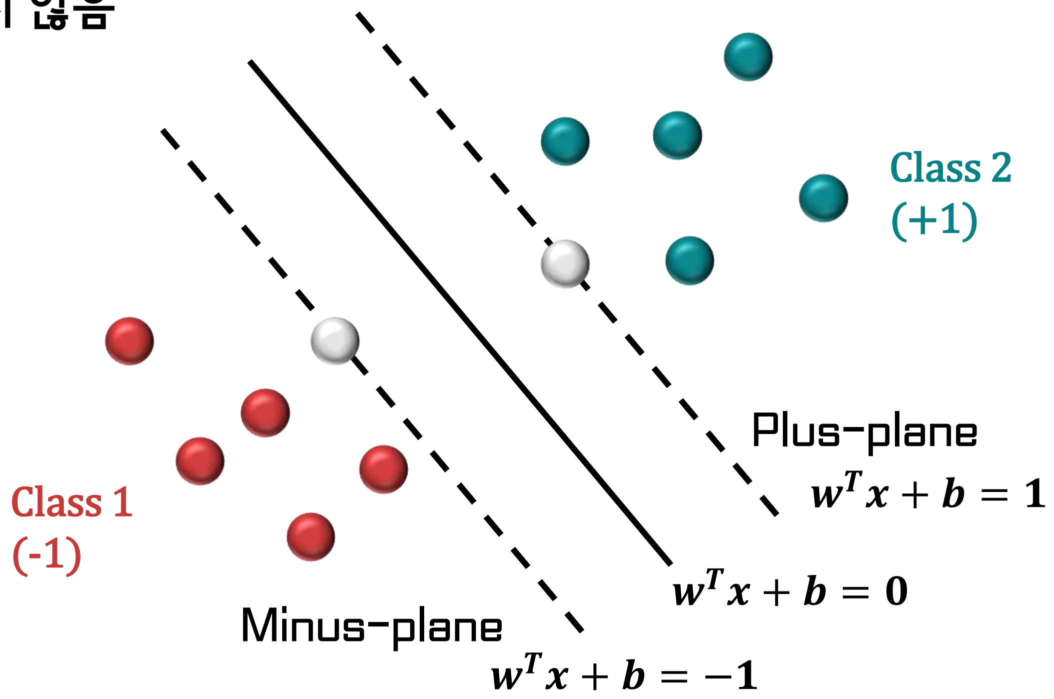


Characteristics of the Solutions

$$\alpha_i\left(y_i\left(w^T x_i + b\right) - 1\right) = 0, i = 1, 2, \dots, n$$

2) $\alpha_i = 0$ and $y_i\left(w^T x_i + b\right) - 1 \neq 0$

- x_i가 plus-plane 또는 minus-plane (margin) 위에 있지 않음
- Hyperplane을 구축하는데 영향을 미치지 않음



Characteristics of the Solutions

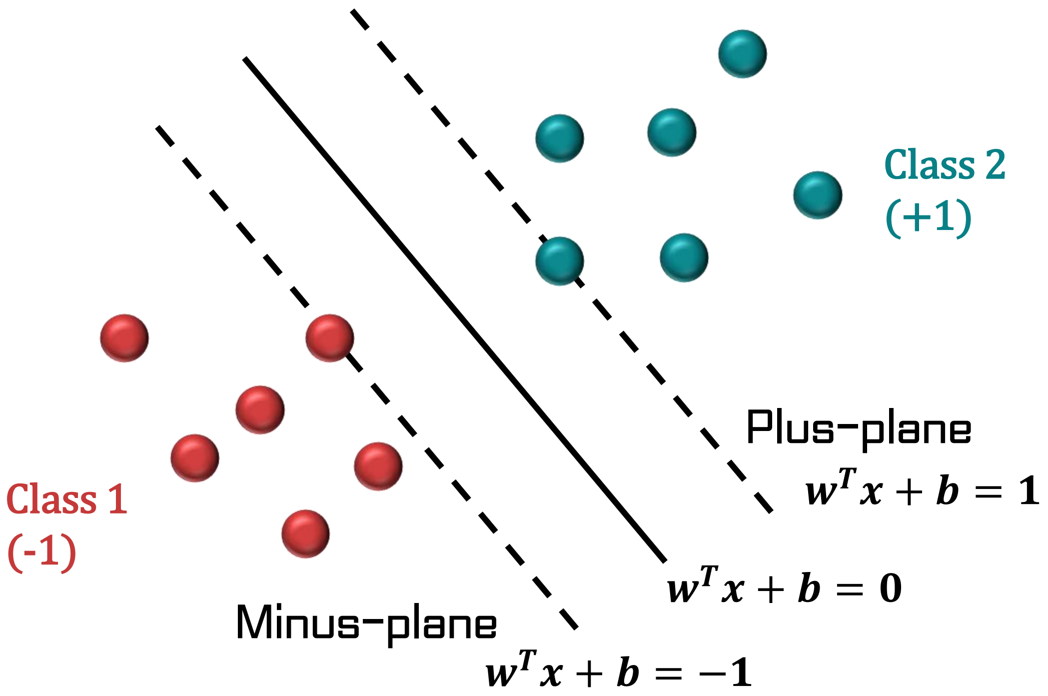
즉, x_i가 support vector인 경우에만 $\alpha_i^* > 0$ 이므로

