

Statistical Machine Learning

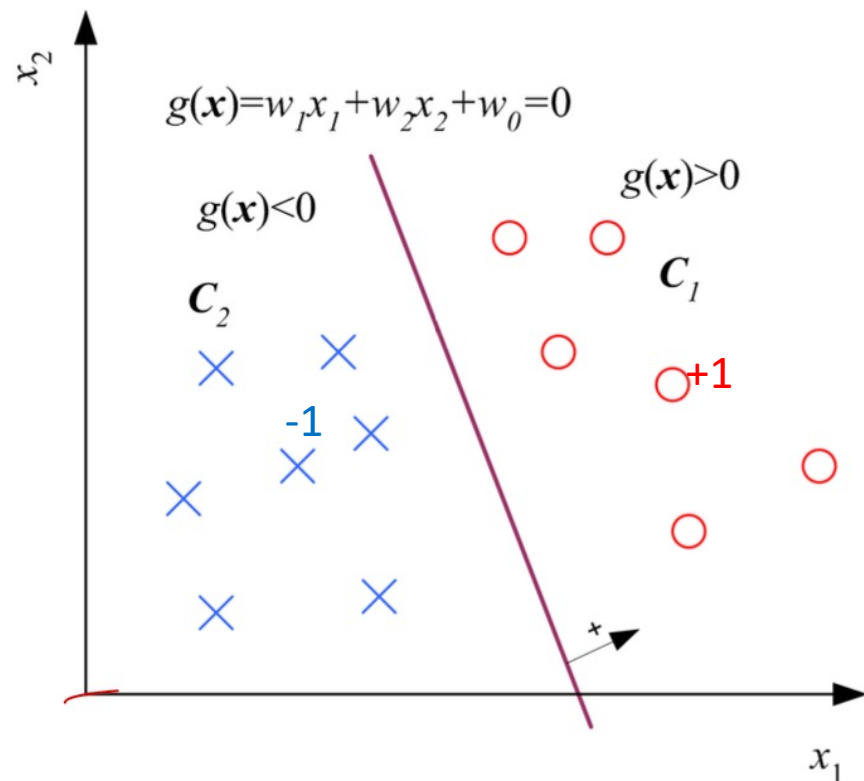
1주차

담당: 15기 염윤석

1. Linear SVM
2. Kernel SVM
3. SVM-Regression

1. Linear SVM - Classification

Linear Discriminant



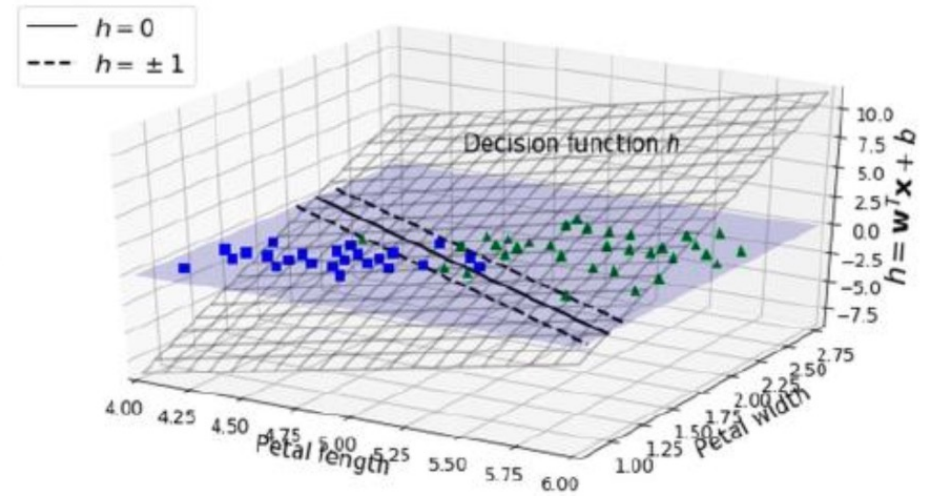
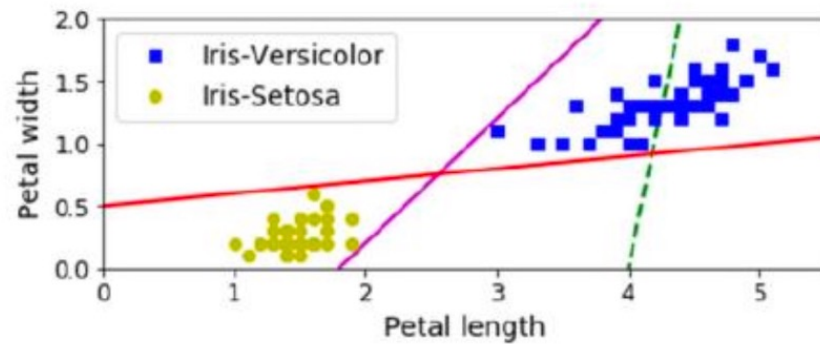
Decision Boundary or separating hyperplane

Decision Boundary : $g(x) = w^T x + w_0 = 0$

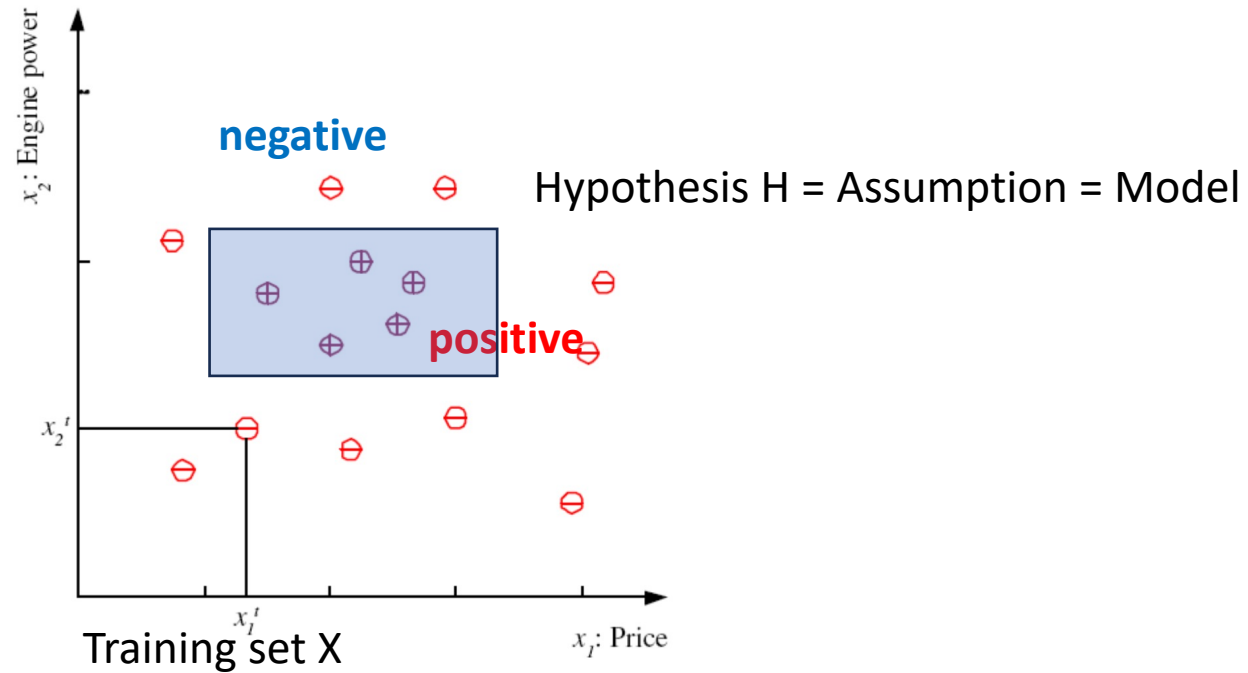
$$X = \{x^t, r^t\} \mid r^t = \begin{cases} +1 \\ -1 \end{cases}$$

