Support Vector Machine

5주차
담당: 14기 박상준

1. Lagrange Multiplier Theorem
2. Linear Support Vector Machine
3. Kernel Support Vector Machine
4. Logistic Regression vs SVM

Lagrange Multiplier Theorem

- Primal Problem

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $\quad g_i(\mathbf{x}) \leq 0, \quad$ for $\quad i = 1, \cdots, m$

$$h_j(\mathbf{x}) = 0, \quad \text{for} \quad j = 1, \cdots, \mathrm{k}$$

Lagrange Multiplier Theorem

- Dual Problem

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) + \sum_{i}^{m} \alpha_i \, g_i(\mathbf{x}) + \sum_{j}^{\mathrm{k}} \gamma_j \, h_i(\mathbf{x})$$

$$\alpha_i \geq 0, \quad \text{for} \quad i = 1, \cdots, m$$

$$\gamma_j \geq 0, \quad \text{for} \quad j = 1, \cdots, \mathrm{k}$$

Karush-Kuhn-Tucker Conditions

1. $\nabla f(\mathbf{x}) + \sum_i^m \alpha_i \nabla g_i(\mathbf{x}) + \sum_j^{\mathrm{k}} \gamma_j \nabla h_i(\mathbf{x}) = 0$ (Stationary)

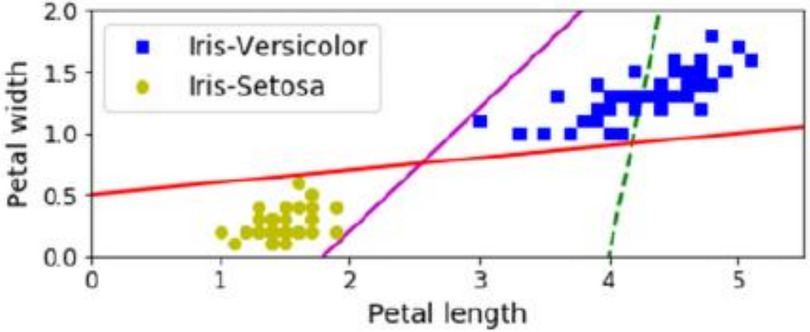
2. $\alpha_i g_i(\mathbf{x}) = 0, \quad$ for $\quad i = 1, \cdots, m$ (Complementary Slackness)

3. $g_i(\mathbf{x}) \leq 0, \quad$ for $\quad i = 1, \cdots, m \quad$ *and* (Primal Feasibility)

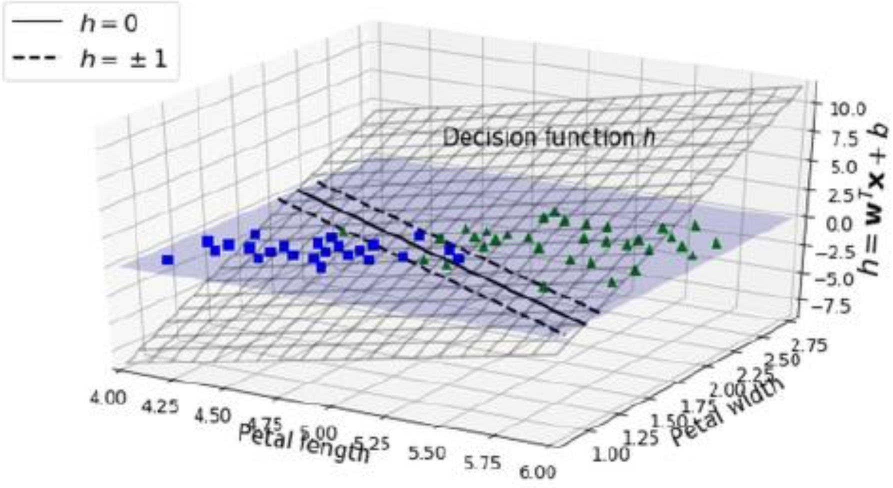
   $h_j(\mathbf{x}) = 0, \quad$ for $\quad j = 1, \cdots, \mathrm{k}$

4. $\alpha_i \geq 0, \quad$ for $\quad i = 1, \cdots, m$ (Dual Feasibility)

