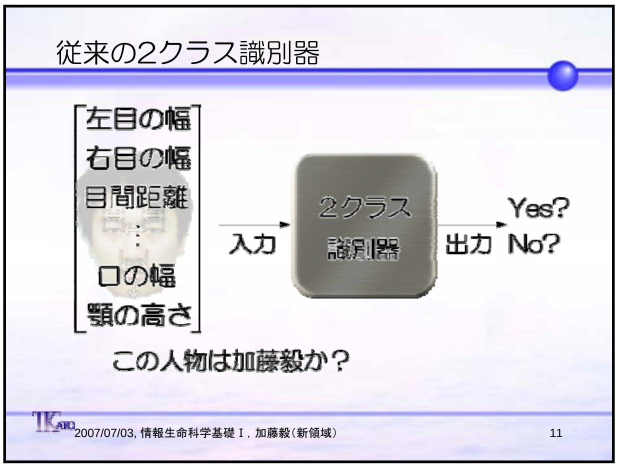
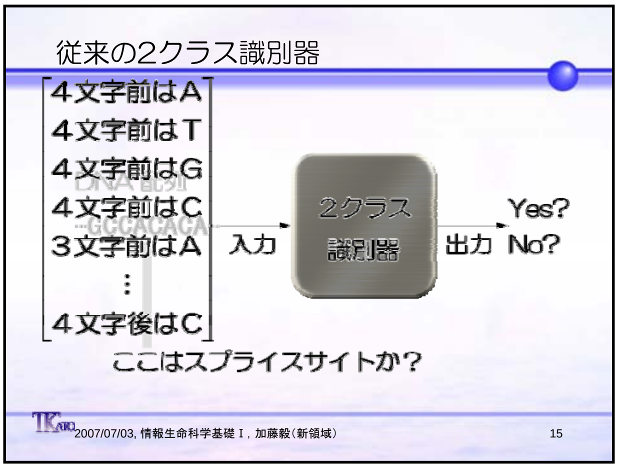
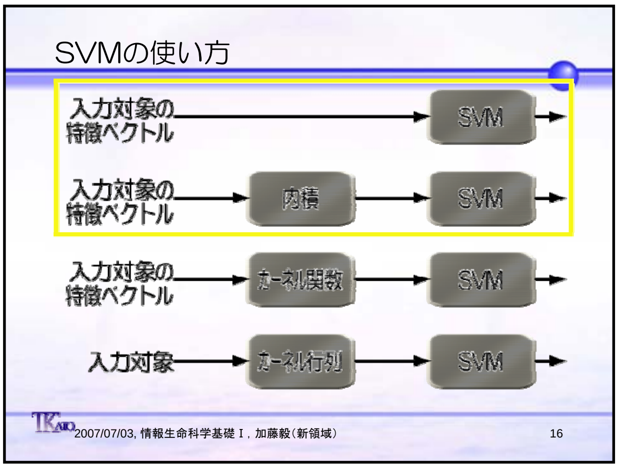

SVMの使い方
東京大学大学院新領域創成科学研究科
情報生命科学専攻
加藤　毅
2007/07/03, 情報生命科学基礎 I，加藤毅（新領域）
1


従来の2クラス識別器
予測対象の
特徴ベクトル
入力
2クラス
識別器
出力
Yes?
No?
2007/07/03, 情報生命科学基礎 I，加藤毅（新領域）
9

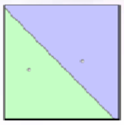

従来の2クラス識別器
左目の幅
右目の幅
目間距離
⋮
口の幅
顎の高さ
入力
2クラス
識別器
出力
Yes?
No?
この人物は加藤毅か？
2007/07/03, 情報生命科学基礎 I，加藤毅（新領域）
11


従来の2クラス識別器
遺伝子1の発現量
遺伝子2の発現量
⋮
遺伝子Nの発現量
マイクロアレイ
入力
2クラス
識別器
出力
Yes?
No?
癌の種類はAMLか？
2007/07/03, 情報生命科学基礎 I，加藤毅（新領域）
13


従来の2クラス識別器
4文字前はA
4文字前はT
4文字前はG
4文字前はC
3文字前はA
⋮
4文字後はC
DNA配列
…GCCACACA…
入力
2クラス
識別器
出力
Yes?
No?
ここはスプライスサイトか？
2007/07/03, 情報生命科学基礎 I，加藤毅（新領域）
15


