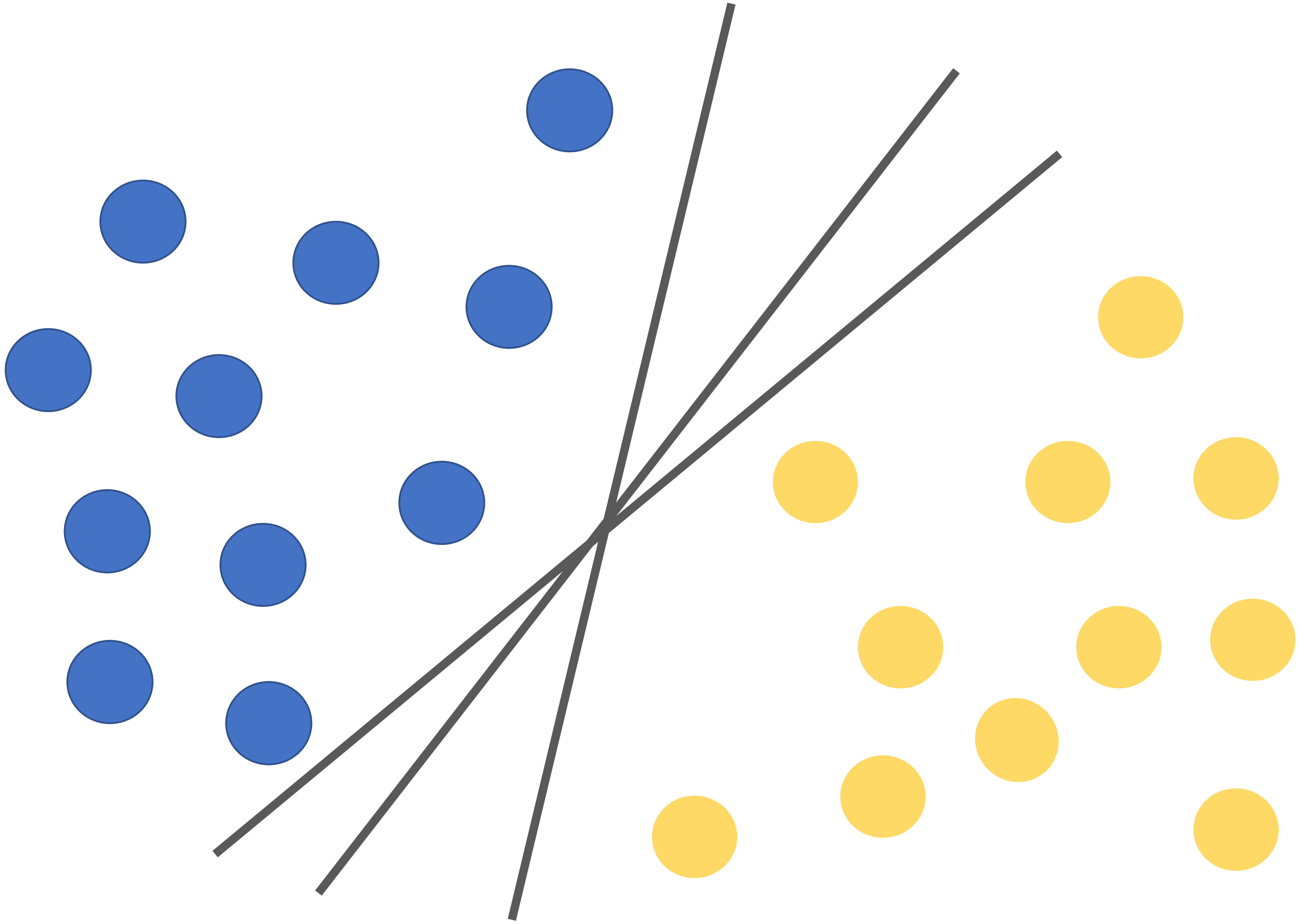
