



再生核ヒルベルト空間を用いた非線形データ解析法

情報・システム研究機構 統計数理研究所
(兼)総合研究大学院大学

福水 健次

Outline

- イントロダクション
- 正定値カーネルと再生核ヒルベルト空間
- 正定値カーネルによるデータ解析 ～ カーネル法 ～
- 独立性、条件付独立性と再生核ヒルベルト空間
- カーネル正準相関分析の統計的性質
- おわりに-

イントロダクション

■ 線形データの解析

□ 古典的なデータ解析

データの行列表現

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

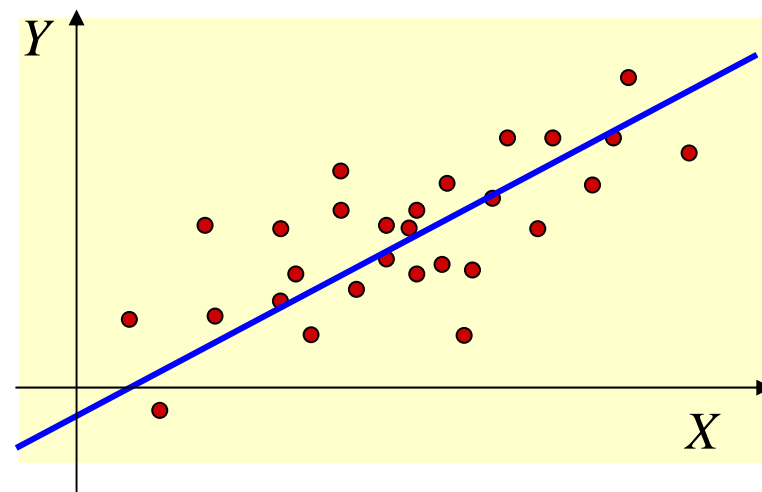
⇒ 線形の処理（主成分分析，正準相関分析，線形回帰...）

■ 線形データ解析の例

□ 線形回帰分析

問題

データを説明するのに最適な
超平面 $y = a^T x + b$ を求めよ

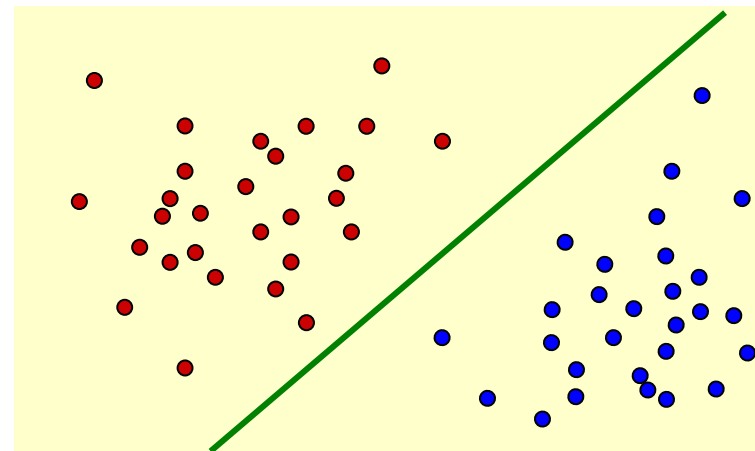


□ 線形識別 (2クラス)

問題

赤: クラス1に属するデータ
青: クラス2に属するデータ

2クラスを識別するのに最適な
超平面 $a^T x + b = 0$ を求めよ



□ 線形回帰の解

- 最適性の基準設定 → 最小2乗誤差

$$\min_a \sum_{i=1}^N |Y^i - a^T X^i|^2 \quad (\text{簡単のため } b \text{ 省略})$$

→ $\min_a [Y^T Y - 2Y^T X a + a^T X^T X a]$

→ $\hat{a} = (X^T X)^{-1} X^T Y$

このとき、2乗誤差 = $Y^T Y - Y^T X (X^T X)^{-1} X^T Y$

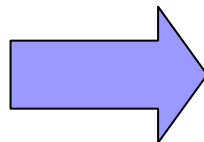
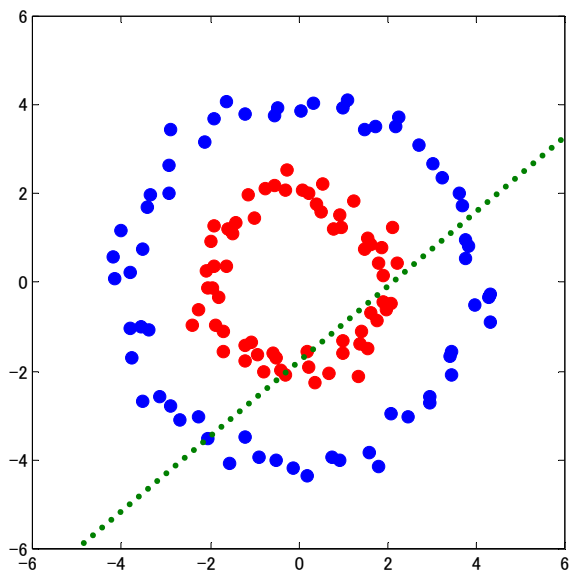
□ 線形なデータ解析を支えるもの