Hyperplane

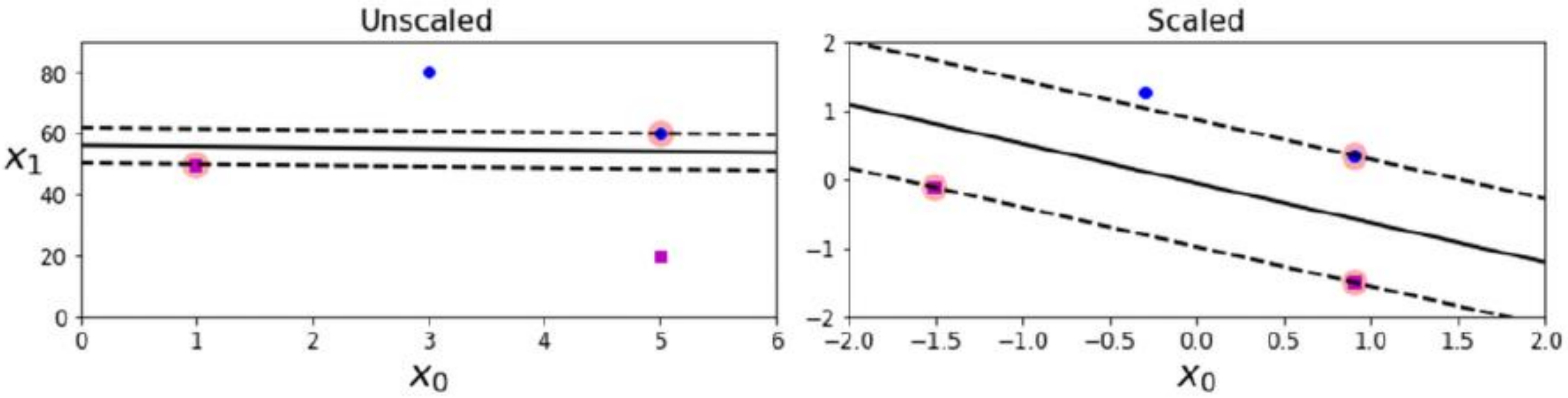


Iris-Versicolor
Iris-Setosa
Petal width
Petal length

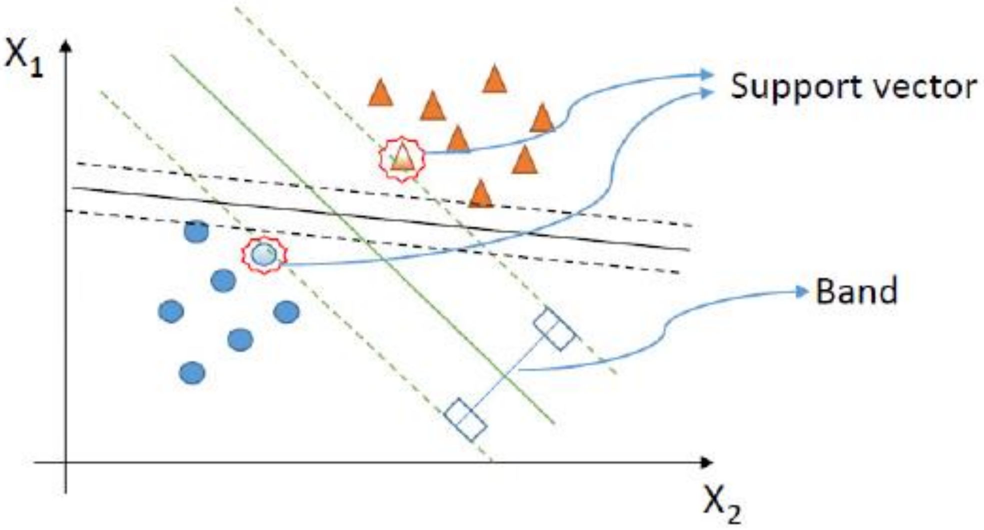


$h = 0$
$h = \pm 1$
Decision function h
$h = \mathbf{w}^T\mathbf{x} + b$
Petal length
Petal width

Scaled? Unscaled?

