

**サポートベクター回帰**

**Support Vector Regression  
SVR**

大阪府立大学 工学研究科  
清水 悠生

# サポートベクター回帰(SVR)とは

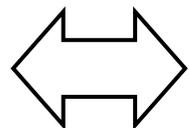
- ✓ サポートベクターマシン(SVM)を回帰手法に応用した手法
- ✓  $\epsilon$ 許容誤差によりノイズの影響を受けにくい
- ✓ カーネル関数による非線形回帰分析も可能
- ✓ 数千オーダーのデータセットに対して非常に強力  
(データ数が増えると学習コストの観点から不利に)

# 線形回帰の行列表現

✓ n個のデータがあるときの線形回帰の行列表現を考える

$$\begin{aligned}y^{(0)} &= b + w_1 x_1^{(0)} + \dots + w_m x_m^{(0)} \\y^{(1)} &= b + w_1 x_1^{(1)} + \dots + w_m x_m^{(1)} \\&\vdots \\y^{(n)} &= b + w_1 x_1^{(n)} + \dots + w_m x_m^{(n)}\end{aligned}$$

$y_1^{(i)}, \dots, y_m^{(i)}$  : i番目の出力データ  
( $i = 0, \dots, n$ )  
 $x_1^{(i)}, \dots, x_m^{(i)}$  : i番目のm(<n)種類の  
入力データ( $i = 0, \dots, n$ )  
 $w_1, \dots, w_m$  : 学習する重み係数  
 $b$  : バイアス


$$\begin{bmatrix} y^{(0)} \\ y^{(1)} \\ \vdots \\ y^{(n)} \end{bmatrix} = b + w_1 \begin{bmatrix} x_1^{(0)} \\ x_1^{(1)} \\ \vdots \\ x_1^{(n)} \end{bmatrix} + \dots + w_m \begin{bmatrix} x_m^{(0)} \\ x_m^{(1)} \\ \vdots \\ x_m^{(n)} \end{bmatrix} = b + \begin{bmatrix} x_1^{(0)} & \dots & x_m^{(0)} \\ x_1^{(1)} & \dots & x_m^{(1)} \\ \vdots & & \vdots \\ x_1^{(n)} & \dots & x_m^{(n)} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}$$

$$= \begin{bmatrix} b + \mathbf{x}^{(0)} \mathbf{w}^T \\ b + \mathbf{x}^{(1)} \mathbf{w}^T \\ \vdots \\ b + \mathbf{x}^{(n)} \mathbf{w}^T \end{bmatrix}$$

# 最小二乗法の誤差関数

- ✓ 最小二乗法の誤差関数 $E(\mathbf{w})$ はこのように定義できる

## 線形回帰(最小二乗法)の誤差関数

$$E(\mathbf{w}) = \sum_{i=0}^n \underbrace{\left( y^{(i)} - (b + \mathbf{x}^{(i)} \mathbf{w}^T) \right)^2}_{\text{誤差の2乗}}$$

## Ridge回帰の誤差関数

$$E(\mathbf{w}) = \sum_{i=0}^n \underbrace{\left( y^{(i)} - (b + \mathbf{x}^{(i)} \mathbf{w}^T) \right)^2}_{\text{誤差の2乗}} + \underbrace{\frac{\alpha}{2} \|\mathbf{w}\|^2}_{\text{正則化項}} \quad \begin{array}{l} \alpha: \text{ハイパーパラメータ} \\ \|\cdot\|: \text{L2ノルム} \end{array}$$

- ✓ 最小二乗法の詳細についてはこちらを参照

<https://yuyumoyuyu.com/2020/12/13/simpleregressionwithols/>

<https://yuyumoyuyu.com/2020/12/20/multipleregression/>

<https://yuyumoyuyu.com/2021/01/03/regularizedleastquares/>

# サポートベクター回帰の誤差関数

- ✓ サポートベクター回帰の誤差関数は Ridge回帰と比べて誤差項に使用する関数のみが異なる

## Ridge回帰の誤差関数

$$E(\mathbf{w}) = \sum_{i=0}^n \underbrace{\left( y^{(i)} - (b + \mathbf{x}^{(i)} \mathbf{w}^T) \right)^2}_{\text{誤差の2乗}} + \underbrace{\frac{\alpha}{2} \|\mathbf{w}\|^2}_{\text{正則化項}}$$

## サポートベクター回帰の誤差関数

$C, \varepsilon (> 0)$ :  
ハイパーパラメータ

$$E(\mathbf{w}) = C \sum_{i=0}^n \underbrace{\max\{0, |y^{(i)} - (b + \mathbf{x}^{(i)} \mathbf{w}^T)| - \varepsilon\}}_{\varepsilon \text{許容誤差}} + \underbrace{\frac{1}{2} \|\mathbf{w}\|^2}_{\text{正則化項}}$$

ハイパーパラメータ  $C$  はSVMに合わせて設定しているが誤差項と正則化項のバランスをとるという点で Ridge回帰の  $\alpha$  と同様の役割を担う