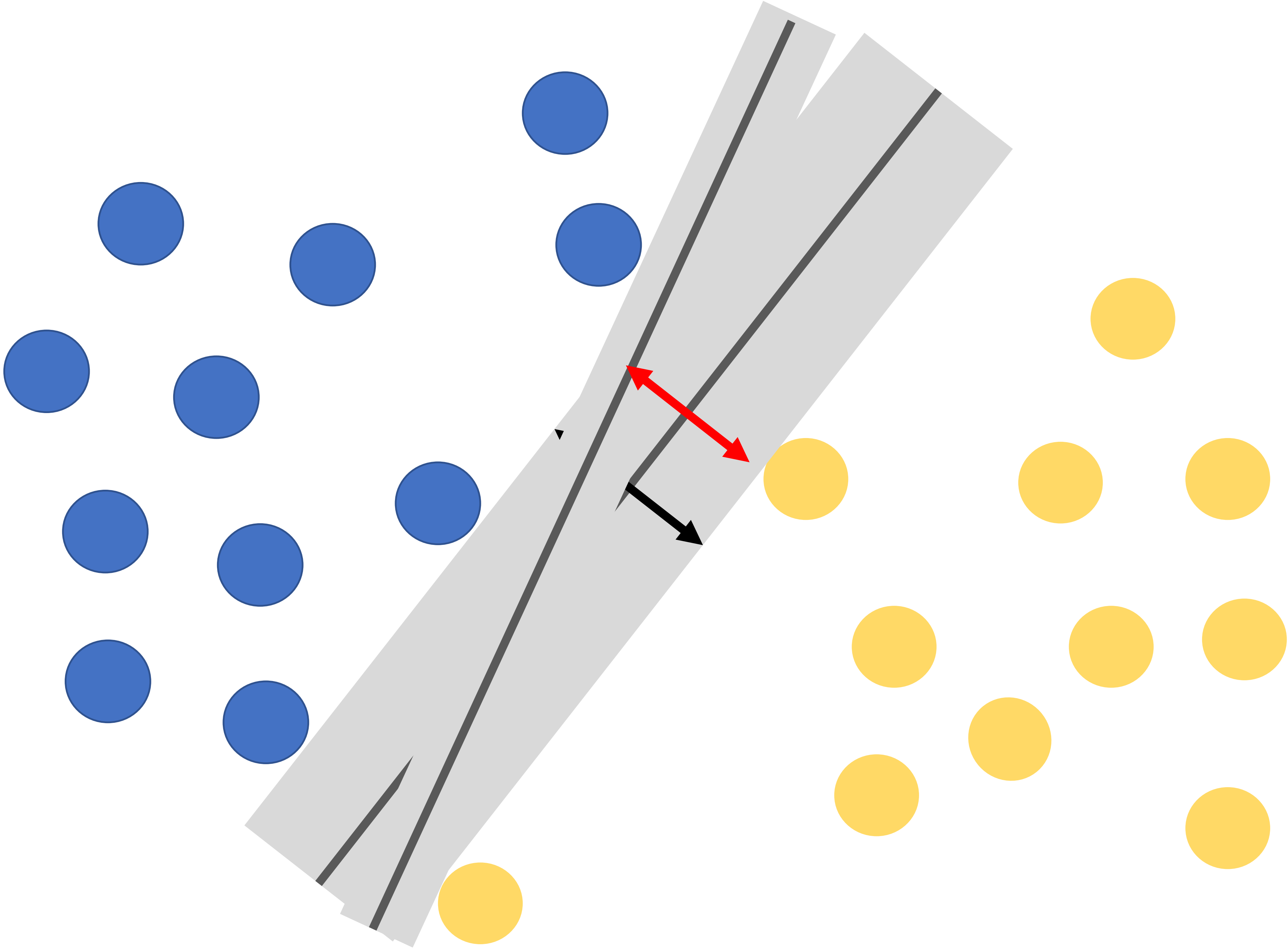
Support Vector Machine(SVM)

