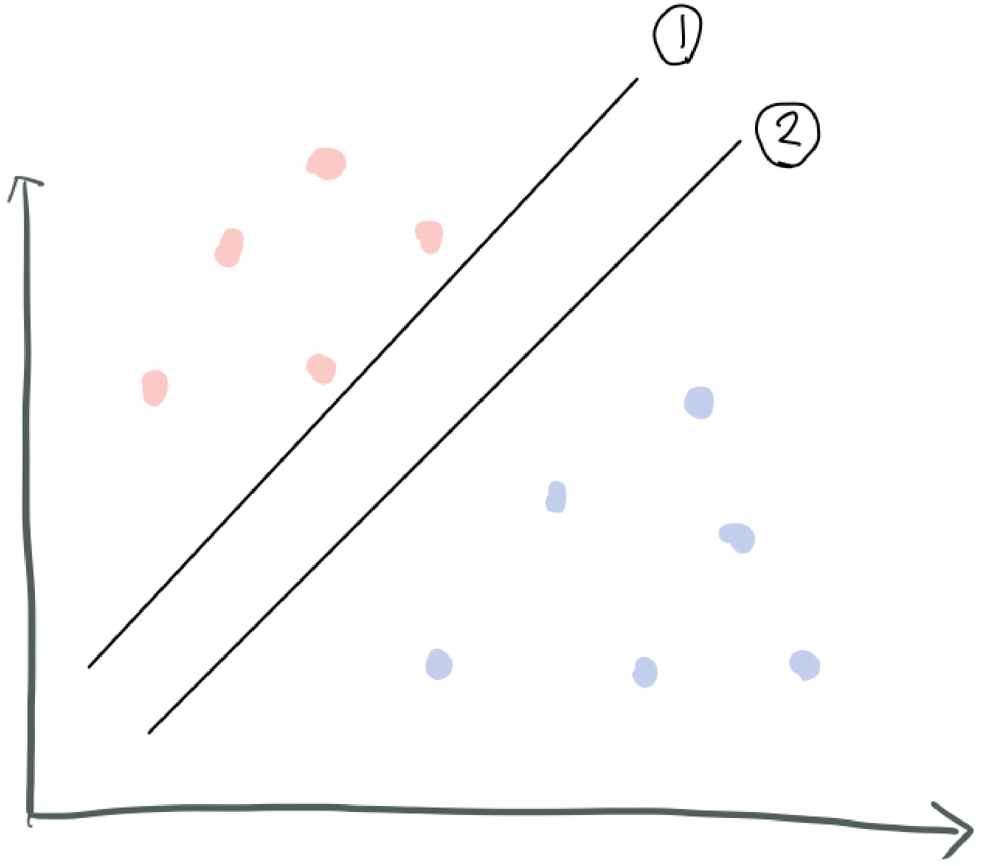
ML

Support Vector Machine

IDEA LAB

Linear SVM

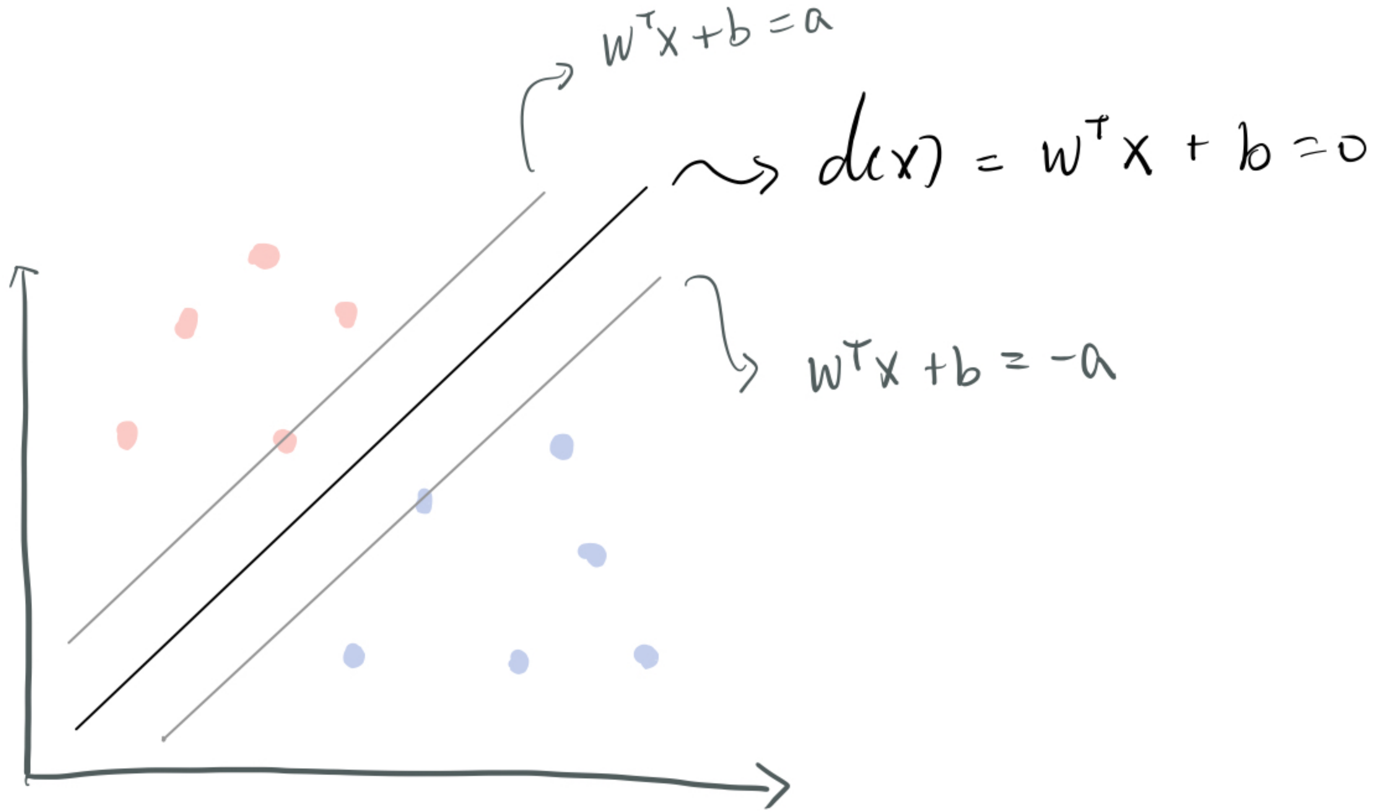


$$d(X) = W^T X + b = 0$$

Decision hyperplane

수학적 특징:

- feature space를 두 영역으로 분할 함
- 임의의 상수 c를 곱하여도 같은 초평면
- W는 초평면의 법선벡터로, 방향을 나타내고 b는 위치를 나타냄
- 임의의 점 X에서 초평면까지의 거리는 $h = \dfrac{|d(x)|}{\|W\|}$



$\oplus$ 샘플에 대해서

$$w^T x + b \geq a$$

$\ominus$ 샘플에 대해

$$w^T x + b \leq -a$$

$$y_i = \begin{cases} +1 & \text{for } \oplus \text{ 샘플} \\ -1 & \text{for } \ominus \text{ 샘플} \end{cases}$$ 로 두면

$$y_i (\bar{w}^T \bar{x}_i + b) \geq a$$

조건부 최적화 문제:

$$t_i(W^T X_i + b) - 1 \geq 0, i = 1, \ldots, N$$

Minimize $J(w) = \frac{1}{2} \|W\|^2$

문제의 특징:

해의 유일성: j(w)는 2차식이므로 convex function. → no local minimum.

어려운 문제: N개의 선형 부등식을 조건으로 가진 2차 함수의 최적화 문제.