# 損失関数のイメージ

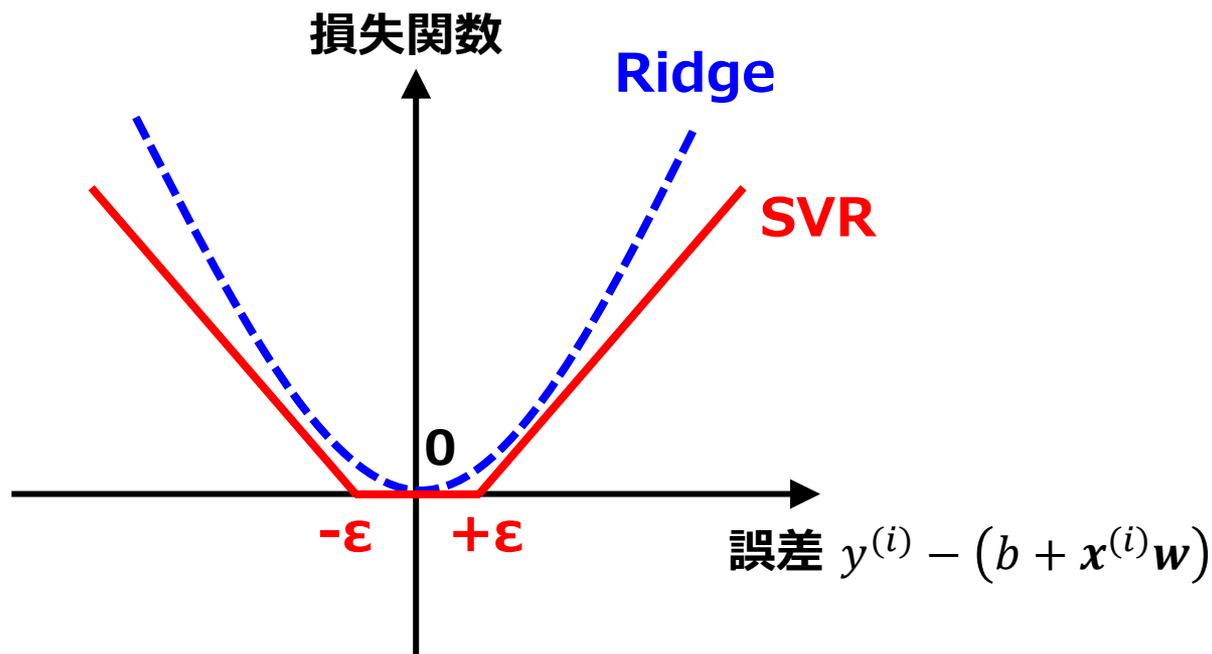
- ✓ サポートベクター回帰では誤差が  $\pm\varepsilon$  以下の場合には誤差項の値は0となる

Ridge回帰

$$E(\mathbf{w}) = \sum_{i=0}^n \left( y^{(i)} - (b + \mathbf{x}^{(i)} \mathbf{w}^T) \right)^2 + \frac{\alpha}{2} \|\mathbf{w}\|^2$$

サポートベクター回帰

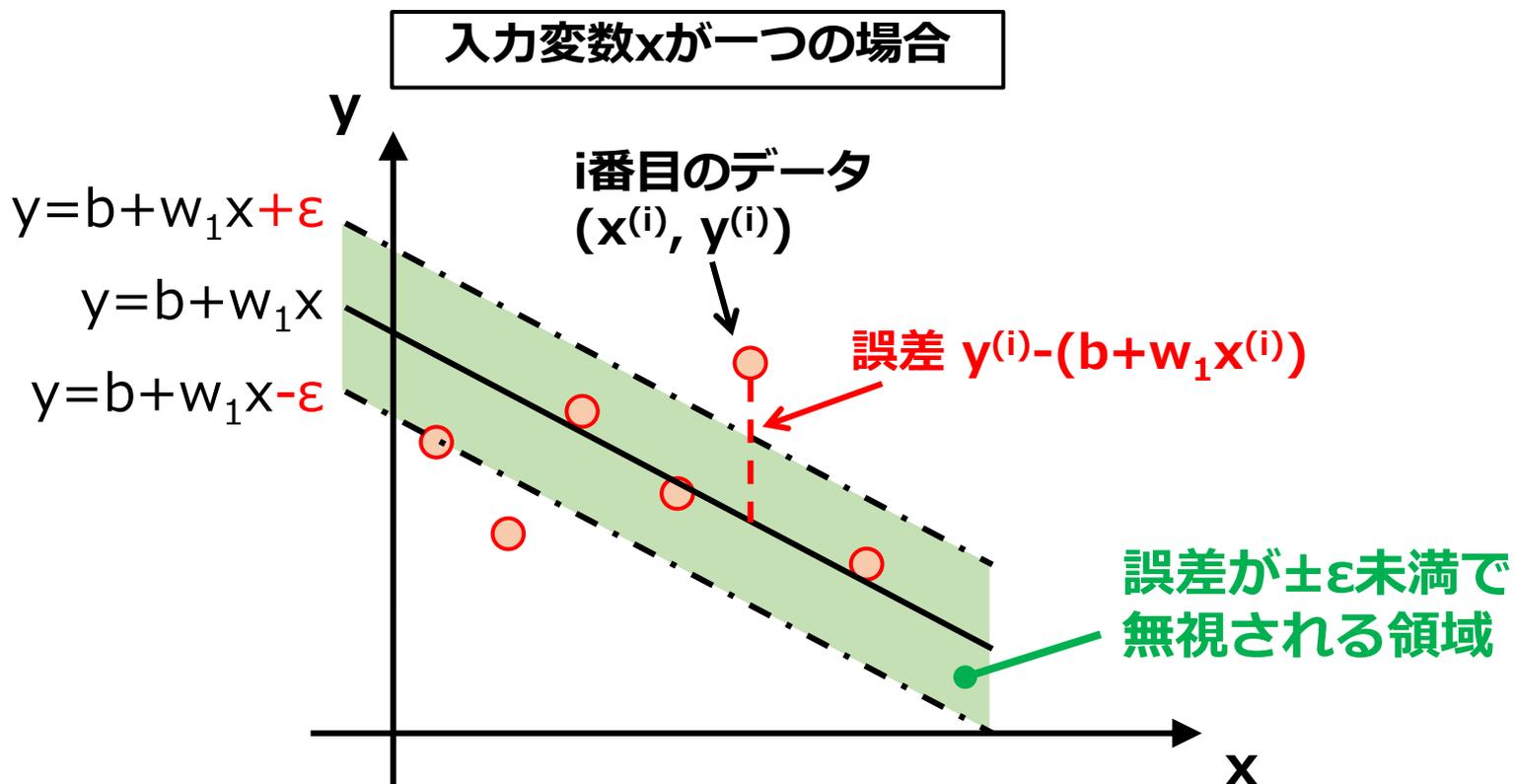
$$E(\mathbf{w}) = C \sum_{i=0}^n \max\{0, |y^{(i)} - (b + \mathbf{x}^{(i)} \mathbf{w}^T)| - \varepsilon\} + \frac{1}{2} \|\mathbf{w}\|^2$$