- 内積(相関)の計算 + 線形代数 + α

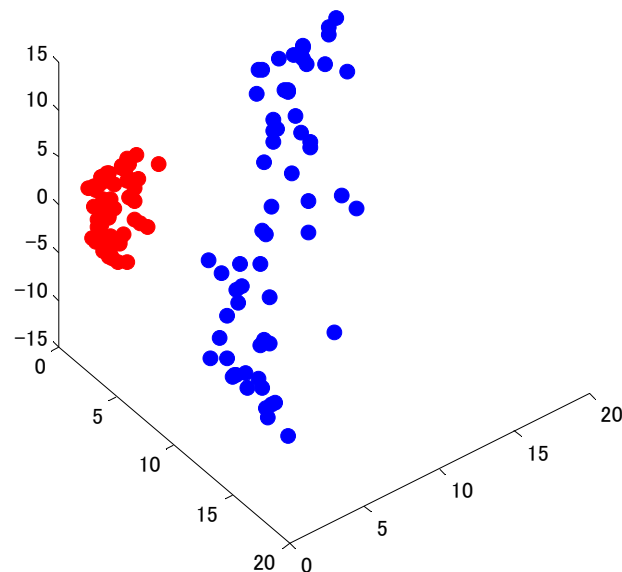
■ 線形で十分か？

■ 非線形データ解析の重要性

線形識別 **不能**



線形識別 **可能**

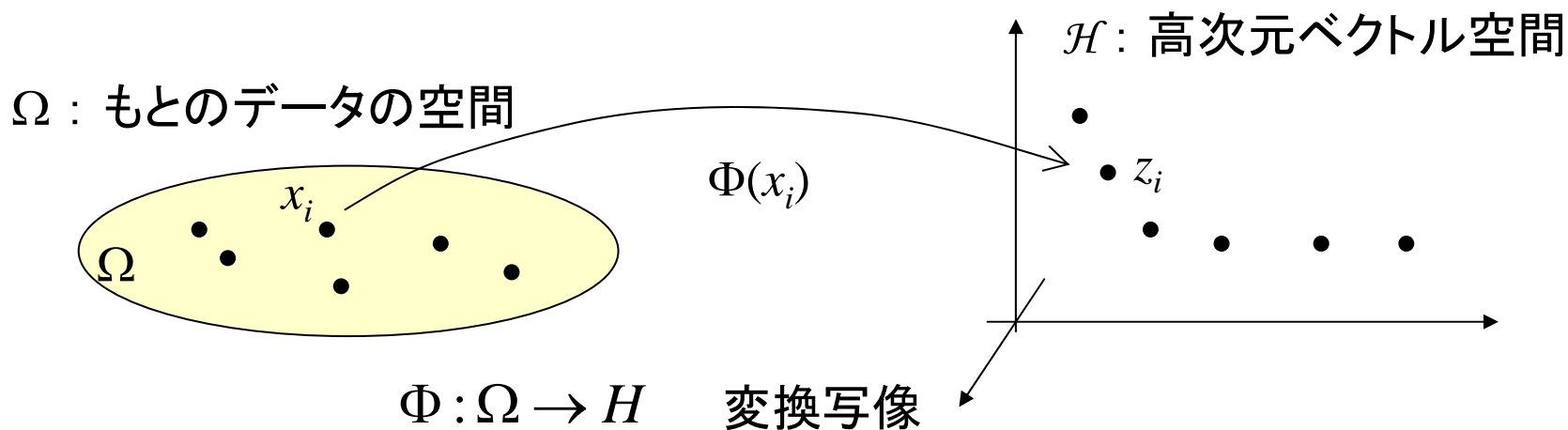


$$(z_1, z_2, z_3) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$$

■ カーネル法の概略

□ 高次元空間への写像を構成する方法論

データを高次元のヒルベルト空間(一般には無限次元)へ写像し、解析しやすいデータに変換する。



□ カーネル法:

- H 内のデータに変換した後で線形の手法を用いる
- H として再生核ヒルベルト空間を用いる

正定値カーネルと再生核ヒルベルト空間

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,

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

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

■ 正定値カーネルの例

□ 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

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

□ Fourierカーネル(複素数値)

$$\Omega = \mathbf{R}^m \quad k(x, y) = e^{\sqrt{-1}\omega^T (x-y)} \quad (\omega \in \mathbf{R}^m)$$

再生核ヒルベルト空間

- 集合 Ω 上の関数を要素に持つヒルベルト空間 \mathcal{H} が再生核ヒルベルト空間 (reproducing kernel Hilbert space, RKHS) であるとは, 任意の $x \in \Omega$ に対して $\phi_x \in \mathcal{H}$ があって, 任意の $f \in \mathcal{H}$ に対し

$$\langle f, \phi_x \rangle = f(x)$$

が成り立つことをいう.