$$w^T x + w_0 \geq +1, \text{ for } r^t = +1$$
$$w^T x + w_0 \leq -1, \text{ for } r^t = -1$$

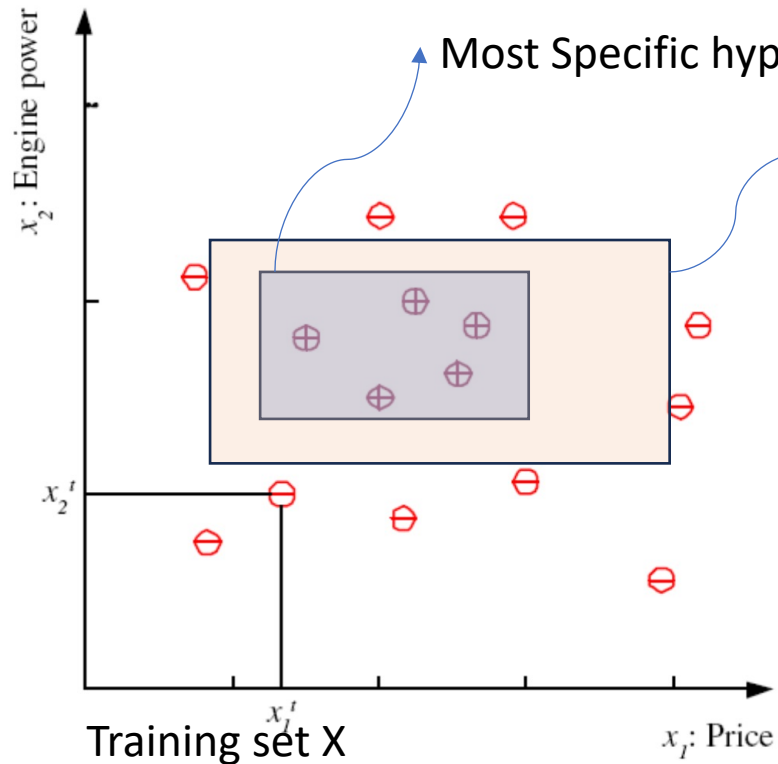
Hyperplane



S, G and the Version Space



S, G and the Version Space

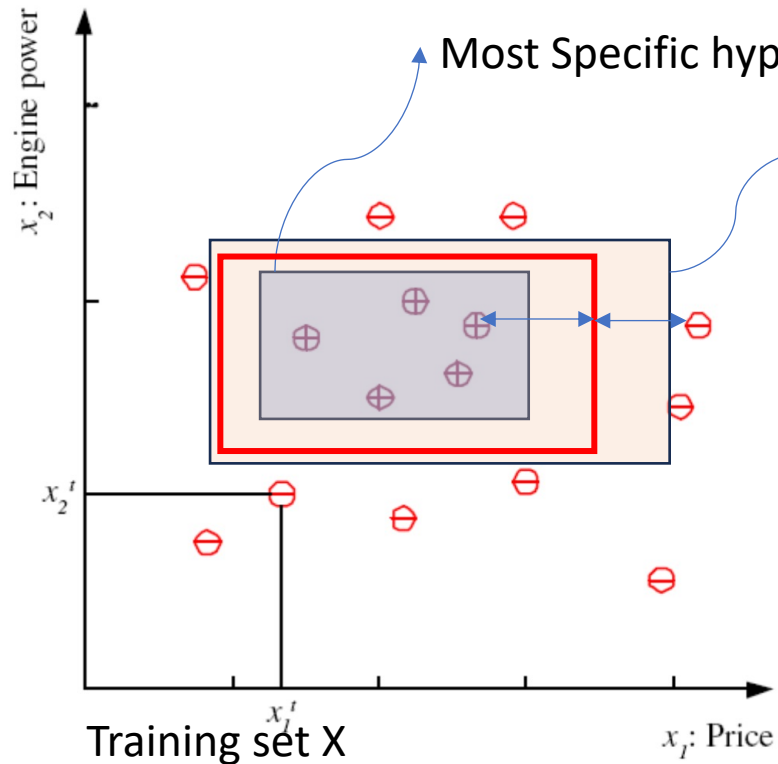


Most General hypothesis, G

“Any hypothesis h in H , between S & G is consistent and make up the Version space”