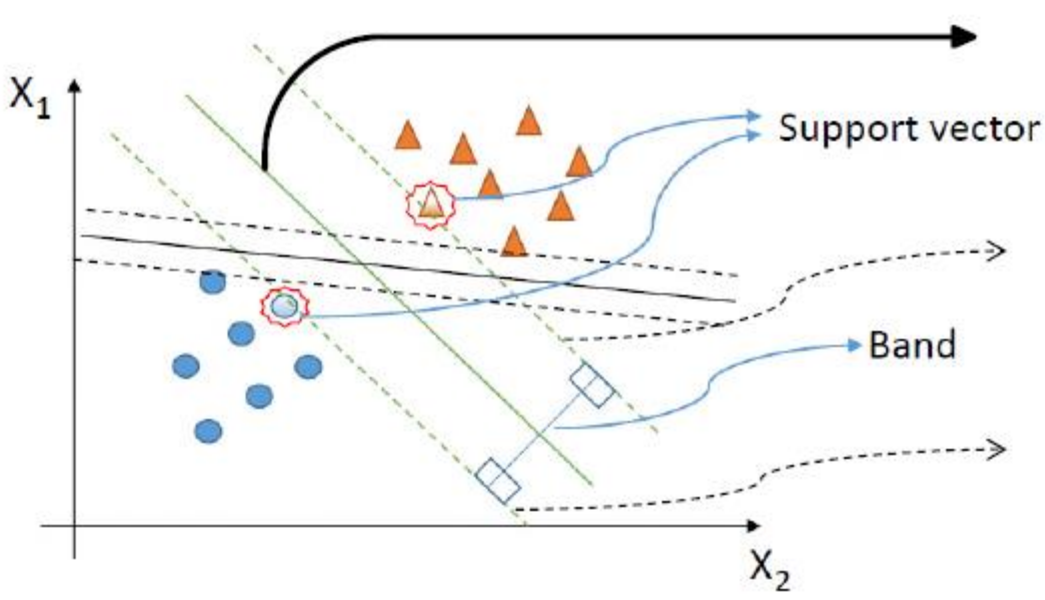


Linear Support Vector Machine



$$y = \{-1, 1\}$$

Linear Support Vector Machine

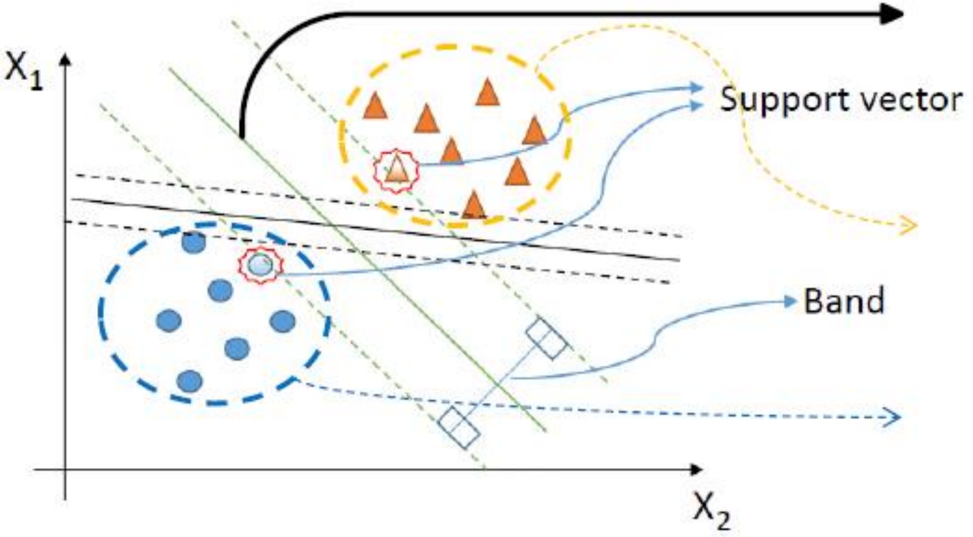


$$f(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i = 0$$

$$\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_+ = M$$

$$\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_- = -M$$

Linear Support Vector Machine



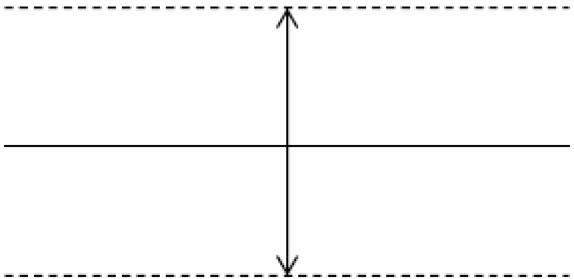
$$f(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i = 0$$

$$\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i \geq M \qquad \text{if } \; y_i = 1$$

$$\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i \leq -M \qquad \text{if } \; y_i = -1$$

Linear Support Vector Machine

- We want to maximize the width of the band.



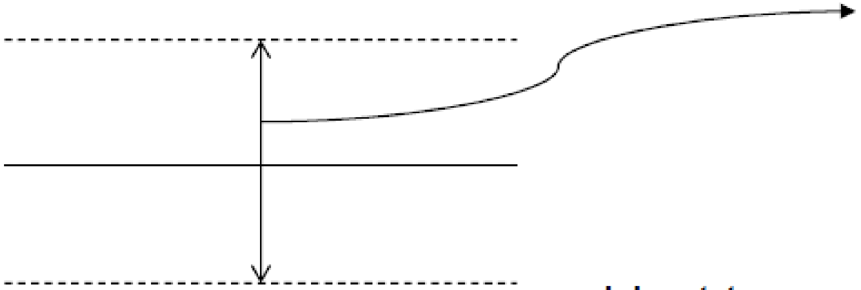
$$\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_+ = M$$

$$\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i = 0$$

$$\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_- = -M$$

Linear Support Vector Machine

- We want to maximize the width of the band.



$$\boldsymbol{\beta}^T(\mathbf{X}_+ - \mathbf{X}_-) = 2M$$

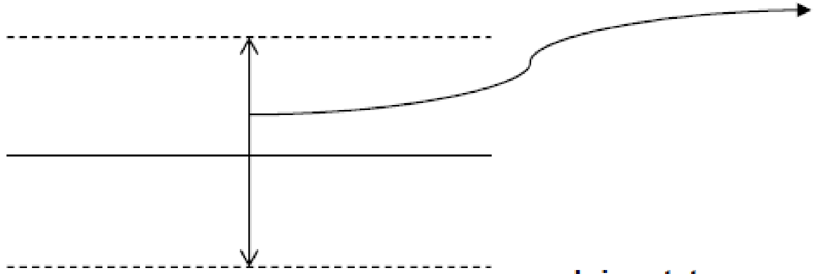
$$\max_{\beta_0,\, \boldsymbol{\beta}} M$$

subject to $y_i(\beta_0 + \boldsymbol{\beta}^T\mathbf{X}_i) \geq \mathrm{M}$, for $i = 1, \cdots, \mathrm{n}$

$$||\boldsymbol{\beta}|| = 1$$

Linear Support Vector Machine

- We want to maximize the width of the band.



$$\frac{\boldsymbol{\beta}^T}{||\boldsymbol{\beta}||}(\mathbf{X}_+ - \mathbf{X}_-) = \frac{2M}{||\boldsymbol{\beta}||}$$

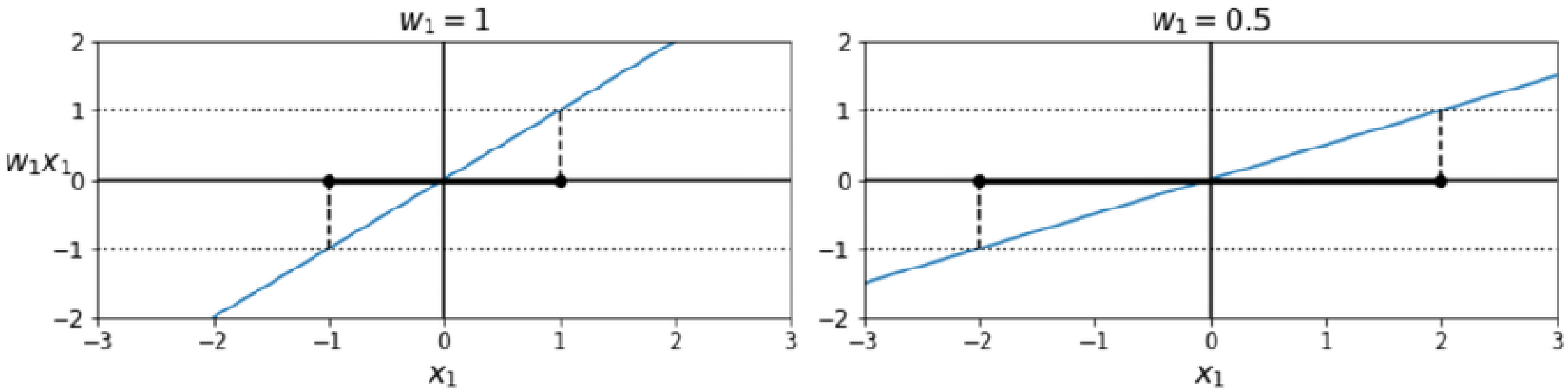
$$\max_{\beta_0, \boldsymbol{\beta}} M$$