SVMの使い方
入力対象の特徴ベクトル
SVM
入力対象の特徴ベクトル
内積
SVM
入力対象の特徴ベクトル
カーネル関数
SVM
入力対象
カーネル行列
SVM
2007/07/03, 情報生命科学基礎 I，加藤毅（新領域）
16

線形識別器



スコア関数

$g(x|w, b) = \langle w, x\rangle + b$

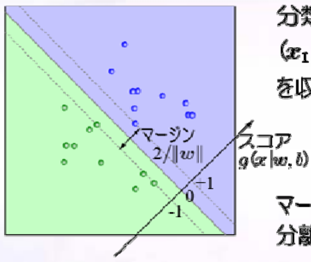
パラメータ $w \in \Re^d,\ b \in \Re^d$

特徴ベクトル $x \in \Re^d$

If スコア>0,　x は + だろう

If スコア<0,　x は − だろう

SVMの学習方法の概要



分類済みの学習用例題集合

$(x_1, y_1), \cdots, (x_\ell, y_\ell)$

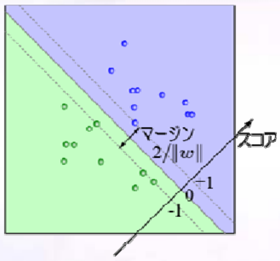
を収集　$x_i \in \Re^d$　$y_i \in \{\pm 1\}$

マージンが最大になる分離面を見付ける

SVMソフトウェア

- SVM-light
  http://svmlight.joachims.org/
- LIBSVM
  http://www.csie.ntu.edu.tw/~cjlin/libsvm/
- kemba-svm.exe
  http://www.cb.k.u-tokyo.ac.jp/asailab/kato/kemba-svm1-ts/

SVM学習の定式化

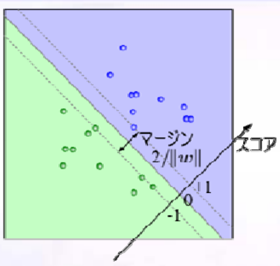


min $1/(\text{マージン})^2$

subj. to
- + の例題は境界面より上
- − の例題は境界面より下

wrt $w \in \Re^d, b \in \Re$

SVM学習の定式化



$$\min \frac{1}{2}\|w\|^2$$

subj. to

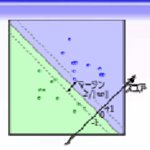
+ $\langle w, x_i\rangle + b \geq +1$

− $\langle w, x_i\rangle + b \leq -1$

$$\Leftrightarrow y_i(\langle w, x_i\rangle + b) \geq +1$$

wrt $w \in \Re^d, b \in \Re$

SVM学習の定式化



$$\min \frac{1}{2}\|w\|^2 \quad \text{wrt } w \in \Re^d, b \in \Re$$

$$\text{subj. to } 1 - y_i(\langle w, x_i\rangle + b) \leq 0$$

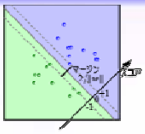
$$\Leftrightarrow \min_{w,b} \max_{\alpha \geq 0} L(w, b, \alpha)$$

定理　最適化問題　$\min f(x)$ wrt $x \in \Re^d$ subj. to $g_i(x) \leq 0\ \ \forall i = 1, \ldots, n$

の解 = $\min_{x} \max_{\alpha \geq 0} L(x, \alpha)$ の解 $(\bar{x}, \bar{\alpha})$

ただし，ラグランジュ関数 $L(x, \alpha) = f(x) + \sum_{i=1}^{n} \alpha_i g_i(x)$

SVM学習の定式化



$$\min \frac{1}{2}\|w\|^2 \quad \text{wrt } w \in \Re^d,\ b \in \Re$$

$$\text{subj. to} \quad 1 - y_i\left(\langle w, x_i\rangle + b\right) \le 0$$

$$\Leftrightarrow \min_{w,b} \max_{\alpha \ge 0} L(w, b, \alpha)$$

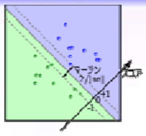
ラグランジェ関数

$$L(w, b, \alpha) = \frac{1}{2}\langle w, w\rangle + \sum_{i=1}^{\ell} \alpha_i \left(1 - y_i\left(\langle w, x_i\rangle + b\right)\right)$$

$$\frac{\partial}{\partial w}L = 0_d,\ \frac{\partial}{\partial b}L = 0 \quad \text{yields} \quad w = \sum_{i=1}^{\ell}\alpha_i y_i x_i,\ \sum_{i=1}^{\ell}\alpha_i y_i = 0$$