Q. But Which one is optimal?

Margin



Margin

: distance between hypothesis and the closest positive and negative instances

→ **Maximize!**

S : False negative에 취약

G : False positive에 취약

Optimal Hyperplane

- Decision Boundary : $g(x) = w^T x + w_0 = 0$

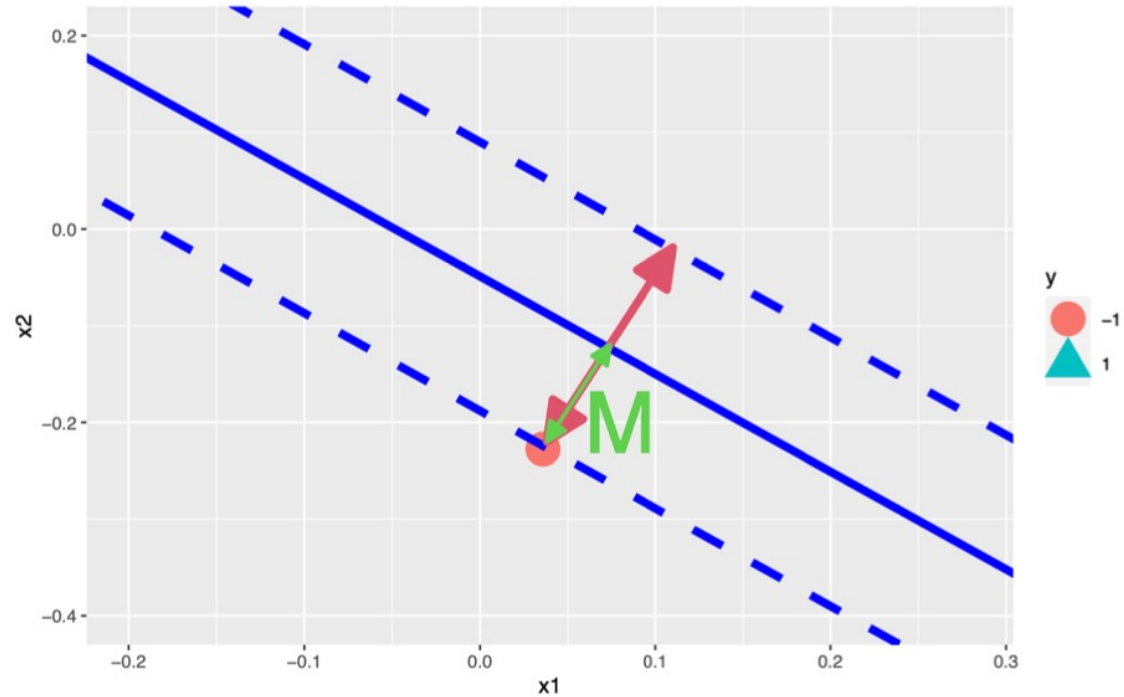
- $X = \{x^t, r^t\} \mid r^t = \begin{cases} +1 \\ -1 \end{cases}$

$\rightarrow r^t(w^T x + w_0) \geq +1$

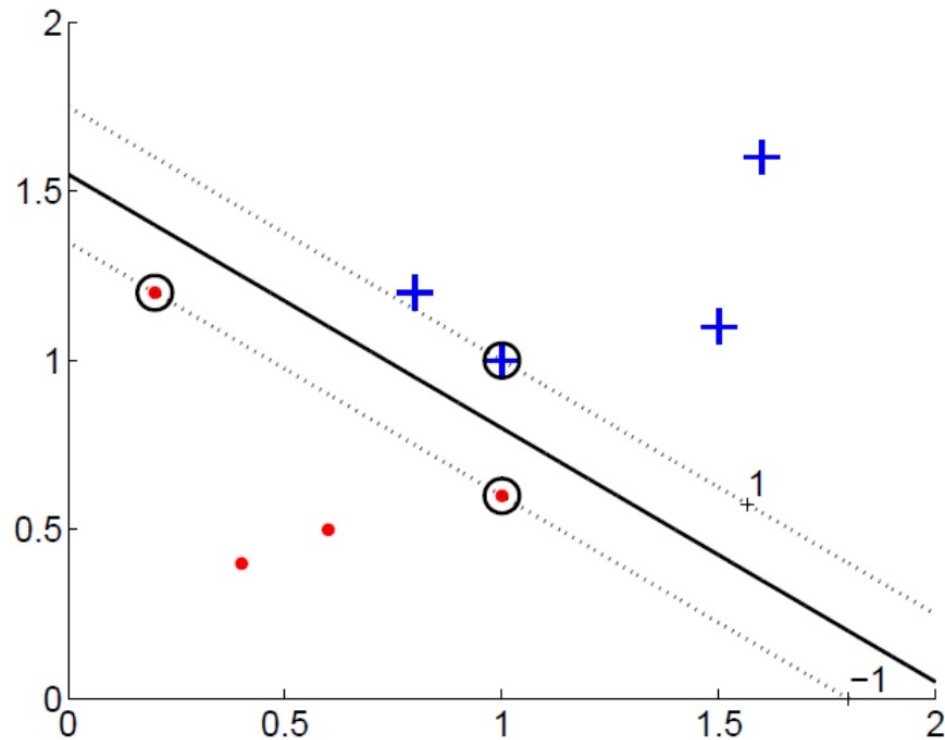
[Margin]

Discriminant부터 양쪽 가장 가까운 instance 까지의 거리

Optimal Hyperplane(Discriminant) maximizes **Margin**



Objective of SVM



- Distance x to the hyperplane $g(x)$

- Margin

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \text{ subject to } r^t (\mathbf{w}^T \mathbf{x}^t + w_0) \geq +1, \forall t$$

Lagrangian multiplier Method

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \text{ subject to } r^t(\mathbf{w}^T \mathbf{x}^t + w_0) \geq +1, \forall t$$

KKT(Karush-Kuhn-Tucker Theorem)

1. Stationarity

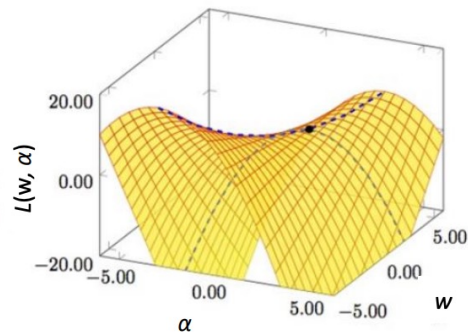
Primal problem

$$\begin{aligned} L_p &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{t=1}^N \alpha^t [r^t(\mathbf{w}^T \mathbf{x}^t + w_0) - 1] \\ &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{t=1}^N \alpha^t r^t(\mathbf{w}^T \mathbf{x}^t + w_0) + \sum_{t=1}^N \alpha^t \end{aligned}$$

2. Primal feasibility

3. Dual feasibility

4. Complementary slackness



Dual problem of SVM

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \text{ subject to } r^t (\mathbf{w}^T \mathbf{x}^t + w_0) \geq +1, \forall t$$

Dual problem

$$\begin{aligned} L_d &= \frac{1}{2} (\mathbf{w}^T \mathbf{w}) - \mathbf{w}^T \sum_t \alpha^t r^t \mathbf{x}^t - w_0 \sum_t \alpha^t r^t + \sum_t \alpha^t \\ &= -\frac{1}{2} (\mathbf{w}^T \mathbf{w}) + \sum_t \alpha^t \\ &= -\frac{1}{2} \sum_t \sum_s \alpha^t \alpha^s r^t r^s (\mathbf{x}^t)^T \mathbf{x}^s + \sum_t \alpha^t \\ \text{subject to } \sum_t \alpha^t r^t &= 0 \text{ and } \alpha^t \geq 0, \forall t \end{aligned}$$

KKT(Karush-Kuhn-Tucker Theorem)