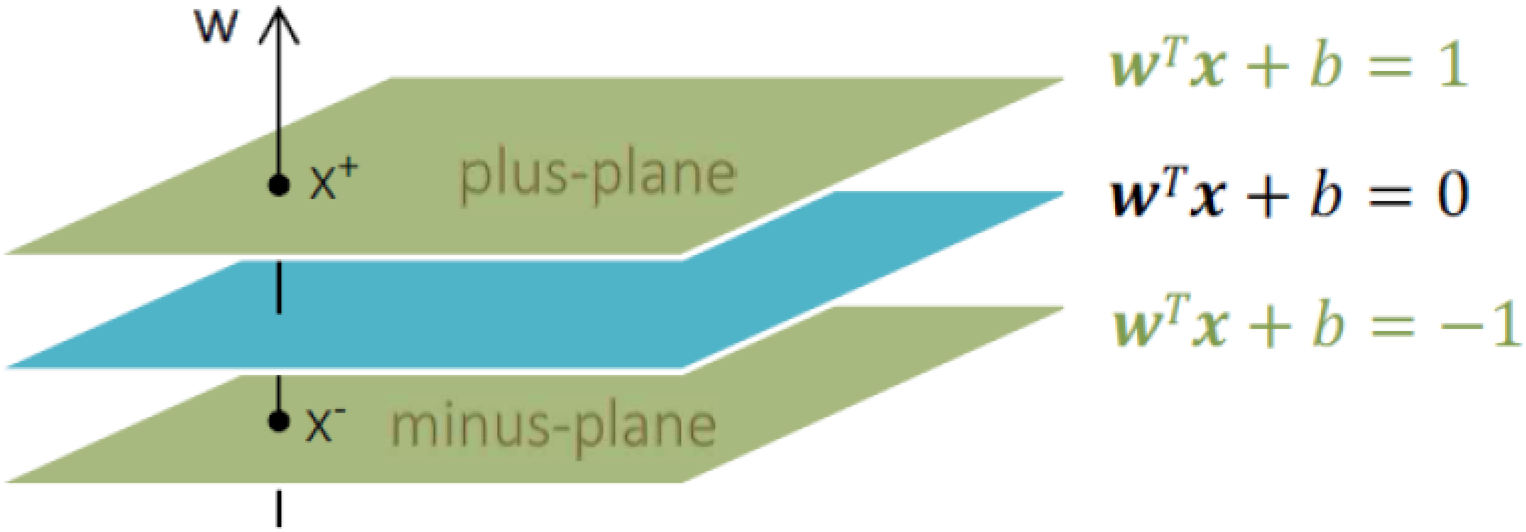
김광호, 서유정, 이현빈

What is SVM?