# 損失関数の役割

- ✓  $\pm\varepsilon$  以下の誤差を無視することで  
ノイズの影響を受けにくい回帰手法となる

**サポートベクター回帰**  $E(\mathbf{w}) = C \sum_{i=0}^n \max\{0, |y^{(i)} - (b + \mathbf{x}^{(i)} \mathbf{w}^T)| - \varepsilon\} + \frac{1}{2} \|\mathbf{w}\|^2$



# 誤差関数の最小化により係数ベクトルを求める

- ✓ 誤差関数の最小化するような係数ベクトルを求めることでサポートベクター回帰の回帰式を決定する

## 回帰式

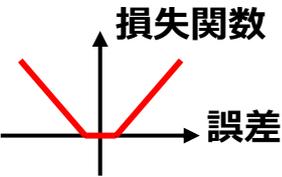
$$f(\mathbf{x}) = \mathbf{x}\mathbf{w}^T + b$$

## 誤差関数

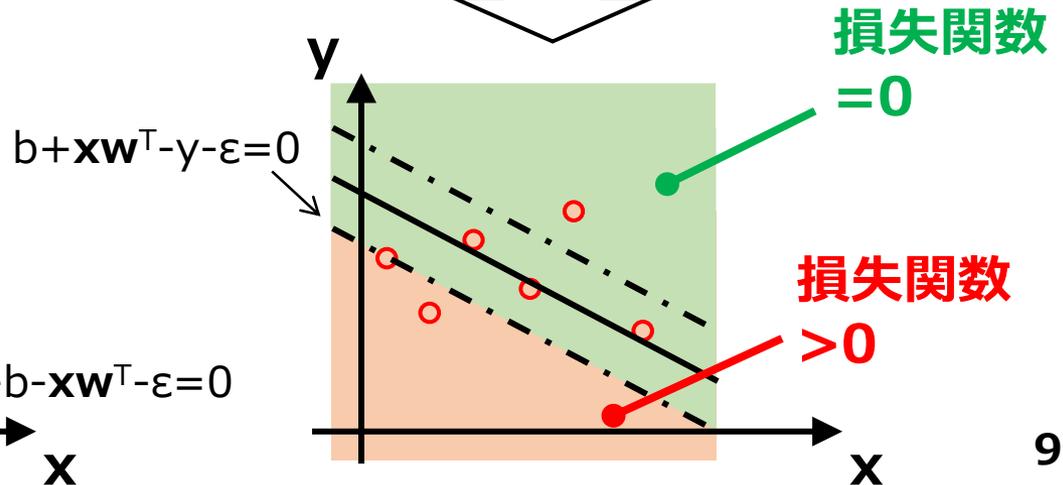
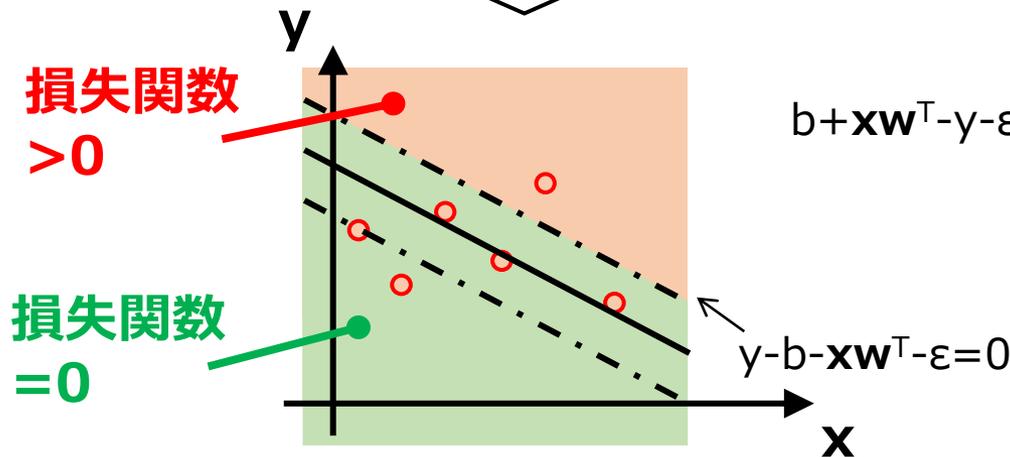
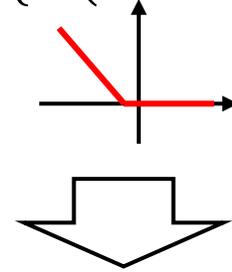
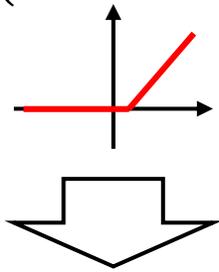
$$E(\mathbf{w}) = C \sum_{i=0}^n \max\{0, |y^{(i)} - (b + \mathbf{x}^{(i)}\mathbf{w}^T)| - \varepsilon\} + \frac{1}{2} \|\mathbf{w}\|^2$$

# 誤差関数の絶対値の場合分け

- ✓ 誤差関数を最小化するため  
まずは損失関数の絶対値を場合分けする

$$\sum_{i=0}^n \max\{0, |y^{(i)} - (b + \mathbf{x}^{(i)} \mathbf{w}^T)| - \varepsilon\}$$


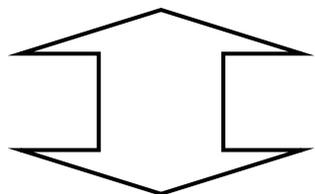
$$= \sum_{i=0}^n \max\{0, (y^{(i)} - (b + \mathbf{x}^{(i)} \mathbf{w}^T)) - \varepsilon\} + \sum_{i=0}^n \max\{0, ((b + \mathbf{x}^{(i)} \mathbf{w}^T) - y^{(i)}) - \varepsilon\}$$