subject to $y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) \geq \mathrm{M}$, for $i = 1, \cdots, \mathrm{n}$

$$||\boldsymbol{\beta}|| = 1$$

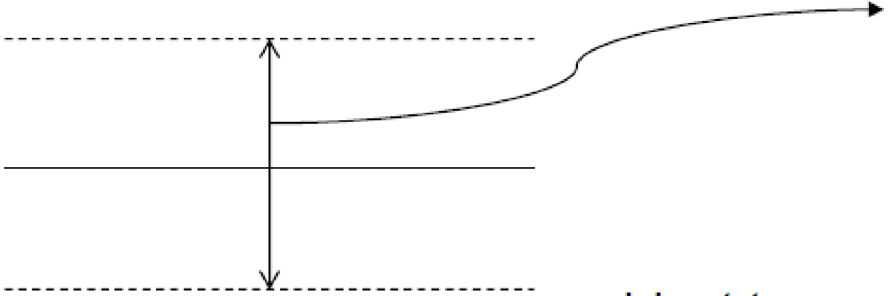
Linear Support Vector Machine

- A smaller weight vector results in a larger margin



Linear Support Vector Machine

- We want to maximize the width of the band.



$$\frac{\boldsymbol{\beta}^T}{||\boldsymbol{\beta}||}(\mathbf{X}_+ - \mathbf{X}_-) = \frac{2M}{||\boldsymbol{\beta}||}$$

$$\min_{\beta_0,\, \boldsymbol{\beta}} ||\boldsymbol{\beta}||$$

subject to $\quad y_i(\beta_0 + \boldsymbol{\beta}^T\mathbf{X}_i) \geq \mathrm{M}, \text{ for } \; i = 1, \cdots, \mathrm{n}$

$$\mathrm{M} = ?$$

Linear Support Vector Machine

- We want to maximize the width of the band.

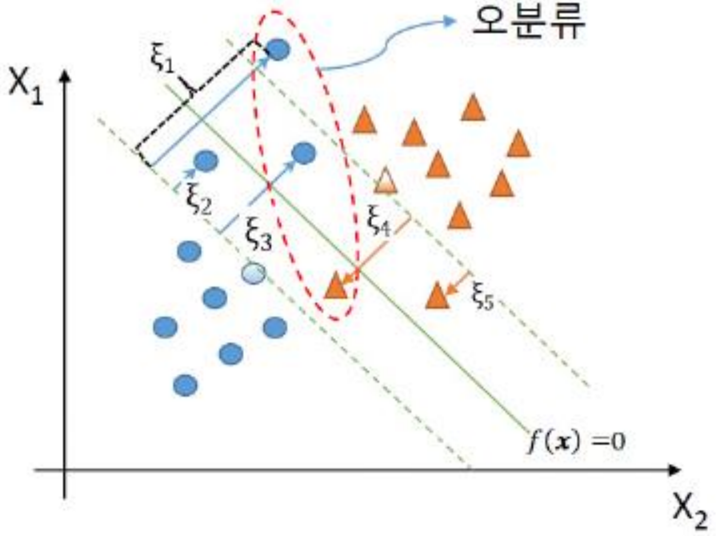
$$\max_{\beta_0, \boldsymbol{\beta}} M \quad \Longleftrightarrow \quad \min_{\beta_0, \boldsymbol{\beta}} ||\boldsymbol{\beta}||^2$$

subject to $y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) \geq \mathrm{M}$, for $i = 1, \cdots, \mathrm{n}$ subject to $y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) \geq 1$, for $i = 1, \cdots, \mathrm{n}$

and $||\boldsymbol{\beta}|| = 1$

Linear Support Vector Machine

- If the data are not perfectly separable, no solution exists.



Linear Support Vector Machine

- Hard Margin Classifier

$$\min_{\beta_0, \boldsymbol{\beta}} ||\boldsymbol{\beta}||^2$$

subject to $y_i(\beta_0 + \boldsymbol{\beta}^T\mathbf{X}_i) \geq 1$, for $i = 1, \cdots, \mathrm{n}$

$\Longrightarrow$

- Soft Margin Classifier

$$\min_{\beta_0, \boldsymbol{\beta}} ||\boldsymbol{\beta}||^2$$

subject to $y_i(\beta_0 + \boldsymbol{\beta}^T\mathbf{X}_i) \geq 1 - \zeta_i$

and $\zeta_i \geq 0,$

and $\sum_{\mathrm{i}}^{\mathrm{n}} \zeta_i \leq \tilde{C},$ for $i = 1, \cdots, \mathrm{n}$