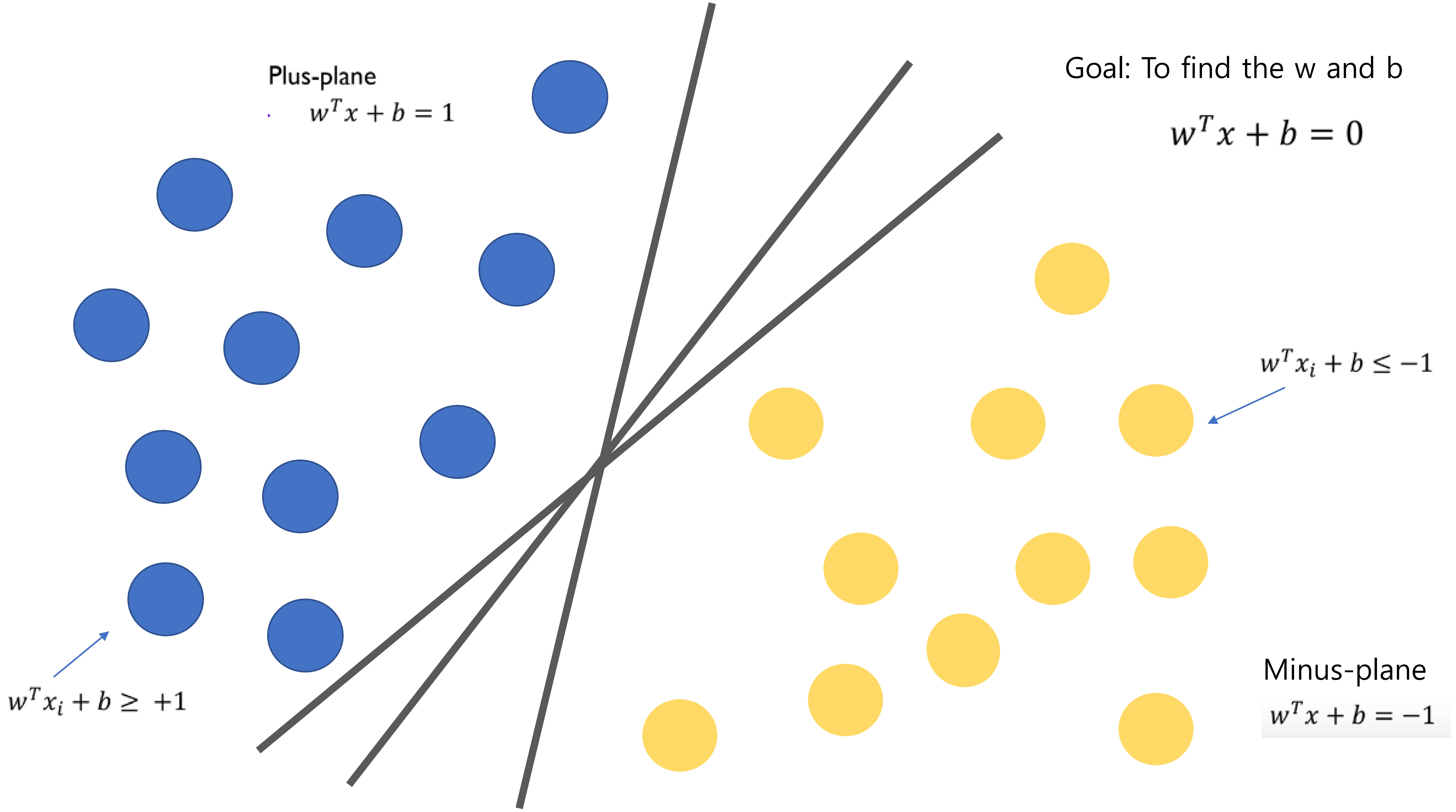
It’s a classification technique







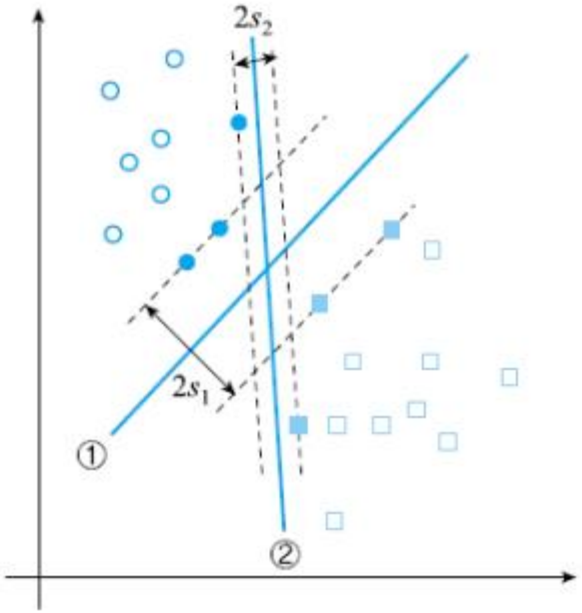
https://imgur.com/afe8W3S


Plus-plane
$w^T x + b = 1$
Goal: To find the w and b
$w^T x + b = 0$
$w^T x_i + b \leq -1$
$w^T x_i + b \geq +1$
Minus-plane
$w^T x + b = -1$

Decision hyperplane

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

$$h = \frac{|d(x)|}{||w||}$$



Training Set : $X = \{(x_1, t_1), (x_2, t_2), \ldots, (x_n, t_n)\}$

Margin: $2h = \frac{2|d(x)|}{||w||} = \frac{2}{||w||}$

How to maximize the margin?

