サポートベクターマシン

中田和秀

東京工業大学 工学院 経営工学系

機械学習入門

http://www.iee.e.titech.ac.jp/~nakatalab/text/lecture.html

概要

ここではサポートベクトルマシン (SVM) について説明を行う。この手法はカーネルトリックを使うことにより、線形的な計算で非線形な判別や回帰ができるという興味深い特徴を持っている。

目次：

記号の使い方：

- $A := B$ は、B で A を定義する、B を A に代入することを意味する
- $[n]$ は n までのインデックスの集合を表し $[n] := \{1, 2, \cdots, n\}$

サポートベクターマシン

- カーネルトリックによって、線形的な計算で非線形的な判別や回帰が可能
- 講義では 2 クラスの判別問題、回帰の順で扱う。

データ： $\{(\boldsymbol{x}_d, y_d)\}_{d\in[D]}$,　　$\boldsymbol{x}_d \in \mathbb{R}^n$, $y_d \in \{1, -1\}$

※式がきれいな形になるため、y_d は ± 1 で考える

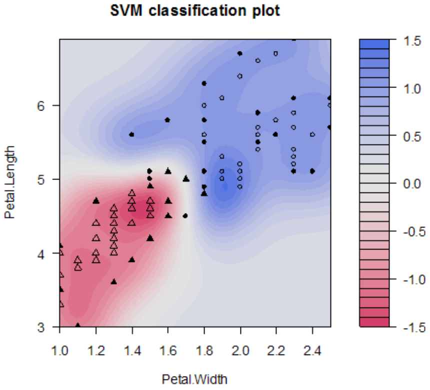


Figure: https://www.cis.doshisha.ac.jp/ mjin/R/31/31.html

線形判別

$\boldsymbol{x}_d$ と超平面 $\{\boldsymbol{x} \mid \boldsymbol{w}^T\boldsymbol{x}+b=0\}$ との距離：

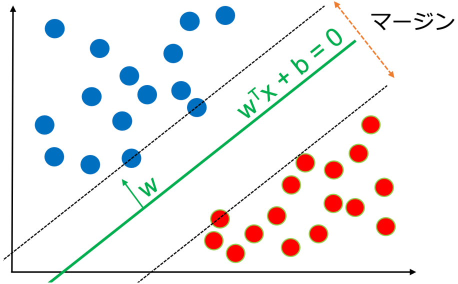
$$\frac{|\boldsymbol{w}^T\boldsymbol{x}_d+b|}{\|\boldsymbol{w}\|}$$

説明：
$\boldsymbol{x}_d$ の超平面上への射影点を $\widetilde{\boldsymbol{x}_d}$ とする

$$\begin{cases} \boldsymbol{w}^T\widetilde{\boldsymbol{x}_d}+b=0 \\ \boldsymbol{x}_d-\widetilde{\boldsymbol{x}_d}=\alpha\dfrac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \end{cases}$$

これを解くと

$$\alpha=\frac{\boldsymbol{w}^T\boldsymbol{x}_d+b}{\|\boldsymbol{w}\|}$$



全てのデータ点に対する最小距離（の２倍）をマージンと呼ぶ

ハードマージン SVM

マージンを最大化する超平面を求める。

$$\max_{\boldsymbol{w},b} \min_{d\in[D]} \frac{y_d(\boldsymbol{w}^T\boldsymbol{x}_d + b)}{\|\boldsymbol{w}\|}$$

これは次の連続最適化問題 [1] と等価（理由は次のスライド）

$$\begin{aligned} \min_{\boldsymbol{w},b} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \\ \text{s.t.} \quad & y_d(\boldsymbol{w}^T\boldsymbol{x}_d + b) \geq 1 \qquad (d \in [D]). \end{aligned}$$

最適解を $\boldsymbol{w}^*, b^*$ とする.

判別法： 超平面 $\{\boldsymbol{x} \mid \boldsymbol{w}^{*T}\boldsymbol{x} + \boldsymbol{b}^* = 0\}$ の右か左かで判別

$$y = \begin{cases} 1 & (\boldsymbol{w}^{*T}\boldsymbol{x} + b^* \geq 0) \\ -1 & (\boldsymbol{w}^{*T}\boldsymbol{x} + b^* \leq 0) \end{cases}$$

[1] s.t. は subject to の略で制約条件を意味する。

等価であることの説明

$$\max_{\boldsymbol{w},b} \min_{d \in [D]} \frac{y_d(\boldsymbol{w}^T \boldsymbol{x}_d + b)}{\|\boldsymbol{w}\|}$$

$$\iff \left\{ \begin{array}{ll} \displaystyle\max_{\boldsymbol{w},b,t} & t \\ \text{s.t.} & \displaystyle\frac{y_d(\boldsymbol{w}^T \boldsymbol{x}_d + b)}{\|\boldsymbol{w}\|} \geq t \qquad (d \in [D]). \end{array} \right.$$

$$\iff \left\{ \begin{array}{ll} \displaystyle\max_{\boldsymbol{w},b,t} & t \\ \text{s.t.} & \displaystyle\frac{y_d(\boldsymbol{w}^T \boldsymbol{x}_d + b)}{\|\boldsymbol{w}\|} \geq t \qquad (d \in [D]), \\ & \|\boldsymbol{w}\| t = 1. \end{array} \right.$$

$$\iff \left\{ \begin{array}{ll} \displaystyle\max_{\boldsymbol{w},b} & 1/\|\boldsymbol{w}\| \\ \text{s.t.} & y_d(\boldsymbol{w}^T \boldsymbol{x}_d + b) \geq 1 \qquad (d \in [D]). \end{array} \right.$$

$$\iff \left\{ \begin{array}{ll} \displaystyle\min_{\boldsymbol{w},b} & \displaystyle\frac{1}{2}\boldsymbol{w}^T \boldsymbol{w} \\ \text{s.t.} & y_d(\boldsymbol{w}^T \boldsymbol{x}_d + b) \geq 1 \qquad (d \in [D]). \end{array} \right.$$

行列を使った表現

$$
\begin{aligned}
\min_{\boldsymbol{w},b} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \\
\text{s.t.} \quad & y_d(\boldsymbol{w}^T\boldsymbol{x}_d + b) \geq 1 \qquad (d \in [D]).
\end{aligned}
$$

$$
\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_D^T \end{pmatrix} \in \mathbb{R}^{D\times n}, \quad \boldsymbol{e} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^{D}, \quad \boldsymbol{Y} = \begin{pmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_D \end{pmatrix} \in \mathbb{R}^{D\times D}.
$$