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 39

SVM学習の双対形式



$$\min \frac{1}{2}\|w\|^2 \quad \text{wrt } w \in \Re^d,\ b \in \Re$$

$$\text{subj. to} \quad 1 - y_i\left(\langle w, x_i\rangle + b\right) \le 0$$

$$\Leftrightarrow \max_{\alpha \ge 0} -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\alpha_i y_i \alpha_j y_j \langle x_i, x_j\rangle + \sum_{i=1}^{\ell}\alpha_i$$

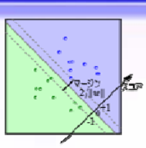
$$\text{subj. to} \quad \sum_{i=1}^{\ell}\alpha_i y_i = 0$$

SVM 学習に
特徴ベクトルは 不要
内積のみ 必要

メモ $w = \sum_{i=1}^{\ell}\alpha_i y_i x_i$

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 44

バイアス b の復元



スコア関数

$$g(x \mid w, b) = \left\langle \sum_{j=1}^{\ell}\alpha_j y_j x_j, x\right\rangle + b = \sum_{j=1}^{\ell}\alpha_j y_j \langle x_j, x\rangle + b$$

定理 最適化問題 $\min f(x)$ wrt. $x \in \Re^d$ subj. to $g_i(x) \le 0 \quad \forall i=1,\dots,n$

の解 = $\min_{x}\max_{\alpha \ge 0} L(x, \alpha)$ の解 $(\bar{x}, \bar{\alpha})$

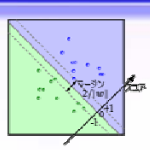
ただし，ラグランジェ関数 $L(x, \alpha) = f(x) + \sum_{i=1}^{n}\alpha_i g_i(x)$

$\frac{\partial}{\partial x}L(x, \alpha)\Big|_{x=\bar{x}, \alpha=\bar{\alpha}} = 0_d$, $\frac{\partial}{\partial \alpha}L(x, \alpha)\Big|_{x=\bar{x}, \alpha=\bar{\alpha}} \le 0_n$, KKT 相補条件 $\bar{\alpha}_i g_i(\bar{x}) = 0,\ \forall i$

$w = \sum_{i=1}\alpha_i y_i x_i$

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 50

バイアス b の復元



スコア関数

$$g(x \mid w, b) = \left\langle \sum_{j=1}^{\ell}\alpha_j y_j x_j, x\right\rangle + b = \sum_{j=1}^{\ell}\alpha_j y_j \langle x_j, x\rangle + b$$

$\alpha_i > 0$ なる i に対し

$$y_i\left(\left\langle \sum_{j=1}^{\ell}\alpha_j y_j x_j, x_i\right\rangle + b\right) - 1 = 0$$

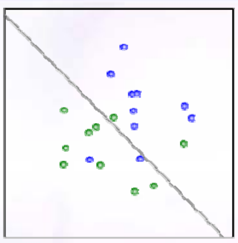
$$\therefore b = -\sum_{j=1}^{\ell}\alpha_j y_j \langle x_j, x_i\rangle + y_i$$

SVM 識別に
特徴ベクトルは 不要
内積のみ 必要

メモ $w = \sum_{i=1}^{\ell}\alpha_i y_i x_i$

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 58

線形分離不可能な場合



$$\min \frac{1}{2}\|w\|^2$$

解なし

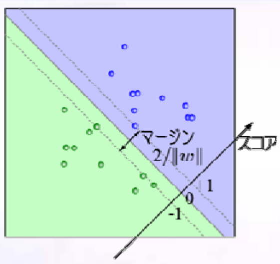
subj. to

$$\begin{cases} \langle w, x_i\rangle + b \ge +1 & (+) \\ \langle w, x_i\rangle + b \le -1 & (-) \end{cases}$$

wrt $w \in \Re^d,\ b \in \Re$

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 61

ソフトマージンSVM



$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}\xi_i$$

正則化定数

subj. to

$$\begin{cases} \langle w, x_i\rangle + b \ge +1 - \xi_i & (+) \\ \langle w, x_i\rangle + b \le -1 + \xi_i & (-) \end{cases}$$