1. Stationarity
2. Primal feasibility
3. Dual feasibility
4. Complementary slackness

Solution of SVM

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \text{ subject to } r^t (\mathbf{w}^T \mathbf{x}^t + w_0) \geq +1, \forall t$$

We want optimal hyperplane $g(x) = w^T x + w_0$

We want optimal w^* & w_0^*

$$w = \sum_t \alpha^t r^t x^t \qquad w_0 = \frac{1}{N} \sum_t r^t - w^T x^t$$

$$g(x) = w_0 + \sum_t \alpha^t r^t x_t^T x$$

Solution of SVM

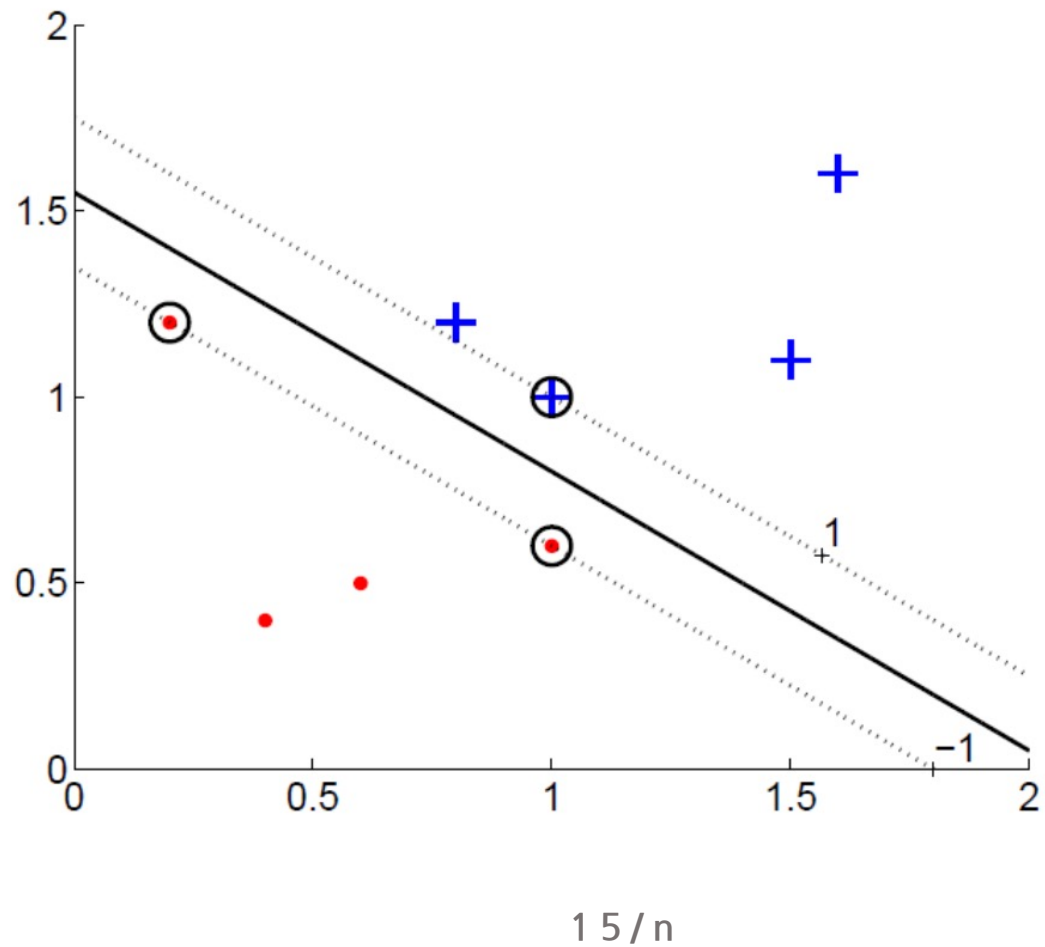
We want optimal hyperplane $g(x) = w^T x + w_0$

$$w = \sum_t \alpha^t r^t x^t \quad w_0 = \frac{1}{N} \sum_t r^t - w^T x^t$$

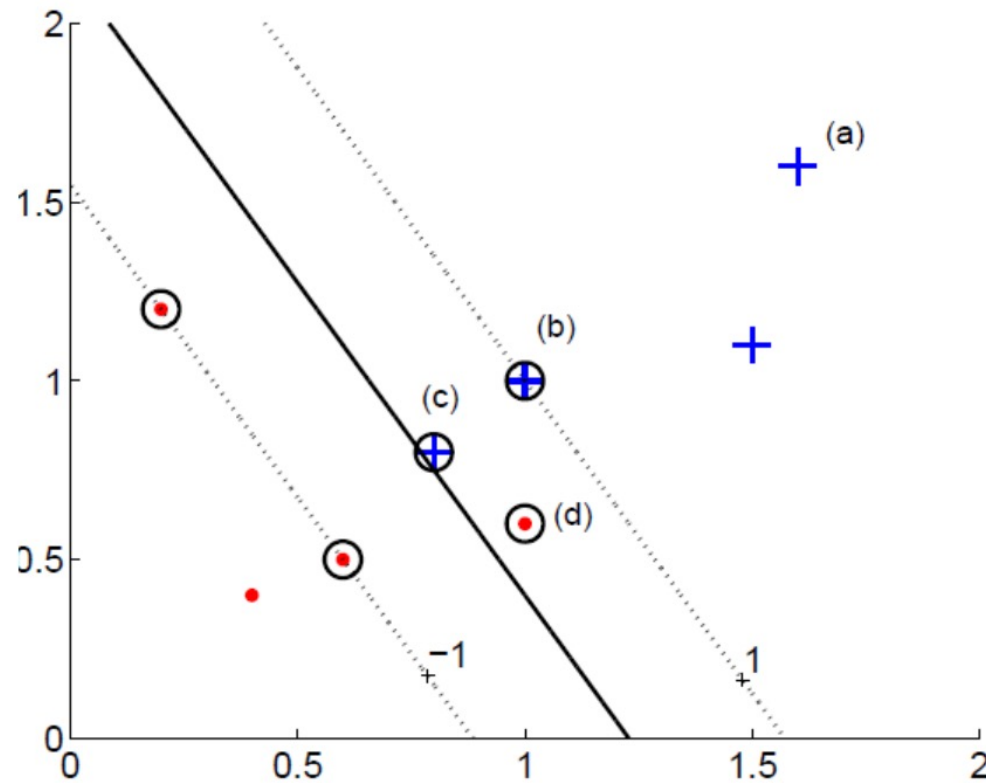
$$g(x) = w_0 + \sum_t \alpha^t r^t x_t^T x_t$$

“Most $\alpha^t = 0$, only a small number have $\alpha^t > 0$ ” : **support vector**

SVM - Classification



What if Non-Separable?