- $w^* = \sum_{i=1}^{n} a_i^* y_i x_i = \sum_{i \in SV} a_i^* y_i x_i$
- x_i만을 사용하여 decision boundary를 구할 수 있음
  - 임의의 support vector 하나를 이용하여 b^*

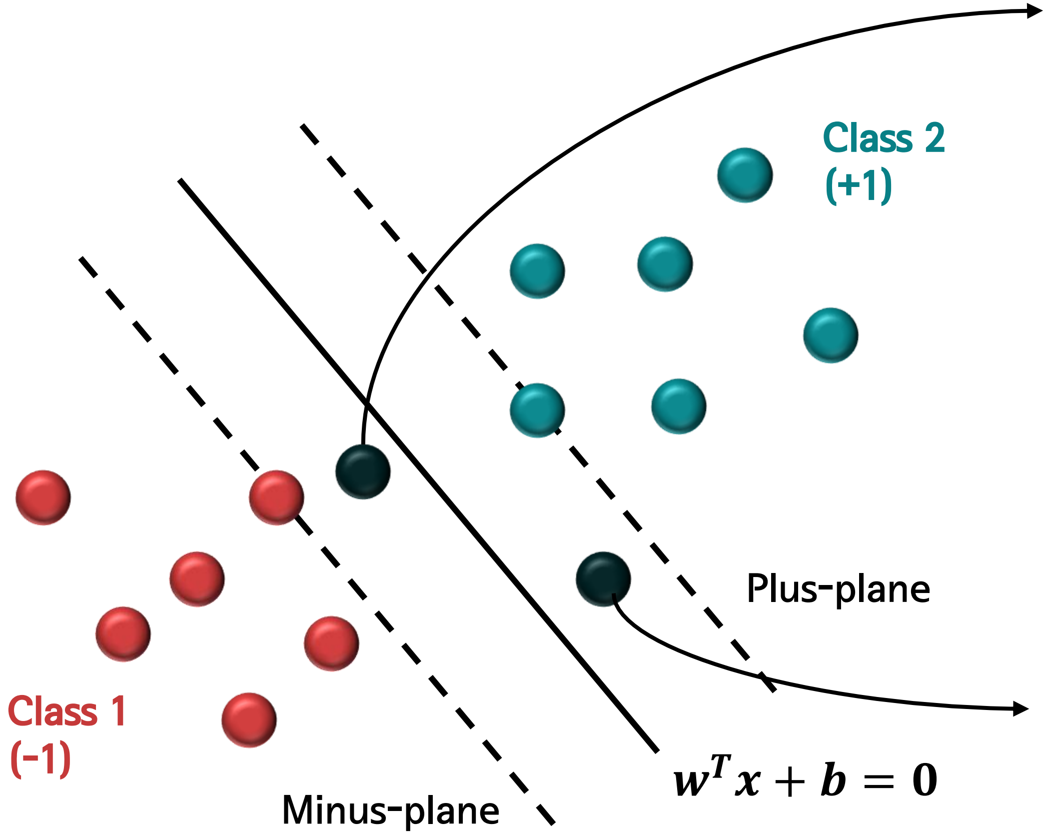
$$w^{*T} + b * = y_{sv}$$

$$w^{*T} + b * = \sum_{i=1}^{n} a_i^* y_i x_i^T x_{sv} + b^* = y_{sv}$$

$$b^* = y_{sv} - \sum_{i=1}^{n} a_i^* y_i x_i^T x_{sv}$$



Classifying New data Points



새로운 데이터가 optimal separating hyperplane보다 밑에 있음

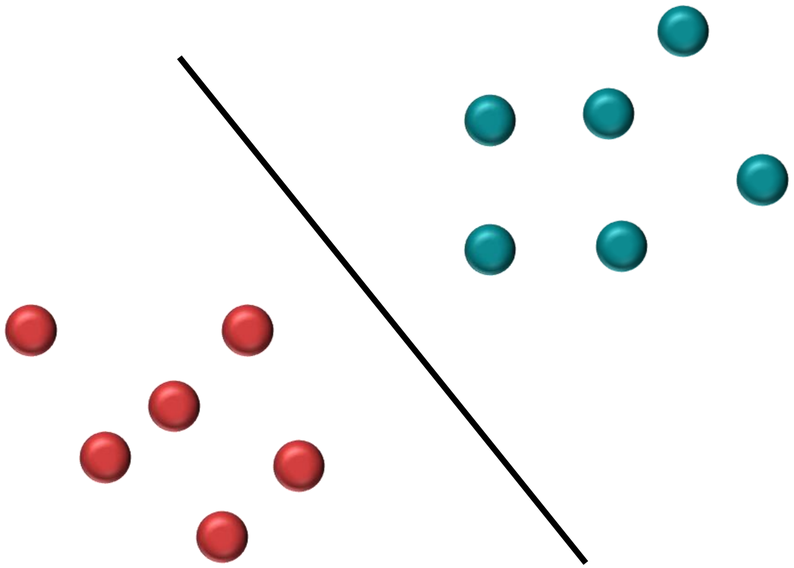
$$w^{*T} x_{new} + b^* < 0 \quad \Rightarrow \quad \hat{y}_{new} = -1$$

$$\hat{y}_{new} = \mathrm{sign}\left(w^{*T} x_{new} + b^*\right) = \mathrm{sign}\left(\sum_{i \in SV} a_i^* y_i x_i^T x_{new} + b^*\right)$$