# スラック変数の導入による主問題の定義

- ✓ max関数が入った誤差関数最小化問題を  
**スラック変数**を導入することで制約あり最小化問題に変換
- ✓ 変換の詳細説明は次ページ

$$\min_{\mathbf{w}, b} C \left( \sum_{i=0}^n \max \left\{ 0, \left( y^{(i)} - (b + \mathbf{x}^{(i)} \mathbf{w}^T) \right) - \varepsilon \right\} + \sum_{i=0}^n \max \left\{ 0, \left( (b + \mathbf{x}^{(i)} \mathbf{w}^T) - y^{(i)} \right) - \varepsilon \right\} \right) + \frac{1}{2} \|\mathbf{w}\|^2$$



スラック変数  $\xi^+, \xi^-$  を導入し  
max関数を消去

$$\min_{\mathbf{w}, b, \xi^+, \xi^-} C \sum_{i=0}^n (\xi_i^+ - \xi_i^-) + \frac{1}{2} \|\mathbf{w}\|^2$$

$$\text{s. t. } \xi_i^+ \geq (y^{(i)} - b - \mathbf{x}^{(i)} \mathbf{w}^T) - \varepsilon, \xi_i^+ \geq 0 \quad (i = 0, 1, \dots, n)$$

$$\xi_i^- \geq (b + \mathbf{x}^{(i)} \mathbf{w}^T - y^{(i)}) - \varepsilon, \xi_i^- \geq 0 \quad (i = 0, 1, \dots, n)$$

ここで

$$\xi^+ = [\xi_0^+, \dots, \xi_n^+]^T$$

$$\xi^- = [\xi_0^-, \dots, \xi_n^-]^T$$

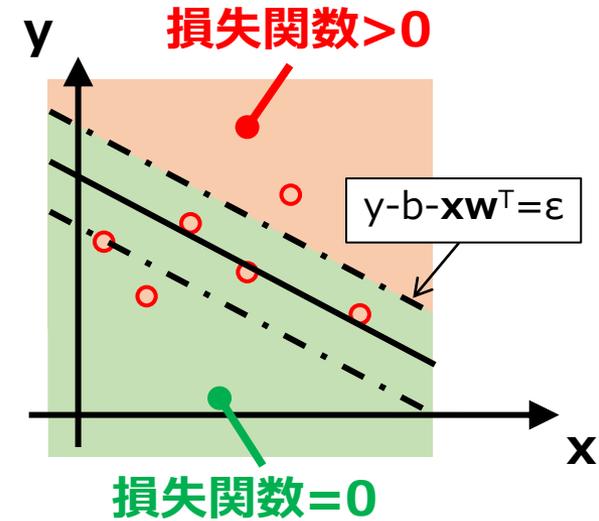
この最適化問題を  
サポートベクター回帰の**主問題**と呼ぶ

# なぜスラック変数によりmax関数が消去できるか(1/2)

点 $(x^{(i)}, y^{(i)})$ の一方の損失関数についてのみ考える  
 最小化問題は次式のように変形できる

$$\min_{w,b} \max \left\{ 0, \left( y^{(i)} - (b + x^{(i)} w^T) \right) - \varepsilon \right\}$$

$$= \begin{cases} \text{If } \left( y^{(i)} - (b + x^{(i)} w^T) \right) \geq \varepsilon \\ \min_{w,b} \left( y^{(i)} - (b + x^{(i)} w^T) \right) - \varepsilon \\ \text{s.t. } \left( y^{(i)} - (b + x^{(i)} w^T) \right) \geq \varepsilon \\ \text{If } \left( y^{(i)} - (b + x^{(i)} w^T) \right) < \varepsilon \\ 0 \end{cases}$$



1つ目の最小化問題はスラック変数を導入することで次式のように変形できる

$$\min_{w,b} \left( y^{(i)} - (b + x^{(i)} w^T) \right) - \varepsilon$$

$$\text{s.t. } y^{(i)} - b - x^{(i)} w^T = \varepsilon + \xi_i^+$$

$$\xi_i^+ \geq 0 \quad \text{不等式制約}$$

$$\xi_i^+ \geq 0 \quad \text{非負制約}$$

$$\min_{w,b} \xi_i^+$$

$$\text{s.t. } \xi_i^+ = y^{(i)} - b - x^{(i)} w^T - \varepsilon$$

$$\xi_i^+ \geq 0$$

スラック変数はそもそも、不等式制約を  
 等式制約と非負制約に変換するために導入する

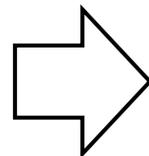
# なぜスラック変数によりmax関数が消去できるか(2/2)

前ページの最小化問題を更にw,bの全ての定義域で成立するように修正する

<p>最小化問題</p>	$\min_{w,b} \xi_i^+$ <p>s. t. <math>\xi_i^+ = y^{(i)} - b - x^{(i)} w^T - \varepsilon</math></p> $\xi_i^+ \geq 0$	$\min_{w,b,\xi_i^+} \xi_i^+$ <p>s. t. <math>\xi_i^+ \geq y^{(i)} - b - x^{(i)} w^T - \varepsilon</math></p> $\xi_i^+ \geq 0$
<p>問題のイメージ</p> <p><math>y^{(i)} - b - x^{(i)} w^T \geq \varepsilon</math> の場合</p>	<p>w,bに関して青点を最小化</p>	<p>w,bに関して青点を最小化</p> <p><math>\xi_i^+</math>に関して青点で最小</p> <p>左図と同じ設定となる</p>
<p>問題のイメージ</p> <p><math>y^{(i)} - b - x^{(i)} w^T &lt; \varepsilon</math> の場合</p>	<p>制約を満たす <math>\xi_i^+</math> が存在しない</p>	<p>w,bによらず0で最小</p> <p><math>\xi_i^+</math>に関して0で最小</p> <p>常に <math>\xi_i^+ = 0</math> となる</p>

