

カーネル法の基礎

福水 健次 (統計数理研究所)

2006年7月6~7日

公開講座「カーネル法の最前線

— SVM, 非線形データ解析, 構造化データ —」

1. イントロダクション

- このセクションの目的
 - カーネル法に関して大まかなイメージを持ってもらう
 - くわしい説明はあとできちんとやる

非線形データ解析としてのカーネル法

■ 非線形データ解析の重要性

□ 古典的な線形データ解析

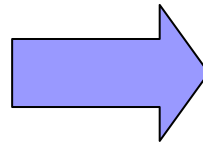
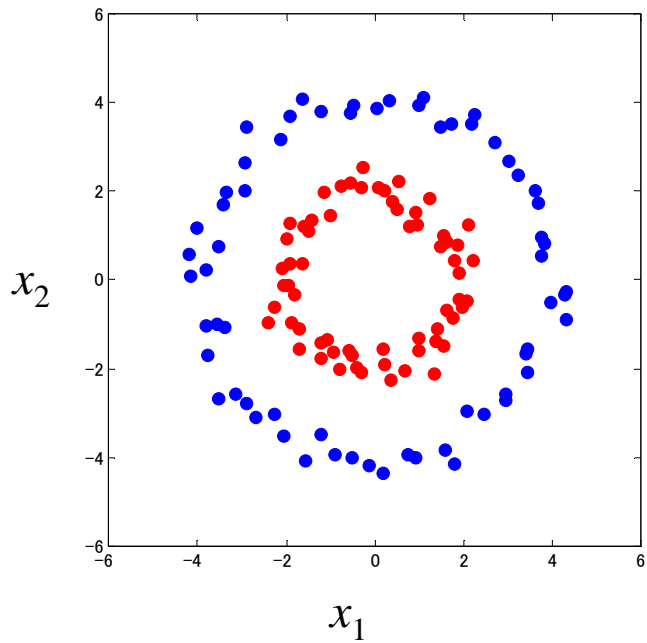
データの行列表現

$$m \text{ 次元 } N \text{ 点のデータ} \quad X = \begin{pmatrix} X_1^1 & \cdots & X_m^1 \\ X_1^2 & \cdots & X_m^2 \\ \vdots & & \vdots \\ X_1^N & \cdots & X_m^N \end{pmatrix}$$

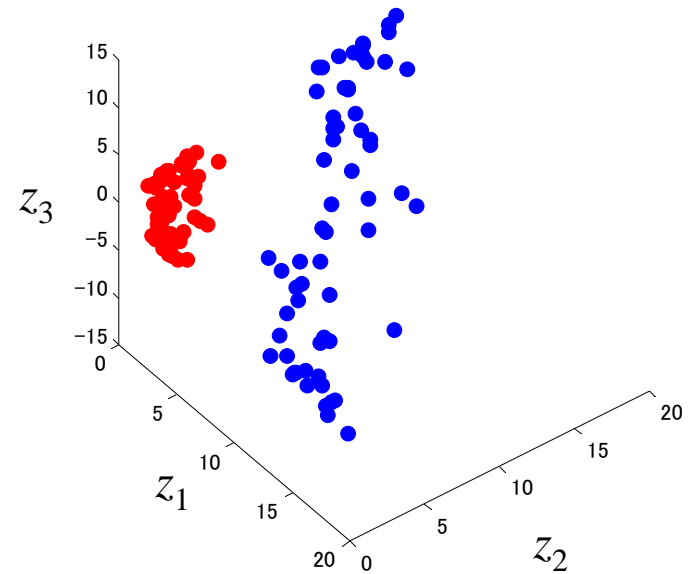
⇒ 線形の処理（主成分分析，正準相関分析，線形回帰...）

□ 線形で十分か？

線形識別不能



線形識別可能



$$(z_1, z_2, z_3) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$$

カーネル法の概略

- 線形手法の非線形化

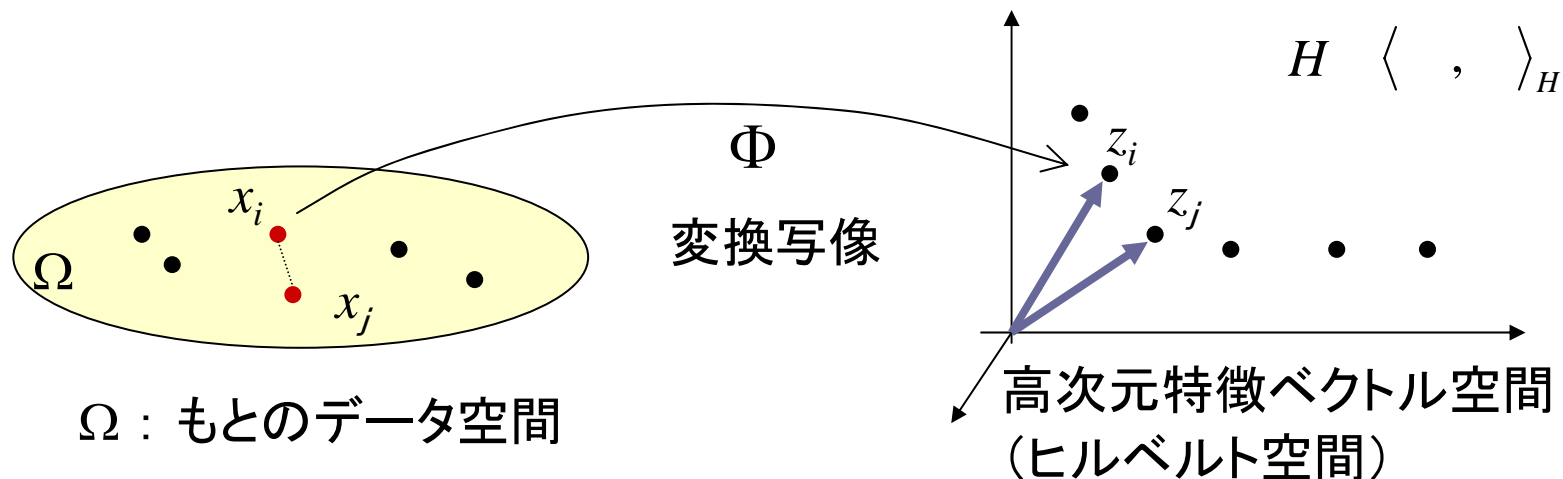
- 線形パラメータの推定 \Rightarrow 非線形関数の推定

線形の関数 $w^T X$  非線形関数 $f(X)$

- 非線形関数の処理をうまく扱えるようにしたのがカーネル法

■ もうひとつの見方： 高次元特徴空間でのデータ解析

- データを高次元空間（一般には無限次元）に変換し，解析しやすくする.
- $\Phi(x)$ はデータ x に対する**特徴ベクトル**と考えることができる
- もとの空間 Ω でなく、(高次元、または無限次元)**特徴空間** H でデータ解析を行う
- もとのデータ空間 Ω は、ベクトル空間でなくてもよい
データ x はツリー、グラフなども扱える → 構造化データ