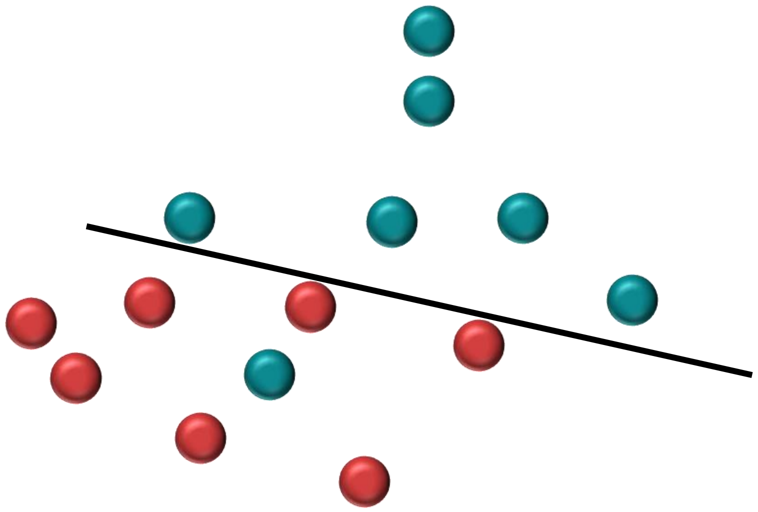
새로운 데이터가 optimal separating hyperplane보다 위에 있음

$$w^{*T} x_{new} + b^* > 0 \quad \Rightarrow \quad \hat{y}_{new} = +1$$

Linearly Non-separable Problems



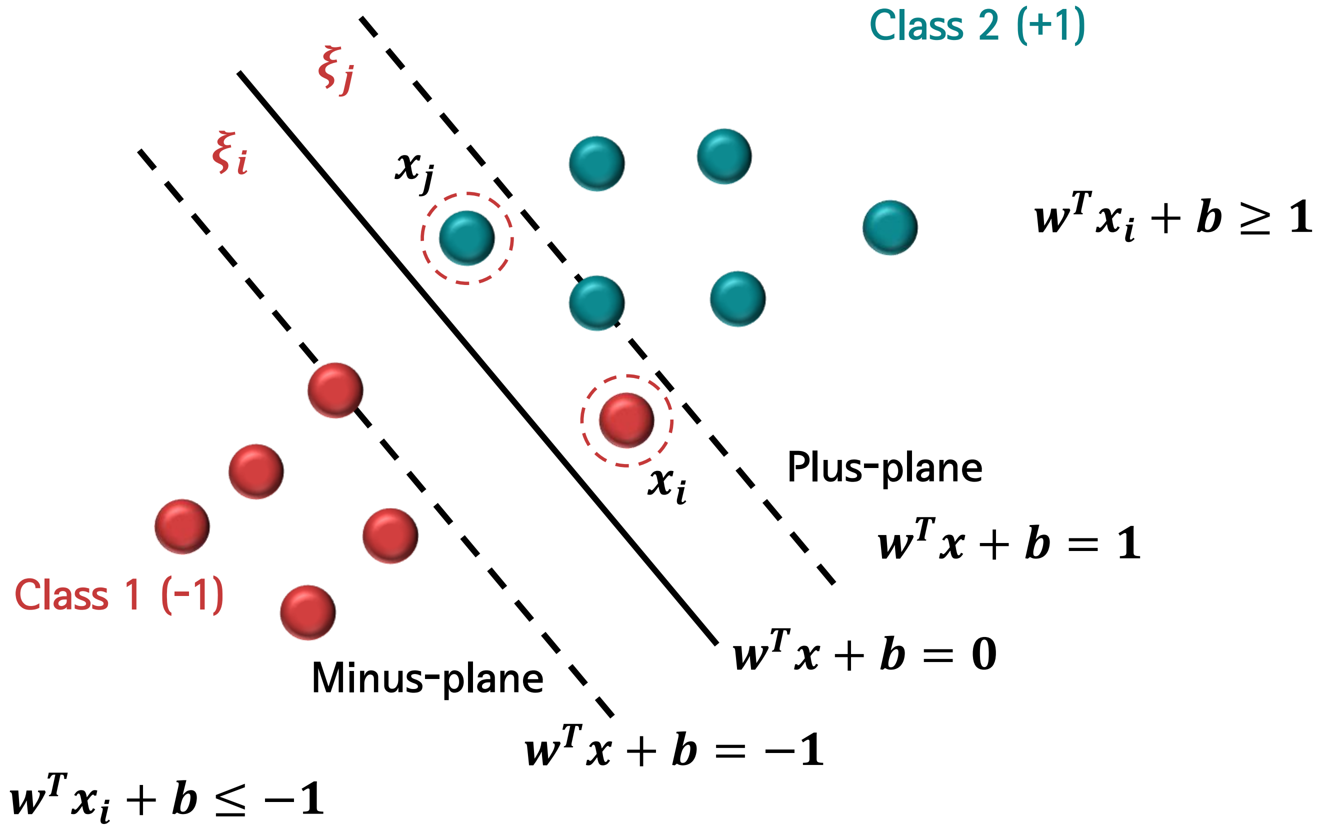
Linearly separable



Linearly non-separable

Linear decision boundary를 이용하여 완벽하게 나누는 것은 불가능 ➔ Error 허용

Linearly Non-separable Problems



Soft Margin

Convex optimization problem

$$\boldsymbol{minimize}\ \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}\xi_i$$

$$\boldsymbol{subject\ to}\ y_i(w^T x_i + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0, \qquad i = 1, 2, \dots, n$$