Linear Support Vector Machine

- Primal Problem

$$\min_{\beta_0, \boldsymbol{\beta}} ||\boldsymbol{\beta}||^2$$

subject to $y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) \geq 1 - \zeta_i$

and $\zeta_i \geq 0,$

and $\sum_{\mathrm{i}}^{\mathrm{n}} \zeta_i \leq \tilde{C},$ for $i = 1, \cdots, \mathrm{n}$

$$\Longleftrightarrow$$

- Primal Problem

$$\min_{\beta_0, \boldsymbol{\beta}, \zeta_i} ||\boldsymbol{\beta}||^2 + \mathrm{C} \sum_{\mathrm{i}}^{\mathrm{n}} \zeta_i$$

subject to $y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) \geq 1 - \zeta_i$

and $\zeta_i \geq 0,$ for $i = 1, \cdots, \mathrm{n}$

C is not a Lagrange multiplier

Linear Support Vector Machine

- Primal Problem

$$\underset{\beta_0, \boldsymbol{\beta}, \zeta_i}{min} \quad ||\boldsymbol{\beta}||^2 + \mathrm{C} \sum_{\mathrm{i}}^{\mathrm{n}} \zeta_i$$

subject to $\quad y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) \geq 1 - \zeta_i$

and $\quad \zeta_i \geq 0, \quad$ for $\; i = 1, \cdots, \mathrm{n}$

$$\Longleftrightarrow$$

- Dual Problem

$$\underset{\beta_0, \boldsymbol{\beta}, \zeta_i}{min} \quad ||\boldsymbol{\beta}||^2 + \mathrm{C} \sum_{\mathrm{i}}^{\mathrm{n}} \zeta_i - \sum_{\mathrm{i}}^{\mathrm{n}} \gamma_i \zeta_i$$

subject to $\quad y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) \geq 1 - \zeta_i$

for $\; i = 1, \cdots, \mathrm{n}$

Linear Support Vector Machine

- Primal Problem

$$\min_{\beta_0,\, \boldsymbol{\beta},\, \zeta_i} \ ||\boldsymbol{\beta}||^2 + \mathrm{C}\sum_{\mathrm{i}}^{\mathrm{n}} \zeta_i - \sum_{\mathrm{i}}^{\mathrm{n}} \gamma_i \zeta_i$$

subject to $y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) \geq 1 - \zeta_i,$ for $i = 1, \cdots, \mathrm{n}$

Linear Support Vector Machine

- Dual Problem

$$\min_{\beta_0,\, \boldsymbol{\beta},\, \zeta_i} \ ||\boldsymbol{\beta}||^2 + \mathrm{C}\sum_{\mathrm{i}}^{\mathrm{n}} \zeta_i - \sum_{\mathrm{i}}^{\mathrm{n}} \gamma_i \zeta_i - \sum_{\mathrm{i}}^{\mathrm{n}} \alpha_i \left[\, y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) - (1 - \zeta_i)\right]$$

- Taking derivative w.r.t β_0, $\boldsymbol{\beta}$, ζ_i

(Stationary)

Linear Support Vector Machine

$$\min_{\beta_0, \boldsymbol{\beta}, \zeta_i} \; ||\boldsymbol{\beta}||^2 + \mathrm{C}\sum_{\mathrm{i}}^{\mathrm{n}} \zeta_i - \sum_{\mathrm{i}}^{\mathrm{n}} \gamma_i \, \zeta_i - \sum_{\mathrm{i}}^{\mathrm{n}} \alpha_i \left[\, y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) - (1 - \zeta_i)\right]$$

$$\text{(Stationary)}\begin{cases} \dfrac{\partial}{\partial \beta_0}\mathcal{L}_p: & \displaystyle\sum_{i}^{n} \alpha_i \, y_i = 0 \\ \dfrac{\partial}{\partial \boldsymbol{\beta}}\mathcal{L}_p: & \boldsymbol{\beta} = \displaystyle\sum_{i}^{n} \alpha_i \, y_i \mathbf{x}_i \\ \dfrac{\partial}{\partial \zeta_i}\mathcal{L}_p: & \alpha_i = C - \gamma_i \end{cases} \qquad \text{(Complementary Slackness)}\begin{cases} \alpha_i[\, y_i f(\mathbf{x}_i) - (1 - \zeta_i)] = 0 \\ \gamma_i \, \zeta_i = 0 \end{cases}$$

Linear Support Vector Machine

$$\min_{\beta_0, \boldsymbol{\beta}, \zeta_i} \ ||\boldsymbol{\beta}||^2 + \mathrm{C}\sum_{\mathrm{i}}^{\mathrm{n}} \zeta_i - \sum_{\mathrm{i}}^{\mathrm{n}} \gamma_i \zeta_i - \sum_{\mathrm{i}}^{\mathrm{n}} \alpha_i \left[y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) - (1 - \zeta_i)\right]$$

$$\Longleftrightarrow \quad \max_{\alpha_i} \ \sum_{\mathrm{i}}^{\mathrm{n}} \alpha_i - \frac{1}{2} \sum_{\mathrm{i}}^{\mathrm{n}} \sum_{\mathrm{j}}^{\mathrm{n}} \alpha_i \alpha_j y_i y_{\mathrm{j}} \mathbf{x}_i{}^T \mathbf{x}_{\mathrm{j}} \qquad \text{QP}$$

subject to $0 \leq \alpha_i \leq C$

and $\sum_{\mathrm{i}}^{\mathrm{n}} \alpha_i y_i = 0,$ for $i = 1, \cdots, \mathrm{n}$

Linear Support Vector Machine

$$\min_{\beta_0,\, \boldsymbol{\beta},\, \zeta_i} \;\; ||\boldsymbol{\beta}||^2 + \mathrm{C}\sum_{\mathrm{i}}^{\mathrm{n}} \zeta_i - \sum_{\mathrm{i}}^{\mathrm{n}} \gamma_i\, \zeta_i - \sum_{\mathrm{i}}^{\mathrm{n}} \alpha_i \left[\, y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) - (1 - \zeta_i)\right]$$

$$\Longrightarrow \qquad \widehat{\boldsymbol{\beta}} = \sum_{i}^{n} \widehat{\alpha_i}\, y_i \mathbf{x}_i$$

$$\widehat{\beta_0} = y_i - \widehat{\boldsymbol{\beta}}^T \mathbf{x}_\mathrm{k} \quad \text{for any support vector } \mathbf{x}_\mathrm{k}$$

$$\widehat{f(\mathbf{x}_i)} = \widehat{\beta_0} + \widehat{\boldsymbol{\beta}}^T \mathbf{x}_\mathrm{k}$$