그림 5.2에 있는 결정 직선의 수학적 특성을 살펴보자. 이 직선의 매개 변수는 $\mathbf{w} = (2,1)^T$이고 $b = -4$이다.

$$2h = \frac{2|d(\mathbf{x})|}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|}$$

Maximize $\dfrac{2}{\|\mathbf{w}\|}$

under the condition of

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1, \forall \mathbf{x}_i \in \omega_1$$

$$\mathbf{w}^T\mathbf{x}_i + b \leq -1, \forall \mathbf{x}_i \in \omega_2$$

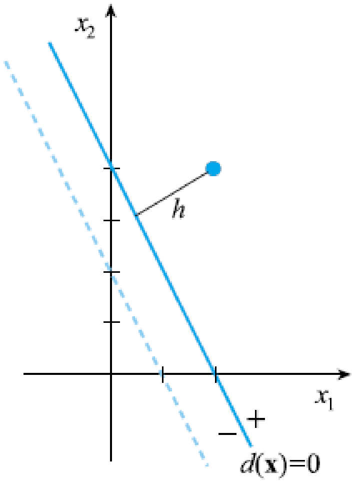


그림 5.2 직선의 수학적 특성

점 $\mathbf{x}=(2,4)^T$에서 직선까지 거리

$$h = \frac{|2\times 2 + 1\times 4 - 4|}{\sqrt{2^2 + 1^2}} = \frac{4}{\sqrt{5}} = 1.78885$$

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

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

converted to
Lagrange Multiplier Method

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

$$\textit{subject to}\ \ \alpha_i \geq 0, i = 1,2, \ldots \ldots, n$$

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda}) = J(\boldsymbol{\theta}) - \sum_{i=1}^{n} \lambda_i f_i(\boldsymbol{\theta})$$

$$\left.\begin{aligned} \frac{\partial L}{\partial \theta_i} &= 0, i = 1, \cdots, k \\ \frac{\partial L}{\partial \lambda_i} &= 0, i = 1, \cdots, n \end{aligned}\right\}$$

To get minimized value

$f_i(\theta)$: Expression condition

$J(\theta)$: Objective function

θ : Parameter Vector

$\theta = (\theta_{1,\ldots,}\theta_k)^T$

KKT (Karush-Kuhn-Tucker) Condition

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = \mathbf{0} \quad \rightarrow \quad \mathbf{w} = \sum_{i=1}^{N} \alpha_i t_i \mathbf{x}_i$$

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = 0 \quad \rightarrow \quad \sum_{i=1}^{N} \alpha_i t_i = 0$$

$$\alpha_i \geq 0, i = 1, \cdots, N$$

$$\alpha_i (t_i (\mathbf{w}^{\mathrm{T}} \mathbf{x}_i + b) - 1) = 0, i = 1, \cdots, N$$

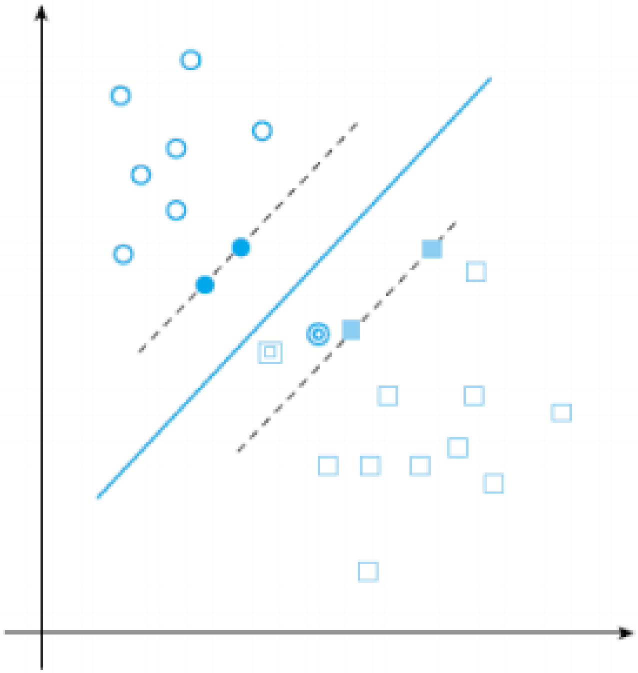
$$\underset{\alpha}{\text{maximize}} \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

