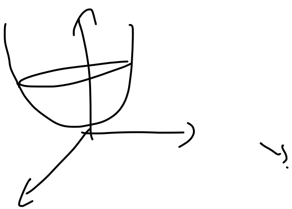
Lagrange Multiplier.

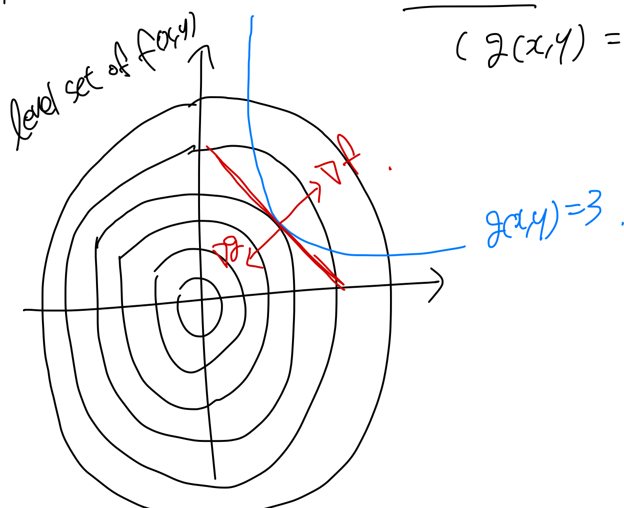
Find min / max with constrained.

min $f(x,y,z)$ where $g(x,y,z) = c$.

ex1. min $f(x,y) = x^2+y^2$ subject to $xy = 3$.

$(g(x,y) = xy)$.





$\nabla f \parallel \nabla g$.

$$\boxed{\nabla f = \lambda \cdot \nabla g.}$$

Lagrange multiplier.

If a surface S is defined by a number of constraints

$$\begin{cases} g_1(x_1, \cdots, x_n) = c_1 \\ g_2(x_1, \cdots, x_n) = c_2 \\ \quad \vdots \\ g_k(x_1, \cdots, x_n) = c_k. \end{cases} \Rightarrow \nabla f = \lambda_1 \cdot \nabla g_1 + \lambda_2 \nabla g_2 + \cdots + \lambda_k \cdot \nabla g_k.$$

Duality

vector space $\qquad\qquad$ dual space of V.

$$V \simeq V^* = L(V, \mathbb{R})$$

$Ax \qquad\qquad A^T f_x$

$$\begin{tikzcd}
x \arrow[r] \arrow[d] & Ax \\
f_x \arrow[r] & A^T f_x \arrow[u]
\end{tikzcd}$$

Laplace transform?

$$\mathcal{L}(f(t)) = \int_0^\infty e^{-st} f(t)\,dt = F(s)$$

미분 방정식

$$af'' + af' + af = 0 \xrightarrow{\mathcal{L}} aF'' + \sim = 0$$

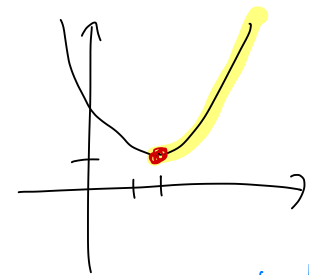
$$\downarrow$$

$$f = a \xleftarrow{\mathcal{L}^{-1}} \boxed{F = a}$$

$$\not\downarrow$$

$f = ??$

$\min \ \frac{1}{2}x^2 - 2x + 3$ subject to $x - 2 \geq 0$. $\longrightarrow \frac{1}{2}(x-2)^2 + 1$.



$L_p : \frac{1}{2}x^2 - 2x + 3 - \lambda(x-2)$.

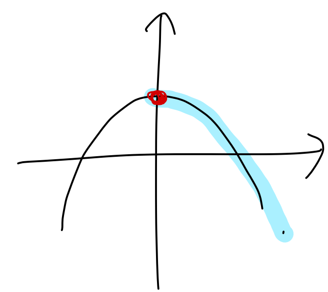
$\frac{\partial L_p}{\partial x} = x - 2 - \lambda = 0$.