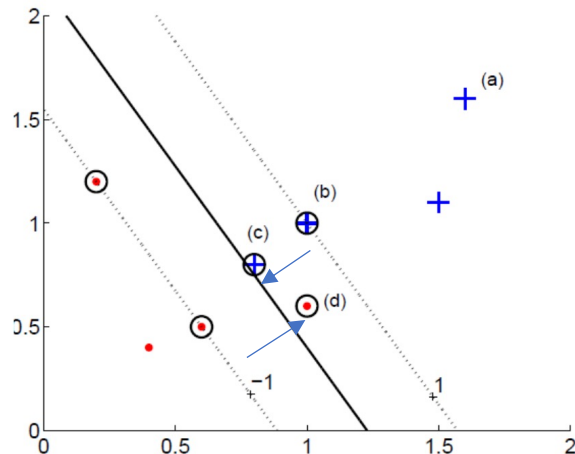
Soft Margin Hyperplane

$$r^t(w^T x + w_0) \geq 1 - \xi^t$$

Slack variable

- $\text{soft error} = \sum_t \xi^t$

$$\min \frac{1}{2} \|w\|^2 + C \sum_t \xi^t \text{ subject to } r^t(w^T x + w_0) \geq 1 - \xi^t, \xi^t \geq 0$$



- New primal problem

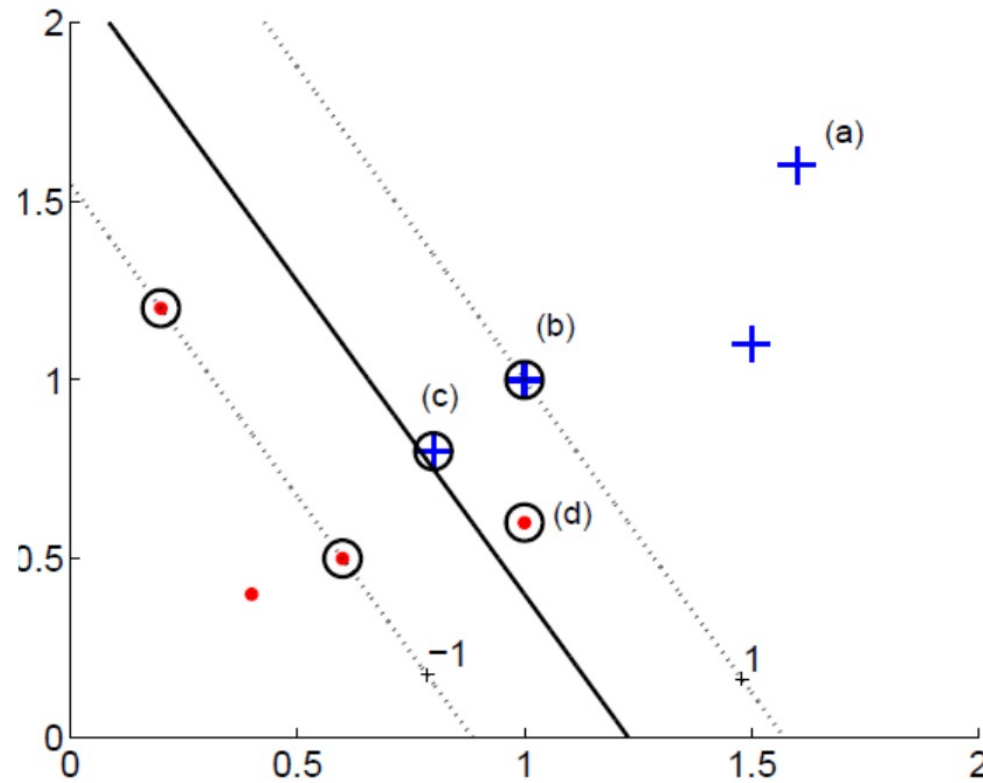
$$L_p = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_t \xi^t - \sum_t \alpha^t [r^t(\mathbf{w}^T \mathbf{x}^t + w_0) - 1 + \xi^t] - \sum_t \mu^t \xi^t$$

- New Dual problem

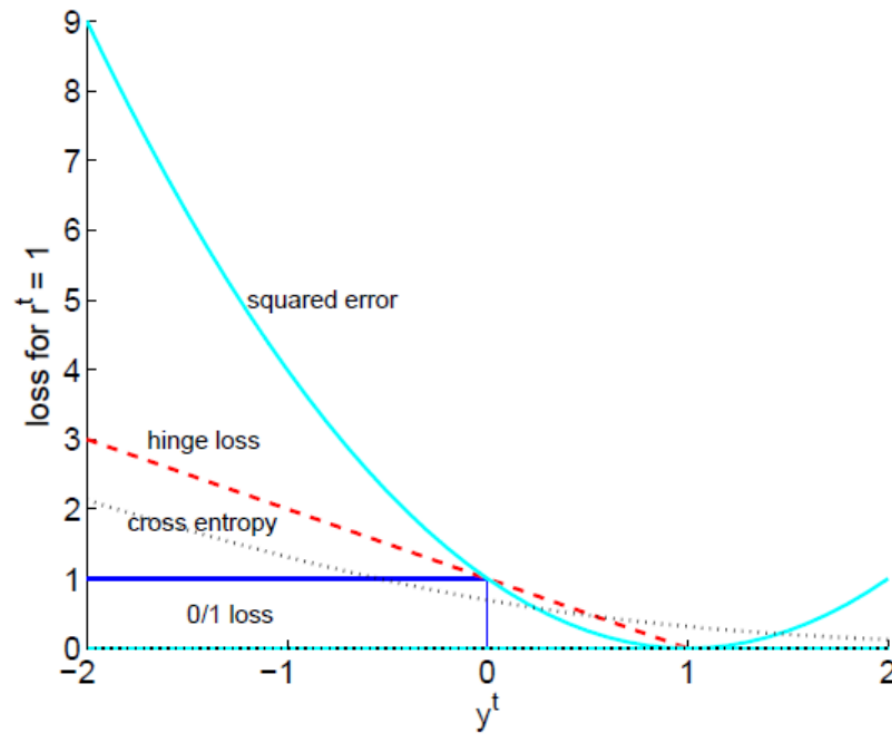
$$L_d(\alpha) = \sum_t \alpha^t - \frac{1}{2} \sum_t \sum_s \alpha^t \alpha^s r^t r^s x_t^T x^s$$