- decision variable: w, b, ξ
- slack variable $\xi_i \geq 0$
  - training error를 허용 (마냥 크게 할 수는 없음)
  - objective function에 penalty를 추가하여 억제

C ↑: training error를 많이 허용하지 않음

C ↓: training error를 많이 허용

- C 는 margin과 training error에 대한 trade-off를 결정하는 hyperparameter
- training data가 linearly separable하지 않아도 해가 존재함

Lagrangian Formulation

Original Problem

$$minimize\ \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}\xi_i$$

$$subject\ to\ y_i(w^T x_i + b) \geq 1 - \xi_i,\ \ \xi_i \geq 0,\ \ i\ = 1, 2, \ldots, n$$

Lagrangian multiplier

$$\max_{\alpha,\gamma}\min_{w,b,\xi}\mathcal{L}(w, b, \alpha, \xi, \gamma)$$

$$= \max_{\alpha,\gamma}\min_{w,b,\xi}\frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i(y_i(w^T x_i + b) - 1) - \sum_{i=1}^{n}\gamma_i\xi_i$$

$$subject\ to\ \alpha_i \cdot \gamma_i \geq 0, i = 1, 2, \ldots, n$$

Lagrangian Primal and Dual

$$\max_{\alpha}\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$subject\ to\ \textstyle\sum_{i=1}^{n}\alpha_i y_i = 0\ and\ 0 \leq \alpha_i \leq C, i = 1, 2, \ldots, n$$

Lagrangian Formulation

Lagrangian dual-Soft margin

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$subject\ to \sum_{i=1}^{n} \alpha_i y_i = 0\ and\ 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n$$

Lagrangian dual-Har margin

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$subject\ to \sum_{i=1}^{n} \alpha_i y_i = 0\ and\ \alpha_i \geq 0, i = 1, 2, \dots, n$$

KKT condition

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

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

$$\frac{\partial \mathcal{L}(w, b, \xi, \alpha, \gamma)}{\partial \xi} = 0 \qquad C - \alpha_i - \gamma_i = 0$$

Complementary slackness

$$\alpha_i\left(y_i\left(w^T x_i + b\right) - 1 + \xi_i\right) = 0, \gamma_i \xi_i = 0, i = 1, 2, \dots, n$$

Characteristics of the Solutions

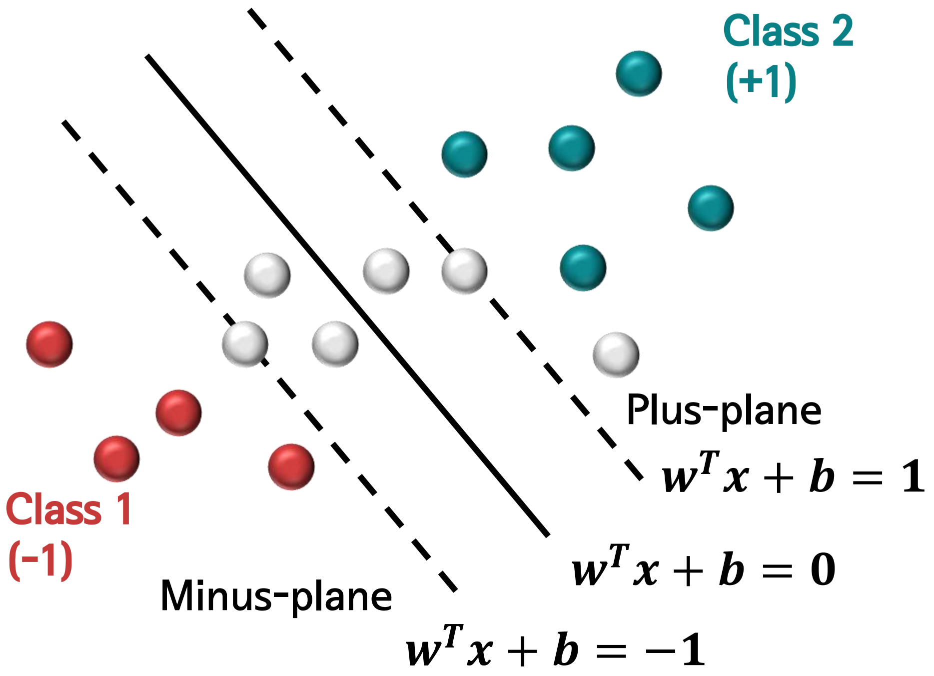
$$\alpha_i\left(y_i\left(w^T x_i + b\right) - 1 + \xi_i\right) = 0,\ \alpha_i = C - \gamma_i,\ \ \gamma_i \xi_i = 0,\ i = 1, 2, \dots, n$$

1) $\alpha_i = 0 \Rightarrow C = \gamma_i$

$\Rightarrow \xi_i = 0$

$\Rightarrow \left(y_i\left(w^T + b\right) - 1\right) \neq 0$

x_i가 plus-plane 또는 minus-plane (margin) 위에 있지 않음



Characteristics of the Solutions

$$\alpha_i\left(y_i\left(w^T x_i + b\right) - 1 + \xi_i\right) = 0,\ \ \alpha_i = C - \gamma_i,\ \ \gamma_i \xi_i = 0,\ \ i = 1, 2, \dots, n$$