$\boxed{x = 2 + \lambda}$.

min L_p $\Updownarrow$ max L_D.

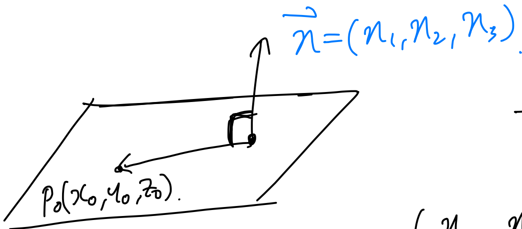
$L_D : \frac{1}{2}(2+\lambda)^2 - 2(2+\lambda) + 3 - \lambda(2+\lambda-2)$.

$= -\frac{1}{2}\lambda^2 + 1$. $\quad \lambda \geq 0$.



KTT condition : $\lambda(x-2) = 0$. 둘 중 하나만 0.

Equation of plane. (= hyper plane).



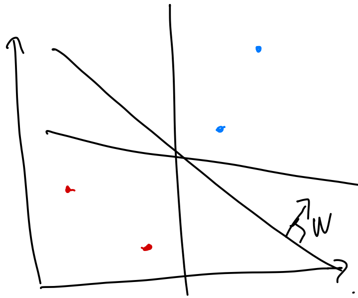
$$\vec{n} \cdot (X - P_0) = 0.$$

$$(n_1, n_2, n_3) \cdot (x - x_0, y - y_0, z - z_0) = 0.$$

$$\boxed{n_1 x + n_2 y + n_3 z + \cancel{d} = 0.}$$

$\vec{n}$ 알면 평면 결정됨.

Hard margin SVM

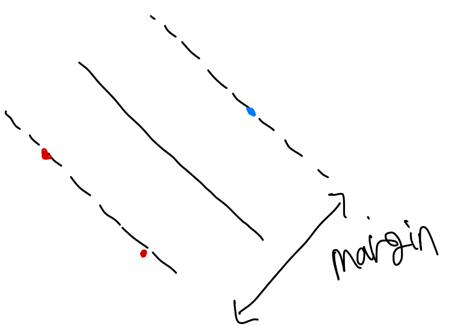


$\boxed{\bar{w}\bar{x}+b=0}$.

Decision Rule.

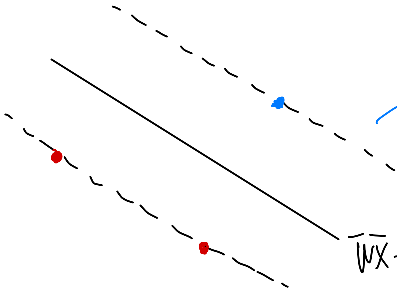
$\bar{w}\cdot\bar{u}+b \geq 0$ then $+$.

$\bar{w}\cdot\bar{u}+b < 0$ then $-$.



가장 margin이 큰

$\bar{w}\bar{x}+b$ 선택.



$\bar{w}\,\bar{x}_+ + b \geq 1$.

$\bar{w}\bar{x} + b = 0$.

$\bar{w}\cdot\bar{x}_- + b \geq -1$.

why 1?

계산하기 편해서.

def y_i s.t $y_i = 1$ for + samples

$y_i = -1$ for − samples.

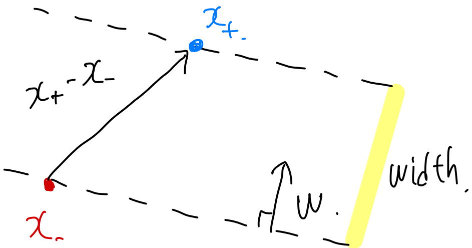
$\Rightarrow \quad y_i(\bar{w}\cdot\bar{x}_i + b) \geq 1$.

$\Rightarrow \quad y_i(\bar{w}\cdot\bar{x}_i + b) - 1 \geq 0$. : constraints.