Hard-SVM 主問題

$$
\begin{aligned}
\min_{\boldsymbol{w},b} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \\
\text{s.t.} \quad & \boldsymbol{Y}(\boldsymbol{X}\boldsymbol{w} + \boldsymbol{e}b) \geq \boldsymbol{e}.
\end{aligned}
$$

双対問題

この最適化問題は凸２次計画問題

- 目的関数は凸２次関数
- 制約条件は線形不等式

Hard-SVM 双対問題

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}} \quad & -\frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{Y}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha} + \boldsymbol{e}^T\boldsymbol{\alpha} \\
\text{s.t.} \quad & \boldsymbol{y}^T\boldsymbol{\alpha} = 0, \\
& \boldsymbol{\alpha} \geq \boldsymbol{0}.
\end{aligned}
$$

双対問題

- 双対問題の双対問題は元の問題（主問題）
- 主問題と双対問題の最適値は等しい
- 一方の問題の最適解からもう一方の問題の最適解を構築可能

双対問題の導出

ラグランジュ関数： $L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \boldsymbol{\alpha}^T(\boldsymbol{e} - \boldsymbol{Y}(\boldsymbol{X}\boldsymbol{w} + \boldsymbol{e}b))$

ラグランジュ双対の内側の最適化問題： $\min_{\boldsymbol{w}, b} L(\boldsymbol{w}, b, \boldsymbol{\alpha})$

制約なしの凸最適化問題なので、最適解の必要十分条件は次のもの。

$$\nabla_{\boldsymbol{w}} L(\boldsymbol{w}^*, b^*, \boldsymbol{\alpha}) = \boldsymbol{w}^* - \boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha} = \boldsymbol{0},$$
$$\nabla_b L(\boldsymbol{w}^*, b^*, \boldsymbol{\alpha}) = \boldsymbol{\alpha}^T\boldsymbol{Y}\boldsymbol{e} = 0.$$

よって、$\boldsymbol{w}^* = \boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha}$ だと分かる。これを代入すると

$$L(\boldsymbol{w}^*, b^*, \boldsymbol{\alpha}) = -\frac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{Y}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T\boldsymbol{e}$$

$\boldsymbol{\alpha}^T\boldsymbol{Y}\boldsymbol{e} = 0$ は $\boldsymbol{y}^T\boldsymbol{\alpha} = 0$ であるため、双対問題は

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & -\frac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{Y}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha} + \boldsymbol{e}^T\boldsymbol{\alpha} \\ \text{s.t.} \quad & \boldsymbol{y}^T\boldsymbol{\alpha} = 0, \\ & \boldsymbol{\alpha} \geq \boldsymbol{0}. \end{aligned}$$

ソフトマージン SVM

- ハードマージン SVM では、線形分離可能でないと解が存在しない。
- ペナルティをつけて、線形分離可能でないデータについても判別を行う

ハードマージン SVM：

$$
\begin{aligned}
\min_{\boldsymbol{w},b} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} \\
\text{s.t.} \quad & y_d(\boldsymbol{w}^T\boldsymbol{x}_d + b) \geq 1 \qquad (d \in [D]).
\end{aligned}
$$

ソフトマージン SVM：

$$
\begin{aligned}
\min_{\boldsymbol{w},b,\boldsymbol{s}} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{d\in[D]} s_d \\
\text{s.t.} \quad & y_d(\boldsymbol{w}^T\boldsymbol{x}_d + b) \geq 1 - s_d \qquad (d \in [D]), \\
& s_d \geq 0 \qquad (d \in [D]).
\end{aligned}
$$

$C > 0$ はペナルティの重みを意味する定数

ソフトマージン SVM の解釈

s_d $(d \in [D])$ に関する制約条件を整理すると

$$
\begin{aligned}
s_d &\geq 1 - y_d(\boldsymbol{w}^T \boldsymbol{x}_d + b) \\
s_d &\geq 0
\end{aligned}
$$

最適化問題では，s_d の和を小さくするため

$$
s_d^* = \max\{1 - y_d(\boldsymbol{w}^T \boldsymbol{x}_d + b),\ 0\}
$$

よって、次の最適化問題となる。

$$
\min_{\boldsymbol{w}, b} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C \sum_{d \in [D]} \max\{1 - y_d(\boldsymbol{w}^T \boldsymbol{x}_d + b),\ 0\}
$$

$L(y, \hat{y}) = \max\{1 - y_d \hat{y}_d,\ 0\}$ とすると，

正則化項つき学習とみなせる

教師あり学習のフレームワークとの関係

予測器

パラメタは $\boldsymbol{w}, b$

$$\hat{y} = \boldsymbol{w}^T \boldsymbol{x} + b$$

誤差関数

ヒンジ損失

$$L(y, \hat{y}) := \max\{1 - y\hat{y}, 0\}$$

$y \in \{-1, 1\}$：目標，　$\hat{y} \in \mathbb{R}$：予測値

正則化項つき学習

$$\min_{\boldsymbol{w}, b} \sum_{d \in [D]} L(y_d, \boldsymbol{w}^T \boldsymbol{x}_d + b) + \frac{1}{2C} \|\boldsymbol{w}\|^2$$

ソフトマージン SVM の行列表現

ソフトマージン SVM：

$$\begin{aligned} \min_{\boldsymbol{w},b,\boldsymbol{s}} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{d\in[D]} s_d \\ \text{s.t.} \quad & y_d(\boldsymbol{w}^T\boldsymbol{x}_d + b) \geq 1 - s_d \qquad (d\in[D]), \\ & s_d \geq 0 \qquad (d\in[D]). \end{aligned}$$

Soft-SVM 主問題

$$\begin{aligned} \min_{\boldsymbol{w},b,\boldsymbol{s}} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}\underline{+C\boldsymbol{e}^T\boldsymbol{s}} \\ \text{s.t.} \quad & \boldsymbol{Y}(\boldsymbol{X}\boldsymbol{w} + \boldsymbol{e}b) \geq \boldsymbol{e}\underline{-\boldsymbol{s}}, \\ & \underline{\boldsymbol{s} \geq \boldsymbol{0}}. \end{aligned}$$

赤字（下線）の部分がハードマージン SVM との違い

双対問題

双対問題を作ると次のようになる。

Soft-SVM 双対問題

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}} \quad & -\frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{Y}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha} + \boldsymbol{e}^T\boldsymbol{\alpha} \\
\text{s.t.} \quad & \boldsymbol{y}^T\boldsymbol{\alpha} = 0, \\
& \mathbf{0} \le \boldsymbol{\alpha} \underline{\le C\boldsymbol{e}}.
\end{aligned}
$$