以上より, max関数付き最小化問題を制約あり最小化問題に変換できる

$$\min_{w,b} \max \left\{ 0, \left( y^{(i)} - (b + x^{(i)} w^T) \right) - \varepsilon \right\}$$



$$\min_{w,b,\xi_i^+} \xi_i^+$$

s. t.  $\xi_i^+ \geq y^{(i)} - b - x^{(i)} w^T - \varepsilon$

$$\xi_i^+ \geq 0$$

# 主問題から双対問題へ

- ✓ サポートベクター回帰の主問題は二次計画問題として解くことが可能
- ✓ が、**双対問題**に変換することで下記のメリットが存在
  - ① 主問題よりも解きやすい場合が存在
  - ② 非線形化がしやすい形式に
- ✓ そのため、主問題をさらに双対問題に変換する

## サポートベクター回帰の主問題

$$\min_{w, b, \xi^+, \xi^-} C \sum_{i=0}^n (\xi_i^+ - \xi_i^-) + \frac{1}{2} \|w\|^2$$

$$\text{s. t. } \xi_i^+ \geq (y^{(i)} - b - \mathbf{x}^{(i)} w^T) - \varepsilon, \xi_i^+ \geq 0 \quad (i = 0, 1, \dots, n)$$

$$\xi_i^- \geq (b + \mathbf{x}^{(i)} w^T - y^{(i)}) - \varepsilon, \xi_i^- \geq 0 \quad (i = 0, 1, \dots, n)$$

# ラグランジュの未定乗数法による変換

- ✓ ラグランジュの未定乗数法により主問題を max-min問題に変換する

$$\min_{\mathbf{w}, b, \xi^+, \xi^-} C \sum_{i=0}^n (\xi_i^+ - \xi_i^-) + \frac{1}{2} \|\mathbf{w}\|^2$$

$$\text{s. t. } \xi_i^+ \geq (y^{(i)} - b - \mathbf{x}^{(i)} \mathbf{w}^T) - \varepsilon, \xi_i^+ \geq 0 \quad (i = 0, 1, \dots, n)$$

$$\xi_i^- \geq (b + \mathbf{x}^{(i)} \mathbf{w}^T - y^{(i)}) - \varepsilon, \xi_i^- \geq 0 \quad (i = 0, 1, \dots, n)$$



## ラグランジュ関数の導入

$$L(\mathbf{w}, b, \xi^+, \xi^-, \alpha^+, \eta^+, \alpha^-, \eta^-) = C \sum_{i=0}^n (\xi_i^+ - \xi_i^-) + \frac{1}{2} \|\mathbf{w}\|^2$$

$$- \sum_{i=0}^n \alpha_i^+ (\xi_i^+ - (y^{(i)} - b - \mathbf{x}^{(i)} \mathbf{w}^T) + \varepsilon) - \sum_{i=0}^n \eta_i^+ \xi_i^+$$

$$- \sum_{i=0}^n \alpha_i^- (\xi_i^- - (b + \mathbf{x}^{(i)} \mathbf{w}^T - y^{(i)}) + \varepsilon) - \sum_{i=0}^n \eta_i^- \xi_i^-$$

以下は全て新たに導入する非負の変数

$$\alpha^+ = [\alpha_0^+, \dots, \alpha_n^+]^T$$

$$\alpha^- = [\alpha_0^-, \dots, \alpha_n^-]^T$$

$$\eta^+ = [\eta_0^+, \dots, \eta_n^+]^T$$

$$\eta^- = [\eta_0^-, \dots, \eta_n^-]^T$$

## 変換したmax-min問題

$$\max_{\alpha^+ \geq 0, \eta^+ \geq 0, \alpha^- \geq 0, \eta^- \geq 0} \min_{\mathbf{w}, b, \xi^+, \xi^-} L(\mathbf{w}, b, \xi^+, \xi^-, \alpha^+, \eta^+, \alpha^-, \eta^-)$$

# 主変数と双対変数

- ✓ もとの主問題の変数を主変数,  
あらたに追加した変数を双対変数と呼ぶ
- ✓ max-min問題から主変数を消すことで双対問題が得られる
- ✓ そのため, ラグランジュ関数を主変数に関して最小化する

$$\underbrace{\max}_{\alpha^+ \geq 0, \eta^+ \geq 0, \alpha^- \geq 0, \eta^- \geq 0} \quad \underbrace{\min}_{w, b, \xi^+, \xi^-} L(\underbrace{w, b, \xi^+, \xi^-}_{\text{主変数}}, \underbrace{\alpha^+, \eta^+, \alpha^-, \eta^-}_{\text{双対変数}})$$

# ラグランジュ関数の主変数による最小化

- ✓ ラグランジュ関数は主変数に対して微分可能であるため各主変数で偏微分して0とすることで関係式を求める

## ラグランジュ関数

$$L(\mathbf{w}, b, \xi^+, \xi^-, \alpha^+, \eta^+, \alpha^-, \eta^-) = C \sum_{i=0}^n (\xi_i^+ - \xi_i^-) + \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=0}^n \alpha_i^+ (\xi_i^+ - (y^{(i)} - b - \mathbf{x}^{(i)} \mathbf{w}^T) + \varepsilon) - \sum_{i=0}^n \eta_i^+ \xi_i^+ - \sum_{i=0}^n \alpha_i^- (\xi_i^- - (b + \mathbf{x}^{(i)} \mathbf{w}^T - y^{(i)}) + \varepsilon) - \sum_{i=0}^n \eta_i^- \xi_i^-$$