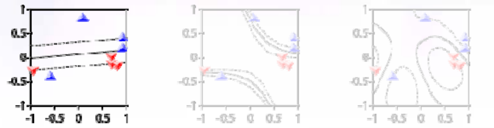
wrt $w \in \Re^d,\ b \in \Re$

$$\xi = [\xi_1, \cdots, \xi_\ell]^T \in \Re_+^\ell$$

全要素が非負の ℓ 次元実数ベクトル

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 62

2次元の例

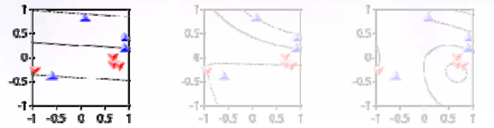


線形カーネル 多項式カーネル RBF カーネル

$C = 1000$

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 63

2次元の例

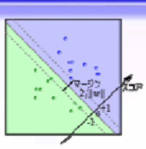


線形カーネル 多項式カーネル RBF カーネル

$C = 1$

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 64

ソフトマージンSVM学習の双対形式



$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{\ell}\xi_i \quad \text{wrt } \boldsymbol{w}\in\Re^d, b\in\Re, \boldsymbol{\xi}\in\Re^{\ell}_{+}$$

$$\text{subj. to} \quad 1 - y_i\left(\langle \boldsymbol{w}, \boldsymbol{x}_i\rangle + b\right) \le \xi_i$$

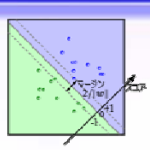
$$\Leftrightarrow \max_{\boldsymbol{\alpha}} -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\alpha_i y_i \alpha_j y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle + \sum_{i=1}^{\ell}\alpha_i$$

$$\text{subj. to} \quad \sum_{i=1}^{\ell}\alpha_i y_i = 0, \quad 0 \le \alpha_i \le C$$

ソフトマージンSVMに
特徴ベクトルは 不要
内積のみ 必要

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 67

ソフトマージンSVMのスコア関数



スコア関数

$$g(\boldsymbol{x}|\boldsymbol{w}, b) = \left\langle \sum_{j=1}^{\ell}\alpha_j y_j \boldsymbol{x}_j, \boldsymbol{x}\right\rangle + b = \sum_{j=1}^{\ell}\alpha_j y_j \langle \boldsymbol{x}_j, \boldsymbol{x}\rangle + b$$

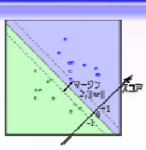
KKT相補条件 $\alpha_i\left(y_i\left(\left\langle \sum_{j=1}^{\ell}\alpha_j y_j \boldsymbol{x}_j, \boldsymbol{x}_i\right\rangle + b\right) - 1 - \xi_i\right) = 0$

$\delta_i \xi_i = 0$ $\qquad \delta_i \ge 0$

$\alpha_i + \delta_i = C$

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 71

ソフトマージンSVMのスコア関数



スコア関数

$$g(\boldsymbol{x}|\boldsymbol{w}, b) = \left\langle \sum_{j=1}^{\ell}\alpha_j y_j \boldsymbol{x}_j, \boldsymbol{x}\right\rangle + b = \sum_{j=1}^{\ell}\alpha_j y_j \langle \boldsymbol{x}_j, \boldsymbol{x}\rangle + b$$

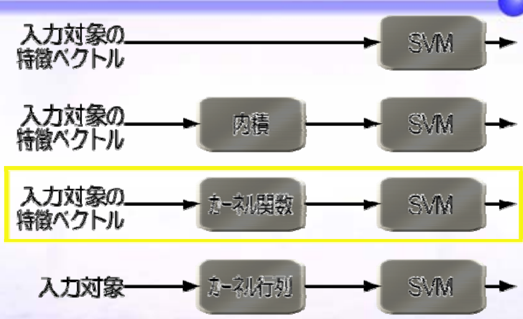
$0 < \alpha_i < C$なるiに対し $\quad y_i\left(\left\langle \sum_{j=1}^{\ell}\alpha_j y_j \boldsymbol{x}_j, \boldsymbol{x}_i\right\rangle + b\right) - 1 = 0$

$$\therefore b = -\sum_{j=1}^{\ell}\alpha_j y_j \langle \boldsymbol{x}_j, \boldsymbol{x}_i\rangle + y_i$$