赤字（下線）の部分がハードマージン SVM の双対問題との違い

双対問題の導出１

ラグランジュ関数：

$$L(\boldsymbol{w}, b, \boldsymbol{s}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\boldsymbol{e}^T\boldsymbol{s} + \boldsymbol{\alpha}^T(\boldsymbol{e} - \boldsymbol{s} - \boldsymbol{Y}(\boldsymbol{X}\boldsymbol{w} + \boldsymbol{e}b)) - \boldsymbol{\beta}^T\boldsymbol{s}$$

ラグランジュ双対の内側の最適化問題： $\min_{\boldsymbol{w},b,\boldsymbol{s}} L(\boldsymbol{w}, b, \boldsymbol{s}, \boldsymbol{\alpha}, \boldsymbol{\beta})$

制約なしの凸最適化問題なので、最適解の必要十分条件は次のもの。

$$\begin{aligned}
\nabla_{\boldsymbol{w}} L(\boldsymbol{w}^*, b^*, \boldsymbol{s}^*, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \boldsymbol{w}^* - \boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha} = \boldsymbol{0} \\
\nabla_b L(\boldsymbol{w}^*, b^*, \boldsymbol{s}^*, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \boldsymbol{\alpha}^T\boldsymbol{Y}\boldsymbol{e} = 0 \\
\nabla_{\boldsymbol{s}} L(\boldsymbol{w}^*, b^*, \boldsymbol{s}^*, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= C\boldsymbol{e} - \boldsymbol{\alpha} - \boldsymbol{\beta} = \boldsymbol{0}
\end{aligned}$$

より、

$$\begin{aligned}
&\boldsymbol{w}^* = \boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha} \\
&\boldsymbol{\alpha}^T\boldsymbol{Y}\boldsymbol{e} = 0 \\
&\boldsymbol{\alpha} + \boldsymbol{\beta} = C\boldsymbol{e}
\end{aligned}$$

が得られる。

双対問題の導出 2

これを、ラグランジュ関数に代入すると、

$$L(\boldsymbol{w}^*, b^*, \boldsymbol{s}^*, \boldsymbol{\alpha}, \boldsymbol{\beta}) = -\frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{Y}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T\boldsymbol{e}$$

である。$\boldsymbol{\alpha}^T\boldsymbol{Y}\boldsymbol{e} = 0$ は $\boldsymbol{y}^T\boldsymbol{\alpha} = 0$ であるため、双対問題は次のようになる。

$$\begin{array}{ll} \max & -\frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{Y}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha} + \boldsymbol{e}^T\boldsymbol{\alpha} \\ \text{s.t.} & \boldsymbol{y}^T\boldsymbol{\alpha} = 0, \\ & \boldsymbol{\alpha} + \boldsymbol{\beta} = C\boldsymbol{e}, \\ & \boldsymbol{\alpha}, \boldsymbol{\beta} \geq \mathbf{0}, \end{array}$$

ここで、$\boldsymbol{\alpha} + \boldsymbol{\beta} = C\boldsymbol{e},\ \boldsymbol{\alpha}, \boldsymbol{\beta} \geq \mathbf{0}$ は $\mathbf{0} \leq \boldsymbol{\alpha} \leq C\boldsymbol{e}$ と書き換えることができる

高次元空間への写像

- 入力空間を高次元の特徴空間に写像して、その特徴空間上で線形判別
- 写像 $\phi : \mathbb{R}^n \to V$ を使って、$\boldsymbol{x}_d$ の代わりに $\phi(\boldsymbol{x}_d)$ に対して線形判別を行う。
- 元の空間 $\mathbb{R}^n$ でみると非線形な判別となる

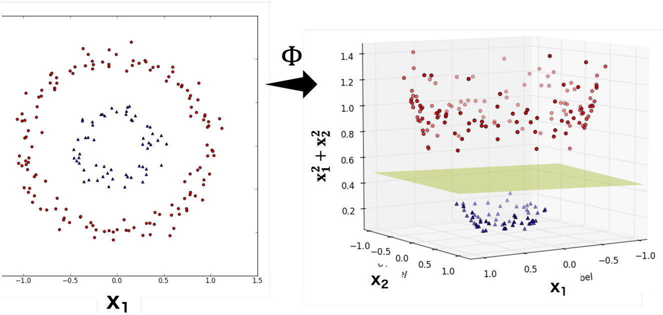


Figure: https://axa.biopapyrus.jp/machine-learning/svm/kernel-svm.html

Soft-SVM

$$\boldsymbol{X} := \begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_D^T \end{pmatrix} \quad \text{の代わりに} \quad \widetilde{\boldsymbol{X}} := \begin{pmatrix} \phi(\boldsymbol{x}_1)^T \\ \phi(\boldsymbol{x}_2)^T \\ \vdots \\ \phi(\boldsymbol{x}_D)^T \end{pmatrix} \quad \text{を使う}$$

Soft-SVM

主問題

$$\begin{aligned} \min_{\boldsymbol{w}, b, \boldsymbol{s}} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\boldsymbol{e}^T\boldsymbol{s} \\ \text{s.t.} \quad & \boldsymbol{Y}(\widetilde{\boldsymbol{X}}\boldsymbol{w} + \boldsymbol{e}b) \geq \boldsymbol{e} - \boldsymbol{s}, \\ & \boldsymbol{s} \geq \boldsymbol{0}. \end{aligned}$$

双対問題

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & -\frac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{Y}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T\boldsymbol{Y}\boldsymbol{\alpha} + \boldsymbol{e}^T\boldsymbol{\alpha} \\ \text{s.t.} \quad & \boldsymbol{y}^T\boldsymbol{\alpha} = 0, \\ & \boldsymbol{0} \leq \boldsymbol{\alpha} \leq C\boldsymbol{e}. \end{aligned}$$

カーネルトリック

- 行列 $\boldsymbol{K} := \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T$ は半正定値
- $K_{ij} := \boldsymbol{\phi}(\boldsymbol{x}_i)^T \boldsymbol{\phi}(\boldsymbol{x}_j)$
- K_{ij} さえ定まれば、$\boldsymbol{\phi}(\boldsymbol{x})$ を陽に計算しなくてよい

カーネル関数 Ker : $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ を先に決め、次の計算で $\boldsymbol{K}$ を構成する。

$$K_{ij} := \mathsf{Ker}(\boldsymbol{x}_i, \boldsymbol{x}_j) \quad (i, j \in [D])$$