"boundary condition":

& $\boxed{y_i(\bar{w}\cdot\bar{x}_i + b) - 1 = 0}$

for boundary samples.



$$y_i(\bar{w}\cdot\bar{x}_i + b) - 1 = 0$$

$$\bar{w}\cdot\bar{x}_+ + b - 1 = 0 \qquad -\bar{w}\cdot\bar{x}_- - b - 1 = 0$$

$$\bar{w}\cdot\bar{x}_+ = 1 - b \qquad \bar{w}\cdot\bar{x}_- = -b - 1$$

$$\text{width} = (x_+ - x_-)\cdot\frac{w}{\|w\|}$$

$$= \frac{1}{\|w\|}(wx_+ - wx_-)$$

$$= \frac{1}{\|w\|}(1 - b - (-b - 1))$$

$$= \frac{2}{\|w\|}$$

We want to maximize width $\frac{2}{\|W\|}$.

$\Leftrightarrow$ minimize $\|W\|$.

$\Leftrightarrow$ minimize $\frac{1}{2}\|W\|^2$.

We want to minimize $\frac{1}{2}\|\overline{w}\|^2$

subject to $y_i(\overline{w}\cdot\overline{x}_i+b)-1$ for all i.

$$\Rightarrow L = \frac{1}{2}\|\overline{w}\|^2 - \sum \lambda_i\left[y_i(\overline{w}\cdot\overline{x}_i+b)-1\right], \quad \lambda_i \geq 0$$

$$L = \frac{1}{2}\|\bar{w}\|^2 - \sum \lambda_i \left[y_i (\bar{w} \cdot \bar{x}_i + b) - 1 \right] . \qquad \lambda_i \geq 0 .$$

$$\frac{\partial L}{\partial w} = \bar{w} - \sum \lambda_i y_i x_i = 0 .$$

$$\Longrightarrow \bar{w} = \sum \lambda_i y_i x_i .$$

$\bar{w}$ is linear sum of samples. → "Support Vector."

(대부분 $\lambda_i = 0$).

$$\frac{\partial L}{\partial b} = \sum \lambda_i y_i = 0 .$$

$$L_P = \frac{1}{2}\|\bar{w}\|^2 - \sum \alpha_i \left[y_i(\bar{w}\cdot\bar{x}_i + b) - 1 \right]. \qquad \alpha_i \geq 0.$$

$$\bar{w} = \sum \alpha_i y_i x_i \quad , \quad \sum \alpha_i y_i = 0.$$

$$L_D = \frac{1}{2}\left(\sum \alpha_i y_i \bar{x}_i\right)\left(\sum \alpha_j y_j \bar{x}_j\right) - \sum \alpha_i \left[y_i(\bar{w}\cdot\bar{x}_i + b) - 1 \right].$$

$$= \frac{1}{2}\left(\sum \alpha_i y_i \bar{x}_i\right)\left(\sum \alpha_j y_j \bar{x}_j\right) - \sum \alpha_i y_i \bar{x}_i \left(\sum \alpha_j y_j x_j\right) - \cancel{\sum \alpha_i y_i b} + \sum \alpha_i.$$

$$b \sum \alpha_i y_i = 0.$$

$$= \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \, \bar{x}_i \cdot \bar{x}_j.$$

$$\boxed{L_D = \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \, \boxed{\bar{x}_i \cdot \bar{x}_j} \quad \text{s.t. } \alpha_i \geq 0, \ \sum \alpha_i y_i = 0.}$$

→ L_D는 α에 대한 convex function.