ソフトマージンSVMに
特徴ベクトルは 不要
内積のみ 必要

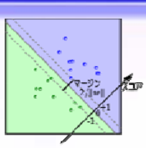
2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 80

SVMの使い方



2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 82

SVM学習のカーネル表現



$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}\xi_i \quad \text{wrt } w\in\Re^d, b\in\Re, \xi\in\Re^{\ell}_{+}$$

$$\text{subj. to} \quad 1 - y_i(\langle w, x_i\rangle + b) \le \xi_i$$

$$\Leftrightarrow \max_{\alpha} \ -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\alpha_i y_i \alpha_j y_j k(x_i, x_j) + \sum_{i=1}^{\ell}\alpha_i$$

$$\text{subj. to} \quad \sum_{i=1}^{\ell}\alpha_i y_i = 0, \quad 0 \le \alpha_i \le C$$

ソフトマージンSVMに
特徴ベクトルは 不要
カーネル値のみ 必要

カーネル関数 $k(x, x') = \phi(x)^\top\phi(x')$, $\phi(x) = x$

主要なカーネル

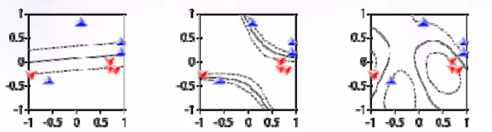
関数 k：入力空間 × 入力空間→実数 が

$k(x, z) = \phi(x)^\top \phi(x')$

なる写像関数 ϕ：入力空間→内積空間 が存在するとき，**カーネル関数**と呼ぶ

	入力空間	定義式
線形カーネル	特徴ベクトル	$k(x, x') = x^\top x'$ $\phi(x) = x$
多項式カーネル	特徴ベクトル	$k(x, x') = (x^\top x' + c^2)^p$
RBF カーネル	特徴ベクトル	$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$

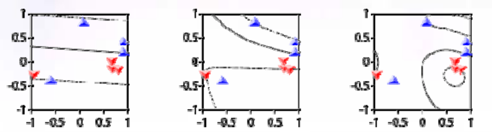
2次元の例



線形カーネル 多項式カーネル RBF カーネル

$C = 1000$

2次元の例



線形カーネル 多項式カーネル RBF カーネル

$C = 1$

多項式カーネルはカーネル？

多項式カーネル 特徴ベクトル $k(x, x') = (x^\top x' \quad)^2$

$$(x^\top x')^2 = x^\top x'(x')^\top x$$
$$= \mathrm{tr}(xx^\top x'(x')^\top)$$
$$= \mathrm{vec}\,(xx^\top)^\top \mathrm{vec}$$

2次の多項式カーネルは
特徴間の情報を
取り込める

$$\phi(x) = \mathrm{vec}\,(xx^\top)$$

$$= \mathrm{vec}\left(\begin{bmatrix} x_1x_1 & \cdots & x_1x_d \\ \vdots & \ddots & \vdots \\ x_dx_1 & \cdots & x_dx_d \end{bmatrix}\right)$$

多項式カーネルはカーネル？

多項式カーネル 特徴ベクトル $k(x, x') = (x^\top x' + c^2)^p$

数学的帰納法

多項式カーネルはカーネル

$p=0$ のとき

$\phi_0(x) = 1$ とおくと $(x^\top x' + c^2)^0 = 1 = \phi_0(x)^\top\phi_0(x')$

$p=k$ のとき $(x^\top x' + c^2)^k = \phi_k(x)^\top\phi_k(x')$ と仮定

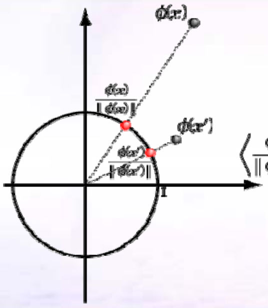
$$(x^\top x' + c^2)^{k+1} = \mathrm{tr}\left(\left(\phi_k(x)\begin{bmatrix} x \\ c \end{bmatrix}^\top\right)^\top\left(\phi_k(x')\begin{bmatrix} x' \\ c \end{bmatrix}^\top\right)\right)$$

$$= \mathrm{vec}\left(\phi_k(x)\begin{bmatrix} x \\ c \end{bmatrix}^\top\right)^\top \mathrm{vec}\left(\phi_k(x')\begin{bmatrix} x' \\ c \end{bmatrix}^\top\right)$$