マーサーの条件

任意の $i, j \in [D]$ に対し、$K_{ij} := \mathsf{Ker}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ で作られる カーネル行列 $\boldsymbol{K} \in \mathbb{R}^{D \times D}$ が半正定値となるとき、写像 ϕ の存在が保証される

カーネルの例

- $\mathsf{Ker}(\boldsymbol{x}, \boldsymbol{x}') := \boldsymbol{x}^T \boldsymbol{x}'$ 線形カーネル
- $\mathsf{Ker}(\boldsymbol{x}, \boldsymbol{x}') := (\boldsymbol{x}^T \boldsymbol{x}' + 1)^d \qquad (d \in \mathbb{N})$ 多項式カーネル
- $\mathsf{Ker}(\boldsymbol{x}, \boldsymbol{x}') := \exp\{-\gamma \|\boldsymbol{x} - \boldsymbol{x}'\|^2\}$ RBF カーネル、ガウシアンカーネル

カーネルの証明

線形カーネル：

$K_{ij} := \boldsymbol{x}_i^T \boldsymbol{x}_j$ のとき、$\boldsymbol{K} = \boldsymbol{X}\boldsymbol{X}^T$ より半正定値行列

多項式カーネル：

$K_{ij} := (\boldsymbol{x}_i^T \boldsymbol{x}_j + 1)^d$ から作られるカーネル行列を $\boldsymbol{K}_d$ とする。

- $\boldsymbol{K}_1 = \boldsymbol{X}\boldsymbol{X}^T + \boldsymbol{e}\boldsymbol{e}^T$ より半正定値行列
- $\boldsymbol{K}_d, \boldsymbol{K}_1$ が半正定値のとき，$\boldsymbol{K}_{d+1} = \boldsymbol{K}_d \circ \boldsymbol{K}_1$ より，$\boldsymbol{K}_{d+1}$ も半正定値

（証明は次スライド）

$\boldsymbol{C} = \boldsymbol{A} \circ \boldsymbol{B}$ は Hadamard（アダマール）積を意味し $C_{ij} := A_{ij} B_{ij}$

RBF カーネル：

証明は無限級数を考えないといけないので難しい

Hadamard 積の半正定値性

$\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ が半正定値のとき、$\boldsymbol{C} := \boldsymbol{A} \circ \boldsymbol{B} \in \mathbb{R}^{n \times n}$ も半正定値

Hadamard 積 $\quad C_{ij} := A_{ij} B_{ij} \quad (i, j \in [n])$

証明：
$\boldsymbol{X} = \boldsymbol{X}^{1/2} \boldsymbol{X}^{1/2}$, $\boldsymbol{Y} = \boldsymbol{Y}^{1/2} \boldsymbol{Y}^{1/2}$ なので、$\boldsymbol{X}^{1/2}, \boldsymbol{Y}^{1/2}$ の i 列目を $\boldsymbol{x}_i, \boldsymbol{y}_i$ とする。

$$\begin{aligned}
\boldsymbol{v}^T (\boldsymbol{X} \circ \boldsymbol{Y}) \boldsymbol{v} &= \sum_{i \in [n]} \sum_{j \in [n]} v_i v_j (\boldsymbol{x}_i^T \boldsymbol{x}_j)(\boldsymbol{y}_i^T \boldsymbol{y}_j) = \sum_{i \in [n]} \sum_{j \in [n]} v_i v_j (\boldsymbol{x}_i^T \boldsymbol{x}_j)(\boldsymbol{y}_j^T \boldsymbol{y}_i) \\
&= \sum_{i \in [n]} \sum_{j \in [n]} v_i v_j \mathsf{Tr}(\boldsymbol{x}_i^T \boldsymbol{x}_j \boldsymbol{y}_j^T \boldsymbol{y}_i) = \sum_{i \in [n]} \sum_{j \in [n]} v_i v_j \mathsf{Tr}(\boldsymbol{y}_i \boldsymbol{x}_i^T \boldsymbol{x}_j \boldsymbol{y}_j^T) \\
&= \mathsf{Tr} \left(\sum_{i \in [n]} \sum_{j \in [n]} v_i v_j \boldsymbol{y}_i \boldsymbol{x}_i^T \boldsymbol{x}_j \boldsymbol{y}_j^T \right) = \mathsf{Tr} \left(\left(\sum_{i \in [n]} v_i \boldsymbol{y}_i \boldsymbol{x}_i^T \right) \left(\sum_{j \in [n]} v_j \boldsymbol{x}_j \boldsymbol{y}_j^T \right) \right) \\
&= \mathsf{Tr} \left(\boldsymbol{M} \boldsymbol{M}^T \right) = \sum_{i \in [n]} \sum_{j \in [n]} (M_{ij})^2 \geq 0.
\end{aligned}$$

写像 ϕ の例

$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$ に対する 3 次の多項式カーネル $\mathsf{Ker}(\boldsymbol{x}, \boldsymbol{y}) := (\boldsymbol{x}^T \boldsymbol{y} + 1)^3$ を考える。

$$\begin{aligned}
\mathsf{Ker}(\boldsymbol{x}, \boldsymbol{y}) &:= (\boldsymbol{x}^T \boldsymbol{y} + 1)^3 \\
&= (x_1 y_1 + x_2 y_2 + 1)^3 \\
&= x_1^3 y_1^3 + 3x_1^2 y_1^2 x_2 y_2 + 3x_1 y_1 x_2^2 y_2^2 + x_2^3 y_2^3 \\
&\quad + 3x_1^2 y_1^2 + 6x_1 y_1 x_2 y_2 + 3x_2^2 y_2^2 + 3x_1 y_1 + 3x_2 y_2 + 1
\end{aligned}$$

よって、$\boldsymbol{\phi}(\boldsymbol{x}) : \mathbb{R}^2 \to \mathbb{R}^{10}$ を

$$\boldsymbol{\phi}(\boldsymbol{x}) := \left(x_1^3,\ \sqrt{3}x_1^2 x_2,\ \sqrt{3}x_1 x_2^2,\ x_2^3,\ \sqrt{3}x_1^2,\ \sqrt{6}x_1 x_2,\ \sqrt{3}x_2^2,\ \sqrt{3}x_1,\ \sqrt{3}x_2,\ 1\right)^T$$

