

サポートベクター回帰 Support Vector Regression SVR

明治大学 理工学部 応用化学科
データ化学工学研究室 金子 弘昌

サポートベクター回帰 (SVR) とは？

- ✓ 線形の回帰分析手法
- ✓ サポートベクターマシン (SVM) を回帰分析に応用
<https://datachemeng.com/supportvectormachine/>
- ✓ 目的変数の誤差に加えて、それぞれ以下の項を最小化することで、過学習を防ぐ
 - リッジ回帰 (RR)・LASSO・Elastic Net (EN) と同じ
 - RR・LASSO・ENとの共通点はこちら
<https://datachemeng.com/rlassoen/>
- ✓ 誤差に不感帯を設けることでノイズの影響を受けにくい
- ✓ カーネルトリックにより非線形の回帰モデルに

基本的にSVRは線形の回帰分析手法

✓線形の回帰分析手法

- たとえば説明変数が2つのとき、目的変数・説明変数をオートスケーリング (<https://datachemeng.com/basicdatapreprocessing/>) したあと、

$$y = x_1 b_1 + x_2 b_2 + f$$
$$= y_C + f$$

$$(y_C = x_1 b_1 + x_2 b_2)$$

y : 目的変数

x_1, x_2 : 説明変数 (記述子)

b_1, b_2 : (標準)回帰係数

y_C : y の、 x で表すことができる部分

f : y の、 x で表すことができない部分
(誤差、残差)

と表わされる

回帰係数 \mathbf{b}

✓回帰係数のベクトル \mathbf{b} を

$$\mathbf{b} = [b_1 \quad b_2 \quad \cdots \quad b_m] \quad m : \text{説明変数(記述子)の数}$$

とする

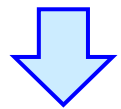
✓あるサンプル (i 番目のサンプル) の目的変数の推定値 $f(\mathbf{x}^{(i)})$ は

$$f(\mathbf{x}^{(i)}) = \mathbf{x}^{(i)} \mathbf{b} \quad \mathbf{x}^{(i)} : \text{あるサンプル} (i \text{ 番目の} \\ \text{サンプル) の説明変数 (記述子)}$$

と表わされる

非線形回帰モデルへ

線形回帰モデル (元の空間) : $f(\mathbf{x}^{(i)}) = \mathbf{x}^{(i)} \mathbf{b}$



高次元空間への写像 (非線形写像) : $\mathbf{x} \rightarrow \phi(\mathbf{x})$

非線形回帰モデル関数 (高次元空間) : $f(\mathbf{x}^{(i)}) = \phi(\mathbf{x}^{(i)}) \mathbf{w} + c$

$$\mathbf{w} = [w_1 \quad w_2 \quad \cdots \quad w_k]$$

c : 定数項

w_i : 重み

k : 高次元空間での次元数

w_i, k は、とりあえずこのように設定しておくだけで、後に考えなくてもよくなるため、気にしなくて問題ない

SVMとSVRとの比較

$$\checkmark \text{SVM} \quad \frac{1}{2} \|\mathbf{w}\|^2 \quad + \quad C \sum_{i=1}^n \xi_i \quad \text{の最小化}$$

マージンの最大化
(重みの最小化)

誤分類するサンプル数の
最小化 (誤差の最小化)

$$\checkmark \text{SVR} \quad \frac{1}{2} \|\mathbf{w}\|^2 \quad + \quad C \sum_{i=1}^n h \left(y^{(i)} - f \left(\mathbf{x}^{(i)} \right) \right) \quad \text{の最小化}$$