■ 再生性

関数空間 H に属する (非線形) 関数の値が内積で計算できる

$$f(x) = \langle f, \Phi(x) \rangle_H$$

■ カーネルトリック

高次元 (無限次元) 関数空間 H において、**内積が容易に計算できる。**

さまざまな線形データ解析手法が H 上で適用可能

Support vector machine, Kernel PCA, Kernel CCA

$$\langle \Phi(x_i), \Phi(x_j) \rangle_H = k(x_i, x_j) \quad \dots \text{正定値カーネル}$$

データ x_i と x_j の**類似度**を定める

2. カーネル法の基礎

- このセクションの目的
 - カーネル法で用いられる基礎的な概念を述べる
 - 正定値カーネルの代表的な例を紹介する
 - 正定値カーネルがヒルベルト空間を定めることを説明する

正定値カーネル

■ 正定値カーネル

Ω : 集合. $k : \Omega \times \Omega \rightarrow \mathbf{R}$

$k(x,y)$ が Ω 上の **正定値カーネル** であるとは, 次の2つを満たすことをいう

1. (対称性) $k(x,y) = k(y,x)$
2. (正定値性) 任意の自然数 n と, 任意の Ω の点 x_1, \dots, x_n に対し,

$$n \times n \text{ 行列} \quad \left(k(x_i, x_j) \right)_{i,j=1}^n = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix}$$

が (半) 正定値. すなわち, 任意の実数 c_1, \dots, c_n に対し,

$$\sum_{i,j=1}^n c_i c_j k(x_i, x_j) \geq 0$$

□ 対称行列 $\left(k(x_i, x_j) \right)_{i,j=1}^n$ のことを, **グラム行列** と呼ぶ

■ 複素数値の正定値カーネル

$$k : \Omega \times \Omega \rightarrow \mathbb{C}$$

の場合にも正定値性は以下のように定義される.

任意の自然数 n と, 任意の Ω の点 x_1, \dots, x_n と, 任意の複素数 c_1, \dots, c_n に対し,

$$\sum_{i,j=1}^n c_i \bar{c}_j k(x_i, x_j) \geq 0 \quad (\bar{c}_j \text{ は複素共役})$$

が成り立つとき, $k(x,y)$ を正定値カーネルという.

- 上の条件から Hermite 性 $k(y, x) = \overline{k(x, y)}$ は自然に従う.
- 本講義では, 特に断らない限り, 実数値カーネルを考える.

正定値カーネルの例

- Euclid内積

$$\Omega = \mathbf{R}^m \quad k(x, y) = x^T y$$

- 多項式カーネル

$$\Omega = \mathbf{R}^m \quad k(x, y) = (x^T y + c)^d \quad (d: \text{自然数}, c \geq 0)$$

- ガウスカーネル (RBFカーネル)

$$\Omega = \mathbf{R}^m \quad k(x, y) = \exp\left(-\frac{1}{\sigma^2} \|y - x\|^2\right) \quad (\sigma > 0)$$

- 正定値であることはあとでチェックする

ヒルベルト空間の復習

■ (実)ヒルベルト空間

ベクトル空間で, 内積 $\langle \cdot, \cdot \rangle$ が与えられており, 完備性(コーシー列が常に収束する)を満たす.

□ ヒルベルト空間は, 無限次元でもよい

□ 例) $L^2(a,b)$: 区間 (a,b) 上の2乗可積分関数全体

$$\text{内積 } \langle f, g \rangle_{L^2(a,b)} = \int_a^b f(x)g(x)dx$$

□ ユークリッド空間 \mathbf{R}^m もヒルベルト空間のひとつ.

□ 複素ヒルベルト空間も定義されるが, 以下では実ヒルベルト空間を考える.

正定値カーネルとヒルベルト空間

定理1 (再生核ヒルベルト空間の構成)

$k(x,y)$: 集合 Ω 上の正定値カーネル



Ω 上の関数からなるヒルベルト空間 H_k が一意に存在して、
次の3つを満たす

(1) $k(\cdot, x) \in H_k$ ($x \in \Omega$ は任意に固定)

(2) 有限和 $f = \sum_{i=1}^n c_i k(\cdot, x_i)$ の形の元は H_k の中で稠密

(3) (再生性) $f(x) = \langle f, k(\cdot, x) \rangle \quad \forall f \in H_k, x \in \Omega$

注) $k(\cdot, x)$ \cdots x を固定した1変数関数

■ 再生核ヒルベルト空間 (Reproducing Kernel Hilbert Space, RKHS)

- 集合 Ω 上の関数を要素に持つヒルベルト空間 H が再生核ヒルベルト空間であるとは、任意の $x \in \Omega$ に対して $\phi_x \in H$ があって、任意の $f \in H$ に対し

$$\langle f, \phi_x \rangle = f(x) \quad (\text{再生性})$$

が成り立つことをいう。 (関数値が内積で与えられる)

ϕ_x のことを再生核という。

- 定理2 H : 集合 Ω 上の関数を要素に持つヒルベルト空間.
 H がRKHS



値を評価する汎関数 $e_x : H \rightarrow \mathbf{R}, f \mapsto f(x)$ が連続.

(注: Rieszの表現定理)

■ 正定値カーネルとRKHS

□ 正定値カーネル \Rightarrow RKHS

正定値カーネル $k(x,y)$ により定まる H_k は再生核を持つ(定理1)

$$\phi_x = k(\cdot, x) \quad \Rightarrow \quad \langle f, \phi_x \rangle = f(x)$$

□ RKHS \Rightarrow 正定値カーネル: 再生核 ϕ_x は正定値カーネルを定める

$$k(y, x) = \phi_x(y) \quad \text{と定義}$$

\Rightarrow

$$k(y, x) = \phi_x(y) = \langle \phi_x, \phi_y \rangle \quad (\text{対称性もわかる})$$

$$\begin{aligned} \sum_{i,j=1}^n c_i c_j k(x_i, x_j) &= \sum_{i,j=1}^n c_i c_j \langle \phi_{x_i}, \phi_{x_j} \rangle = \left\langle \sum_{i=1}^n c_i \phi_{x_i}, \sum_{j=1}^n c_j \phi_{x_j} \right\rangle \\ &= \left\| \sum_{i=1}^n c_i \phi_{x_i} \right\|^2 \geq 0 \quad (\text{正定値性}) \end{aligned}$$