と定めると、次の関係が成り立つ。

$$\boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{\phi}(\boldsymbol{y}) = (\boldsymbol{x}^T \boldsymbol{y} + 1)^3 = \mathsf{Ker}(\boldsymbol{x}, \boldsymbol{y})$$

$\boldsymbol{\phi}(\boldsymbol{x}) \in \mathbb{R}^{10}$ に対する線形な判別は、$\boldsymbol{x} \in \mathbb{R}^2$ に対する非線形な判別（3 次多項式）

カーネル付き双対問題

Soft-SVM 双対問題（カーネル付き）

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & -\frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{Y} \boldsymbol{K} \boldsymbol{Y} \boldsymbol{\alpha} + \boldsymbol{e}^T \boldsymbol{\alpha} \\ \text{s.t.} \quad & \boldsymbol{y}^T \boldsymbol{\alpha} = 0, \\ & \mathbf{0} \le \boldsymbol{\alpha} \le C\boldsymbol{e}. \end{aligned}$$

ただし、$K_{ij} = \mathsf{Ker}(\boldsymbol{x}_i, \boldsymbol{x}_j) \quad (i, j \in [D])$

- 最小化問題にすると目的関数は凸２次関数
- 制約条件は線形の不等式条件
- 凸２次計画問題

- 汎用解法：　有効制約法や内点法
- SVM 用の解法：　SMO(Sequential Minimal Optimization)　　後述する

判別方法

予測器

$\boldsymbol{x}$ の判別：

$$y = \begin{cases} 1 & (\boldsymbol{w}^{*T}\boldsymbol{\phi}(\boldsymbol{x}) + b^* \geq 0) \\ -1 & (\boldsymbol{w}^{*T}\boldsymbol{\phi}(\boldsymbol{x}) + b^* \leq 0) \end{cases}$$

ただし、

$$(\boldsymbol{w}^*)^T\boldsymbol{\phi}(\boldsymbol{x}) = \sum_{d\in[D]} y_d \alpha_d^* \mathsf{Ker}(\boldsymbol{x}_d, \boldsymbol{x})$$

$$b^* = y_j - \sum_{d\in[D]} y_d \alpha_d^* \mathsf{Ker}(\boldsymbol{x}_d, \boldsymbol{x}_j) \qquad (0 < \alpha_j^* < C \text{ に対して})$$

説明

$$(\boldsymbol{w}^*)^T\boldsymbol{\phi}(\boldsymbol{x}) = (\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{\alpha}^*)^T\boldsymbol{\phi}(\boldsymbol{x}) = \left(\sum_{d\in[D]}\boldsymbol{\phi}(\boldsymbol{x}_d)y_d\alpha_d^*\right)^T\boldsymbol{\phi}(\boldsymbol{x})$$
$$= \sum_{d\in[D]} y_d\alpha_d^*\boldsymbol{\phi}(\boldsymbol{x}_d)^T\boldsymbol{\phi}(\boldsymbol{x}) = \sum_{d\in[D]} y_d\alpha_d^*\mathsf{Ker}(\boldsymbol{x}_d,\boldsymbol{x})$$

最適性条件

$$\boldsymbol{\alpha}^T\nabla_{\boldsymbol{\alpha}}L(\boldsymbol{w},b,\boldsymbol{s},\boldsymbol{\alpha},\boldsymbol{\beta}) = \boldsymbol{\alpha}^T(\boldsymbol{e}-\boldsymbol{s}-\boldsymbol{Y}(\boldsymbol{X}\boldsymbol{w}+\boldsymbol{e}b)) = 0$$
$$\iff \quad \alpha_d(1-s_d-y_d(\boldsymbol{\phi}(\boldsymbol{x}_d)^T\boldsymbol{w}+b)) = 0 \qquad (d\in[D])$$
$$\boldsymbol{\beta}^T\nabla_{\boldsymbol{\beta}}L(\boldsymbol{w},b,\boldsymbol{s},\boldsymbol{\alpha},\boldsymbol{\beta}) = \boldsymbol{\beta}^T\boldsymbol{s} = 0$$
$$\iff \quad \beta_d s_d = 0 \qquad (d\in[D])$$

$0<\alpha_j^*<C$ のとき、$\alpha_j^*,\beta_j^*>0$ である。

このとき、$1-s_j-y_j(\boldsymbol{\phi}(\boldsymbol{x}_j)^T\boldsymbol{w}^*+b^*)=0,\ \ s_j=0$ より，

$$b^* = y_j - \boldsymbol{\phi}(\boldsymbol{x}_j)^T\boldsymbol{w}^* = y_j - \sum_{d\in[D]} y_d\alpha_d^*\mathsf{Ker}(\boldsymbol{x}_d,\boldsymbol{x}_j)$$

サポートベクター

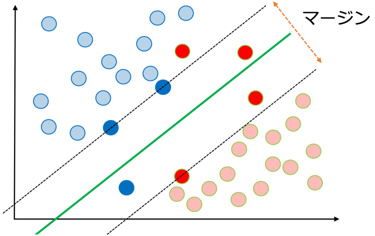
$$(\boldsymbol{w}^*)^T \boldsymbol{\phi}(\boldsymbol{x}) = \sum_{d \in [D]} y_d \alpha_d^* \mathsf{Ker}(\boldsymbol{x}_d, \boldsymbol{x})$$

$$b^* = y_j - \sum_{d \in [D]} y_d \alpha_d^* \mathsf{Ker}(\boldsymbol{x}_d, \boldsymbol{x}_j)$$