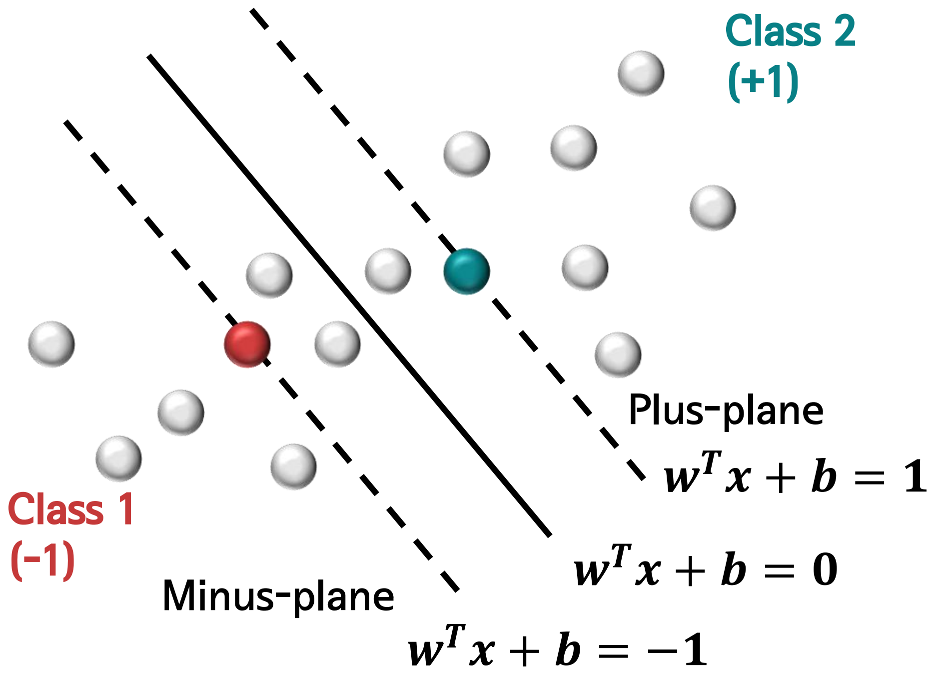
2) $0 < \alpha_i < C \Rightarrow \ \gamma_i > 0$

$\Rightarrow \xi_i = 0$

$\Rightarrow \left(y_i\left(w^T + b\right) - 1\right) = 0$

x_i가 plus-plane 또는 minus-plane (margin) 위에 있음

(support vector)



Characteristics of the Solutions

$$\alpha_i\left(y_i\left(w^T x_i + b\right) - 1 + \xi_i\right) = 0,\ \ \alpha_i = C - \gamma_i,\ \ \gamma_i \xi_i = 0,\ \ i = 1, 2, \dots, n$$

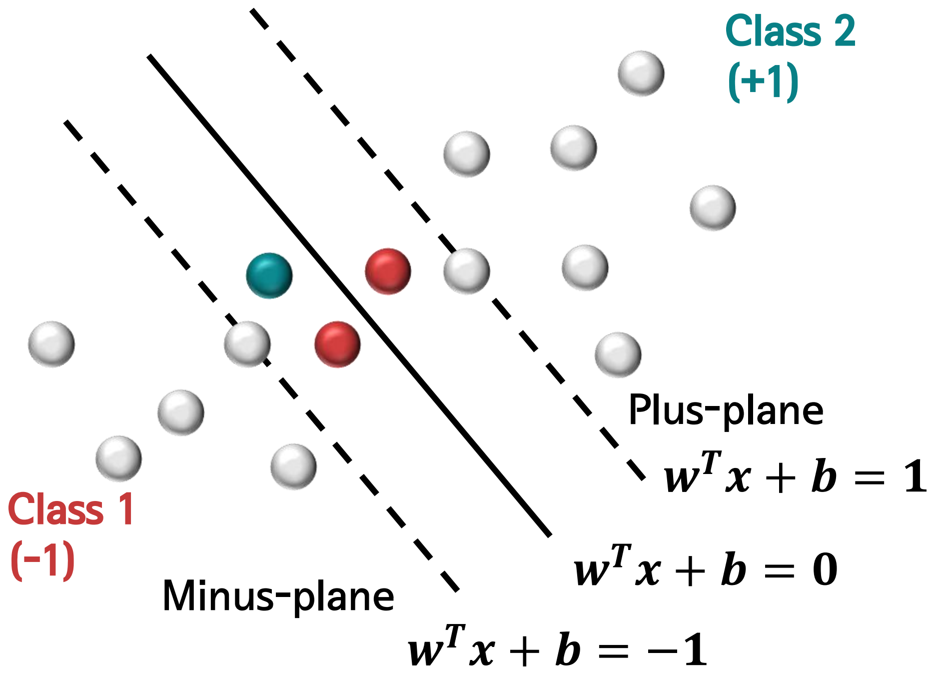
3) $\alpha_i = C \Rightarrow \ \gamma_i = 0$