Kernel Support Vector Machine

$$\min_{\beta_0, \boldsymbol{\beta}, \zeta_i} \ ||\boldsymbol{\beta}||^2 + C\sum_i^n \zeta_i - \sum_i^n \gamma_i \zeta_i - \sum_i^n \alpha_i [y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i) - (1 - \zeta_i)]$$

$$\Longleftrightarrow \quad \max_{\alpha_i} \ \sum_i^n \alpha_i - \frac{1}{2}\sum_i^n \sum_j^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

subject to $0 \le \alpha_i \le C$

and $\sum_i^n \alpha_i y_i = 0$, for $i = 1, \cdots, n$

Kernel Support Vector Machine

- Kernel function

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

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)

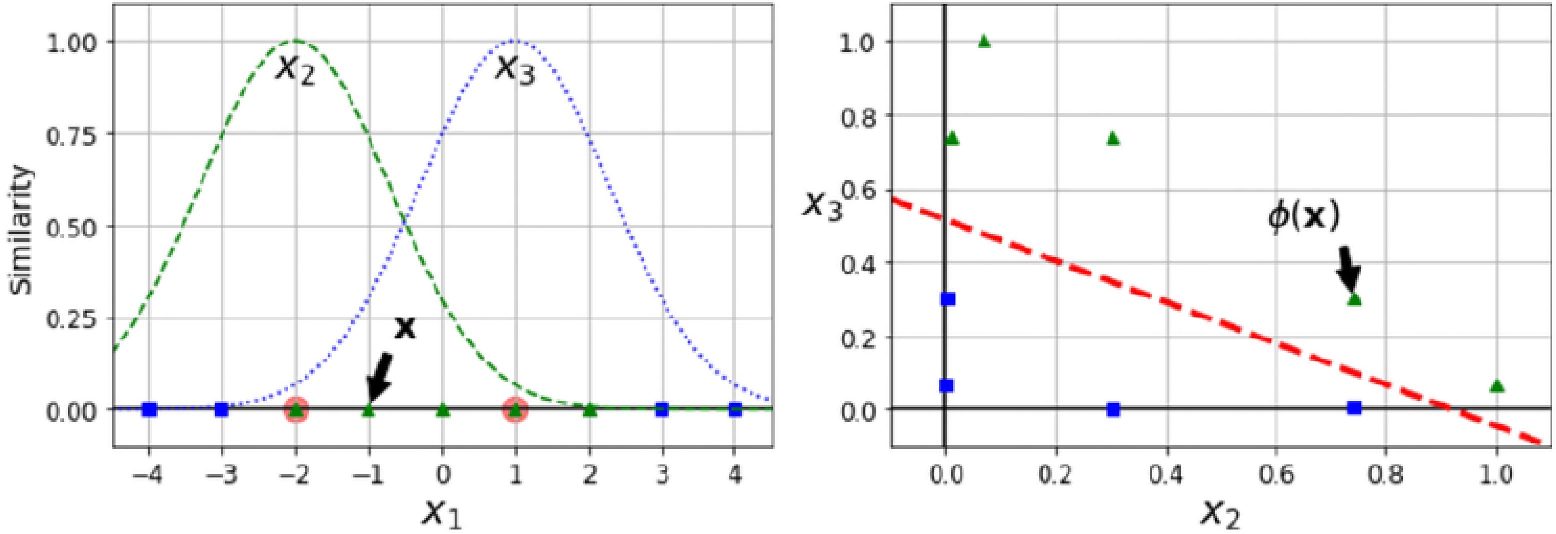
$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma + \gamma\, \mathbf{x}_i^T \mathbf{x}_j)^p$$

polynomial Kernel

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

Sigmoid Kernel

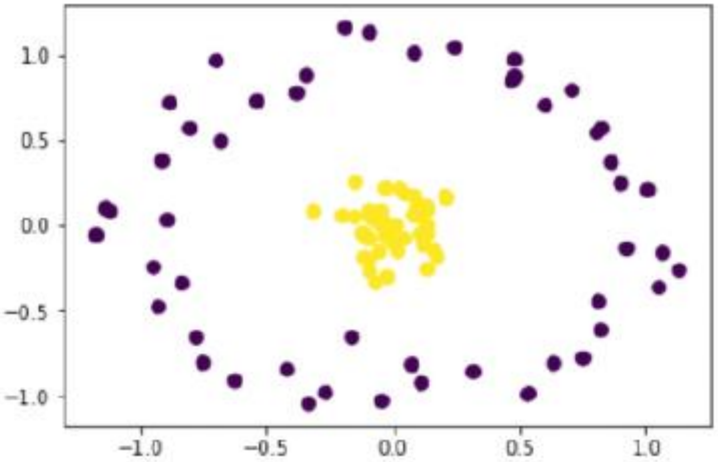
Kernel Support Vector Machine



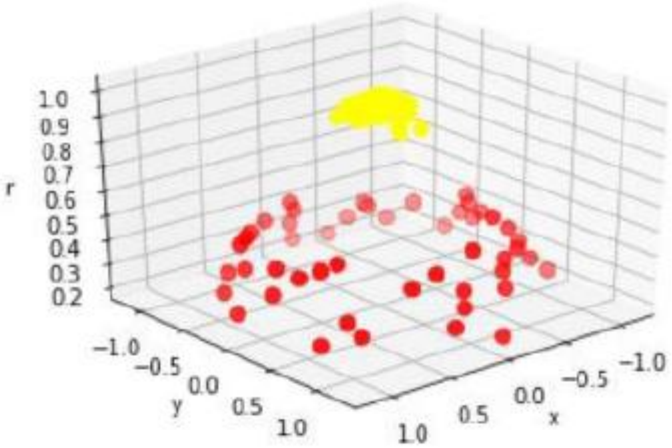
x_1

$$x_2 = \exp(-(x_1 - 2)^2)$$
$$x_3 = \exp(-(x_1 + 1)^2)$$

Kernel Support Vector Machine



(x_1, x_2)