判別に当たって、$\alpha_d^* \neq 0$ となるデータ点 $(\boldsymbol{x}_d, y_d)$ のみが影響を与える。

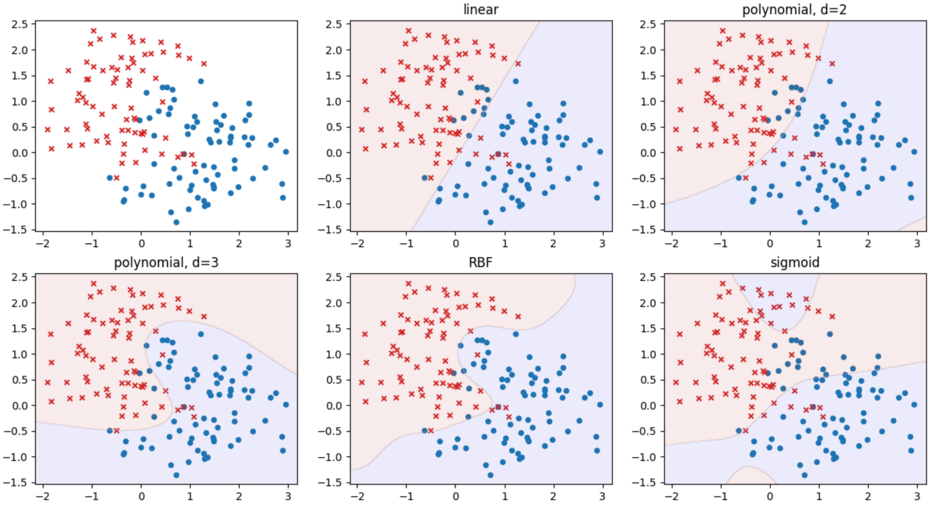
これらのデータ点をサポートベクターと呼ぶ。

- マージンを規定している点
- ハードマージン制約を違反している点

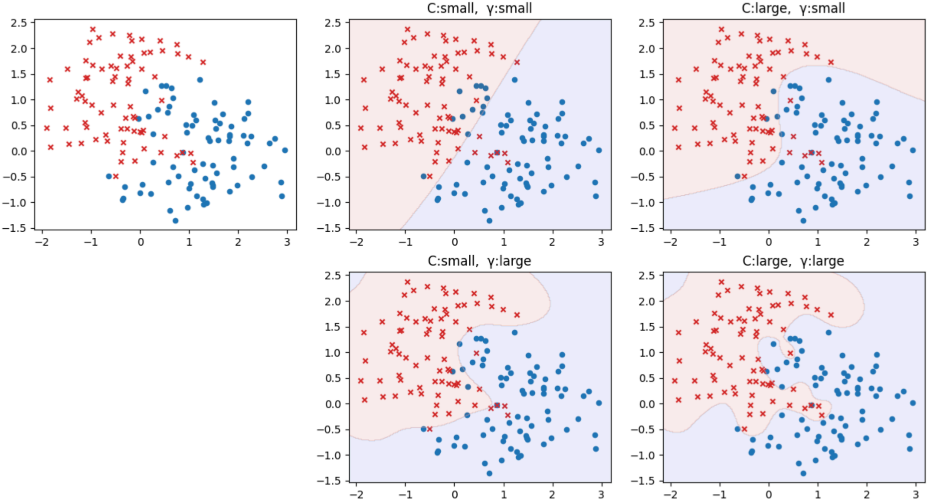


疎性（影響のないデータ点）があることが、
計算効率の上で大きな役割

様々なカーネルの判別面



ガウシアンカーネルの判別面



SVM の解法

Soft-SVM 双対問題の解法： SMO (Sequential Minimal Optimization)

Soft-SVM 双対問題（カーネル付き）

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}} \quad & -\frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{Y}\boldsymbol{K}\boldsymbol{Y}\boldsymbol{\alpha} + \boldsymbol{e}^T\boldsymbol{\alpha} \\
\text{s.t.} \quad & \boldsymbol{y}^T\boldsymbol{\alpha} = 0, \\
& \boldsymbol{0} \le \boldsymbol{\alpha} \le C\boldsymbol{e}.
\end{aligned}
$$

$\boldsymbol{\alpha}$ の 2 変数以外を固定した最適化問題を考える。$\boldsymbol{\alpha}' := \begin{pmatrix} \alpha_i \\ \alpha_j \end{pmatrix}$

2 変数の 2 次計画問題

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}'} \quad & -\frac{1}{2}(\boldsymbol{\alpha}')^T \boldsymbol{Q}\boldsymbol{\alpha}' + \boldsymbol{q}^T\boldsymbol{\alpha}' + \text{定数} \\
\text{s.t.} \quad & (\boldsymbol{y}')^T\boldsymbol{\alpha}' = s, \\
& \boldsymbol{0} \le \boldsymbol{\alpha}' \le C\boldsymbol{e}.
\end{aligned}
$$

$\boldsymbol{Q} \in \mathbb{R}^{2\times 2}$, $\boldsymbol{q} \in \mathbb{R}^2$, $\boldsymbol{y}' \in \mathbb{R}^2$, $s \in \mathbb{R}$ は定数

SMO

2 変数の 2 次計画問題の最適解は解析的に求めることが可能。

$(\boldsymbol{y}')^T\boldsymbol{\alpha}' = s$ より、「$\alpha_j := -y_iy_j\alpha_i +$ 定数」を目的関数と不等式制約に代入。
$q_2, q_1, l, h \in \mathbb{R}$ を定数とする 1 変数の凸 2 次関数最小化になる。

$$\begin{aligned} \min_{\alpha_i} \quad & \frac{1}{2}q_2\alpha_i^2 + q_1\alpha_i + \text{定数} \\ \text{s.t.} \quad & l \le \alpha_i \le h. \end{aligned}$$

$q_2 \neq 0$ のとき、最適解は次のようになる。

$$\alpha_i^* := \begin{cases} -\dfrac{q_1}{q_2} & (l \le -\dfrac{q_1}{q_2} \le h) \\ l & (-\dfrac{q_1}{q_2} < l) \\ h & (h < -\dfrac{q_1}{q_2}) \end{cases}$$

SMO は 2 変数のみの最適化を繰り返して最適解を求める方法

2 変数の選び方 (1)

最適性条件

$$\boldsymbol{\alpha}^T(\boldsymbol{e} - \boldsymbol{s} - \boldsymbol{Y}(\widetilde{\boldsymbol{X}}\boldsymbol{w} + \boldsymbol{e}b)) = 0,$$
$$\boldsymbol{\alpha} \geq \boldsymbol{0}, \quad \boldsymbol{e} - \boldsymbol{s} - \boldsymbol{Y}(\widetilde{\boldsymbol{X}}\boldsymbol{w} + \boldsymbol{e}b) \leq \boldsymbol{0},$$
$$\boldsymbol{\beta}^T\boldsymbol{s} = 0, \quad \boldsymbol{\beta} \geq \boldsymbol{0}, \quad \boldsymbol{s} \geq \boldsymbol{0},$$
$$\boldsymbol{\alpha} + \boldsymbol{\beta} = C\boldsymbol{e},$$
$$\boldsymbol{w} = \widetilde{\boldsymbol{X}}^T\boldsymbol{Y}\boldsymbol{\alpha},$$
$$\boldsymbol{y}^T\boldsymbol{\alpha} = 0.$$