→ depend only inner-product value.

$$\min \ L_p = \frac{1}{2}\|\bar{w}\|^2 - \sum \alpha_i [y_i(\bar{w} \cdot \bar{x}_i + b) - 1] \quad s.t \ \alpha_i \geq 0$$

$$\Big\Downarrow \text{ To dual problem}$$

$$\max \ L_D = \sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \bar{x}_i \cdot \bar{x}_j$$

$$s.t \ \sum \alpha_i y_i = 0 \ , \ \alpha_i \geq 0.$$

From KKT condition,

$$\alpha_i \left(y_i(\bar{w} \cdot \bar{x}_i + b) - 1 \right) = 0.$$

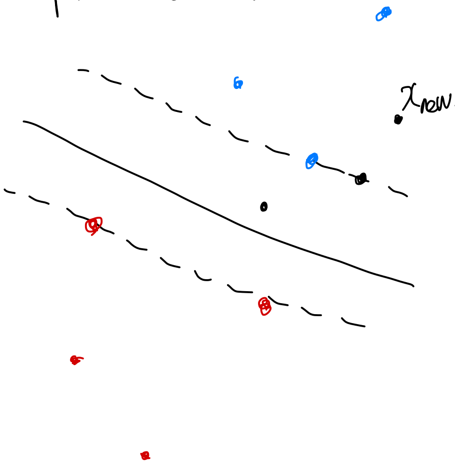
$$\begin{cases} \alpha_i = 0 \Rightarrow y_i(\bar{w} \cdot \bar{x}_i + b) - 1 \neq 0. \\ \alpha_i \neq 0 \Rightarrow \underline{y_i(\bar{w} \cdot \bar{x}_i + b) - 1 = 0}. \end{cases}$$

boundary 위에 있는 데이터.
(= support vector).

b can be computed by $y_i(\bar{w} \cdot \bar{x}_i + b) - 1 = 0$. for support vector x_i.

margin $\|W\|^2 = \sum \alpha_i$.

prediction.

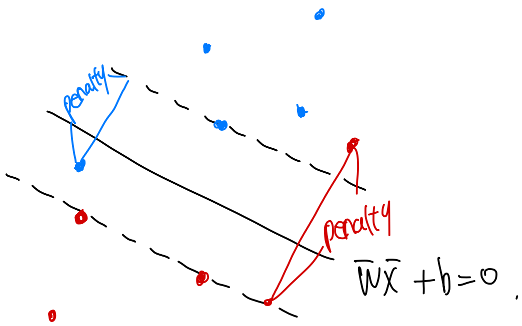


$$\bar{w}\cdot\bar{x}_{new} + b = \sum \alpha_i y_i \bar{x}_i \cdot \bar{x}_{new} + b \quad \left.\begin{array}{l} > 1 \\ = 1 \\ 0 \sim 1 \end{array}\right) +.$$

선형으로 분류가 불가능할 때?

- 선형으로 하되 오류 인정 & 패널티 부여 → Soft margin SVM
- 비선형으로. (차수 높여서) → kernel trick.

Soft margin SVM.



$\Rightarrow$ we want to minimize

$$\frac{1}{2}\|w\|^2 + C \cdot \sum \xi_i .$$

s.t $\quad y_i(\bar{w} \cdot \bar{x}_i + b) \geq 1 - \xi_i \quad , \quad \xi_i \geq 0 .$

given : $\bar{X}$, $\bar{Y}$

parameter : $\bar{w}$, b , ξ_i

hyper-parameter : C.

Lagrangian multiplier.

$$L_p = \frac{1}{2}\|W\|^2 + C \cdot \sum \eta_i - \sum \underline{\alpha_i} \left[y_i(\bar{w} \cdot \bar{x}_i + b) - 1 + \eta_i \right] - \sum \underline{\mu_i} \eta_i .$$

$$\text{s.t. } \alpha_i \geq 0, \ \mu_i \geq 0.$$

$$\begin{cases} \dfrac{\partial L}{\partial W} = 0 \Rightarrow \bar{w} = \sum \alpha_i y_i \bar{x}_i . \\[2ex] \dfrac{\partial L}{\partial b} = 0 \Rightarrow \sum \alpha_i y_i = 0 . \\[2ex] \dfrac{\partial L}{\partial \eta_i} = 0 \Rightarrow C - \alpha_i - \mu_i = 0 \quad \text{for all } i. \end{cases}$$

$$L_p = \frac{1}{2}\|\bar{w}\|^2 + C\cdot\sum \xi_i - \sum \alpha_i\left[y_i(\bar{w}\cdot\bar{x}_i + b) - 1 + \xi_i\right] - \sum \mu_i \xi_i$$