- ϕ_x のことを再生核という.

正定値カーネルと再生核ヒルベルト空間

■ 定理

$k(x,y)$: 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 \mathcal{H}_k が一意に存在して、
次の3つを満たす

(1) $k(\cdot, x) \in \mathcal{H}_k$ ($x \in \Omega$ は任意に固定)

(2) 有限和 $f = \sum_{i=1}^n c_i k(\cdot, x_i)$ の形の関数全体は \mathcal{H}_k の中で稠密

(3) (再生性) $f(x) = \langle f, k(\cdot, x) \rangle \quad \forall f \in \mathcal{H}_k, x \in \Omega$

注) $k(\cdot, x) \cdots x$ を固定した1変数関数

■ 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 \mathcal{H}_k は再生核を持つ(定理の(3))

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

- 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

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

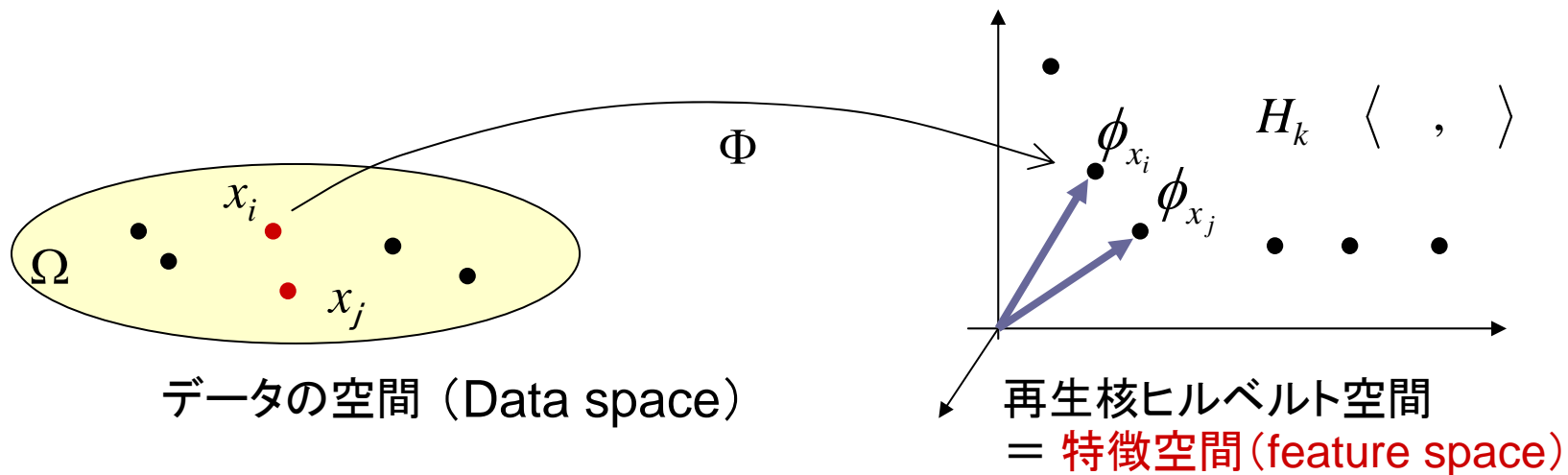
$$\begin{aligned} \sum_{i,j=1}^n c_i c_j k(x_i, x_j) &= \sum_{i,j=1}^n c_i c_j \langle \phi_{x_i}, \phi_{x_j} \rangle = \left\langle \sum_{i=1}^n c_i \phi_{x_i}, \sum_{j=1}^n c_j \phi_{x_j} \right\rangle \\ &= \left\| \sum_{i=1}^n c_i \phi_{x_i} \right\|^2 \geq 0 \quad (\text{正定値性}) \end{aligned}$$

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

正定値カーネルによるデータ解析

■ カーネルによるデータ変換

- データの非線形変換 $x \mapsto \Phi(x) = k(\cdot, x) = \phi_x$



- 内積計算: $\langle \Phi(x), \Phi(y) \rangle = \langle k(\cdot, x), k(\cdot, y) \rangle = k(x, y) \dots$ カーネルトリック
- 特徴空間における線形アルゴリズム
→ データの空間での非線形アルゴリズム
カーネルPCA, カーネルCCA,
SVM, プライン平滑化 etc

カーネル法の例 – カーネルCCA –

■ 正準相関分析 (canonical correlation analysis, CCA)

2つの確率ベクトル X, Y に対して、相関が最大となる方向を見つける

$$\max_{\substack{a \in \mathbf{R}^m \\ b \in \mathbf{R}^n}} \frac{\text{Cov}[a^T X, b^T Y]}{(\text{Var}[a^T X] \text{Var}[b^T Y])^{1/2}} = \max_{\substack{a \in \mathbf{R}^m \\ b \in \mathbf{R}^n}} \frac{a^T C_{XY} b^T}{\sqrt{a^T C_{XX} a} \sqrt{b^T C_{YY} b}}$$
$$C_{YX} = E[(Y - E[Y])(X - E[X])^T]$$