$\phi_{k+1}(x) = \mathrm{vec}\left(\phi_k(x)\begin{bmatrix} x \\ c \end{bmatrix}^\top\right)$ とおくと

$$= \phi_{k+1}(x)^\top\phi_{k+1}(x')$$

カーネルの正規化



$k(x,x') = \phi(x)^\top \phi(x')$
とする

$$\left\langle \frac{\phi(x)}{\|\phi(x)\|}, \frac{\phi(x')}{\|\phi(x')\|} \right\rangle = \frac{\langle \phi(x), \phi(x') \rangle}{\|\phi(x)\| \|\phi(x')\|} = \frac{k(x,x')}{\sqrt{k(x,x)k(x',x')}}$$

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 118

RBFカーネルはカーネル?

RBFカーネル　特徴ベクトル　$k(x,x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$

補題　カーネルの重みつき線形和はカーネル （重み > 0）

$k_a(x,x') = \langle \phi_a(x), \phi_a(x') \rangle$
$k_b(x,x') = \langle \phi_b(x), \phi_b(x') \rangle$ を仮定

$$\begin{aligned} k(x,x') &= c_a k_a(x,x') + c_b k_b(x,x') \qquad c_a, c_b > 0 \in \mathbb{R} \\ &= c_a \phi_a(x)^\top \phi_a(x') + c_b \phi_b(x)^\top \phi_b(x') \\ &= \begin{bmatrix} c_a^{1/2}\phi_a(x) \\ c_b^{1/2}\phi_b(x) \end{bmatrix}^\top \begin{bmatrix} c_a^{1/2}\phi_a(x') \\ c_b^{1/2}\phi_b(x') \end{bmatrix} \\ &= \phi(x)^\top \phi(x') \end{aligned}$$

$\phi(x) = \begin{bmatrix} c_a^{1/2}\phi_a(x) \\ c_b^{1/2}\phi_b(x) \end{bmatrix}$ とおくと

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 125

RBFカーネルはカーネル?

RBFカーネル　特徴ベクトル　$k(x,x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$

$$\exp\left(\frac{x^\top x'}{\sigma^2}\right) = 1 + \frac{1}{\sigma^2}(x^\top x') + \frac{1}{2!(\sigma^2)^2}(x^\top x')^2 + \frac{1}{3!(\sigma^2)^3}(x^\top x')^3 + \cdots$$

はカーネル　正規化すると

$$\frac{\exp(x^\top x'/(\sigma^2))}{\sqrt{\exp(x^\top x/(\sigma^2))\exp(x'^\top x'/(\sigma^2))}} = \exp\left(-\frac{-2x^\top x' + \|x\|^2 + \|x'\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$$

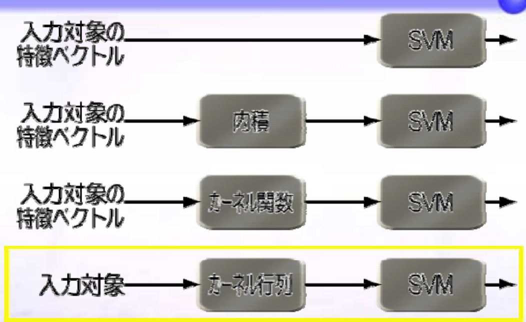
RBFカーネルは
カーネル

正規化カーネル　$k(x,x')/\sqrt{k(x,x)k(x',x')}$

補題　カーネルの重みつき線形和はカーネル

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 139

SVMの使い方

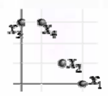


2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 142

カーネル行列

ユークリッド座標による表現

$$X = [x_1 \ x_2 \ x_3 \ x_4] = \begin{bmatrix} 3 & 2 & 0 & 1 \\ 0 & 1 & 3 & 3 \end{bmatrix}$$



カーネル行列による表現　$k(x,x') = x^\top x'$ の場合

$$K = \begin{bmatrix} 9 & 6 & 0 & 3 \\ 6 & 5 & 3 & 5 \\ 0 & 3 & 9 & 9 \\ 3 & 5 & 9 & 10 \end{bmatrix}$$

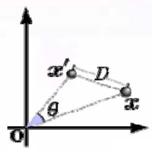
2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 144