重みの最小化

誤差の最小化

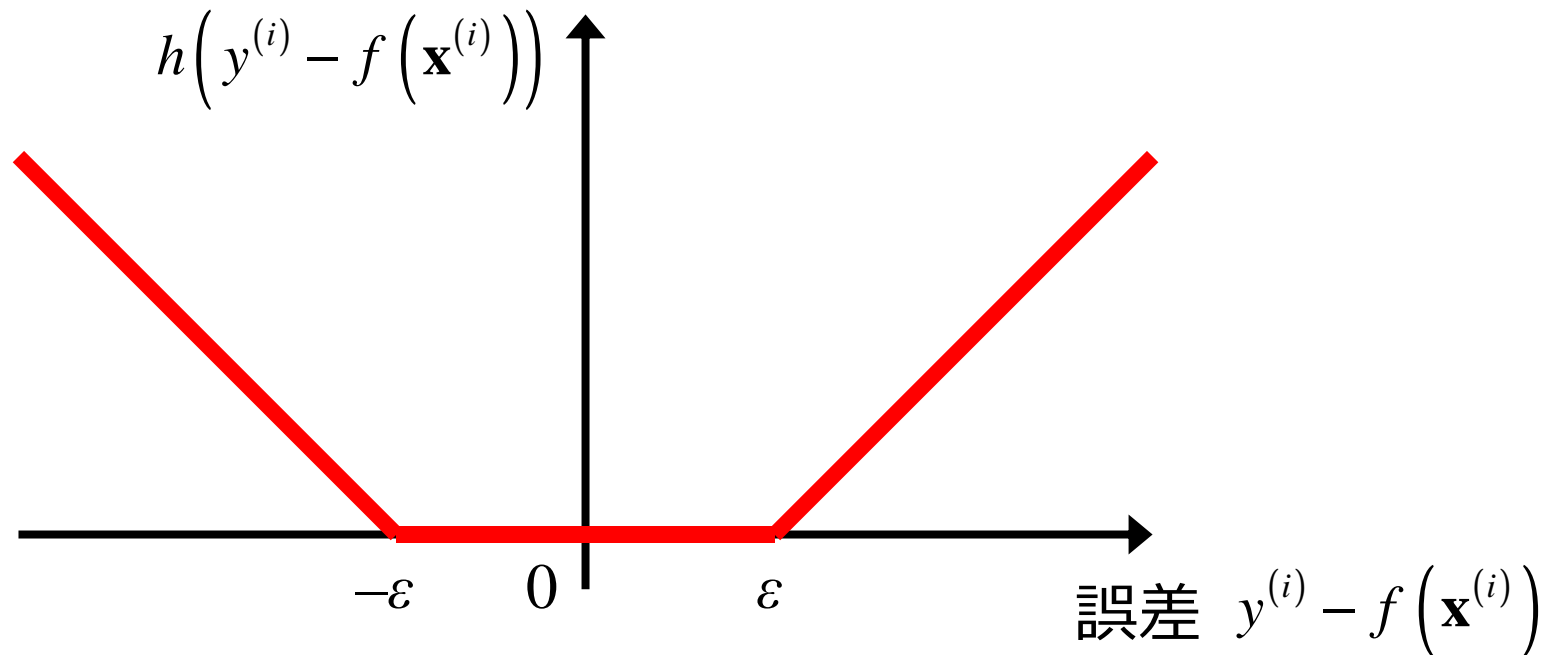
C : 2つの項のバランスを決める係数

h : 誤差関数

n : モデル構築用サンプル数

SVRの誤差関数

$$h\left(y^{(i)} - f\left(\mathbf{x}^{(i)}\right)\right) = \max\left(0, \left|y^{(i)} - f\left(\mathbf{x}^{(i)}\right)\right| - \varepsilon\right)$$