□ 正定値カーネル \longleftrightarrow 再生核ヒルベルト空間

再生性とカーネルトリック

Ω : データが含まれる空間. $k(x,y)$: 集合 Ω 上の正定値カーネル

H_k : k により定まる再生核ヒルベルト空間

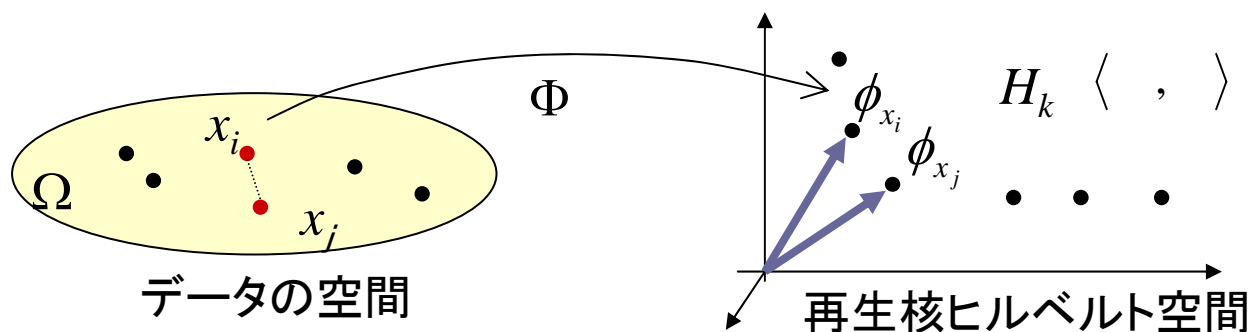
$\Phi: \Omega \rightarrow H_k$, $\Phi(x) = k(\cdot, x) = \phi_x$ により変換写像を定める

□ 関数の値: 内積計算でOK

$$f(x) = \langle f, \Phi(x) \rangle_H \quad \forall f \in H_k \quad \dots \text{再生性}$$

□ 内積計算: カーネル k の計算でOK (c.f. L^2 空間)

$$\langle \Phi(x), \Phi(y) \rangle = \langle k(\cdot, x), k(\cdot, y) \rangle = k(x, y) \quad \dots \text{カーネルトリック}$$



再生核ヒルベルト空間の性質

- 関数の値が扱える
 - L^2 空間などでは関数の値は定まらない
(測度0の集合上の値を変更しても同じベクトル)
- 再生性
 - 関数の値が内積で計算できる … 非線形関数が線形に扱える！
 - 内積がカーネルの値で計算できる … カーネルトリック
- 連続性／微分可能性 … L^2 よりずっと扱いやすい関数
 - $\Omega = \mathbb{R}^m$, 正定値カーネル $k(x, y)$ が連続だとすると,
定義されるRKHS H_k の関数はすべて連続関数

$$\begin{aligned} \therefore |f(x) - f(y)|^2 &= \left| \langle f, k(\cdot, x) - k(\cdot, y) \rangle \right|^2 \leq \|f\|^2 \|k(\cdot, x) - k(\cdot, y)\|^2 \\ &= \|f\|^2 (k(x, x) - 2k(x, y) + k(y, y)) \rightarrow 0 \quad (x \rightarrow y) \end{aligned}$$

実は, $k(x, y)$ が微分可能だと, すべての関数が微分可能

セクション2のまとめ

■ 正定値カーネルの定義

- グラム行列の(半)正定値性

■ 正定値カーネルの代表的な例

- 多項式カーネル $k(x, y) = (x^T y + c)^d$
- ガウスカーネル $k(x, y) = \exp\left(-\frac{1}{\sigma^2} \|y - x\|^2\right)$

■ 再生核ヒルベルト空間

- 正定値カーネルは, 特別な内積を持つ関数空間を定める
- 再生核ヒルベルト空間は都合のよい性質を持つ
 - 再生性, 関数の値が定まる, (場合によっては)連続性, 微分可能性

3. 線形アルゴリズムの非線形化 としてのカーネル法

- このセクションの目的
 - 線形なデータ解析法をカーネルによって非線形化する方法を紹介する
 - 具体的なカーネル化アルゴリズム
 - カーネルPCA
 - カーネルCCA
 - スプライン平滑化
 - サポートベクターマシン

PCAからカーネルPCAへ

■ 主成分分析 (PCA, 復習)

m 次元データ X_1, \dots, X_N

単位ベクトル a 方向の分散最大化

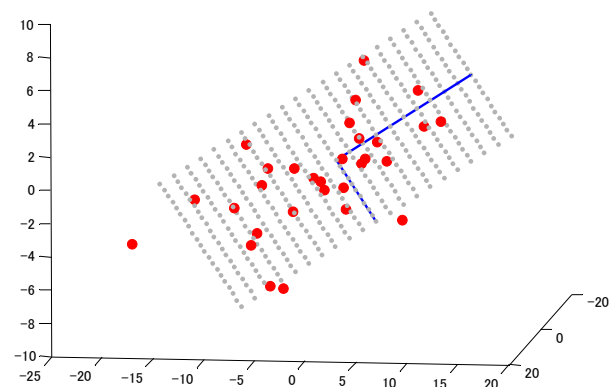
$$\text{Var}[a^T X] = \frac{1}{N} \sum_{i=1}^N \left(a^T (X_i - \frac{1}{N} \sum_j X_j) \right)^2 = a^T V a$$

$$V = \frac{1}{N} \sum_{i=1}^N \tilde{X}_i \tilde{X}_i^T \quad \text{分散共分散行列}$$

V の単位固有ベクトル u_1, u_2, \dots, u_m ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$)

第 p 主成分の軸 = u_p

データ X_j の第 p 主成分 = $u_p^T X_j$



■ PCAの非線形化

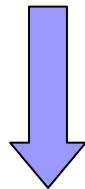
$$\max_{\|a\|=1} \text{Var}[a^T X] = \frac{1}{N} \sum_{i=1}^N \left(a^T X_i - \frac{1}{N} \sum_j a^T X_j \right)^2$$

非線形化



$$\max_{\|f\|=1} \text{Var}[f(X)] = \frac{1}{N} \sum_{i=1}^N \left(f(X_i) - \frac{1}{N} \sum_j f(X_j) \right)^2$$

f を探す関数空間として
RKHS H_k をとると, . . .



$$\max_{f \in H_k, \|f\|=1} \text{Var}[\langle f, \Phi(X) \rangle] = \frac{1}{N} \sum_{i=1}^N \left(\langle f, \Phi(X_i) \rangle - \frac{1}{N} \sum_j \langle f, \Phi(X_j) \rangle \right)^2$$