## 主変数による偏微分

$$\frac{\partial}{\partial \mathbf{w}} L = \mathbf{w} - \sum_{i=0}^n \alpha_i^+ \mathbf{x}^{(i)} - \sum_{i=0}^n \alpha_i^- (-\mathbf{x}^{(i)}) = \mathbf{0} \Leftrightarrow \mathbf{w} = \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) \mathbf{x}^{(i)}$$

$$\frac{\partial}{\partial b} L = \sum_{i=0}^n \alpha_i^+ - \sum_{i=0}^n \alpha_i^- (-1) = \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) = 0$$

$$\frac{\partial}{\partial \xi_i^+} L = C - \alpha_i^+ - \eta_i^+ = 0, (i = 0, 1, \dots, n)$$

$$\frac{\partial}{\partial \xi_i^-} L = C - \alpha_i^- - \eta_i^- = 0, (i = 0, 1, \dots, n)$$

# 主変数を削除する

- ✓ 前ページで求めた関係式をラグランジュ関数に適用し  
双対変数のみの関数に（主変数に対して最小化する）
- ✓ 計算過程は次のページ

$$\begin{aligned} & \min_{\mathbf{w}, b, \xi^+, \xi^-} L(\mathbf{w}, b, \xi^+, \xi^-, \alpha^+, \eta^+, \alpha^-, \eta^-) \\ &= -\frac{1}{2} \sum_{i=0}^n \sum_{j=0}^n (\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-) \mathbf{x}^{(i)} \mathbf{x}^{(j)T} - \sum_{i=0}^n (-\alpha_i^+ + \alpha_i^-) y^{(i)} - \sum_{i=0}^n (\alpha_i^+ + \alpha_i^-) \varepsilon \end{aligned}$$

**$\eta^+, \eta^-$ が計算過程で削除できるため  
 $\alpha^+, \alpha^-$ のみの関数となる！**

# (前ページの計算過程)

## 偏微分から得られた4つの関係式

$$\mathbf{w} = \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) \mathbf{x}^{(i)},$$

$$\sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) = 0$$

$$C - \alpha_i^+ - \eta_i^+ = 0,$$

$$C - \alpha_i^- - \eta_i^- = 0$$

$$(i = 0, 1, \dots, n)$$

$$\begin{aligned}
 L(\mathbf{w}, b, \xi^+, \xi^-, \alpha^+, \eta^+, \alpha^-, \eta^-) &= C \sum_{i=0}^n (\xi_i^+ - \xi_i^-) + \frac{1}{2} \|\mathbf{w}\|^2 \\
 &\quad - \sum_{i=0}^n \alpha_i^+ (\xi_i^+ - (y^{(i)} - b - \mathbf{x}^{(i)} \mathbf{w}^T) + \varepsilon) - \sum_{i=0}^n \eta_i^+ \xi_i^+ - \sum_{i=0}^n \alpha_i^- (\xi_i^- - (b + \mathbf{x}^{(i)} \mathbf{w}^T - y^{(i)} + \varepsilon)) - \sum_{i=0}^n \eta_i^- \xi_i^- \\
 &= \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=0}^n \{(C - \alpha_i^+ - \eta_i^+) \xi_i^+ + (C - \alpha_i^- - \eta_i^-) \xi_i^-\} - \sum_{i=0}^n (-\alpha_i^+ + \alpha_i^-) y^{(i)} - \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) b \\
 &\quad - \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) \mathbf{x}^{(i)} \mathbf{w}^T - \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) \varepsilon \\
 &= \frac{1}{2} \sum_{i=0}^n \sum_{j=0}^n (\alpha_i^+ - \alpha_i^-) (\alpha_j^+ - \alpha_j^-) \mathbf{x}^{(i)} \mathbf{x}^{(j)T} - \sum_{i=0}^n (-\alpha_i^+ + \alpha_i^-) y^{(i)} \\
 &\quad - \sum_{i=0}^n \sum_{j=0}^n (\alpha_i^+ - \alpha_i^-) (\alpha_j^+ - \alpha_j^-) \mathbf{x}^{(i)} \mathbf{x}^{(j)T} - \sum_{i=0}^n (\alpha_i^+ + \alpha_i^-) \varepsilon \\
 &= -\frac{1}{2} \sum_{i=0}^n \sum_{j=0}^n (\alpha_i^+ - \alpha_i^-) (\alpha_j^+ - \alpha_j^-) \mathbf{x}^{(i)} \mathbf{x}^{(j)T} - \sum_{i=0}^n (-\alpha_i^+ + \alpha_i^-) y^{(i)} - \sum_{i=0}^n (\alpha_i^+ + \alpha_i^-) \varepsilon
 \end{aligned}$$

まとめる

関係式を適用

まとめる

# 双対問題が完成

- ✓ 双対変数の非負条件と偏微分で得られた関係式から下記の範囲制約が求まる

$$0 \leq \alpha_i^+, 0 \leq \eta_i^+ = C - \alpha_i^+ \Rightarrow 0 \leq \alpha_i^+ \leq C$$

$$0 \leq \alpha_i^-, 0 \leq \eta_i^- = C - \alpha_i^- \Rightarrow 0 \leq \alpha_i^- \leq C$$

- ✓ よって、双対問題は次式のように求まる
- ✓ 双対問題を二次計画問題として解くと  $\mathbf{a}^+, \mathbf{a}^-$  が求められる

## サポートベクター回帰の双対問題

$$\begin{aligned} \max_{\alpha^+, \alpha^-} \quad & -\frac{1}{2} \sum_{i=0}^n \sum_{j=0}^n (\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-) \mathbf{x}^{(i)} \mathbf{x}^{(j)T} \\ & + \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) y^{(i)} - \sum_{i=0}^n (\alpha_i^+ + \alpha_i^-) \varepsilon \end{aligned}$$

$$\text{s. t.} \quad \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) = 0, 0 \leq \alpha_i^+ \leq C, 0 \leq \alpha_i^- \leq C (i = 0, 1, \dots, n)$$