データ $X_1, \dots, X_N; Y_1, \dots, Y_N$ に対して

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

$$\text{ただし } \hat{C}_{XY} = \frac{1}{N} \sum_{i=1}^N \left(X^i - \frac{1}{N} \sum_{j=1}^N X^j \right) \left(Y^i - \frac{1}{N} \sum_{j=1}^N Y^j \right)^T$$

■ カーネルCCA

Ω_X, Ω_Y 上の正定値カーネル k_X, k_Y .

$$\begin{array}{ccc} X_1, \dots, X_N & \longrightarrow & \phi_1^X, \dots, \phi_N^X \in \mathcal{H}_X & \phi_i^X = k_X(\cdot, X_i), \\ Y_1, \dots, Y_N & \text{変換} & \phi_1^Y, \dots, \phi_N^Y \in \mathcal{H}_Y & \phi_i^Y = k_Y(\cdot, Y_i) \end{array}$$

$$\tilde{\phi}_i^X = k_X(\cdot, X_i) - \frac{1}{N} \sum_{j=1}^N k_X(\cdot, X_j), \quad \tilde{\phi}_i^Y = k_Y(\cdot, Y_i) - \frac{1}{N} \sum_{j=1}^N k_Y(\cdot, Y_j)$$

とにおいて関数データに対するCCAを書き下すと

$$\max_{\substack{f \in \mathcal{H}_X \\ g \in \mathcal{H}_Y}} \frac{\frac{1}{N} \sum_{i=1}^N \langle f, \tilde{\phi}_i^X \rangle_{\mathcal{H}_X} \langle g, \tilde{\phi}_i^Y \rangle_{\mathcal{H}_Y}}{\sqrt{\frac{1}{N} \sum_{i=1}^N \langle f, \tilde{\phi}_i^X \rangle_{\mathcal{H}_X}^2 + \varepsilon_N \|f\|_{\mathcal{H}_X}^2} \sqrt{\frac{1}{N} \sum_{i=1}^N \langle g, \tilde{\phi}_i^Y \rangle_{\mathcal{H}_Y}^2 + \varepsilon_N \|g\|_{\mathcal{H}_Y}^2}}$$

■ グラム行列による表現

$$f \in \text{Span} \left\{ \tilde{\phi}_i^X \right\}_{i=1}^N \quad g \in \text{Span} \left\{ \tilde{\phi}_i^Y \right\}_{i=1}^N \quad \text{で十分}$$

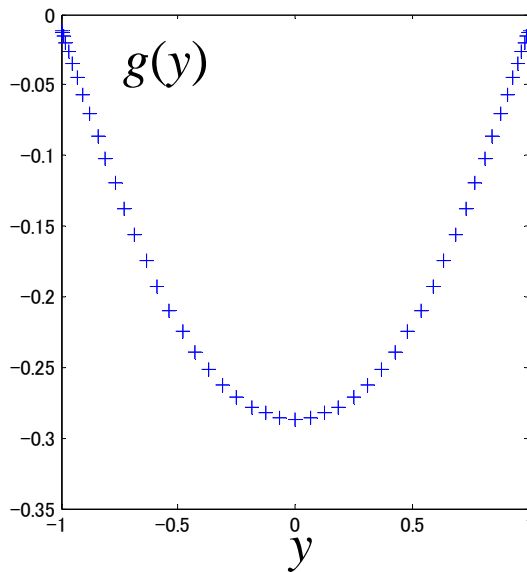
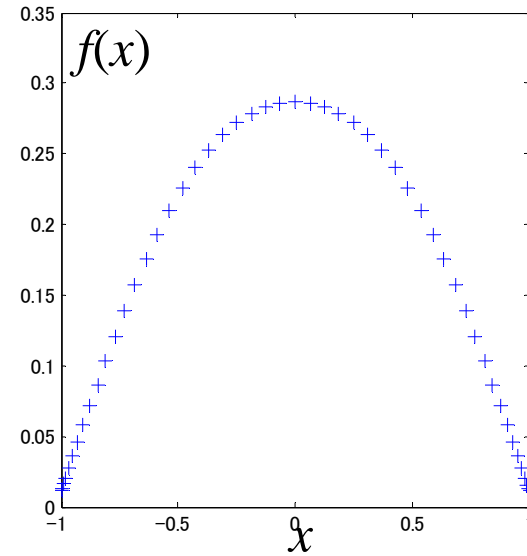
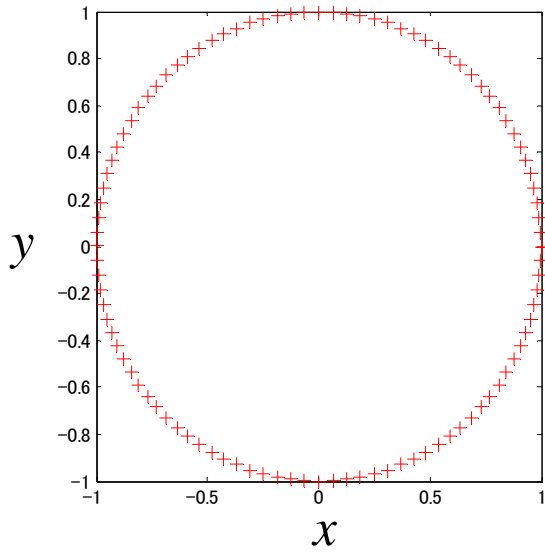