誤差の不感帯・・・ ε チューブと呼ぶ

$-\varepsilon \leq \text{誤差} \leq \varepsilon$ のとき、誤差 = 0 となる

スラック変数

✓SVMと同様にスラック変数 ξ, ξ^* を導入

$$y^{(i)} \leq f(\mathbf{x}^{(i)}) + \varepsilon + \xi_i$$

$$y^{(i)} \geq f(\mathbf{x}^{(i)}) - \varepsilon - \xi_i^*$$

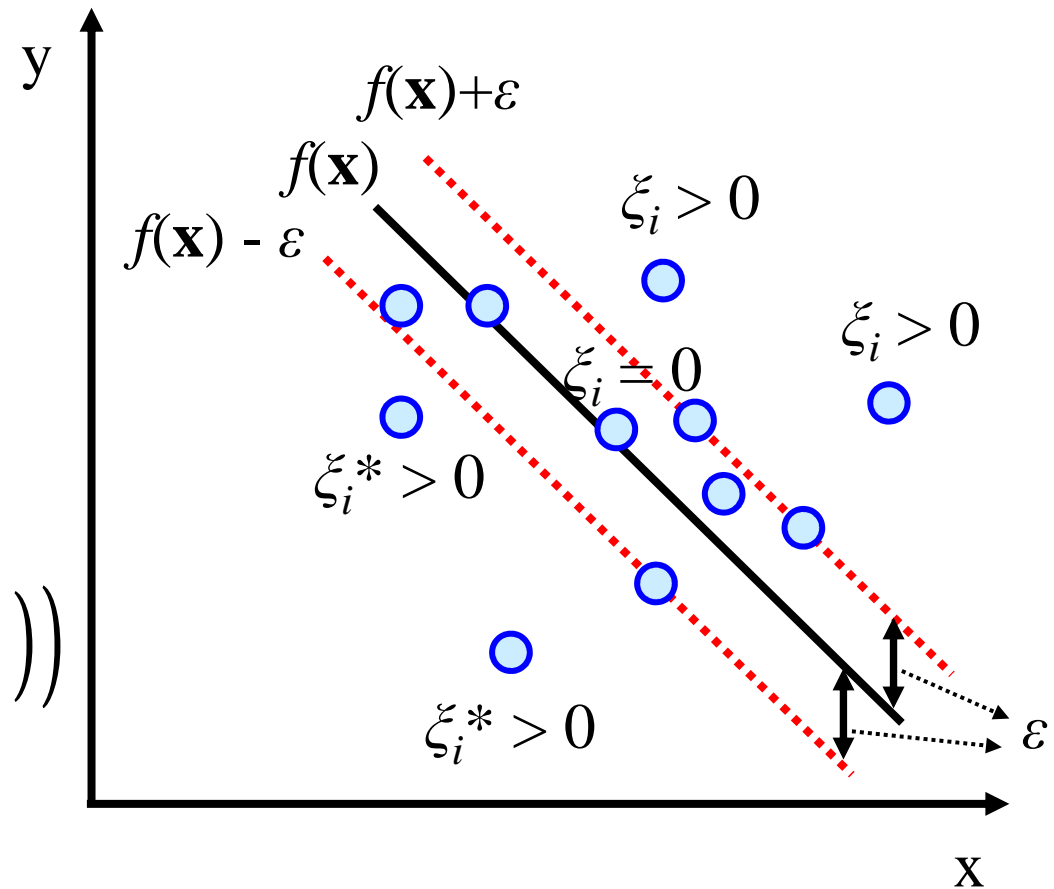
ただし、 $\xi_i \geq 0$

$$\xi_i^* \geq 0$$

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n h\left(y^{(i)} - f(\mathbf{x}^{(i)})\right)$$



$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$



RR・LASSO・EN との関係

- ✓SVRでもRR・LASSO・ENと同様にして、
誤差だけでなく重み(回帰係数)も一緒に最小化することで、
過学習を防ぐ

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad \text{の最小化}$$

重みの最小化

誤差の最小化

Lagrangeの未定乗数法

✓ラグランジュ乗数 $\alpha_i, \alpha_i^*, \beta_i, \beta_i^*$ ($i=1, 2, \dots, n$) を導入

$$G = C \sum_{i=1}^n (\xi_i + \xi_i^*) + \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n (\beta_i \xi_i + \beta_i^* \xi_i^*) \\ - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i + f(\mathbf{x}^{(i)}) - y_i) - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* - f(\mathbf{x}^{(i)}) + y_i)$$

$\mathbf{w}, c, \xi_i, \xi_i^*$ に関して G を最小化し、 $\alpha_i, \alpha_i^*, \beta_i, \beta_i^*$ に関して G を最大化



$\mathbf{w}, c, \xi_i, \xi_i^*$ に関して G が極小



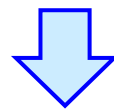
G を $\mathbf{w}, c, \xi_i, \xi_i^*$ それぞれで偏微分して 0 とする