$$\bar{w} = \sum \alpha_i y_i \bar{x}_i$$

$$\sum (C - \alpha_i - \mu_i) = 0$$

$$L_D = \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \bar{x}_i\cdot\bar{x}_j + \cancel{C\sum \xi_i} - \sum_i\sum_j \alpha_i\alpha_j y_i y_j \bar{x}_i\cdot\bar{x}_j - b\cancel{\sum \alpha_i y_i}^{\,0}$$

$$+ \sum \alpha_i - \cancel{\sum \alpha_i \xi_i} - \cancel{\sum \mu_i \xi_i}$$

$$\Rightarrow \boxed{L_D = \sum \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \bar{x}_i\cdot\bar{x}_j}$$

(Same as hard margin SVD.)

조건만 추가됨.

s.t $\underline{C - \alpha_i - \mu_i = 0}$.

$\Leftrightarrow \quad 0 \le \alpha_i \le C$.

$$\min \quad L_p = \frac{1}{2}\|\bar{w}\|^2 + C \cdot \sum \eta_i - \sum \alpha_i [y_i(\bar{w} \cdot \bar{x}_i + b) - 1 + \eta_i] - \sum \mu_i \eta_i .$$

$$\text{s.t } \alpha_i \geq 0, \quad \mu_i \geq 0 .$$

$\Downarrow$ To dual problem.

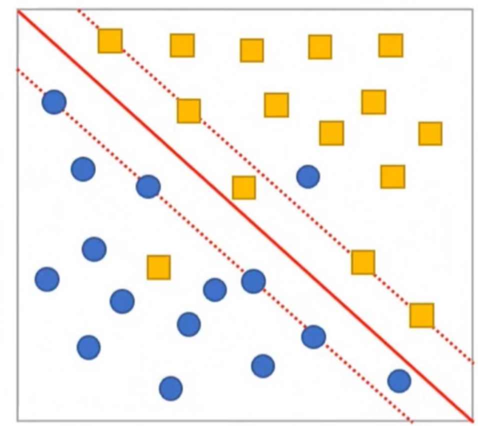
$$\max \quad L_p = \sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \bar{x}_i \cdot \bar{x}_j .$$

$$\text{s.t } \sum \alpha_i y_i = 0 \quad , \quad 0 \leq \alpha_i \leq C .$$

From KKT condition.

$$\begin{cases} \alpha_i (y_i (\bar{w} \cdot \bar{x}_i + b) - 1 + \eta_i) = 0 . \\ \mu_i \eta_i = 0 . \end{cases}$$

and $C - \alpha_i - \mu_i = 0 . \Rightarrow 0 \le \alpha_i \le C$



case 1)

$\alpha_i = 0$: non-Support vector. $(\eta_i = 0)$.

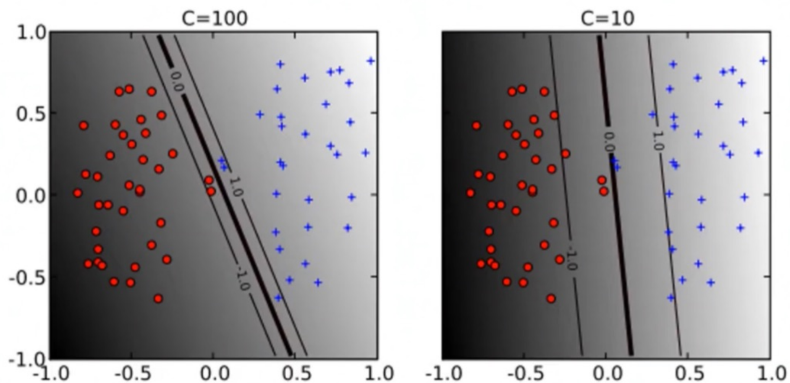
case 2)

$0 < \alpha_i < C$: support-vector on margin. $\begin{pmatrix} y_i(\bar{w} \cdot \bar{x}_i + b) = 1 . \\ \eta_i = 0 \end{pmatrix}$.

case 3)

$\alpha_i = C$: Support vector outside the margin. $(\eta_i > 0)$.

$$\min \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$



$C\uparrow \quad \xi_i \downarrow \quad \rightarrow$ margin $\downarrow$.