$$f = \sum_{i=1}^N a_i \tilde{\phi}_i^X, \quad g = \sum_{i=1}^N b_i \tilde{\phi}_i^Y \quad \text{により表すと}$$

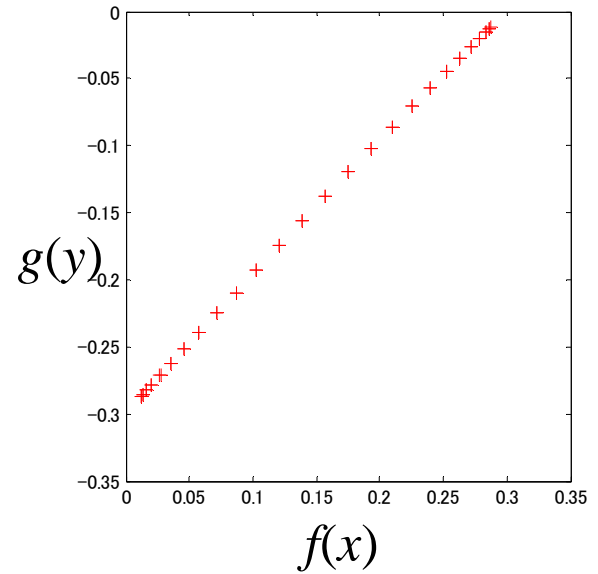
$$\max_{\substack{a \in \mathbf{R}^N \\ b \in \mathbf{R}^N}} \frac{a^T \tilde{K}_X \tilde{K}_Y b}{\sqrt{a^T \left(\tilde{K}_X^2 + N \varepsilon_N \tilde{K}_X \right) a} \sqrt{b^T \left(\tilde{K}_Y^2 + N \varepsilon_N \tilde{K}_Y \right) b}}$$

→ 固有値問題として解ける

$$\begin{aligned} \text{ただし } \tilde{K}_{X,ij} &= k(X_i, X_j) - \frac{1}{N} \sum_{a=1}^N k(X_i, X_a) \\ &\quad - \frac{1}{N} \sum_{a=1}^N k(X_a, X_j) + \frac{1}{N^2} \sum_{a,b=1}^N K(X_a, X_b) \end{aligned}$$



Gaussian kernel
 $\exp(-\|z - \tilde{z}\|^2 / \sigma^2)$
 is used.



データ解析と関数空間

■ 関数空間内での線形データ解析

- 線形データ解析のために必要な性質
 - 内積(相関)の計算 → ヒルベルト空間
 - 「関数の値」が決まる ← 処理結果の利用
- 定理: 集合 Ω 上の関数を要素に持つヒルベルト空間 \mathcal{H} に対し、任意の $x \in \Omega$ に対して

$$H \rightarrow \mathbf{R}, \quad f \mapsto f(x)$$

が連続汎関数であるとき、 \mathcal{H} は再生核ヒルベルト空間である

- RKHSは、データ解析を行うのに自然な関数空間
c.f. L^2 空間

■ RKHSを使う利点

□ 数理的な観点から

- k が C^r 級ならば \mathcal{H}_k に属する関数は C^r 級
- $E_p[k(X,X)] < \infty$ ならば $\mathcal{H}_k \subseteq L^2(P)$ かつ包含写像は連続

□ 情報处理的観点から

- 関数空間での内積計算が容易

$$f = \sum_i a_i k(\cdot, X_i), g = \sum_j b_j k(\cdot, X_j) \iff \langle f, g \rangle = \sum_{ij} a_i b_j k(X_i, X_j)$$

- 従来の線形手法の非線形拡張が容易： 相関行列 \rightarrow グラム行列
- 非ベクトルデータの処理

もとのデータ空間はベクトル空間でなくてもよい

\rightarrow グラフデータ、ツリーデータ、ヒストグラムなどの処理

独立性、条件付独立性と 再生核ヒルベルト空間

独立性とRKHS

■ 特性関数による独立性の特徴づけ(確率論の復習)