$\Rightarrow \xi_i > 0$

$\Rightarrow \left(y_i\left(w^T + b\right) - 1\right) = \alpha_i \xi_i \neq 0$

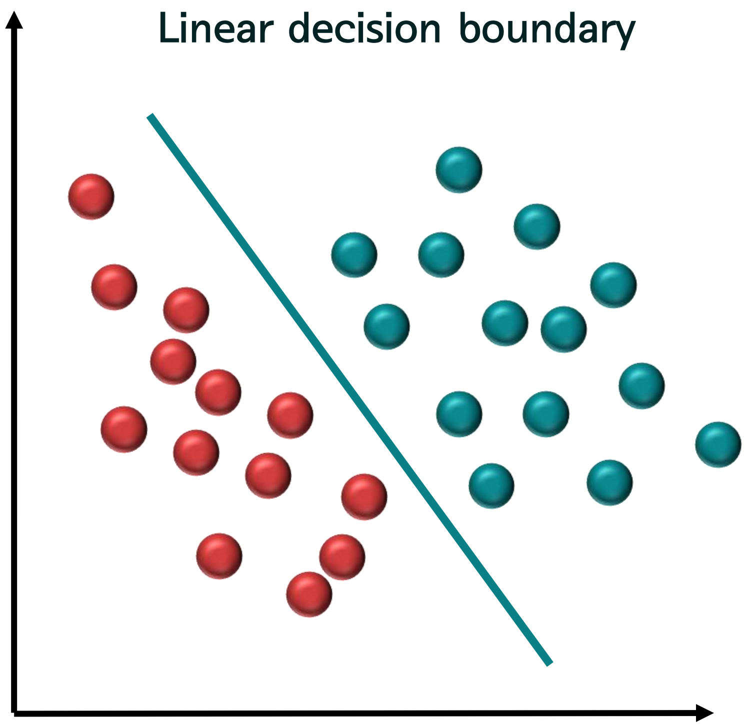
x_i가 plus-plane 또는 minus-plane (margin) 사이에 있음

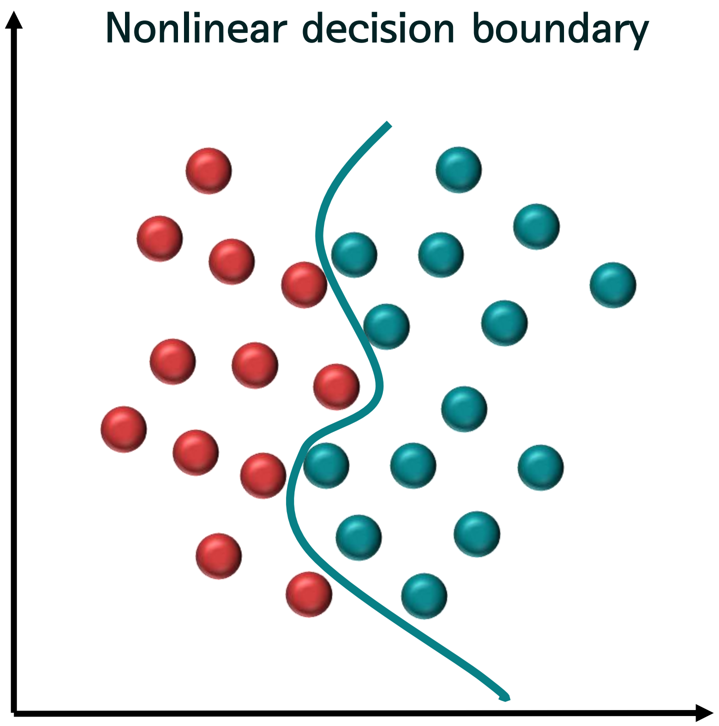
(support vector)



Nonlinear Decision Boundary

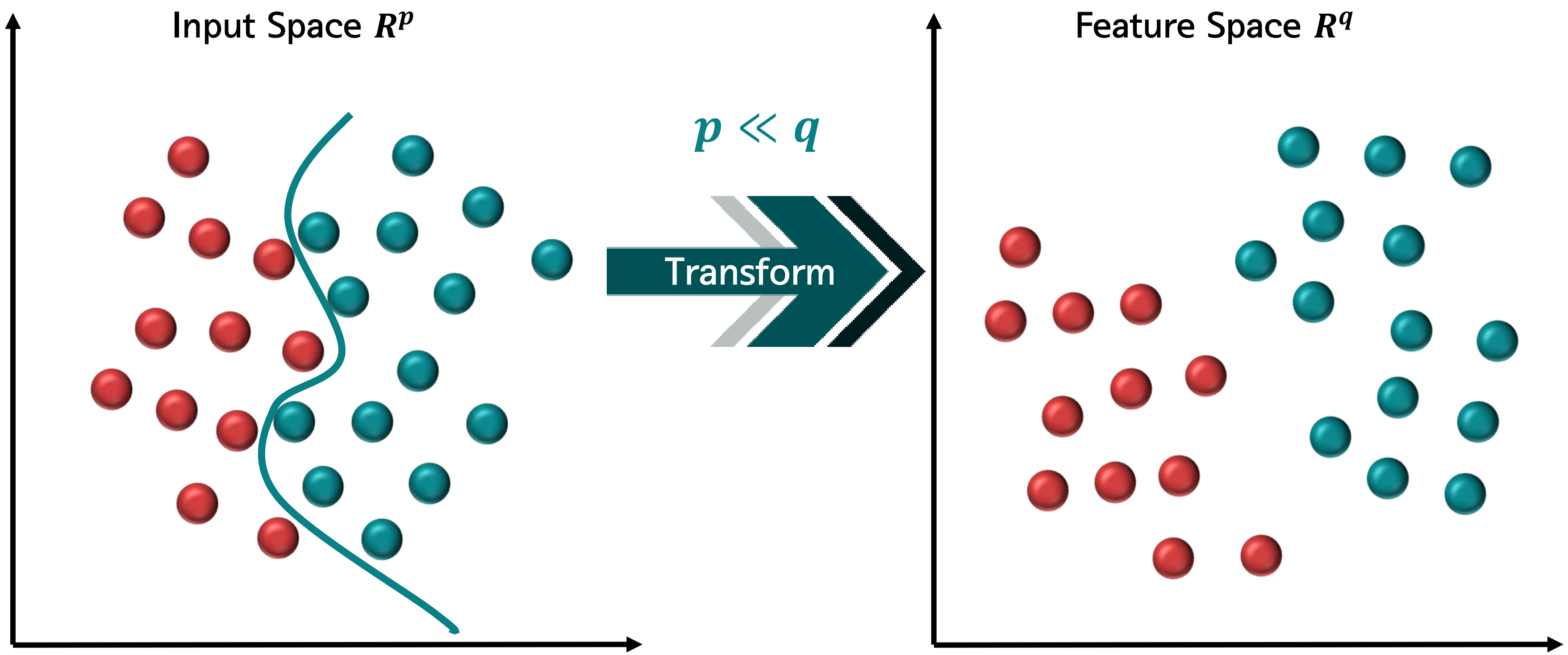
데이터에 비선형성이 있을 때는 어떻게 분류할까?