→Lagrange multiplier 도입

Linear SVM

Lagrange function

$$L(w, b, \alpha) = \frac{1}{2}\|W\|^2 - \sum_{i=1}^{N} \alpha_i (t_i(W^T X_i + b) - 1)$$

Karush-Kuhn-Tucker (KKT) 조건을 이용

KKT조건:

1. $\frac{\partial L(W,b,\alpha)}{\partial W} = 0 \quad \rightarrow \quad W = \sum_{i=1}^{N} \alpha_i t_i X_i$

2. $\frac{\partial L(W,b,\alpha)}{\partial b} = 0 \quad \rightarrow \quad \sum_{i=1}^{N} \alpha_i t_i = 0$

3. $\alpha_i \geq 0, \quad i = 1, \ldots, N$

4. $\alpha_i (t_i(W^T X_i + b) - 1) = 0, \quad i = 1, \ldots, \mathrm{N}$

t_i와 X_i 는 주어진 데이터 → α_i 만 구하면 됨.

$t_i(W^T X_i + b) = 1$ 인 i 들이 <u>Support Vectors</u>

대입하고 정리하면,

조건

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

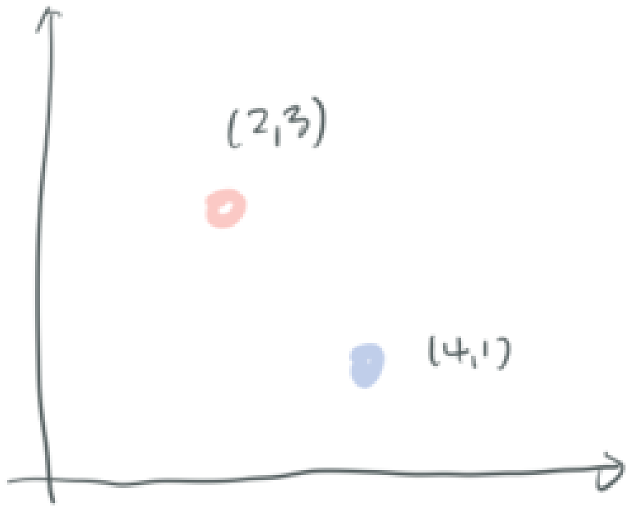
$\alpha_i \geq 0, \quad i = 1, \dots, N$

하에,

$\tilde{L}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j t_i t_j X_i^T X_j$ 를 최대화.

→ 하나의 등식 조건과, N 개의 부등식 조건을 가진 2차 (quadratic) 목적 함수의 최대화 문제.

<u>목적함수에서 두 벡터의 내적으로 계산.</u>



$x_1 = (2,3)^T,\ t_1 = 1$

$x_2 = (4,1)^T,\ t_2 = -1$

$\alpha_1 t_1 + \alpha_2 t_2 = 0$

$\alpha_1 \geq 0,\ \alpha_2 \geq 0$ 일 때,

$$\tilde{L}(\alpha) = (\alpha_1 + \alpha_2) - \frac{1}{2}(\alpha_1\alpha_1 t_1 t_1 x_1^T x_1 + \alpha_1\alpha_2 t_1 t_2 x_1^T x_2 + \alpha_2\alpha_1 t_2 t_1 x_2^T x_1 + \alpha_2\alpha_2 t_2 t_2 x_2^T x_2)$$ 를 최대화

$$\alpha_1 - \alpha_2 = 0$$

$\alpha_1 \geq 0, \quad \alpha_2 \geq 0$ 일 때,

$\tilde{L}(\alpha) = (\alpha_1 + \alpha_2) - \frac{1}{2}(13\alpha_1^2 + 4\alpha_2^2 - 22\alpha_1\alpha_2)$ 을 최대화

$$\alpha_1 = \alpha_2$$

$$\tilde{L}(\alpha) = -4\alpha^2 + 2\alpha$$

$$= -4\left\{\left(\alpha - \frac{1}{4}\right)^2 - \frac{1}{16}\right\}$$

$\therefore \ \alpha_1 = \alpha_2 = \frac{1}{4}$ 에서 최대값.

$$W = \sum_{i=1}^{N} \alpha_i t_i x_i = \frac{1}{4}(2,3)^T - \frac{1}{4}(4,1)^T = \left(-\frac{1}{2}, \frac{1}{2}\right)^T$$