確率ベクトル X と Y が独立

$$\Leftrightarrow E_{XY} \left[e^{\sqrt{-1}\omega^T X} e^{\sqrt{-1}\eta^T Y} \right] = E_X \left[e^{\sqrt{-1}\omega^T X} \right] E_Y \left[e^{\sqrt{-1}\eta^T Y} \right] \quad \text{for all } \omega \text{ and } \eta.$$

$$\Leftrightarrow \text{Cov} \left[e^{\sqrt{-1}\omega^T X}, e^{\sqrt{-1}\eta^T Y} \right] = 0 \quad \text{for all } \omega \text{ and } \eta.$$

$e^{\sqrt{-1}\omega^T x}$, $e^{\sqrt{-1}\eta^T y}$ は様々な関数による非線形相関を調べるテスト関数

■ RKHSによる独立性の特徴づけ

$\mathcal{H}_X, \mathcal{H}_Y$: 可測空間 Ω_X, Ω_Y 上のRKHS (有界で可測な正定値カーネル)

X : Ω_X 上に値を持つ確率変数, Y : Ω_Y 上に値を持つ確率変数

X と Y が独立

$$\Leftrightarrow E_{XY}[f(X)g(Y)] = E_X[f(X)]E_Y[g(Y)] \quad \text{for all } f \in \mathcal{H}_X, g \in \mathcal{H}_Y$$

という特徴づけは可能か？

□ 確率決定的

\mathcal{H} : 可測空間 (Ω, \mathcal{B}) 上の有界で可測な正定値カーネルによるRKHS

Π : (Ω, \mathcal{B}) 上の確率測度全体

\mathcal{H} が確率決定的であるとは

$$\Pi \rightarrow \mathcal{H}^*, \quad P \mapsto (f \mapsto E_P[f(X)])$$

が単射であることをいう。

定理1

直積 $\mathcal{H}_X \otimes \mathcal{H}_Y$ ($k_X k_Y$ で定まるRKHS)が確率決定的であるとき、
 X と Y が独立

$$\Leftrightarrow E_{XY}[f(X)g(Y)] = E_X[f(X)]E_Y[g(Y)]$$

for all $f \in \mathcal{H}_X, g \in \mathcal{H}_Y$

定理2 (Bach and Jordan, 2002)

(\mathbf{R}^m, B_m) に対して、ガウスRBFカーネル

$$k(z, \tilde{z}) = \exp\left(-\frac{\|z - \tilde{z}\|^2}{\sigma^2}\right)$$

は確率決定的である。

□ 独立成分分析への応用

独立な m 個の確率変数の線形変換像から、もとの確率変数を復元

相互共分散作用素

■ 相互共分散作用素の定義

X, Y : 可測空間 Ω_X, Ω_Y に値をとる確率変数

$\mathcal{H}_X, \mathcal{H}_Y$: Ω_X, Ω_Y 上のRKHS (有界で可測な正定値カーネル)

有界作用素 $\Sigma_{YX} : \mathcal{H}_X \rightarrow \mathcal{H}_Y$ を次式により定義することができる。

$$\langle g, \Sigma_{YX} f \rangle_{\mathcal{H}_Y} = E_{XY}[f(X)g(Y)] - E_X[f(X)]E_Y[g(Y)] \quad (= \text{Cov}[f(X), g(Y)])$$

for all $f \in \mathcal{H}_X, g \in \mathcal{H}_Y$

c.f. $a^T C_{YX} b = \text{Cov}[a^T X, b^T Y]$

系3 $\mathcal{H}_X, \mathcal{H}_Y$: ガウスカーネルによるRKHS

$$X \text{ と } Y \text{ が独立} \quad \Leftrightarrow \quad \Sigma_{YX} = 0$$

条件付共分散作用素

■ 共分散作用素の分解

相互共分散作用素 $\Sigma_{YX} : \mathcal{H}_X \rightarrow \mathcal{H}_Y$ は以下のように分解できる

$$\Sigma_{YX} = \Sigma_{YY}^{1/2} V_{YX} \Sigma_{XX}^{1/2}$$

ただし V_{YX} は有界作用素、 $\|V_{YX}\| \leq 1$ かつ

$$\overline{R(V_{YX})} = \overline{R(\Sigma_{YY})}, \quad N(V_{YX}) = N(\Sigma_{XX})$$

■ 条件付共分散作用素

$$\begin{aligned} \Sigma_{YY|X} &\equiv \Sigma_{YY} - \Sigma_{YY}^{1/2} V_{YX} V_{XY} \Sigma_{YY}^{1/2} \\ &\left(= \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \right) \end{aligned}$$

条件付独立性とRKHS

■ 条件付共分散作用素と2乗誤差

定理4

$$\left\langle g, \Sigma_{YY|X} g \right\rangle_{\mathcal{H}_Y} = \inf_{f \in \mathcal{H}_X} E_{XY} |(g(Y) - E[g(Y)]) - (f(X) - E[f(X)])|^2$$

c.f. 線形回帰の2乗誤差

$$\min_{a \in \mathbf{R}^m, b \in \mathbf{R}} E_{YX} |Y - (a^T X + b)|^2 = C_{YY} - C_{YY} C_{XX}^{-1} C_{YY}$$