$C\downarrow \quad \xi_i \uparrow \quad \rightarrow$ margin $\uparrow$.

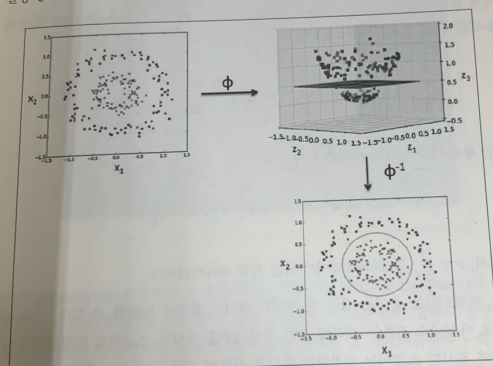
kernel trick. (non-linear boundary).

idea:

고차원의 공간에 투영하는 것이다. 다음의 그림에서 볼 수 있는 것처럼, 2차원 데이터를 투영을 통해 분류들이 분리 가능하게 되는 3차원의 피처 공간으로 변환할 수 있다.

$$\phi(x_1, x_2) = (z_1, z_2, z_3) = (x_1, x_2, x_1^2 + x_2^2)$$

이것은 선형 하이퍼 평면을 사용한 플롯 내에 보이는 두 개의 분류를 분리할 수 있게 해준다. 여기서 선형 하이퍼 평면은 원래의 피처 공간으로 되돌려서 투영하면 비선형 결정 경계가 된다.



Soft margin SVM이랑 똑같은 상황!

$(\bar{x}_i \to \Phi(\bar{x}_i))$.

we want to minimize $\frac{1}{2}\|W\|^2 + C \cdot \sum \xi_i$

s.t $y_i(\bar{w} \cdot \Phi(\bar{x}_i) + b) \geq 1 - \xi_i$, $\xi_i \geq 0$ for all i.

$\Phi : \mathbb{R}^d \longrightarrow \mathbb{R}^D$ $(d < D)$

$\bar{x}_i \longmapsto \Phi(\bar{x}_i)$

$$\min \quad L_p = \frac{1}{2}\|W\|^2 + C\sum \xi_i - \sum \alpha_i \left[y_i (\bar{w} \cdot \Phi(\bar{x}_i) + b) - 1 + \xi_i \right] - \sum \mu_i \xi_i$$

$$\text{s.t} \quad \alpha_i \geq 0, \ \mu_i \geq 0.$$

$$\Downarrow$$

define

$$K(\bar{x}_i, \bar{x}_j).$$

$$\max \quad L_D = \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \boxed{\Phi(\bar{x}_i) \cdot \Phi(\bar{x}_j)}.$$

$$\text{s.t} \quad \sum \alpha_i y_i = 0, \quad 0 \leq \alpha_i < C.$$

Mercer's Theorem.

$$\begin{cases} K(x, x') = K(x', x) \\ K(x, x) \geq 0 \end{cases} \longrightarrow K(x, x') = \Phi(x) \cdot \Phi(x')$$

만족하는 Φ 존재.

(K is valid kernel).

Ex. Let $K(x, x') = (1 + x^T x')^2$. $(x, x' \in \mathbb{R}^2)$.

$$(1 + x^T x')^2 = (1 + x_1 x_1' + x_2 x_2')^2$$
$$= 1 + x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2x_1 x_1' + 2x_2 x_2' + 2x_1 x_1' x_2 x_2'$$

This is inner product of

$$\Phi(x) = (1, x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2).$$

$$\Phi(x') = (1, x_1'^2, x_2'^2, \sqrt{2}x_1', \sqrt{2}x_2', \sqrt{2}x_1' x_2').$$

$$\Phi : \mathbb{R}^2 \longrightarrow \mathbb{R}^6.$$
$$(x_1, x_2) \longmapsto (1, x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2).$$

- linear kernel $K(x,x') = x^T x'$.