G を \mathbf{w} で偏微分して0 $\mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(\mathbf{x}^{(i)})^T$

G を c で偏微分して0 $\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$

G を ξ_i で偏微分して0 $\alpha_i + \beta_i = C \quad (i = 1, 2, \dots, n)$

G を ξ_i^* で偏微分して0 $\alpha_i^* + \beta_i^* = C \quad (i = 1, 2, \dots, n)$



これらを使って
 G を変形すると...

$$G = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$$
$$-\varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i$$

K : カーネル関数 $K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \phi(\mathbf{x}^{(i)}) \phi(\mathbf{x}^{(j)})^T$

カーネルトリックについてはこちら

<https://datachemeng.com/supportvectormachine/>

カーネル関数の例

✓ 線形カーネル

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

✓ ガウシアンカーネル (使われることが多い)

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

✓ 多項式カーネル

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

a を求める

$$G = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$$

$$- \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i$$

$\alpha_i, \alpha_i^*, \beta_i, \beta_i^*$ は Lagrange 定数 $\Rightarrow \alpha_i \geq 0, \alpha_i^* \geq 0, \beta_i \geq 0, \beta_i^* \geq 0$

p.10より $\alpha_i + \beta_i = C \quad (i = 1, 2, \dots, n) \Rightarrow \alpha_i \leq C, \alpha_i^* \leq C$

$$\alpha_i^* + \beta_i^* = C \quad (i = 1, 2, \dots, n)$$

$$G = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \\ - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i$$

制約 $\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$, $0 \leq \alpha_i \leq C$, $0 \leq \alpha_i^* \leq C$ のもとで、

G を α_i, α_i^* に対して最大化する二次計画問題を解くと、
 α_i, α_i^* が求まる

$$f(\mathbf{x}^{(j)}) = \phi(\mathbf{x}^{(j)})\mathbf{w} + c \quad \mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(\mathbf{x}^{(i)})^T \quad (\text{p.10})$$

$$\Rightarrow f(\mathbf{x}^{(j)}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(\mathbf{x}^{(j)}) \phi(\mathbf{x}^{(i)})^T + c$$

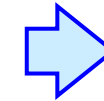
$$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \phi(\mathbf{x}^{(i)}) \phi(\mathbf{x}^{(j)})^T \quad \text{より、}$$

$$f(\mathbf{x}^{(j)}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}^{(j)}, \mathbf{x}^{(i)}) + c$$

・・・ α_i, α_i^* は求まっているため、SVRの回帰式も求まる

サポートベクターとは

ε チューブ内 (誤差の絶対値が ε 未満) のサンプル



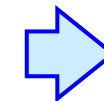
$$\alpha_i - \alpha_i^* = 0$$



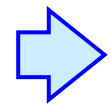
SVRの回帰式に寄与しない

ε チューブ上 (誤差の絶対値が ε) のサンプル

ε チューブ外 (誤差の絶対値が ε 以上) のサンプル



$$\alpha_i - \alpha_i^* \neq 0$$



これらのサンプルでSVRの回帰式がつけられる



サポートベクター

c の計算

$$f(\mathbf{x}^{(j)}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}^{(j)}, \mathbf{x}^{(i)}) + c$$