$(x_1, x_2, \exp(-(x_1{}^2 + x_2{}^2)))$

Kernel Support Vector Machine

$$\min_{\beta_0, \boldsymbol{\beta}, \zeta_i} \ ||\boldsymbol{\beta}||^2 + C\sum_{i}^{n} \zeta_i - \sum_{i}^{n} \gamma_i \zeta_i - \sum_{i}^{n} \alpha_i \left[y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i) - (1 - \zeta_i) \right]$$

$$\Longleftrightarrow \quad \max_{\alpha_i} \ \sum_{i}^{n} \alpha_i - \frac{1}{2}\sum_{i}^{n}\sum_{j}^{n} \alpha_i \alpha_j y_i y_j \mathrm{h}(\mathbf{x}_i)^T h(\mathbf{x}_j)$$

subject to $\quad 0 \le \alpha_i \le C$

and $\quad \sum_{i}^{n} \alpha_i y_i = 0, \qquad$ for $\ i = 1, \cdots, \mathrm{n}$

Kernel Trick

$$\min_{\beta_0, \boldsymbol{\beta}, \zeta_i} \ ||\boldsymbol{\beta}||^2 + C\sum_{i}^{n} \zeta_i - \sum_{i}^{n} \gamma_i \zeta_i - \sum_{i}^{n} \alpha_i \left[y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i) - (1 - \zeta_i) \right]$$

$$\Longleftrightarrow \quad \max_{\alpha_i} \ \sum_{i}^{n} \alpha_i - \frac{1}{2}\sum_{i}^{n}\sum_{j}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

subject to $\quad 0 \leq \alpha_i \leq C$

and $\quad \sum_{i}^{n} \alpha_i y_i = 0, \qquad$ for $i = 1, \cdots, n$

Kernel Trick

- 특성함수의 생성 어려움 + 고차원 확장시 차원의 저주 문제 발생.

- 2차 다항커널 : 입력변수 x_1과 x_2이고 i번째 관측치와 j번째 관측치일때,

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (1 + \boldsymbol{x}_i^T \boldsymbol{x}_j)^2 \tag{7.11}$$
$$= (1 + x_{i,1}x_{j,1} + x_{i,2}x_{j,2})^2$$
$$= 1 + 2x_{i,1}x_{j,1} + 2x_{i,2}x_{j,2} + (x_{i,1}x_{j,1})^2 + (x_{i,2}x_{j,2})^2 + 2x_{i,1}x_{j,1}x_{i,2}x_{j,2}$$

- 이때 다음과 같이 정의하면,

$$h_1(x_1, x_2) = 1,\ h_2(x_1, x_2) = \sqrt{2}x_1,\ h_3(x_1, x_2) = \sqrt{2}x_2,\ h_4(x_1, x_2) = x_1^2,\ h_5(x_1, x_2) = x_2^2,\ h_6(x_1, x_2) = \sqrt{2}x_1x_2$$

$$\boldsymbol{h}(x_1, x_2) = (h_1(x_1, x_2), h_2(x_1, x_2), \cdots, h_6(x_1, x_2))^T;$$

- 식 (7.11)은 $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (1 + \boldsymbol{x}_i^T \boldsymbol{x}_j)^2 = \boldsymbol{h}(\boldsymbol{x}_i)^T \boldsymbol{h}(\boldsymbol{x}_j)$ 로 변형 가능.

- 특성함수를 정의하지 않고 커널 함수를 이용.

- 즉, β이 $\boldsymbol{h}(\boldsymbol{x}_i)^T \boldsymbol{h}(\boldsymbol{x}_j)$의 형태이면, $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$를 직접 이용하여 추정.

Kernel Trick

- 특성변수 x로 부터 basis함수 $\boldsymbol{h}(\boldsymbol{x})$로 차원을 증대시키면 커널 SVM 목적함수.

$$L_k = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \boldsymbol{h}(\boldsymbol{x}_i)^T \boldsymbol{h}(\boldsymbol{x}_j) \qquad (7.12)$$

- 선형 SVM 식 (7.11)은 $\hat{f}(\boldsymbol{x}) = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}^T \boldsymbol{x} = \hat{\beta}_0 + \sum_{i=1}^{n} \hat{\alpha}_i y_i \boldsymbol{x}_i^T \boldsymbol{x}$ 로 변형 가능.