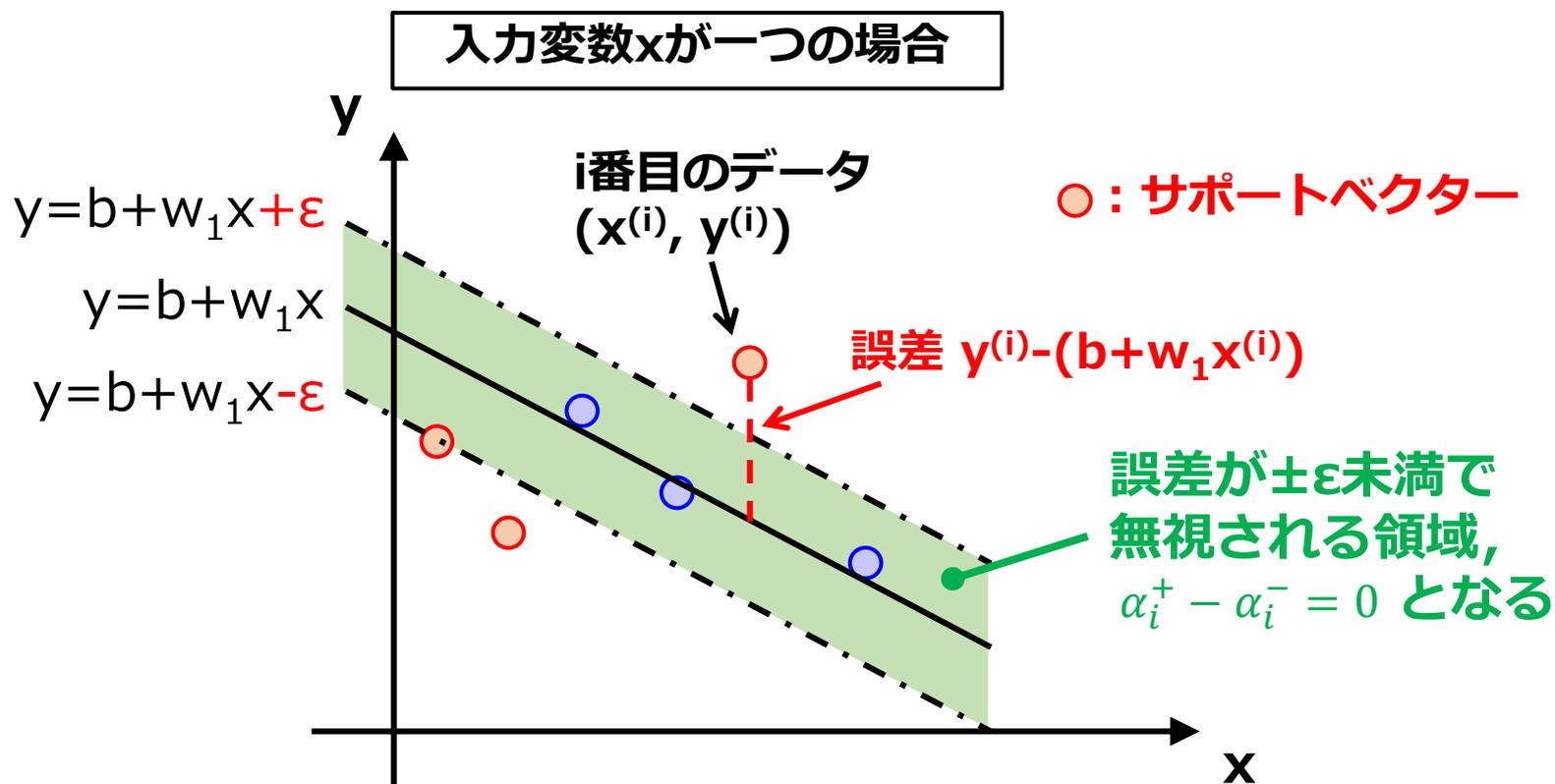
# サポートベクター回帰の回帰関数

- ✓ 偏微分から得られた関係式より,  
サポートベクター回帰の回帰関数は次式のように表される
- ✓ 二次計画問題の解  $\alpha^+, \alpha^-$  と入力データ  $\mathbf{x}^{(i)}$  によって  
サポートベクター回帰の回帰関数が求まる！
- ✓ バイアス  $b$  の導出は2つ後のスライド参照

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{x}\mathbf{w}^T + b \\ &= \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) \mathbf{x}^{(i)} \mathbf{x}^T + b \end{aligned}$$