(再生性)

$f, \Phi(X_i)$ は H_k のベクトル $\Rightarrow H_k$ における「線形」な問題
特徴ベクトル $\Phi(X_1), \dots, \Phi(X_N)$ に対する H_k 内でのPCA

■ カーネルPCA (Schölkopf et al 98)

カーネル k を設定

特徴ベクトル $\Phi(X_1), \dots, \Phi(X_N)$ に対するPCA

$$\max_{f \in H_k, \|f\|=1} \text{Var}[\langle f, \Phi(X) \rangle] = \frac{1}{N} \sum_{i=1}^N (\langle f, \tilde{\phi}_i \rangle)^2$$

$$\text{ただし } \tilde{\phi}_i = \Phi(X_i) - \frac{1}{N} \sum_{j=1}^N \Phi(X_j)$$

$f = \sum_{i=1}^N \alpha_i \tilde{\phi}_i$ としてよい (\because 直交する方向は分散に寄与しない)

➡ 分散 $= \frac{1}{N} \sum_{a=1}^N \left\langle \sum_{j=1}^N \alpha_j \tilde{\phi}_j, \tilde{\phi}_a \right\rangle^2 = \frac{1}{N} \alpha^T \tilde{K}^2 \alpha$ ただし $\tilde{K}_{ij} = \langle \tilde{\phi}_i, \tilde{\phi}_j \rangle$

主成分は

$$\begin{cases} \max_{\alpha} \alpha^T \tilde{K}^2 \alpha \\ \text{制約条件 } \alpha^T \tilde{K} \alpha = 1 \end{cases} \iff \|f\|_{H_k}^2 = \langle \sum_i \alpha_i \tilde{\phi}_i, \sum_i \alpha_i \tilde{\phi}_i \rangle = \alpha^T \tilde{K} \alpha$$

■ カーネルPCAのアルゴリズム

\tilde{K} (中心化グラム行列)の計算

$$\tilde{K}_{ij} = \langle \tilde{\phi}_i, \tilde{\phi}_j \rangle = \left\langle \Phi(X_i) - \frac{1}{N} \sum_a \Phi(X_a), \Phi(X_j) - \frac{1}{N} \sum_b \Phi(X_b) \right\rangle$$

(カーネルトリック) $\rightarrow = k(X_i, X_j) - \frac{1}{N} \sum_{a=1}^N k(X_i, X_a) - \frac{1}{N} \sum_{a=1}^N k(X_a, X_j) + \frac{1}{N^2} \sum_{a,b=1}^N k(X_a, X_b)$

$$= (Q_N K Q_N)_{ij} \quad \text{ただし } Q_N = I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T, \quad \mathbf{1}_N = (1, \dots, 1)^T$$

分散最大の方向 = \tilde{K} の最大固有値の固有ベクトル方向

固有値分解 $\tilde{K} = \sum_{a=1}^N \lambda_a u^a u^{aT}$

第 p 軸(ノルム1) $\alpha^{(p)} = \frac{1}{\sqrt{\lambda_p}} u^p$ ($\alpha^T \tilde{K} \alpha = 1$ に注意)



$$\text{データ } X_j \text{ の第 } p \text{ 主成分} = \left\langle \sum_{i=1}^N \alpha_i^{(p)} \tilde{\phi}_i, \tilde{\phi}_j \right\rangle = \sqrt{\lambda_p} u_j^p$$

一般論：非線形化としてのカーネル法

■ 線形のアルゴリズム

データが \mathbb{R}^m のベクトル \rightarrow 線形回帰、主成分分析、正準相関分析 etc

- $a^T X$ として、パラメータ a を求める
- 相関、分散共分散行列による計算

■ カーネル法の2つの見方

- 非線形化

$$a^T X \quad \longrightarrow \quad f(X)$$

- 特徴ベクトルへの線形アルゴリズム

特徴ベクトル $\Phi(X_1), \dots, \Phi(X_N)$

■ カーネルによる非線形化

□ $a^T X$ の代わりに $f(X)$ (f はRKHS内の関数)

□ $f(X_i) = \langle f, \Phi(X_i) \rangle_{H_k}$ (再生性)

□ 多くの場合 $f = \sum_{i=1}^N \alpha_i \Phi(X_i) = \sum_{i=1}^N \alpha_i k(\cdot, X_i)$ でOK (representer定理)

$$\langle f, \Phi(X_i) \rangle_{H_k} = \sum_j \alpha_j \langle \Phi(X_j), \Phi(X_i) \rangle = \sum_j \alpha_j k(X_j, X_i)$$

(カーネルトリック)