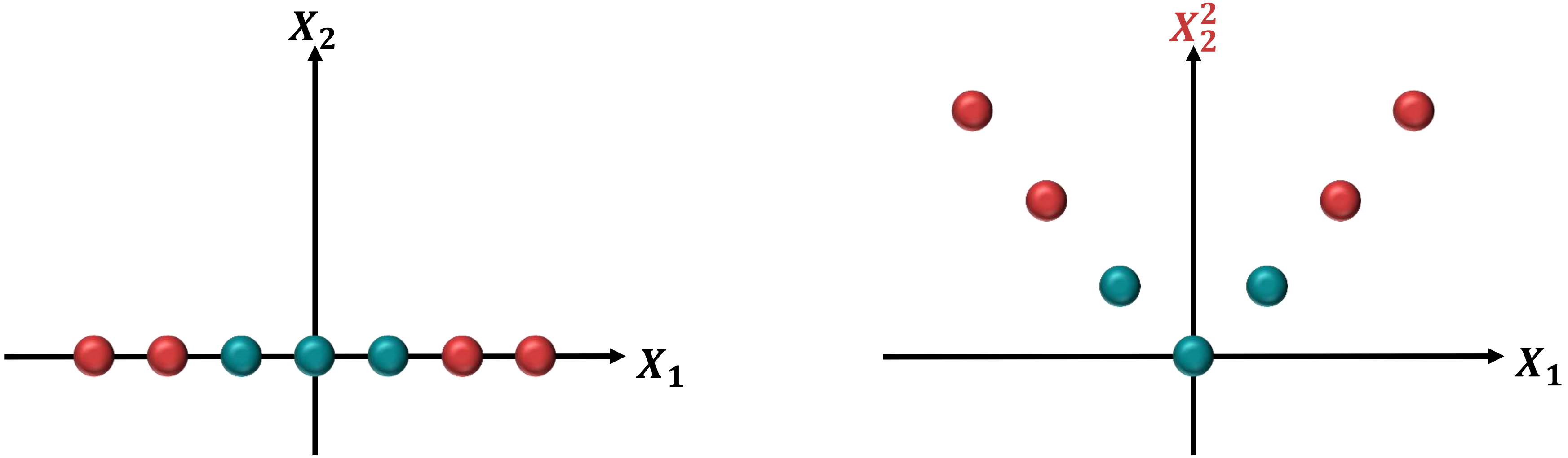


Transforming Data

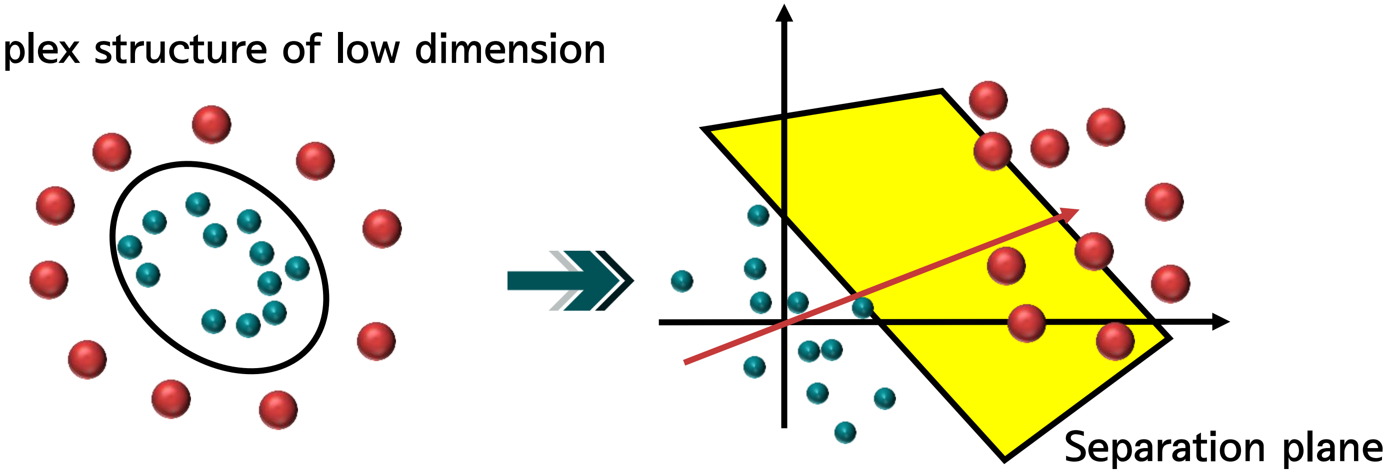
$$x = (x_1, x_2, \dots, x_n) \rightarrow \phi(x) = z = (z_1, z_2, \dots, z_n)$$