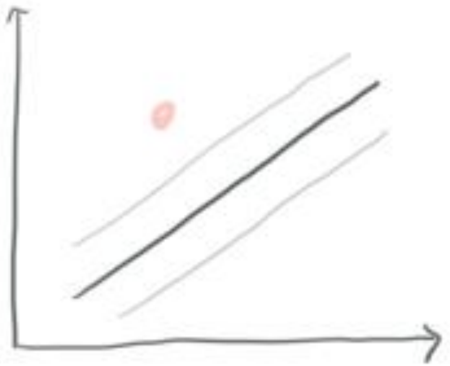
$b \rightarrow \quad \alpha_i (t_i (w^T x_i + b) - 1) = 0$

$$\frac{1}{4}\left[\left(-\frac{1}{2}, \frac{1}{2}\right)\begin{pmatrix}2\\3\end{pmatrix} + b - 1\right] = 0$$

$$b - \frac{1}{2} = 0 \rightarrow b = \frac{1}{2}$$

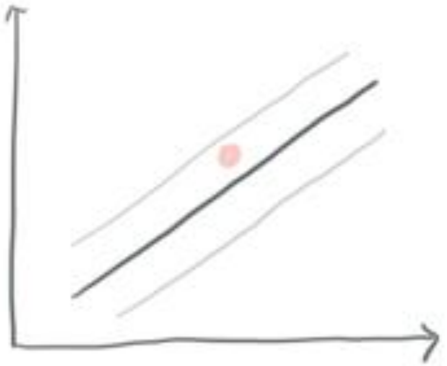
$$\therefore \ d(x) = w^T x + b = \left(-\frac{1}{2}, \frac{1}{2}\right)\begin{pmatrix}x_1\\x_2\end{pmatrix} + \frac{1}{2}$$

$$= -\frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}$$



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

$\xi = 0$



$0 \leq t(w^T x + b) < 1$

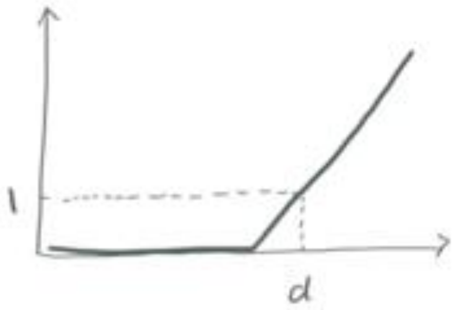
$0 < \xi \leq 1$

$t(w^T x + b) < 0$

$\xi > 1$

$$t(w^T x + b) \geq 1 - \xi$$

SVM Loss



$$J(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

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

$0 \le \alpha_i \le C$, $i = 1, \cdots, N$ 일 때,

$\tilde{L}(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j t_i t_j x_i^T x_j$ 를 최대화

Kernel function: 다른 공간으로 사상 된 두 벡터의 내적

SVM Kernel function의 성질:

L 공간 상의 두 벡터 X와 Y를 매개 변수로 갖는 커널 함수를 $K(X,Y)$라 할 때,

$K(X,Y) = \phi(X) \cdot \phi(Y)$를 만족하는 맵핑 함수 $\phi(\cdot)$이 존재해야 한다.

즉 커널 함수의 값과, H공간 상으로 맵핑된 두 점 $\phi(X)$와 $\phi(Y)$의 내적이 같아야 한다.

하나의 kernel function에 대응하는 맵핑 함수는 여러 개 존재 가능.

$$\tilde{L}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j t_i t_j \underline{X_i^T X_j}$$

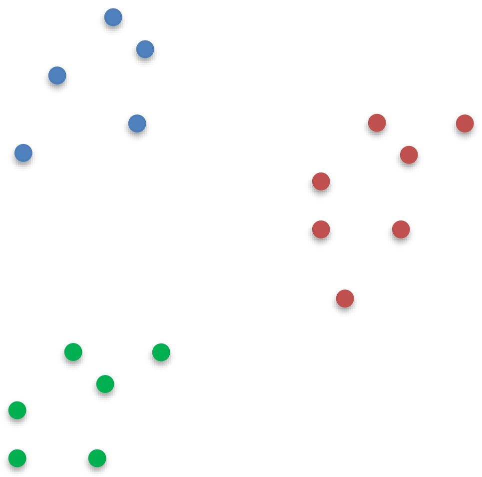
실제 계산은 저차원 L 공간에서 이루어지지만 분류 작업은 고차원 공간인 H 공간에서 수행되는 효과

$$\tilde{L}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j t_i t_j \underline{\phi(X_i)^T \phi(X_j)}$$