□ グラム行列による計算

■ 特徴ベクトルへの線形アルゴリズムとしての見方

□ データ X_1, \dots, X_N \rightarrow 特徴ベクトル $\Phi(X_1), \dots, \Phi(X_N)$

□ RKHS 内での線形アルゴリズム

$\langle f, \Phi(X_i) \rangle$ として f を求める

□ 以下同様

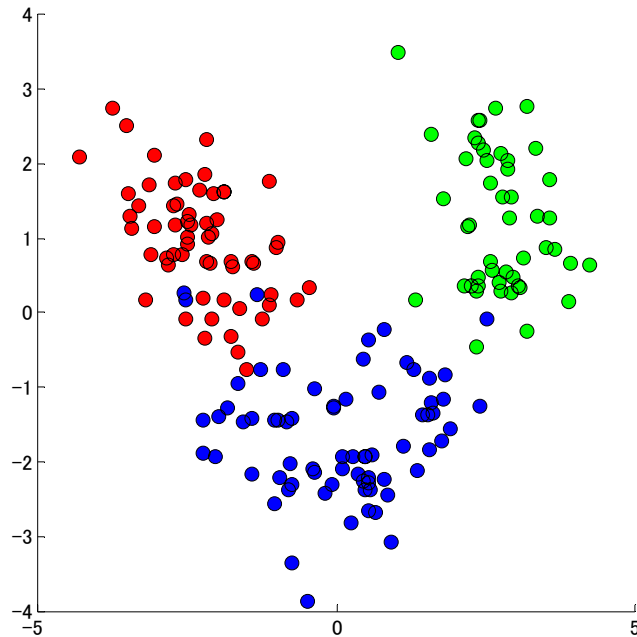
カーネルPCAの応用

■ カーネルPCAの実験例

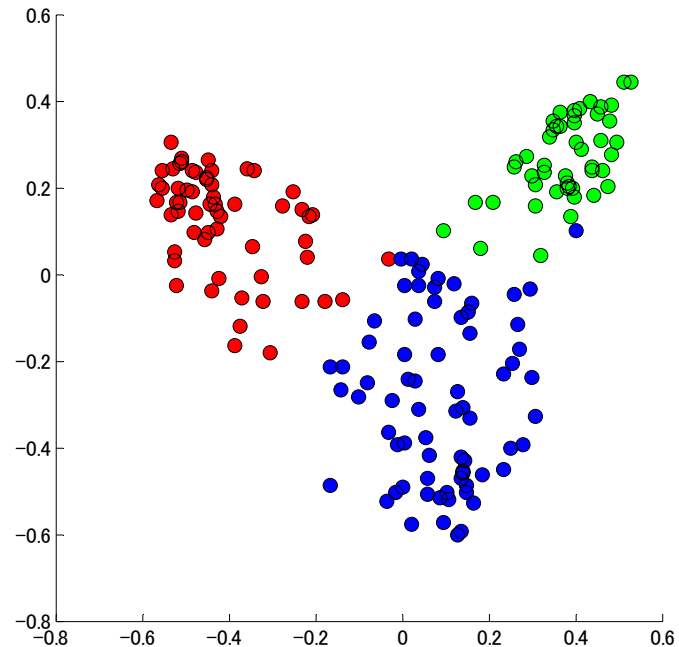
‘Wine’ データ (UCI Machine Learning Repository)

13次元, 178データ, 3種類のワインの属性データ

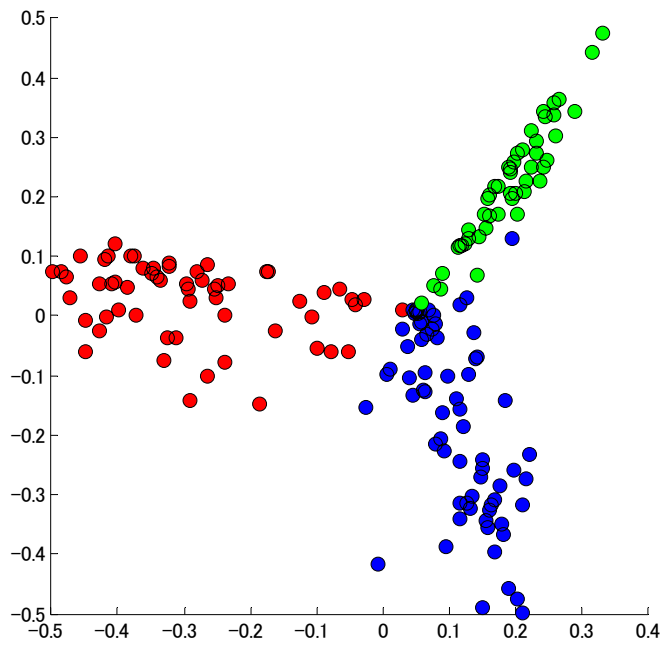
2つの主成分を取った (3クラスの色は参考に付けたもの)



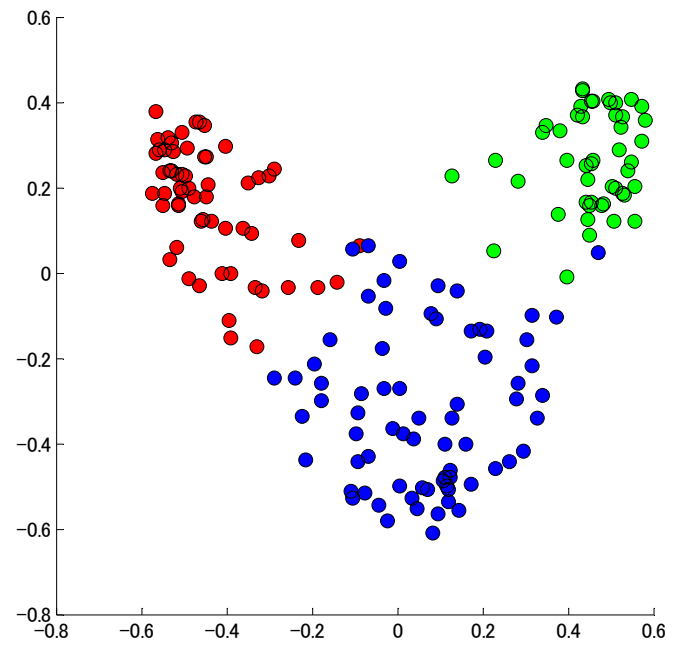
PCA (線形)



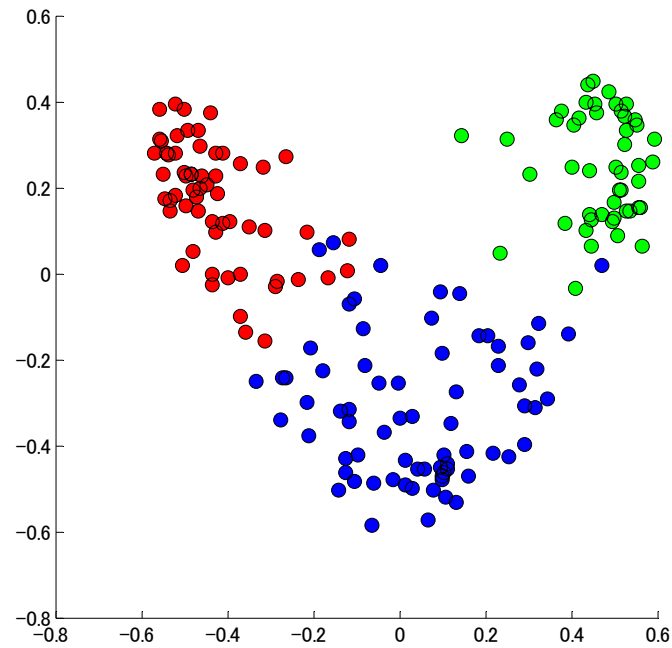
KPCA (RBF, $\sigma = 3$)



KPCA(RBF, $\sigma=2$)



KPCA(RBF, $\sigma=4$)



KPCA(RBF, $\sigma=5$)