Example of Transforming Data



Complex structure of low dimension



Simple structure of high dimension

Example of Transforming Data

$$\boldsymbol{\phi}: \boldsymbol{x} \mapsto \boldsymbol{z} = \boldsymbol{\phi}(\boldsymbol{x})$$

$$\boldsymbol{\phi}: (\boldsymbol{x_1}, \boldsymbol{x_2}) \mapsto (\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_1^2}, \boldsymbol{x_2^2}, \boldsymbol{x_1}, \boldsymbol{x_2})$$

2D (original space) → 6D (feature space)

- original space를 feature space로 변환하여 학습에 활용
- original space에서 nonlinear decision boundary ➔ feature space에서 linear decision boundary
- 고차원 feature space에서 관측치 분류가 더 쉬울 수 있음
    - 고차원 feature space를 효율적으로 계산할 수 있는 방법 존재 (Kernel mapping, kernel trick)

Kernel Mapping

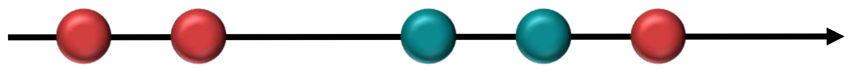
$$\hat{y}_{new} = \text{sign}\left(\sum_{i \in SV} a_i^* y_i x_i^T x_{new} + b^*\right)$$

$$\hat{y}_{new} = \text{sign}\left(\sum_{i \in SV} a_i^* y_i \phi(x_i)\phi(x_{new}) + b^*\right)$$

$$= \text{sign}\left(\sum_{i \in SV} a_i^* y_i K\langle \phi(x_i), \phi(x_{new})\rangle + b^*\right)$$

Example of Nonlinear SVM using Kernel Function

x	y
1	+1
2	+1
4	-1
5	-1
6	+1



- 1-dimensional data 분류 문제
- C=100
- Linearly non-separable problem ➔ kernel function 사용

$$\boldsymbol{K}\langle \boldsymbol{x}, \boldsymbol{y}\rangle = (\boldsymbol{xy} + \mathbf{1})^2$$

Example of Nonlinear SVM using Kernel Function

Solution

x	y
1	+1
2	+1
4	-1
5	-1
6	+1

$$\max_{\alpha} \sum_{i=1}^{5} \alpha_i - \frac{1}{2} \sum_{i=1}^{5} \alpha_i \alpha_j y_i y_j \left(x_i x_j + 1\right)^2$$

$$subject\ to \sum_{i=1}^{5} \alpha_i y_i = 0\ and\ 0 \leq \alpha_i \leq 100$$

α_1	α_2	α_3	α_4	α_5
0	2.5	0	7.333	4.833

$$f(x) = \sum_{i \in SV} a_i^* y_i K\langle \phi(x_i), \phi(x_{new}) \rangle + b$$

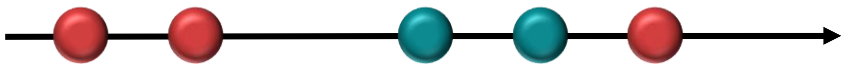
Example of Nonlinear SVM using Kernel Function

Solution

$$\begin{aligned} f(x) &= 2.5(+1)(2x+1)^2 + 7.333(-1)(5x+1)^2 + 4.833(+1)(6x+1)^2 + b \\ &= 0.667x^2 - 5.333x + b \end{aligned}$$

$$x = 2 \Rightarrow f(2) = 0.667 * 2^2 - 5.333 * 2 + b = 1$$
$$b = 8.993$$

$$f(x) = 0.667x^2 - 5.333x + 8.993$$



Choosing Kernel Functions

SVM 사용시 kernel을 결정하는 것은 어려운 문제 (기준은 없음)

사용하는 kernel에 따라 feature space의 특징이 달라지기 때문에 데이터의 특성에 맞는 kernel 결정해야 함

일반적으로 RBF kernel, sigmoid kernel, low degree polynomial kernel 등이 사용됨

- linear kernel $\boldsymbol{K}\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle = \langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle$
- polynomial kernel $\boldsymbol{K}\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle = (\mathbf{a}\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle + \boldsymbol{b})^{\boldsymbol{d}}$
- sigmoid kernel $\boldsymbol{K}\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle = \boldsymbol{tanh}(\mathbf{a}\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle + \boldsymbol{b})$
- RBF kernel (Gaussian kernel, Radial basis function) $\boldsymbol{K}\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle = \boldsymbol{exp}\left(\frac{-\|\boldsymbol{x_1} - \boldsymbol{x_2}\|_2^2}{2\boldsymbol{\sigma}^2}\right)$

End of slide