$$\text{subject to } 0 \leq \alpha^t \leq C, \sum_t \alpha^t r^t = 0$$

Soft Margin Hyperplane



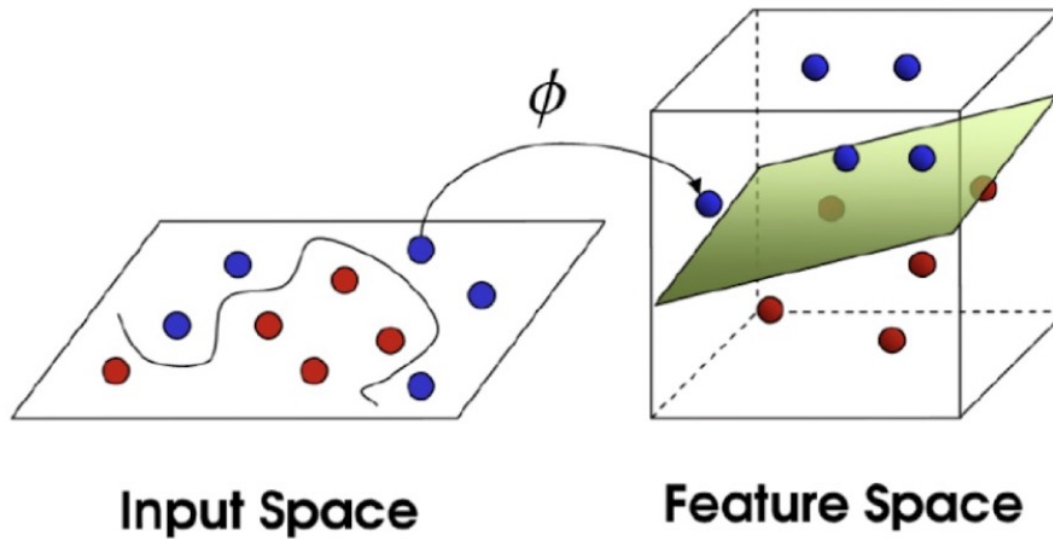
Hinge Loss



$$: \begin{cases} 0 & \text{if } y^t r^t \geq 1 \\ 1 - y^t r^t & \text{otherwise} \end{cases}$$

2. Kernel SVM

Extension to non-linearity



$$x = \{x_1, x_2\} \rightarrow z = \{1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2\}$$

$$z = \phi(x)$$

Feature mapping

$$\phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_n(\mathbf{x}))$$

Kernel Trick

$$z = \{1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2\} = [z_1 \ z_2 \ z_3 \ z_4 \ z_5 \ z_6]$$

$$g(z) = w^T z + w_0$$

$$z = \varphi(x)$$

$$g(x) = w^T \varphi(x) + w_0$$

In linear SVM...

New feature space

$$g(x) = w_0 + \sum_t \alpha^t r^t x_t^T x \quad \rightarrow \quad g(z) = w_0 + \sum_t \alpha^t r^t z_t^T z$$

$$g(x) = w_0 + \sum_t \alpha^t r^t \varphi(x^t)^T \varphi(x)$$

Using Kernel Trick : $K(x^t, x)$

Kernel Trick

$$f(\mathbf{x}) = \beta_0 + \sum_{i=1}^n y_i \alpha_i K(\mathbf{x}_i, \mathbf{x})$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$$

Linear Kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j))$$

*Gaussian Kernel
(Radial Basis function)*

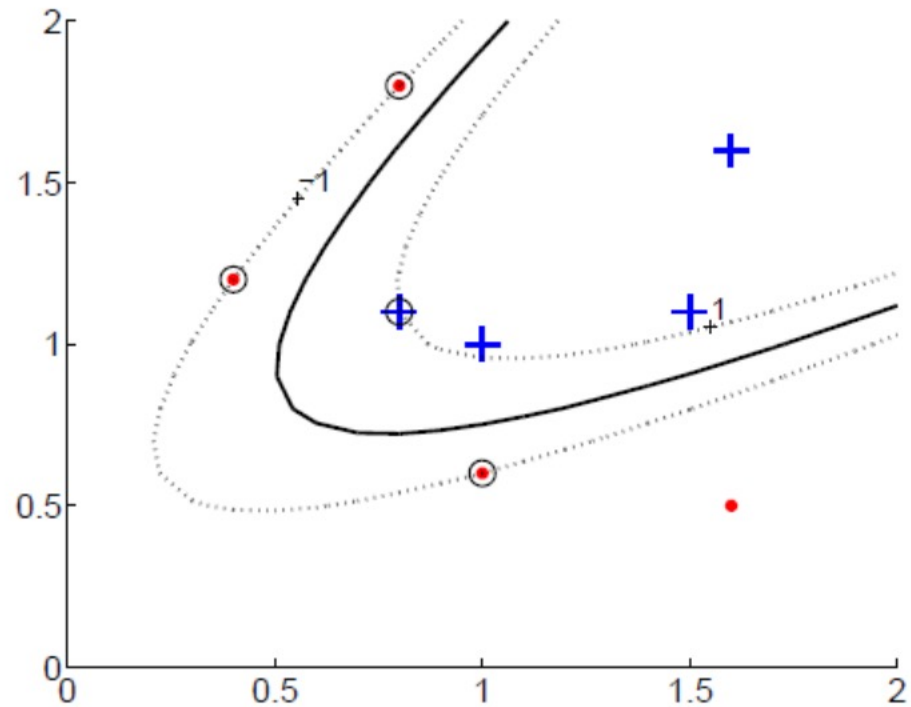
$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma + \gamma \mathbf{x}_i^T \mathbf{x}_j)^p$$

polynomial Kernel

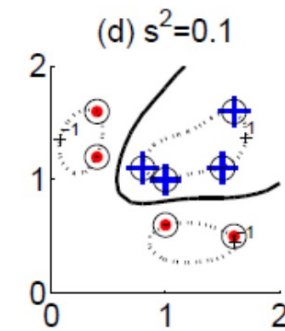
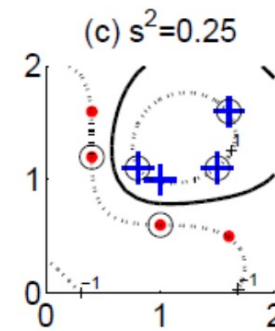
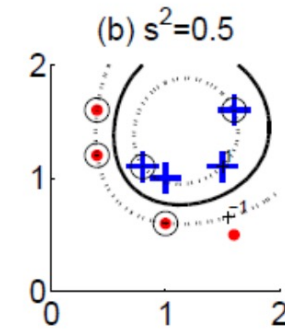
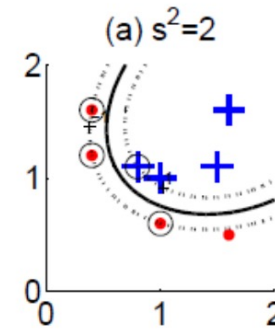
$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(k_1 \mathbf{x}_i^T \mathbf{x}_j + k_2)$$