■ カーネルPCAの特徴

- 非線形な方向でのデータのばらつきが扱える.
- 結果はカーネルの選び方に依存するので、解釈には注意が必要
ガウスカーネルの分散パラメータなど

どうやって選ぶか？ → 必ずしも明確でない、 目的に依存

- 前処理として使われることが多い
後の処理の結果を改良するための非線形特徴抽出

例えば、

カーネルPCA + 識別機(SVM) によるクラス識別問題

→ 最終的な識別の正答率を向上させるカーネル、パラメータがよい

カーネルPCAの雑音除去への応用

■ (カーネル)PCAによる雑音除去

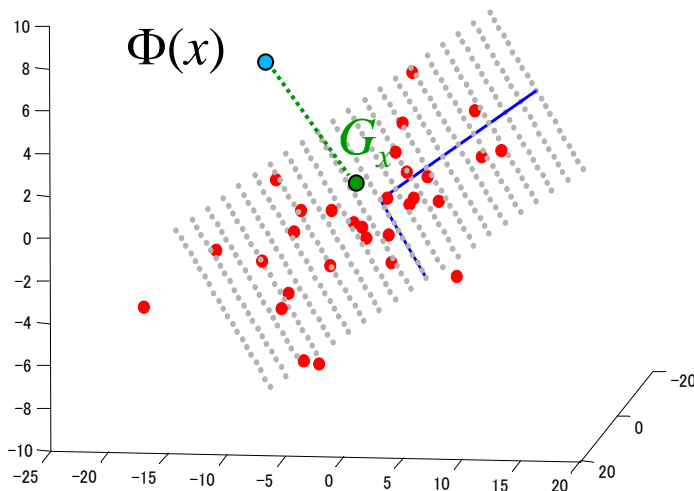
高次元の空間のなかで、 d 個の主成分の軸 F_1, F_2, \dots, F_d が張る d 次元部分空間へ、データ $\Phi(x)$ を射影した点を G_x とおく。

雑音除去された特徴ベクトル

G_x に最も近い埋め込み点を探す。

$$\hat{x} = \arg \min_{x'} \|\Phi(x') - G_x\|^2$$

カーネル $k(x_1, x_2)$ を使って表せる

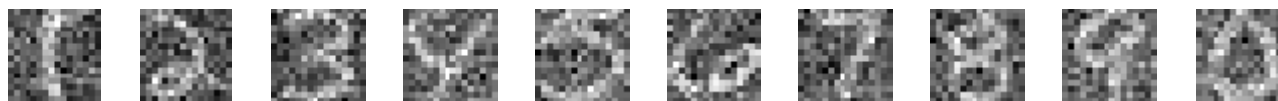


■ USPS 手書き数字データベース

16x16画素(256次元) 7291データ



元の画像 (データとしては使用していない)



ノイズつきの画像



ノイズ除去された画像 (linear PCA)



ノイズ除去された画像 (kernel PCA, Gaussカーネル)

Matlab stprtool (by V. Franc) により作成

CCAからカーネルCCAへ

■ 正準相関分析 (Canonical Correlation Analysis, CCA; 復習)

CCA ... 2種類の多次元データの相関を探る

m 次元データ X_1, \dots, X_N

n 次元データ Y_1, \dots, Y_N

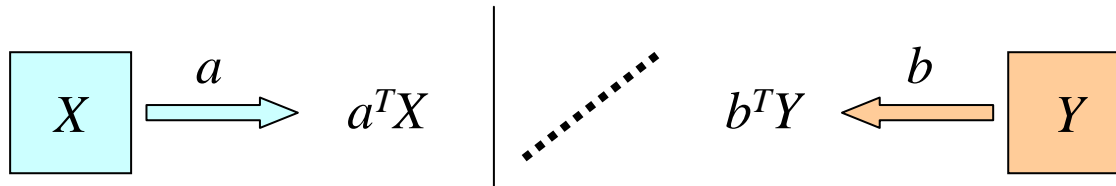
X を a , Y を b 方向に射影したときに相関を最大にする (a, b) を求める

正準相関

$$\rho = \max_{\substack{a \in \mathbf{R}^m \\ b \in \mathbf{R}^n}} \text{Corr}[a^T X, b^T Y] = \max_{\substack{a \in \mathbf{R}^m \\ b \in \mathbf{R}^n}} \frac{\frac{1}{N} \sum_i (a^T \tilde{X}_i)(b^T \tilde{Y}_i)}{\sqrt{\frac{1}{N} \sum_i (a^T \tilde{X}_i)^2} \sqrt{\frac{1}{N} \sum_i (b^T \tilde{Y}_i)^2}}$$

$$= \max_{\substack{a \in \mathbf{R}^m \\ b \in \mathbf{R}^n}} \frac{a^T V_{XY} b}{\sqrt{a^T V_{XX} a} \sqrt{b^T V_{YY} b}}$$

ただし $V_{XY} = \frac{1}{N} \sum_i \tilde{X}_i \tilde{Y}_i^T$ など



$$\rho = \max a^T V_{XY} b \quad \text{subj. to} \quad a^T V_{XX} a = b^T V_{YY} b = 1$$



Lagrange乗数法

$$\max a^T V_{XY} b + \frac{\mu}{2} (a^T V_{XX} a - 1) + \frac{\nu}{2} (b^T V_{YY} b - 1)$$