# サポートベクターとは

- ✓ 誤差の絶対値が  $\varepsilon$  以上のサンプルによってSVRの回帰式は形成され, これらのサンプルをサポートベクターと呼ぶ

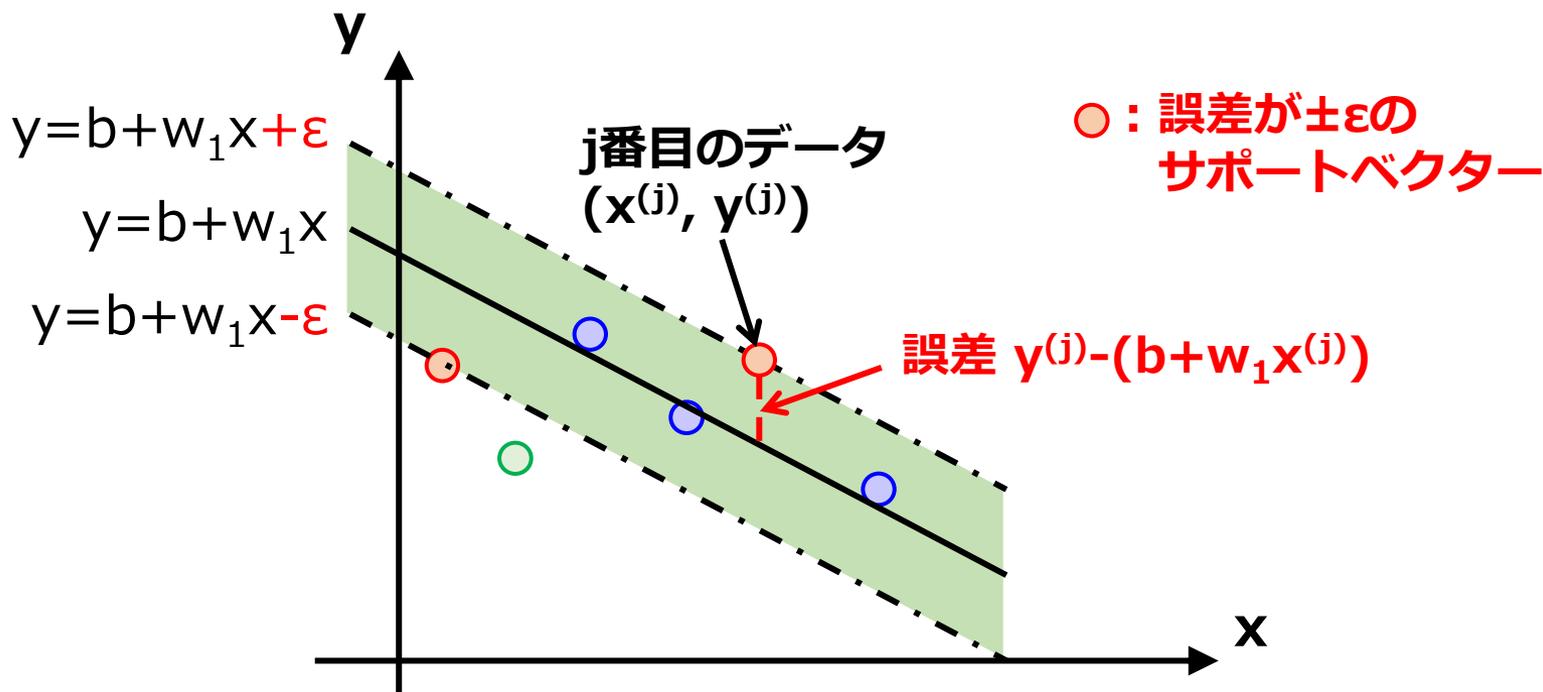


# バイアスの計算

- ✓ 双対問題にはバイアス  $b$  が存在しないため導出できない
- ✓ バイアス  $b$  は誤差が  $\pm\varepsilon$  であるサポートベクターから計算

$$b = y^{(j)} - \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) \mathbf{x}^{(i)} \mathbf{x}^{(j)T} - \varepsilon \quad \text{誤差が} +\varepsilon \text{の} (\mathbf{x}^{(j)}, y^{(j)}) \text{の場合}$$

$$b = y^{(j)} - \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) \mathbf{x}^{(i)} \mathbf{x}^{(j)T} + \varepsilon \quad \text{誤差が} -\varepsilon \text{の} (\mathbf{x}^{(j)}, y^{(j)}) \text{の場合}$$



# 非線形写像

- ✓ 入力変数  $\mathbf{x}$  を何らかの特徴空間に写像する関数  $\phi$  を考えると、双対問題・回帰関数は次式のようになる

## 双対問題

$$\max_{\alpha^+, \alpha^-} -\frac{1}{2} \sum_{i=0}^n \sum_{j=0}^n (\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-) \phi(\mathbf{x}^{(i)}) \phi(\mathbf{x}^{(j)})^T$$
$$+ \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) y^{(i)} - \sum_{i=0}^n (\alpha_i^+ + \alpha_i^-) \varepsilon$$

$$\text{s. t.} \quad \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) = 0, 0 \leq \alpha_i^+ \leq C, 0 \leq \alpha_i^- \leq C (i = 0, 1, \dots, n)$$

## 回帰関数

$$f(\mathbf{x}) = \phi(\mathbf{x}) \mathbf{w}^T + b = \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) \phi(\mathbf{x}^{(i)}) \phi(\mathbf{x})^T + b$$

- ✓ 非線形写像についてはこちら↓の最後のほう参照

<https://yuyumoyuyu.com/2020/12/20/multipleregression/>

# カーネル関数による非線形写像

- ✓ 前頁で  $\phi$  は内積  $\phi\phi^T$  の形でしか現れないため  $\phi\phi^T$  をカーネル関数  $K$  に置き換える
- ✓ 適切なカーネル関数を選ぶことで  $\phi(\mathbf{x})$  を直接計算せずとも内積を算出することが可能に

## カーネル関数の定義

$$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \phi(\mathbf{x}^{(i)})\phi(\mathbf{x}^{(j)})^T$$

## 双対問題の目的関数

$$\begin{aligned} \max_{\alpha^+, \alpha^-} \quad & -\frac{1}{2} \sum_{i=0}^n \sum_{j=0}^n (\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-) K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \\ & + \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) y^{(i)} - \sum_{i=0}^n (\alpha_i^+ + \alpha_i^-) \varepsilon \end{aligned}$$

## 回帰関数

$$f(\mathbf{x}) = \sum_{i=0}^n (\alpha_i^+ - \alpha_i^-) K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + b$$

# カーネル関数の例

- ✓ カーネル関数は例えば以下のようなものが存在
- ✓ RBFカーネルが最もよく用いられる

## RBFカーネル (ガウシアンカーネル)

$$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \exp\left(-\frac{\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|^2}{2\sigma^2}\right) = \exp\left(-\gamma\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|^2\right)$$

## 線形カーネル

$$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \mathbf{x}^{(i)} \mathbf{x}^{(j)T}$$

## 多項式カーネル

$$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \left(1 + \lambda \mathbf{x}^{(i)} \mathbf{x}^{(j)T}\right)^d$$

## シグモイドカーネル

$$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \tanh\left(a \mathbf{x}^{(i)} \mathbf{x}^{(j)T} + b\right)$$

■ : ハイパーパラメータ