Sigmoid Kernel

Kernel SVM



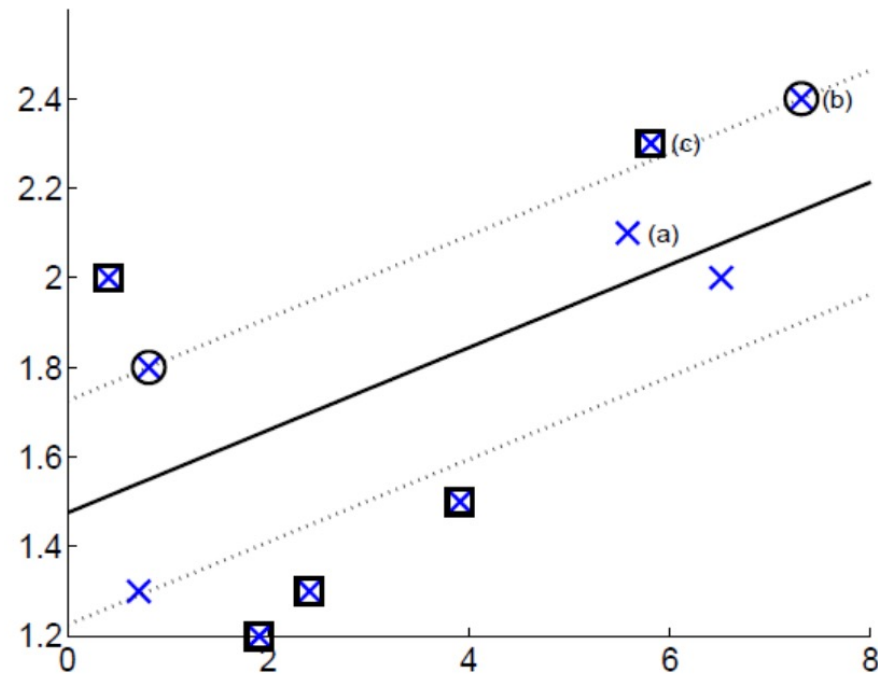
Polynomial Kernel



Gaussian(Radial-Basis function) Kernel

3. SVM - Regression

SVM- Regression



SVM- Regression

Let Assume linear model

$$f(x) = w^T x + w_0$$

- Error function(loss)

$$e = \begin{cases} 0 & \text{if } |r^t - f(x^t)| < \varepsilon \\ |r^t - f(x^t)| - \varepsilon & \end{cases}$$

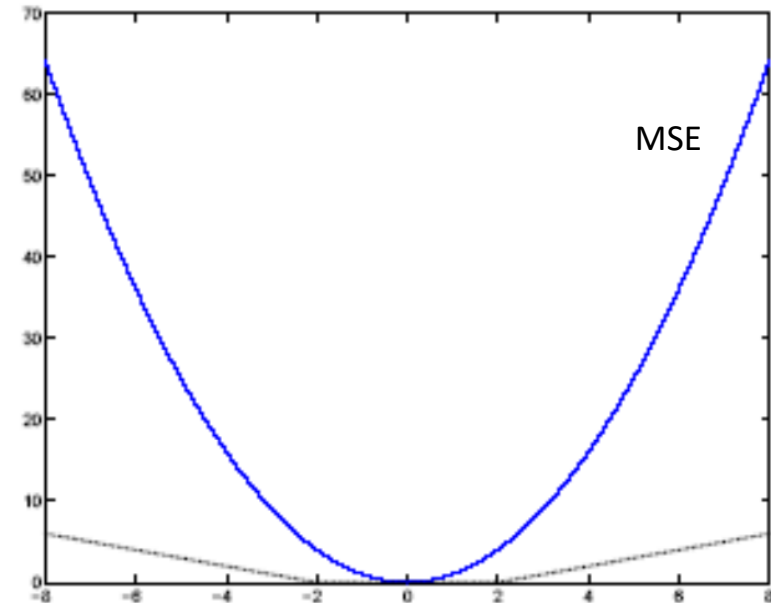
최대한 Margin 내로 들어오도록 학습 \rightarrow Margin 밖에 있는 Error를 최소화

Lagragian Method $\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_t (\xi_+^t + \xi_-^t)$

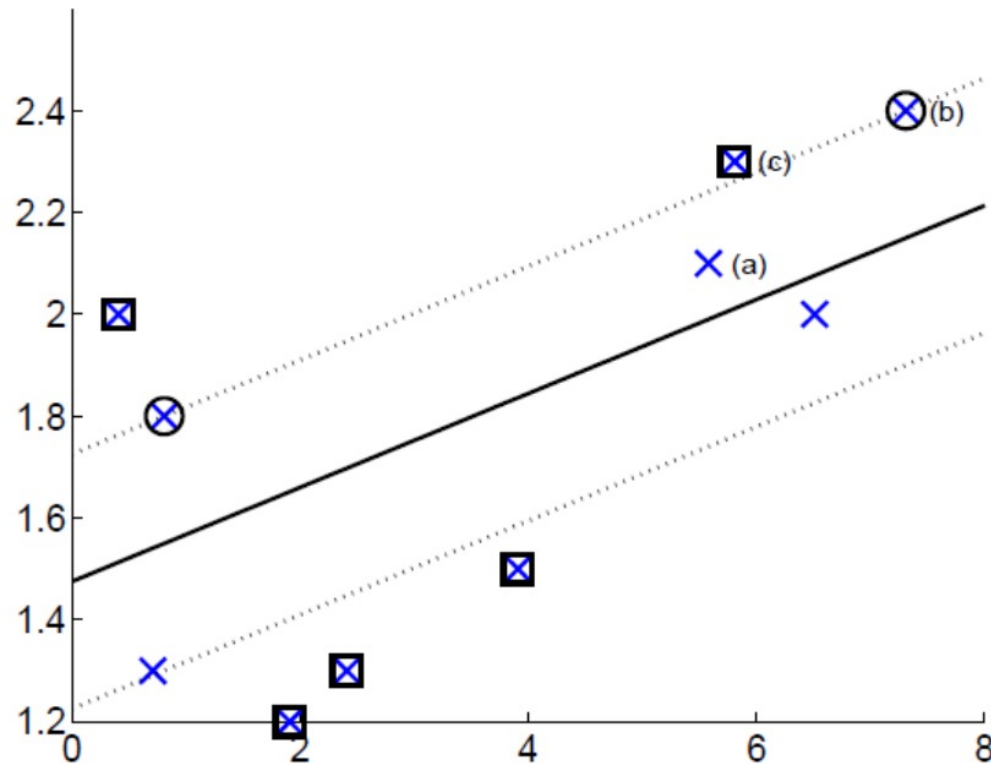
$$r^t - (\mathbf{w}^T \mathbf{x} + w_0) \leq \varepsilon + \xi_+^t$$