ε チューブ上 (誤差の絶対値が ε) のサンプル... 誤差 = $\pm\varepsilon$

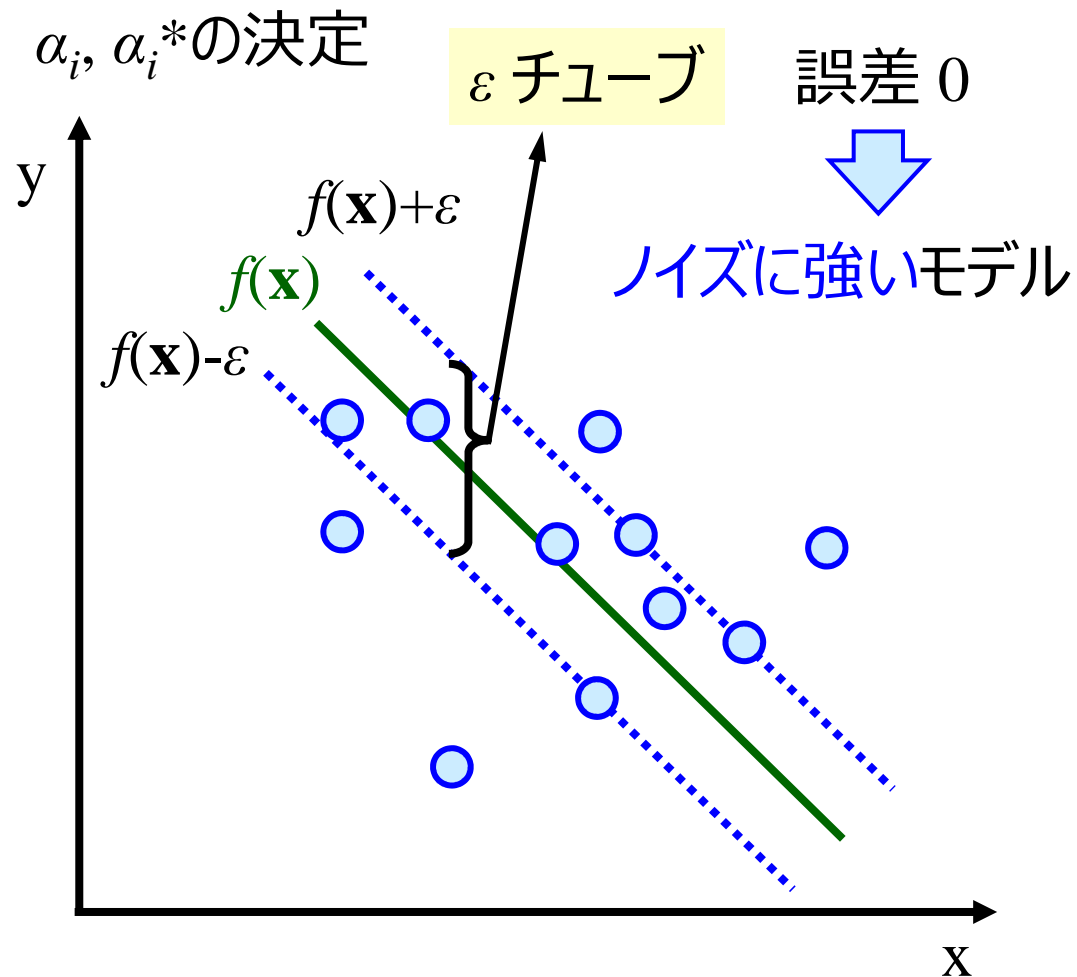
$$\text{上側} \quad c = y^{(i)} - \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}^{(j)}, \mathbf{x}^{(i)}) - \varepsilon$$

$$\text{下側} \quad c = y^{(i)} - \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}^{(j)}, \mathbf{x}^{(i)}) + \varepsilon$$

ε チューブ上のサンプルで c を計算し、それらの平均値を最終的な c として使用

SVRのまとめ・特徴

$$\text{SVRの回帰式} \quad f(\mathbf{x}^{(j)}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}^{(j)}, \mathbf{x}^{(i)}) + c$$



α_i, α_i^* の範囲

$$0 \leq \alpha_i \leq C$$

$$0 \leq \alpha_i^* \leq C$$

モデルの複雑度を調整

カーネル関数 K

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

非線形回帰モデル

C, ε , γ の決め方

✓ グリッドサーチによって、クロスバリデーションの後の r^2 の値がもっとも高い C , ε , γ (ガウシアンカーネル) の組み合わせとする

- C の候補の例: $2^{-5}, 2^{-4}, \dots, 2^9, 2^{10}$
- ε の候補の例: $2^{-10}, 2^{-9}, \dots, 2^{-1}, 2^0$
- γ の候補の例: $2^{-20}, 2^{-19}, \dots, 2^9, 2^{10}$

✓ C , ε , γ を高速に最適化する手法については、こちら
<https://datachemeng.com/fastoptsvrhyperparams/>