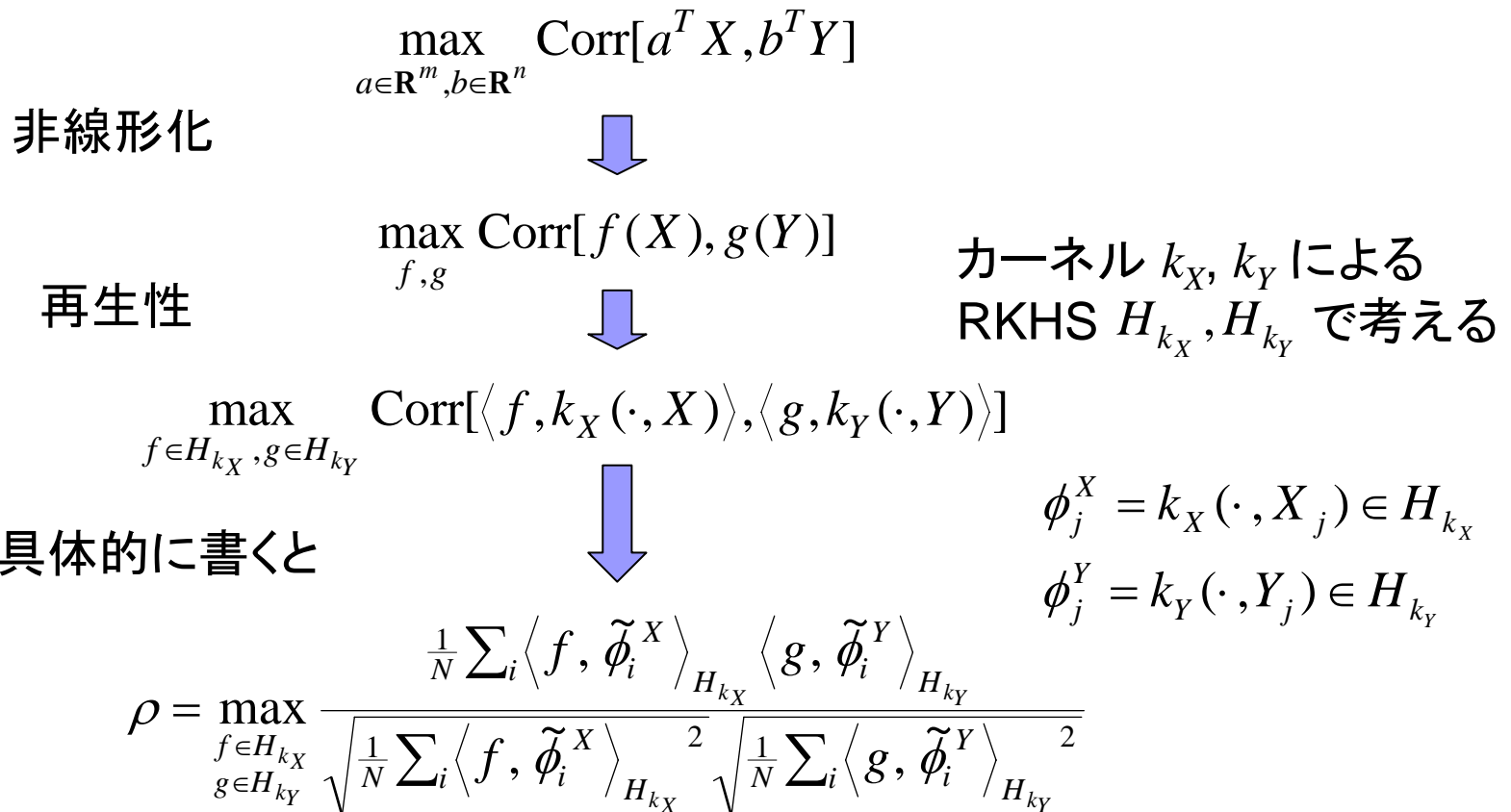
一般化固有値問題

$$\begin{pmatrix} \mathbf{O} & V_{XY} \\ V_{YX} & \mathbf{O} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \rho \begin{pmatrix} V_{XX} & \mathbf{O} \\ \mathbf{O} & V_{YY} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

$$[\mu = \nu = -\rho]$$

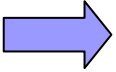
最大固有値, 固有ベクトルを求めればよい

■ カーネルCCA (Akaho 2001, Bach and Jordan 2002)



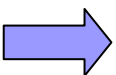
特徴ベクトル $\phi_1^X, \dots, \phi_N^X; \phi_1^Y, \dots, \phi_N^Y$ に対するCCA

カーネルPCA同様 $f = \sum_{l=1}^N \alpha_l \tilde{\phi}_l^X$, $g = \sum_{l=1}^N \beta_l \tilde{\phi}_l^Y$ としてよい.



 (カーネルトリック)

$$\rho = \max_{\substack{\alpha \in \mathbf{R}^N \\ \beta \in \mathbf{R}^N}} \frac{\alpha^T \tilde{K}_X \tilde{K}_Y \beta}{\sqrt{\alpha^T \tilde{K}_X^2 \alpha} \sqrt{\beta^T \tilde{K}_Y^2 \beta}} \quad \tilde{K}_X, \tilde{K}_Y : \text{中心化グラム行列}$$



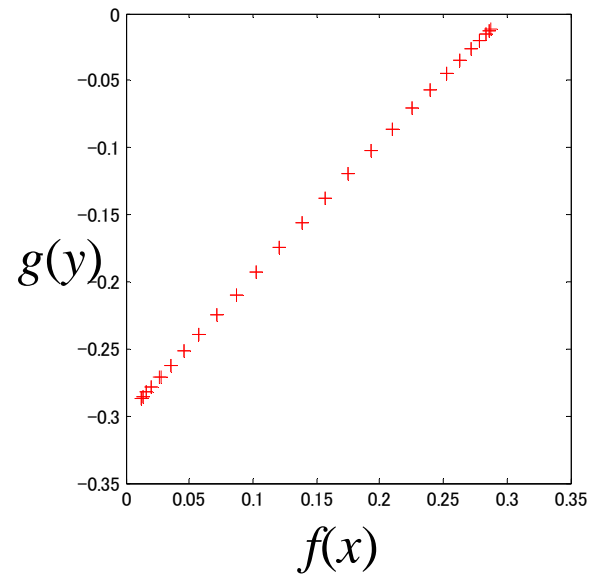
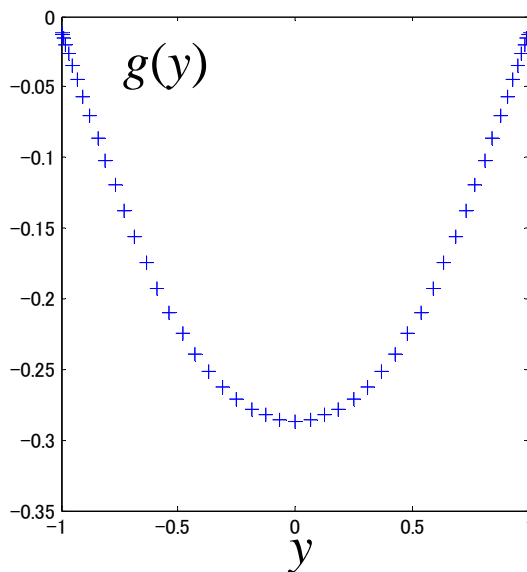
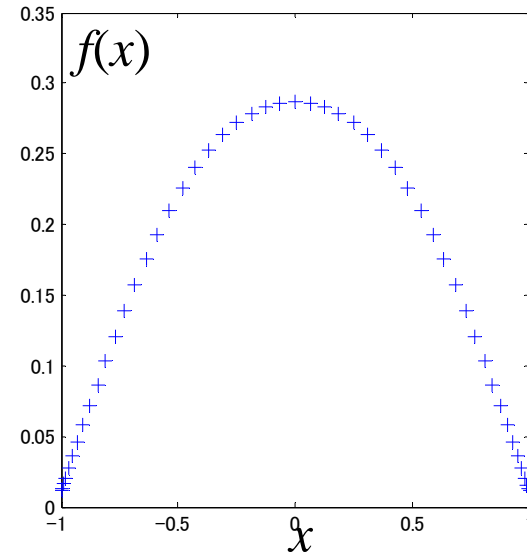
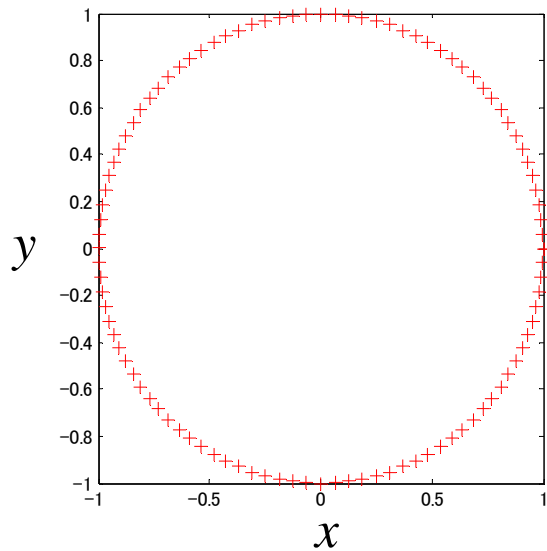
一般化固有値問題として解ける