- L_k 최소화한 모수 추정치를 $\hat{\beta}_0^*$와 $\hat{\boldsymbol{\beta}}^*$ 라 할 때 커널 SVM의 예측치

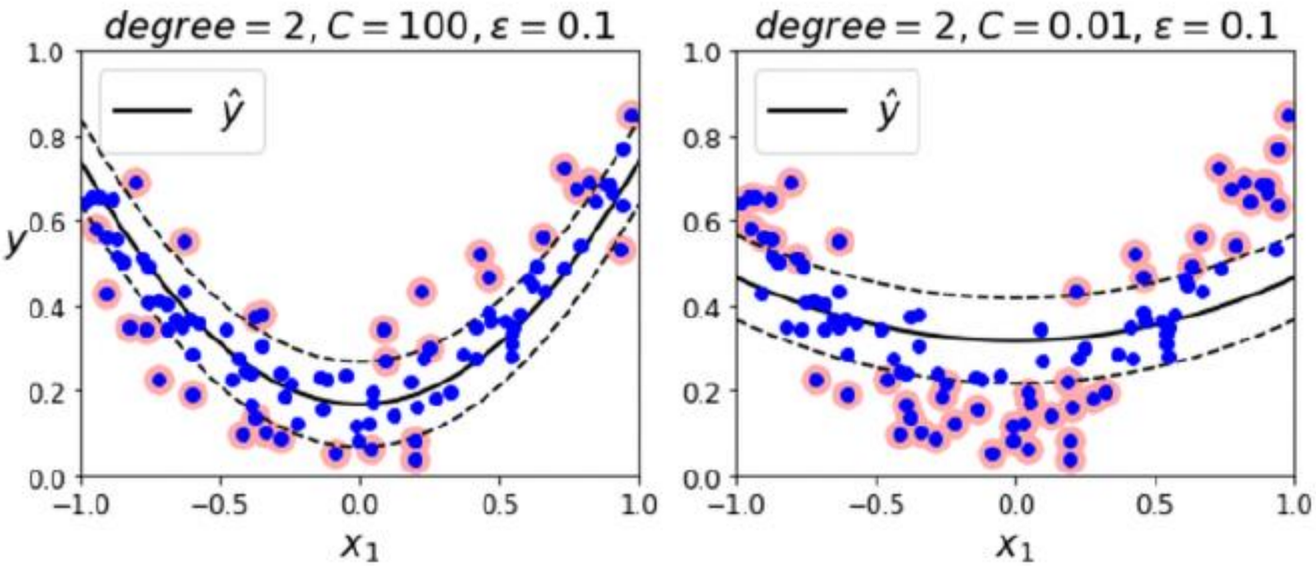
$$\hat{f}(\boldsymbol{x}) = \hat{\beta}_0^* + \sum_{i=1}^{n} \hat{\alpha}_i^* y_i \boldsymbol{h}(\boldsymbol{x}_i)^T \boldsymbol{h}(\boldsymbol{x}) \qquad (7.13)$$

- 식(7.12)와 식(7.13) 모두 $\boldsymbol{h}(\boldsymbol{x}_i)^T \boldsymbol{h}(\boldsymbol{x}_j)$의 형태임.

- 식(7.12)에 $\boldsymbol{h}(\boldsymbol{x}_i)^T \boldsymbol{h}(\boldsymbol{x}_j)$ 대신 커널 함수 $K(\boldsymbol{x}_i, \boldsymbol{x})$를 대체하여 $\hat{\beta}_0^*$와 $\hat{\boldsymbol{\beta}}^*$ 를 추정.

- 식(7.13)도 $\boldsymbol{h}(\boldsymbol{x}_i)^T \boldsymbol{h}(\boldsymbol{x}_j)$를 이용하여 동일한 커널 SVM을 구함.

Support Vector Regression



Logistic Regression vs SVM

- Logistic Regression
  - How to Estimate? $\underset{\boldsymbol{\beta}}{argmax}\ L(\boldsymbol{\beta})$

$$L(\boldsymbol{\pi}; \mathbf{X}) = \prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

$$l(\boldsymbol{\pi}; \mathbf{X}) = \sum_{i=1}^{n} [y_i \log \pi_i + (1 - y_i) \log(1 - \pi_i)]$$

Logistic Regression vs SVM

- Logistic Regression

- For $y_i \in \{-1, 1\}$, MLE for LR minimizes

$$\frac{1}{n}\sum_{i=1}^{n} \underbrace{\log\left(1 + e^{-y_i f(\mathbf{x}_i)}\right)}_{\text{Logistic Loss}}$$

which converges to

$$E[L_{\text{logit}}\{M\}] = E\left\{\log\left(1 + e^{-Y f(\mathbf{X})}\right)\right\}$$

where $M = Y f(\mathbf{X})$ denotes the margin of $f(\mathbf{x}) = \beta + \beta^T \mathbf{x}$.

Logistic Regression vs SVM

- SVM

  - Linear SVM solves

$$\min_{\beta_0,\boldsymbol{\beta},\xi_i} \frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\beta} + C\sum_{i=1}^{n}\xi_i$$

$$\text{subject to} \quad y_i(\beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i) \geq 1 - \xi_i, \quad i = 1, \cdots, n$$

$$\xi_i \geq 0, \qquad i = 1, \cdots, n.$$

  - It is equivalent to solve

$$\min_{\beta_0,\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n}[1 - y_i(\beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i)]_+ + \frac{\lambda}{2}\boldsymbol{\beta}^T\boldsymbol{\beta}$$

Logistic Regression vs SVM

- LR solves

$$\min_{\beta_0,\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n} \log\left\{1+e^{-y_i f(\mathbf{x}_i)}\right\}$$

which converges to the expectation of

$$L_{\text{logit}}(M) = \log\left\{1+e^{-M}\right\}.$$

as $n \to \infty$. (Logit Risk)

- SVM

$$\min_{\beta_0,\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n} [1 - y_i(\beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i)]_+ + \lambda \boldsymbol{\beta}^T \boldsymbol{\beta}$$

which converges to the expectation of

$$L_{\text{hinge}}(M) = [1-M]_+$$

as $n \to \infty$ and $\lambda \to 0$. (Hinge Risk)

Logistic Regression vs SVM