■ 条件付分散による表現

定理5 (1) 任意の $g \in \mathcal{H}_Y$ に対して $E[g(Y) | X = \cdot] \in L^2(P_X)$

(2) $\mathcal{H}_X \oplus \mathbf{R}$ は $L^2(P_X)$ で稠密

➡
$$\left\langle g, \Sigma_{YY|X} g \right\rangle_{\mathcal{H}_Y} = E_X [\text{Var}[g(Y) | X]]$$

■ 条件付独立性の特徴付け

確率変数 Y, U, V に対し、

$$Y \perp V | U \quad \text{あるいは同値な条件} \quad Y \perp X | U \quad (X = (U, V))$$

の特徴づけを考える。

定理6 (Fukumizu et al. 2004)

U, V, Y : 可測空間 $\Omega_U, \Omega_V, \Omega_Y$ に値をとる確率変数

$\mathcal{H}_U, \mathcal{H}_V, \mathcal{H}_Y$: $\Omega_U, \Omega_V, \Omega_Y$ 上のRKHS (有界で可測なカーネルによる)

$X = (U, V), \quad \mathcal{H}_X = \mathcal{H}_U \otimes \mathcal{H}_V$ (直積)

$\mathcal{H}_X, \mathcal{H}_U$ は定理5の(1), (2) を満たすとする。このとき、

$$\Sigma_{YY|U} \geq \Sigma_{YY|X} \quad (\geq \text{は自己共役作用素の半順序})$$

さらに \mathcal{H}_Y が確率決定的であるならば

$$\Sigma_{YY|U} = \Sigma_{YY|X} \quad \Leftrightarrow \quad Y \perp X | U$$

条件付共分散作用素による次元削減

■ 回帰問題における次元削減

- 回帰問題: 確率変数 Y が X にどう依存するかを調べる

$p(Y | X)$ を推定する

- 回帰問題における次元削減

X : m 次元確率ベクトル

$$p(Y | X) = \tilde{p}(Y | B^T X) \quad (\Leftrightarrow Y \perp X | B^T X)$$

を満たす $m \times d$ ($d < m$) 行列 B を求める。回帰の低次元表現。

■ RKHSを用いた解法

$$\min_B \text{Tr} \left[\Sigma_{YY|B^T X} \right]$$

カーネル次元削減法
(Fukumizu et al. 2004)

カーネルCCAの統計的性質

統計的推測

■ 統計的推測の原理

- データの背後には、それを発生させる確率分布がある
- 有限個のデータを用いて、その確率分布あるいはそれから決まる量を推測する

■ カーネル法の統計的推測としての見方

- データの変換ではなく、確率変数の変換 $k(\cdot, X)$
 - RKHSに値をとる確率変数
- $k(\cdot, X)$ から決まる量(関数)を、有限個のデータから推測する

■ 統計的推測法の妥当性

- 一致性: 有限データからの推定量が、確率分布から決まる真の量に収束するか?

カーネルCCAの作用素による表現

■ 期待値によるカーネルCCA

$$\max_{f, g} \langle g, \Sigma_{YX} f \rangle \quad \text{subj. to} \quad \langle f, \Sigma_{XX} f \rangle = 1, \quad \langle g, \Sigma_{YY} g \rangle = 1$$

■ 最適解

ξ_*, ζ_* : $V_{YX} = \Sigma_{YY}^{-1/2} \Sigma_{YX} \Sigma_{XX}^{-1/2}$ の最大特異値に対する単位固有ベクトル

$$\rightarrow \begin{cases} f_* = \Sigma_{XX}^{-1/2} \xi_* \\ g_* = \Sigma_{YY}^{-1/2} \zeta_* \end{cases}$$

■ 有限サンプルに対するカーネルCCA

- 経験相互共分散作用素

$$\left\langle g, \hat{\Sigma}_{YX}^{(N)} f \right\rangle = \frac{1}{N} \sum_{i=1}^N f(X_i) g(Y_i) - \frac{1}{N} \sum_{i=1}^N f(X_i) \frac{1}{N} \sum_{i=1}^N g(Y_i)$$

- カーネルCCA

$$\max_{f, g} \left\langle g, \hat{\Sigma}_{YX}^{(N)} f \right\rangle \quad \text{subj. to} \quad \left\langle f, \left(\hat{\Sigma}_{XX}^{(N)} + \varepsilon_N I \right) f \right\rangle = 1,$$
$$\left\langle g, \left(\hat{\Sigma}_{YY}^{(N)} + \varepsilon_N I \right) g \right\rangle = 1,$$

- 解

$$\hat{\xi}_N, \hat{\zeta}_N : \quad \hat{V}_{YX}^{(N)} := \left(\hat{\Sigma}_{YY}^{(N)} + \varepsilon_N I \right)^{-1/2} \hat{\Sigma}_{YX}^{(N)} \left(\hat{\Sigma}_{XX}^{(N)} + \varepsilon_N I \right)^{-1/2}$$