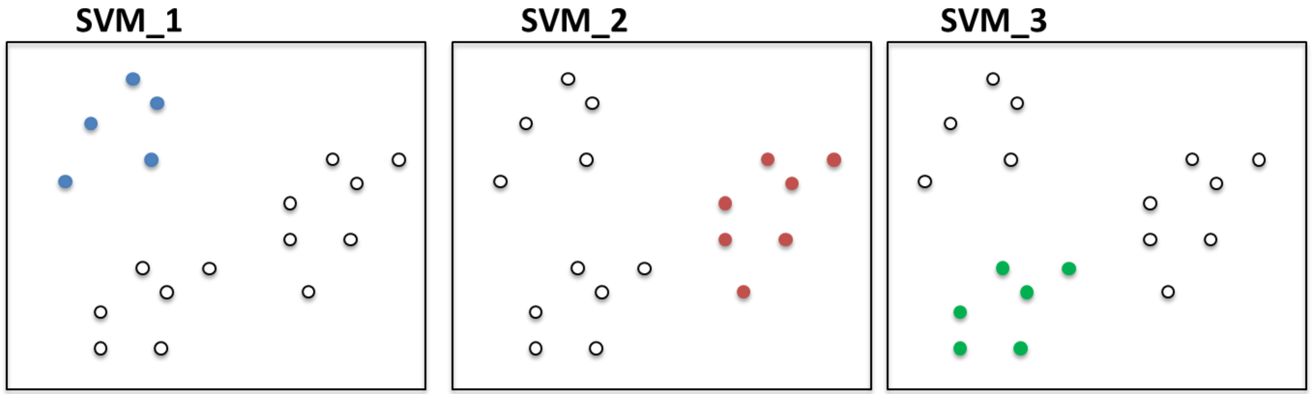
$$K(X_i, X_j)$$

- Polynomial $K(X,Y) = (X \cdot Y + 1)^p$
- RBF (Radial Basis Function) $K(X,Y) = e^{-\|X-Y\|^2/2\sigma^2}$
- Hyperbolic tangent $K(X,Y) = \tanh(\alpha X \cdot Y + \beta)$

SVM은 기본적으로 이진분류기 !

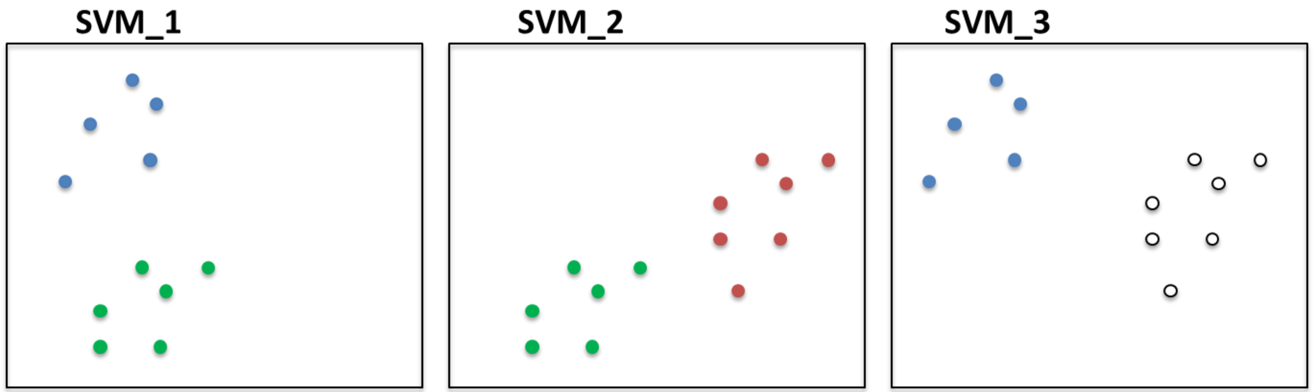


OAA(One Against All) 혹은 One-versus-rest(OVR)



→ $d(i)$값을 비교하여 결정

OVO(One-versus-one)



→ Majority voting