$$(\mathbf{w}^T \mathbf{x} + w_0) - r^t \leq \varepsilon + \xi_-^t$$

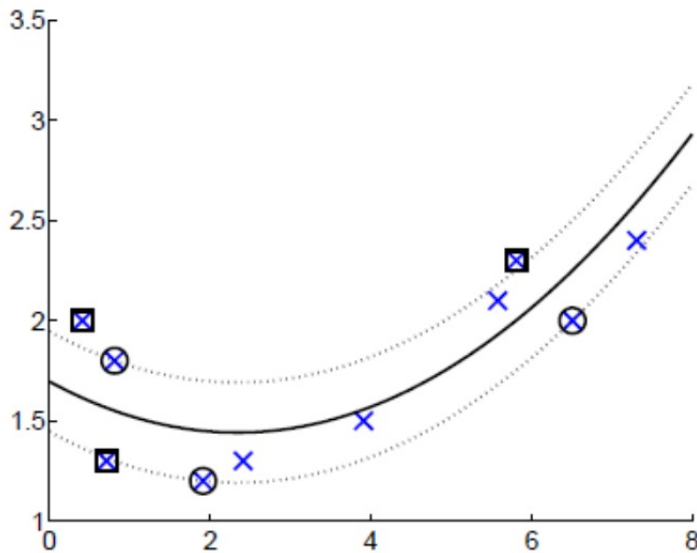
$$\xi_+^t, \xi_-^t \geq 0$$



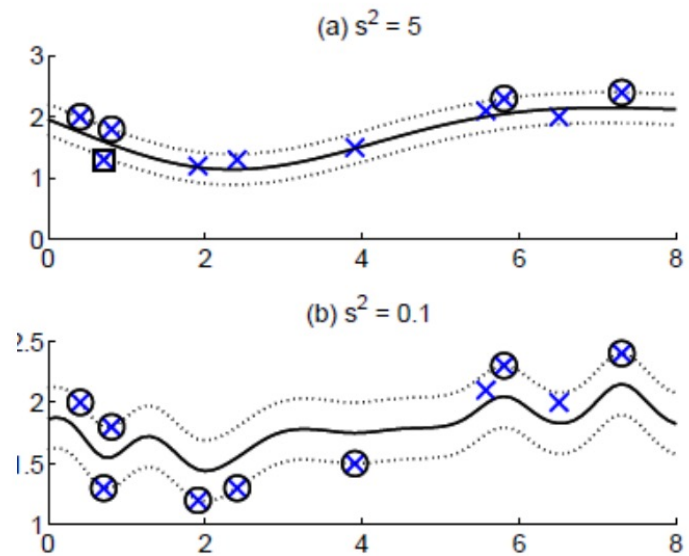
SVM- Regression



SVM Kernel Regression

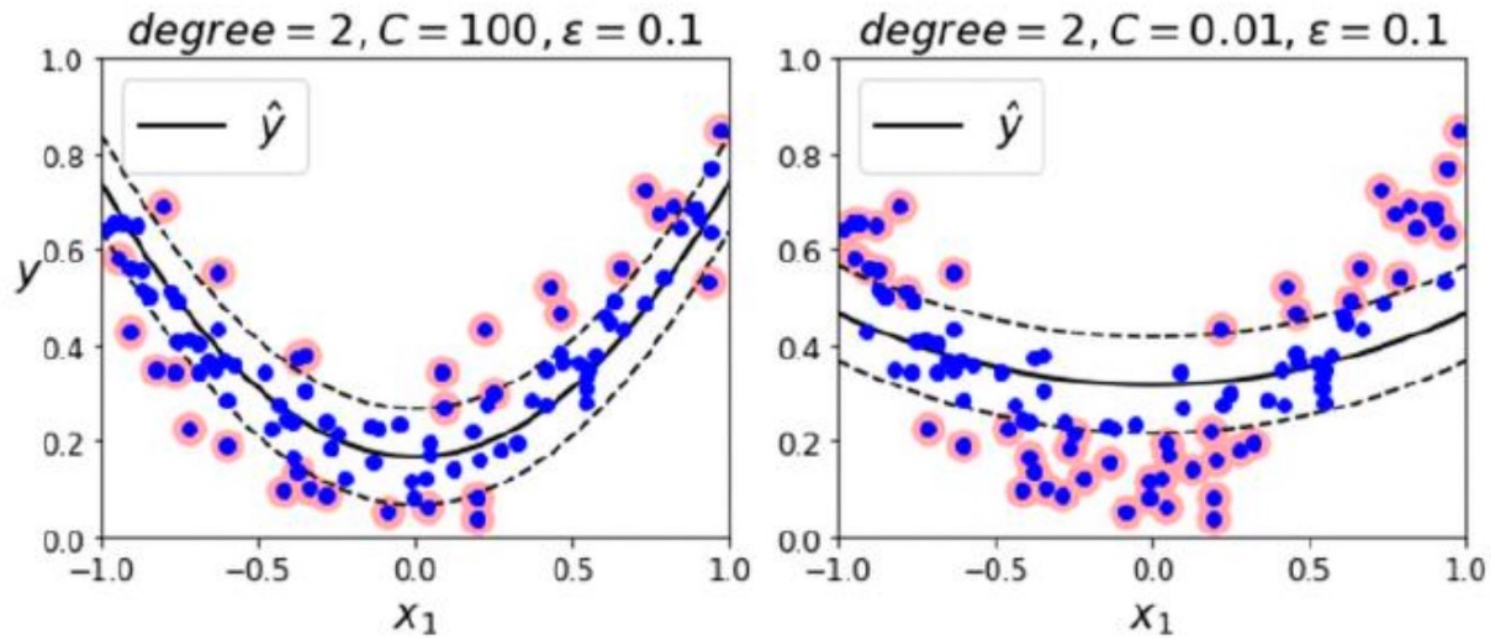


Polynomial Kernel



Gaussian Kernel

SVM- Regression



[6주차 프로젝트 발표 공지사항]

- 발표 형식 : ppt 제작 및 발표
 - 발표 ppt pdf 제출 : KUBIG github > 1. 방학분반 > 머신러닝 > 3. 프로젝트 > 팀명(팀원들 이름)
- 발표 시간 : 6주차 세션 후, 진행
 - 팀당 10분 내외로 준비
 - 발표 후, 질의응답



수고하셨습니다!

해당 세션자료는 KUBIG Github에서 보실 수 있습니다!