の最大特異値に対する単位固有ベクトル

$$\rightarrow \begin{cases} \hat{f}_N = \left(\hat{\Sigma}_{XX}^{(N)} + \varepsilon_N I \right)^{-1/2} \hat{\xi}_N \\ \hat{g}_N = \left(\hat{\Sigma}_{YY}^{(N)} + \varepsilon_N I \right)^{-1/2} \hat{\zeta}_N \end{cases}$$

■ 解の一致性(収束性)

□ $\hat{f}_N \rightarrow f_*$, $\hat{g}_N \rightarrow g_*$

あるいは

$\hat{\xi}_N \rightarrow \xi_*$, $\hat{\zeta}_N \rightarrow \zeta_*$

は成り立つか？

- どのような関数空間のノルムのもとで収束するか？
- 収束するための正則化係数 ε_N の条件は？

カーネルCCAの一致性

Theorem 1

If V_{YX} is compact, and the regularization coefficient satisfies

$$\varepsilon_N \rightarrow 0, \quad N^{1/3} \varepsilon_N \rightarrow \infty \quad (N \rightarrow \infty)$$

then,

$$\left| \left\langle \hat{\xi}_N, \xi^* \right\rangle_{\mathcal{H}_X} \right| \rightarrow 1, \quad \left| \left\langle \hat{\zeta}_N, \zeta^* \right\rangle_{\mathcal{H}_Y} \right| \rightarrow 1$$

in probability as $N \rightarrow \infty$.

Note: convergence of RKHS is very strong.

If the kernels are bounded, it is stronger than the uniform norm.

Theorem 2

Suppose V_{YX} is compact, and the regularization coefficient satisfies

$$\varepsilon_N \rightarrow 0, \quad N^{1/3} \varepsilon_N \rightarrow \infty \quad (N \rightarrow \infty).$$

If ξ_* and ζ_* are included in $\mathcal{R}(\Sigma_{XX})$ and $\mathcal{R}(\Sigma_{YY})$, respectively, then

$$\left\| \left(\hat{f}_N - E_X[\hat{f}_N(X)] \right) - \left(f - E_X[f(X)] \right) \right\|_{L_2(P_X)} \rightarrow 0$$

$$\left\| \left(\hat{g}_N - E_Y[\hat{g}_N(Y)] \right) - \left(g - E_Y[g(Y)] \right) \right\|_{L_2(P_Y)} \rightarrow 0$$

in probability as $N \rightarrow \infty$

■ 証明の核になる命題

命題

$$\left\| \hat{\Sigma}_{YX}^{(N)} - \Sigma_{YX} \right\|_{HS} = O_p(N^{-1/2}), \quad (N \rightarrow \infty)$$

□ Hilbert-Schmidt ノルム $\|A\|_{HS}^2 = \sum_{i=1}^{\infty} \|A\varphi_i\|^2$ ($\{\varphi_i\}_{i=1}^{\infty} : \text{CONS}$)

□ 大数の法則 $\frac{1}{N} \sum_{i=1}^N f(X^i)g(Y^i) - \frac{1}{N} \sum_{i=1}^N f(X^i) \frac{1}{N} \sum_{i=1}^N g(Y^i)$
 $\rightarrow E[f(X)g(Y)] - E[f(X)]E[g(X)]$

$$\longleftrightarrow \langle g, \hat{\Sigma}_{YX}^{(N)} f \rangle \rightarrow \langle g, \Sigma_{YX} f \rangle$$

□ 中心極限定理 $\longrightarrow \langle g, (\hat{\Sigma}_{YX}^{(N)} - \Sigma_{YX}) f \rangle = O_p(N^{-1/2})$

□ 上の命題は収束のオーダーに関する強い結果

おわりに

- 再生核ヒルベルト空間を用いたデータ解析
 - RKHS: 関数の値が意味を持つヒルベルト空間
 - 内積計算が容易
 - 属する関数の連続性や微分可能性
 - ベクトルでないデータに対しても相関(内積)計算を可能にする
- 独立性、条件付独立性の特徴づけ
 - 相互共分散作用素を用いた線形手法の非線形化
- カーネル法の統計的性質
 - 漸近的な性質、一致性
 - カーネルCCAの一致性のための正則化係数の十分条件

■ 参考文献

- K. Fukumizu, F.R. Bach, and M.I. Jordan (2004)
Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *Journal of Machine Learning Research*. 5:73-99.
- 福水健次(2003)「再生核ヒルベルト空間を用いた回帰問題における次元削減法」 日本数学会2003年秋季大会予稿集
- K. Fukumizu, F.R. Bach, and Arthur Gretton (2005)
Consistency of kernel canonical correlation analysis. ISM Research Memo 942.
- K. Fukumizu, F.R. Bach, and M.I. Jordan (2005)
Consistency of kernel dimensionality reduction. *In preparation*.

URL: <http://www.ism.ac.jp/~fukumizu/>