上の 4 行を要素ごとに書くと、$d \in [D]$ に対し

$$\alpha_d(1 - s_d - y_d(\boldsymbol{\phi}(\boldsymbol{x}_d)^T\boldsymbol{w} + b)) = 0,$$
$$\alpha_d \geq 0, \quad 1 - s_d - y_d(\boldsymbol{\phi}(\boldsymbol{x}_d)^T\boldsymbol{w} + b) \leq 0,$$
$$\beta_d s_d = 0,$$
$$\beta_d \geq 0, \quad s_d \geq 0,$$
$$\alpha_d + \beta_d = C.$$

2 変数の選び方 (2)

$$t_d := y_d - \boldsymbol{\phi}(\boldsymbol{x}_d)^T \boldsymbol{w} = y_d - \sum_{i \in [D]} y_i \alpha_i \mathsf{Ker}(\boldsymbol{x}_i, \boldsymbol{x}_d)$$ とする。

このとき、$y_d t_d = 1 - y_d \boldsymbol{\phi}(\boldsymbol{x}_d)^T \boldsymbol{w}$ となる。

$d \in [D]$ において、

- $\alpha_d = 0$ のとき、 $\beta_d \neq 0$ より $s_d = 0$ なので $y_d t_b \leq y_d b$
  $\cdot\, y_d = +1$ ならば $t_b \leq b$
  $\cdot\, y_d = -1$ ならば $b \leq t_b$
- $\alpha_d = C$ のとき、 $1 - y_d(\boldsymbol{\phi}(\boldsymbol{x}_d)^T \boldsymbol{w} + b) = s_d \geq 0$ より $y_d t_b \geq y_d b$
  $\cdot\, y_d = +1$ ならば $b \leq t_b$
  $\cdot\, y_d = -1$ ならば $t_b \leq b$
- $0 < \alpha_d < C$ のとき、 $y_d t_b = y_d b$
  $\cdot\, t_b \leq b \leq t_b$

b の範囲 $l_d \leq b \leq h_d$ が得られる $l_d \in \{t_d, -\infty\}$, $h_d \in \{t_d, \infty\}$,

2 変数の選び方 (3)

- もし $\max_{d \in [D]} l_d \leq \min_{d \in [D]} h_d$ ならば、その間の b^* を選べば最適解。

$$\max_{d \in [D]} l_d \leq b^* \leq \min_{d \in [D]} h_d$$

- そうでない場合、一番不等式を満たしていない２つを選ぶ。

$$i := \operatorname*{argmax}_{d \in [D]} l_d,$$

$$j := \operatorname*{argmin}_{d \in [D]} h_d$$

回帰の場合

データ $\{\boldsymbol{x}_d, y_d\}_{d\in[D]} \quad \boldsymbol{x}_d \in \mathbb{R}^n,\ y_d \in \mathbb{R}$

予測器

パラメタは $\boldsymbol{w}, b$

$$\hat{y} = \boldsymbol{w}^T\boldsymbol{x} + b$$

誤差関数

ϵ 感度損失：

$$L(y, \hat{y}) := \max\{|y - \hat{y}| - \epsilon, 0\}$$

$y \in \mathbb{R}$：目標,　$\hat{y} \in \mathbb{R}$：予測値

正則化項つき学習

$$\min_{\boldsymbol{w},b} \sum_{d\in[D]} L(y_d, \boldsymbol{w}^T\boldsymbol{x} + b) + \frac{1}{2C}\boldsymbol{w}^T\boldsymbol{w}$$

ヒンジ損失と ϵ 感度損失

ヒンジ損失

ϵ 感度損失

SVR

正則化項つき学習：

$$\min_{\boldsymbol{w},b} \sum_{d\in[D]} L(y_d, \boldsymbol{w}^T\boldsymbol{x}+b) + \frac{1}{2C}\boldsymbol{w}^T\boldsymbol{w}$$

ただし、 $L(y,\hat{y}) := \max\{|y-\hat{y}|-\epsilon, 0\}$

$-L(y,\hat{y}) - \epsilon \leq \hat{y} - y \leq \epsilon + L(y,\hat{y})$ という関係があるため、

最適化問題：

$$\begin{aligned} \min_{\boldsymbol{w},b,\boldsymbol{s},\boldsymbol{t}} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{d\in[D]}(s_d^+ + s_d^-) \\ \text{s.t.} \quad & y_d - \epsilon - s_d^- \leq \boldsymbol{x}_d^T\boldsymbol{w} + b \leq y_d + \epsilon + s_d^+ \quad && (d\in[D]), \\ & s_d^+, s_d^- \geq 0 && (d\in[D]). \end{aligned}$$

※ s_d^+, s_d^- を一つにまとめることも可能である

SVR の行列表現

Soft-SVR 主問題

$$
\begin{aligned}
\min_{\boldsymbol{w},b,\boldsymbol{s}_1,\boldsymbol{s}_2} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\boldsymbol{e}^T(\boldsymbol{s}_1 + \boldsymbol{s}_2) \\
\text{s.t.} \quad & \boldsymbol{y} - \epsilon\boldsymbol{e} - \boldsymbol{s}_2 \leq \boldsymbol{X}\boldsymbol{w} + \boldsymbol{e}b \leq \boldsymbol{y} + \epsilon\boldsymbol{e} + \boldsymbol{s}_1, \\
& \boldsymbol{s}_1, \boldsymbol{s}_2 \geq \boldsymbol{0}.
\end{aligned}
$$

C はペナルティの重みを意味する定数

Soft-SVR 双対問題

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2} \quad & -\frac{1}{2}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)^T\boldsymbol{X}\boldsymbol{X}^T(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) - \epsilon\boldsymbol{e}^T(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2) - \boldsymbol{y}^T(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) \\
\text{s.t.} \quad & \boldsymbol{e}^T(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) = 0, \\
& \boldsymbol{0} \leq \boldsymbol{\alpha}_1 \leq C\boldsymbol{e}, \quad \boldsymbol{0} \leq \boldsymbol{\alpha}_2 \leq C\boldsymbol{e}.
\end{aligned}
$$

双対問題の導出１

ラグランジュ関数： $L(\boldsymbol{w}, b, \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$

$$
\begin{aligned}
=&\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\boldsymbol{e}^T(\boldsymbol{s}_1 + \boldsymbol{s}_2) + \boldsymbol{\alpha}_1^T(\boldsymbol{X}\boldsymbol{w} + \boldsymbol{e}b - \boldsymbol{y} - \epsilon\boldsymbol{e} - \boldsymbol{s}_1) \\
&+ \boldsymbol{\alpha}_2^T(\boldsymbol{y} - \epsilon\boldsymbol{e} - \boldsymbol{s}_2 - \boldsymbol{X}\boldsymbol{w} - \boldsymbol{e}b) - \boldsymbol{\beta}_1^T\boldsymbol{s}_1 - \boldsymbol{\beta}_2^T\boldsymbol{s}_2
\end{aligned}
$$