$$\begin{pmatrix} O & \tilde{K}_X \tilde{K}_Y \\ \tilde{K}_Y \tilde{K}_X & O \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \rho \begin{pmatrix} (\tilde{K}_X + N\varepsilon_N I_N)^2 & O \\ O & (\tilde{K}_Y + N\varepsilon_N I_N)^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

注) 自明な解を避けるため正則化係数 ε_N が必要
 適当なオーダーで $\varepsilon_N \rightarrow 0$ とすると, 収束性が保証される
 (Fukumizu et al. 2006)

■ カーネルCCAの実験例

ガウスクーネル



リッジ回帰とスプライン平滑化

■ リッジ回帰(復習)

$$(X^1, Y^1), \dots, (X^N, Y^N) \quad X^i \in \mathbf{R}^m, Y^i \in \mathbf{R}$$

問題 $\min \sum_{i=1}^N (Y^i - f_w(X^i))^2 + \lambda \|w\|^2$ を達成する線形関数 $f_w(x) = w^T x$ を求めよ

データ行列 $X = \begin{pmatrix} X_1^1 & \dots & X_m^1 \\ X_1^2 & \dots & X_m^2 \\ \vdots & & \vdots \\ X_1^N & \dots & X_m^N \end{pmatrix}, \quad Y = \begin{pmatrix} Y^1 \\ Y^2 \\ \vdots \\ Y^N \end{pmatrix}$ を使うと,

最適解は $\hat{w} = (X^T X + \lambda I_N)^{-1} X^T Y$

最適な関数は $f_{\hat{w}}(x) = Y^T X (X^T X + \lambda I_N)^{-1} x$

- リッジ回帰は、 $X^T X$ が特異になるときに特に有効
- Bayes的な解釈もできる(縮小推定)

■ 関数ノルムによる平滑化

リッジ回帰 $\min \sum_{i=1}^N (Y^i - f_w(X^i))^2 + \lambda \|w\|^2$

カーネル k
を用意



非線形化 $\min_{f \in H} \sum_{i=1}^N (Y^i - f(X^i))^2 + \lambda \|f\|_H^2 \quad \dots \text{平滑化}$

H : 正定値カーネル k により定まる再生核ヒルベルト空間
 $\|f\|_H$ が小さい \rightarrow 関数が滑らか

解の構成

Representer定理より $f(x) = \sum_{i=1}^N \alpha_i k(x, X^i)$ の形なので、

$$\min_{\alpha} \|Y - K\alpha\|^2 + \lambda \alpha^T K \alpha \quad \text{を解けばよい}$$

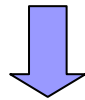
解 $f(x) = Y^T (K + \lambda I_N)^{-1} g(x)$

$$K_{ij} = k(X^i, X^j) \quad \text{グラム行列}$$
$$g_i(x) = k(x, X^i)$$

■ 正則化による関数の推定

2乗誤差によるカーブフィッティング

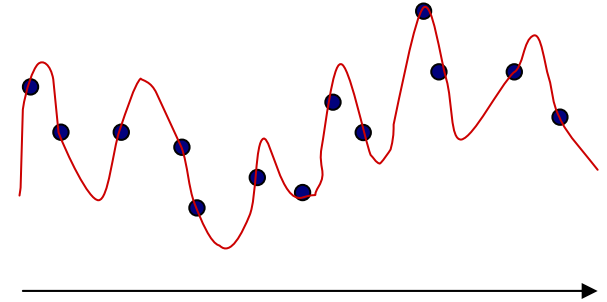
$$\min_f \sum_{i=1}^N (Y^i - f(X^i))^2$$



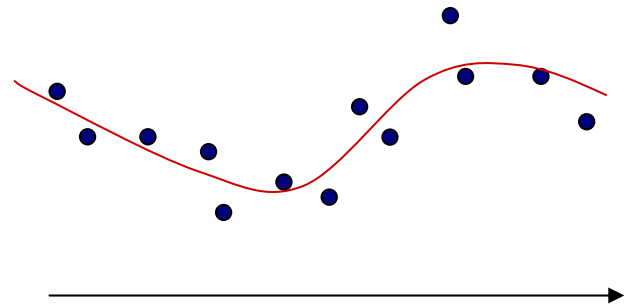
正則化

$$\min_f \sum_{i=1}^N (Y^i - f(X^i))^2 + \boxed{\Psi(f)}$$

誤差 0 となる f は
無数にある



解が一意になるように正則化項を付加



■ スプライン平滑化：関数の滑らかさによる正則化

□ Cubic spline

$$\min_f \sum_{i=1}^N (Y^i - f(X^i))^2 + \lambda \int \left| \frac{df(x)}{dx} \right|^2 dx$$

微分による正則化項 …… 正定値カーネルと密接な関係

2日目に詳しく述べる

■ サポートベクトルマシン(SVM)

SVMもRKHSノルムによる正則化と考えられる

$$\min_{f,b} \sum_{i=1}^N (1 - Y^i (f(X^i) + b))_+ + \lambda \|f\|_H^2$$

ただし2乗誤差ではない

午後の部(赤穂氏担当)を参照

セクション3のまとめ

■ カーネルによる非線形化

□ カーネル化 (kernelization)

- 線形関数 $w^T X \Rightarrow$ 非線形関数 $f(X) \Rightarrow$ 再生性 $f(X) = \langle f, \Phi(X) \rangle$
- RKHS内の特徴ベクトルに対する線形手法とも考えられる

□ データとパラメータの内積を使って表される線形手法なら容易に拡張が可能 (射影, 相関, 分散共分散)

- ### □ 例: カーネルPCA, カーネルCCA, サポートベクターマシン, スプライン平滑化, カーネルFisher判別分析, カーネルPLS など (PLS: partial least square)

■ 非線形アルゴリズムの特徴

- 線形ではとらえられない性質が調べられる.
- とらえることのできる非線形性はカーネルの選び方に影響を受ける