- Polynomial kernel $K(x,x') = (1 + x^T x')^Q$

- Gaussian RBF $K(x,x') = e^{-r\|x-x'\|^2}$. $\left(e^{-\frac{\|x-x'\|^2}{\sigma^2}} \rightarrow r = \frac{1}{\sigma^2} \right)$

let $r=1$, $K(x,x') = e^{-x^2} \cdot e^{-x'^2} \cdot e^{2x^T x'}$

$$= \sum_{k=0}^{\infty} \frac{2^k (x)^k \cdot (x')^k}{k!}.$$

$\left(e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \cdots \right)$

$\Rightarrow$ infinite-dimension expansion.

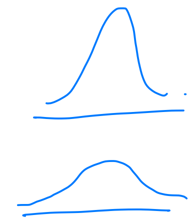
hyper-parameter optimize (RBF)

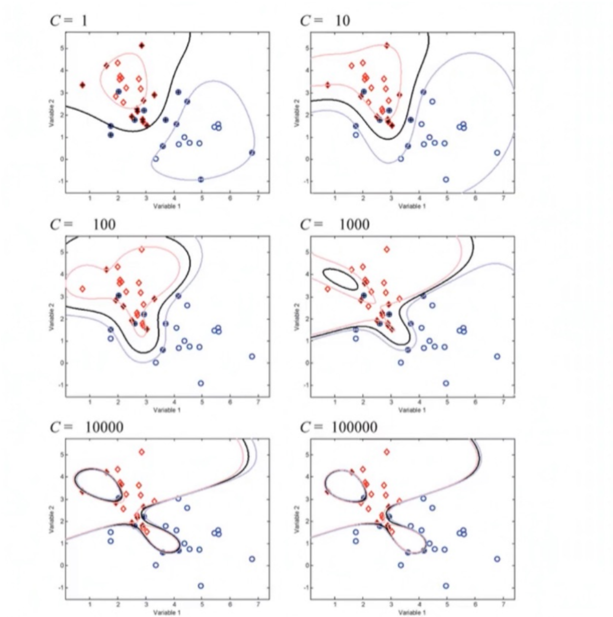
C, γ.

$\gamma = \frac{1}{\sigma^2}$

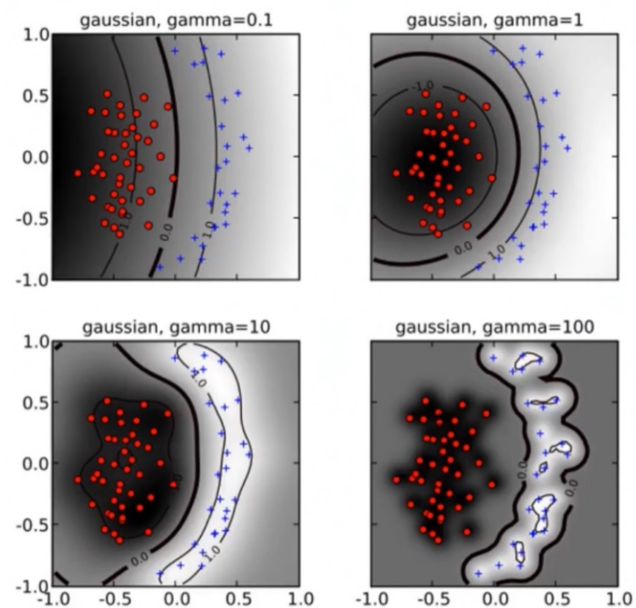
$\gamma \uparrow \rightarrow \sigma \downarrow$

$\gamma \downarrow \rightarrow \sigma \uparrow$





http://pubs.rsc.org/-/content/articlehtml/2010/an/b918972f



Q & A.

유튜브 강의 & 사진 참고

Pilsung Kang
School of Industrial Management Engineering
Korea University