ラグランジュ双対問題： $\displaystyle\max_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2 \geq \mathbf{0}} \min_{\boldsymbol{w}, b, \boldsymbol{s}_1, \boldsymbol{s}_2} L(\boldsymbol{w}, b, \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$

内側の最小化問題は制約なしの凸最適化問題である。

$$
\begin{aligned}
\nabla_{\boldsymbol{w}} L(\boldsymbol{w}, b, \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= \boldsymbol{w} + \boldsymbol{X}^T\boldsymbol{\alpha}_1 - \boldsymbol{X}^T\boldsymbol{\alpha}_2 = \mathbf{0} \\
\nabla_b L(\boldsymbol{w}, b, \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= \boldsymbol{e}^T\boldsymbol{\alpha}_1 - \boldsymbol{e}^T\boldsymbol{\alpha}_2 = 0 \\
\nabla_{\boldsymbol{s}_1} L(\boldsymbol{w}, b, \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= C\boldsymbol{e} - \boldsymbol{\alpha}_1 - \boldsymbol{\beta}_1 = \mathbf{0} \\
\nabla_{\boldsymbol{s}_2} L(\boldsymbol{w}, b, \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= C\boldsymbol{e} - \boldsymbol{\alpha}_2 - \boldsymbol{\beta}_2 = \mathbf{0}
\end{aligned}
$$

より、次の関係が得られる。

$$
\begin{aligned}
&\boldsymbol{w} = -\boldsymbol{X}^T(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2), \quad \boldsymbol{e}^T(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) = 0 \\
&\boldsymbol{\alpha}_1 + \boldsymbol{\beta}_1 = C\boldsymbol{e}, \qquad \boldsymbol{\alpha}_2 + \boldsymbol{\beta}_2 = C\boldsymbol{e}
\end{aligned}
$$

双対問題の導出２

これを、ラグランジュ関数に代入すると、

$$
\begin{aligned}
&L(\boldsymbol{w}, b, \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \\
&= -\frac{1}{2}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)^T \boldsymbol{X}\boldsymbol{X}^T(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) - \epsilon \boldsymbol{e}^T(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2) - \boldsymbol{y}^T(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)
\end{aligned}
$$

である。よって、双対問題は次のようになる。

$$
\begin{aligned}
\max \quad & -\frac{1}{2}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)^T \boldsymbol{X}\boldsymbol{X}^T(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) - \epsilon \boldsymbol{e}^T(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2) - \boldsymbol{y}^T(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) \\
\text{s.t.} \quad & \boldsymbol{e}^T(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) = 0, \\
& \boldsymbol{\alpha}_1 + \boldsymbol{\beta}_1 = C\boldsymbol{e}, \\
& \boldsymbol{\alpha}_2 + \boldsymbol{\beta}_2 = C\boldsymbol{e}, \\
& \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \geq \mathbf{0},
\end{aligned}
$$

$$
\begin{aligned}
&\boldsymbol{\alpha}_1 + \boldsymbol{\beta}_1 = C\boldsymbol{e},\ \boldsymbol{\alpha}_1, \boldsymbol{\beta}_1 \geq \mathbf{0} \ \rightarrow \ \mathbf{0} \leq \boldsymbol{\alpha}_1 \leq C\boldsymbol{e} \\
&\boldsymbol{\alpha}_2 + \boldsymbol{\beta}_2 = C\boldsymbol{e},\ \boldsymbol{\alpha}_2, \boldsymbol{\beta}_2 \geq \mathbf{0} \ \rightarrow \ \mathbf{0} \leq \boldsymbol{\alpha}_2 \leq C\boldsymbol{e}
\end{aligned}
$$

と書き換えることができる

カーネルの導入

Soft-SVR 双対問題

$$\begin{array}{ll} \max\limits_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2} & -\frac{1}{2}(\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2)^T \boldsymbol{X}\boldsymbol{X}^T(\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2) - \epsilon \boldsymbol{e}^T(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2) - \boldsymbol{y}^T(\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2) \\ \text{s.t.} & \boldsymbol{e}^T(\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2)=0, \\ & \boldsymbol{0} \le \boldsymbol{\alpha}_1 \le C\boldsymbol{e}, \quad \boldsymbol{0} \le \boldsymbol{\alpha}_2 \le C\boldsymbol{e}. \end{array}$$

SMV と同様にカーネル化が可能

Soft-SVR 双対問題（カーネル付き）

$$\begin{array}{ll} \max\limits_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2} & -\frac{1}{2}(\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2)^T \boldsymbol{K}(\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2) - \epsilon \boldsymbol{e}^T(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2) - \boldsymbol{y}^T(\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2) \\ \text{s.t.} & \boldsymbol{e}^T(\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2)=0, \\ & \boldsymbol{0} \le \boldsymbol{\alpha}_1 \le C\boldsymbol{e}, \quad \boldsymbol{0} \le \boldsymbol{\alpha}_2 \le C\boldsymbol{e}. \end{array}$$

カーネル行列 $K_{ij} := \mathrm{Ker}(\boldsymbol{x}_i, \boldsymbol{x}_j)$

SVR の例

丸印はサポートベクター

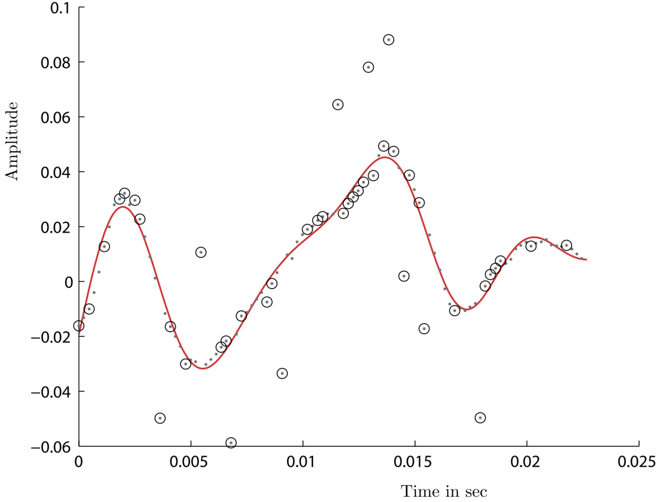


Figure: Machine Learning A Bayesian and Optimization Perspective, Fig. 11.12