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

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

Constrained optimization is transformed to quadratic optimization

Slack variable SVV



Case 1. Vector is outside the seperation band.
Is subject to $t(\mathbf{w}^{T}\mathbf{x}+b)<0$

Case 2. Vector is within its separation band.
Is subject to $0 \leq t(\mathbf{w}^{T}\mathbf{x}+b)<1$

Case 3. Vector is placed on the opposite side of the separation band.
Is subject to $t(\mathbf{w}^{T}\mathbf{x}+b)<0$

Given $x = (x_1, x_2)^T, y = (y_1, y_2)^T$

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^2 = x_1{}^2 y_1{}^2 + 2x_1 y_1 x_2 y_2 + x_2{}^2 y_2{}^2$$

$$\begin{aligned} K(\mathbf{x}, \mathbf{y}) &= (\mathbf{x} \cdot \mathbf{y})^2 \\ &= x_1{}^2 y_1{}^2 + 2x_1 y_1 x_2 y_2 + x_2{}^2 y_2{}^2 \\ &= (x_1{}^2, \sqrt{2} x_1 x_2, x_2{}^2)^{\mathrm{T}} \cdot (y_1{}^2, \sqrt{2} y_1 y_2, y_2{}^2)^{\mathrm{T}} \\ &= \mathbf{\Phi}_1(\mathbf{x}) \cdot \mathbf{\Phi}_1(\mathbf{y}) \end{aligned}$$

$$\left.\begin{aligned} K(\mathbf{a}, \mathbf{b}) &= ((0,0)^{\mathrm{T}} \cdot (1,0)^{\mathrm{T}})^2 = 0 \\ \mathbf{\Phi}_1(\mathbf{a}) \cdot \mathbf{\Phi}_1(\mathbf{b}) &= ((0,0,0)^{\mathrm{T}} \cdot (1,0,0)^{\mathrm{T}}) = 0 \end{aligned}\right\} \rightarrow \quad K(\mathbf{a}, \mathbf{b}) = \mathbf{\Phi}_1(\mathbf{a}) \cdot \mathbf{\Phi}_1(\mathbf{b})$$

$$\left.\begin{aligned} K(\mathbf{c}, \mathbf{d}) &= ((0,1)^{\mathrm{T}} \cdot (1,1)^{\mathrm{T}})^2 = 1 \\ \mathbf{\Phi}_1(\mathbf{c}) \cdot \mathbf{\Phi}_1(\mathbf{d}) &= ((0,0,1)^{\mathrm{T}} \cdot (1,\sqrt{2},1)^{\mathrm{T}}) = 1 \end{aligned}\right\} \rightarrow \quad K(\mathbf{c}, \mathbf{d}) = \mathbf{\Phi}_1(\mathbf{c}) \cdot \mathbf{\Phi}_1(\mathbf{d})$$

Kernel Substitution

When a certain mathematical expression includes vector's dot product, calculate the dot product by substituting it to kernel function.

$$
\begin{aligned}
d(\mathbf{x}) &= \mathbf{w}^{\mathrm{T}}\mathbf{\Phi}(\mathbf{x}) + b \\
&= \left(\sum_{\mathbf{x}_k \in Y} \alpha_k t_k \mathbf{\Phi}(\mathbf{x}_k) \right)^{\mathrm{T}} \mathbf{\Phi}(\mathbf{x}) + b \\
&= \sum_{\mathbf{x}_k \in Y} \alpha_k t_k \mathbf{\Phi}(\mathbf{x}_k) \cdot \mathbf{\Phi}(\mathbf{x}) + b \\
&= \sum_{\mathbf{x}_k \in Y} \alpha_k t_k K(\mathbf{x}_k, \mathbf{x}) + b
\end{aligned}
$$