カーネル行列

ユークリッド座標による表現

二乗ユークリッド距離　$D^2 = \|x - x'\|^2$

余弦定理　$\cos\theta = \frac{x^\top x'}{\|x\|\|x'\|}$



カーネル　による表現

$$D^2 = k(x,x) + k(x',x') - 2k(x,x')$$

$$\cos\theta = \frac{k(x,x')}{\sqrt{k(x,x)k(x',x')}}$$

カーネル行列は
特徴ベクトルに代わる
もう一つの表現

2007/07/03, 情報生命科学基礎Ｉ，加藤毅（新領域） 151

カーネル行列

定義 $A \in \Re^{n\times n}$ を対称行列とする

$$c^\top A c \geq 0 \quad \forall c \in \Re^n$$

ならば A は **半正定値**

$A \succeq O \Leftrightarrow A$ の**すべての固有値が非負**

定理 $K \in \Re^{n\times n}$ を半正定値行列とする

For $\forall x_1, \cdots, x_n \in \mathcal{X}$

$$K_{ij} = \phi(x_i)^\top \phi(x_j)$$

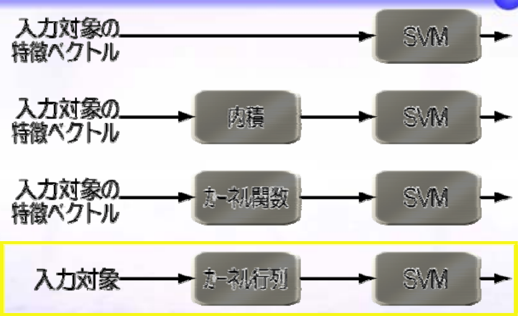
なる写像

$$\phi(x) = \begin{bmatrix} \phi_1(x) & \cdots & \phi_d(x) \end{bmatrix}^\top$$

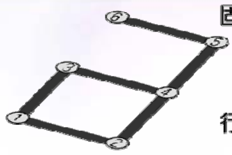
が存在

全固有値非負 ⇔ カーネル行列

SVMの使い方



準備: 固有値分解と行列指数



固有値分解 $A = \underbrace{U}_{\text{固有ベクトルの行列}}\underbrace{D}_{\text{固有値の行列}}U^\top$

$$U^\top U = I_n$$

$$D = \mathrm{diag}\{D_{ii}\}_{i=1}^n$$

行列指数

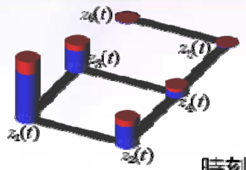
$$\exp(A) = U\,\mathrm{diag}\{\exp D_{ii}\}_{i=1}^n U^\top$$

$$\lim_{s\to\infty}\left(I+\frac{1}{s}A\right)^s = \lim_{s\to\infty} U\left(I+\frac{1}{s}D\right)\left(I+\frac{1}{s}D\right)\cdots\left(I+\frac{1}{s}D\right)U^\top$$

$$= U\left(\lim_{s\to\infty}\left(I+\frac{1}{s}D\right)^s\right)U^\top = U\left(\mathrm{diag}\left\{\lim_{s\to\infty}\left(1+\frac{1}{s}D_{ii}\right)^s\right\}_{i=1}^n\right)U^\top$$

$$= U\left(\mathrm{diag}\{\exp D_{ii}\}_{i=1}^n\right)U^\top = \exp(A)$$

拡散カーネル exp(-βL)の解釈



$$z(0) \sim N(0_n, I_n)$$

毎時刻 割合 α 隣接ノードに送る

$$z_i(t) = z_i(t-1) + \alpha \sum_{j: j\sim i} (z_j(t-1) - z_i(t-1))$$

行列表現

$$z(t) = (I_n - \alpha L) z(t-1) = \quad T(t) z(0)$$

オペレータ

~~$T(t) = (I_n - \alpha L)^t$~~

１ステップ Δt に変更

$$T(t) = \left(I_n - \frac{\alpha}{1/\Delta t} L\right)^{t/\Delta t}$$

時刻 t の共分散行列

$$\mathrm{Cov}(z(t)) = T(2t) \qquad (\Delta t \to 0 \text{ とすると})$$

$$= \left(\lim_{1/\Delta t\to\infty}\left(I_n - \frac{1}{1/\Delta t}\alpha L\right)^{1/\Delta t}\right)^{2t}$$

$$= (\exp(-\alpha L))^{2t} = \exp(\quad \beta L)$$

（ $\beta = 2\alpha t$ とおくと）