



カーネル法による 非線形データ解析入門

福水健次

情報・システム研究機構 統計数理研究所

March 3, 2006. @ ROIS Cross-talk

あらまし

1. イントロ： 線形から非線形へ
2. カーネル法： 高次元の内積計算
3. カーネル法の具体例： カーネルPCAとカーネルCCA
4. グラフに対するデータ解析
5. まとめ



Introduction

線形から非線形へ

1. イントロ： 線形から非線形へ
2. カーネル法： 高次元の内積計算
3. カーネル法の具体例： カーネルPCAとカーネルCCA
4. グラフに対するデータ解析
5. まとめ

はじめに

■ データ解析

実験／観測などで得られたデータから、有用な情報を抽出するための方法

- 情報の集約

低次元表現、圧縮表現

- 関係の抽出

相関、依存性

- 可視化による分析

2, 3次元表現

- 予測・発見

線形なデータ解析

■ 線形データ解析

- データが m 次元数値ベクトルとして表現されている

$$X^1 = (1.52, -0.98, 5.01, -0.24, -3.31)$$

$$X^2 = (2.31, 1.58, -2.76, 4.37, 3.09)$$

$$X^3 = (-0.01, -2.28, 4.71, -2.51, -0.83)$$

⋮

- データに線形な処理を施してデータ解析を行う

例:

- 主成分分析
- 回帰分析
- 相関分析 etc.

例1: 主成分分析 (PCA, principal component analysis)

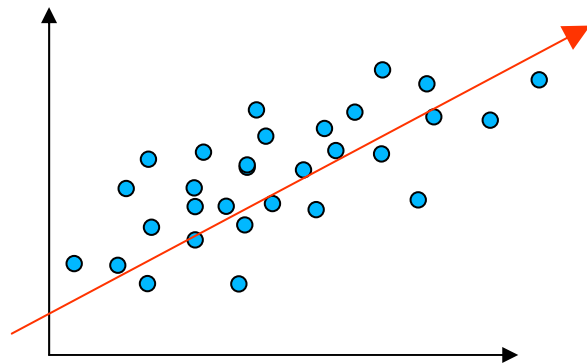
□ 主成分分析 – ばらつきの大きい方向にデータを射影して眺める –

復習

a^T は a の転置

$$a^T b = a_1 b_1 + \dots + a_m b_m$$

は内積を表す



m 次元データ
⇒ 低次元の主成分による表現

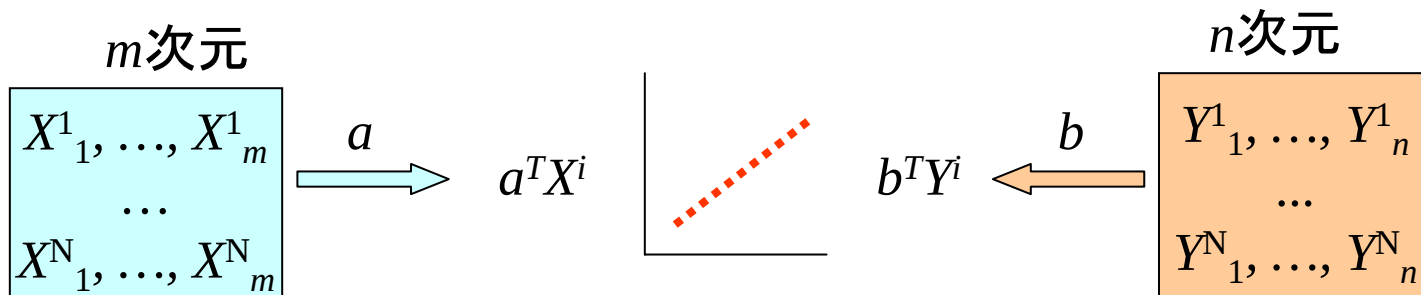
$$\max_{\|a\|=1} \frac{1}{N} \sum_{i=1}^N \left(a^T \left(X^i - \frac{1}{N} \sum_j X^j \right) \right)^2 = \max_{\|a\|=1} a^T V_{XX} a \quad \text{なる } a \text{ を求める}$$

$$V_{XX} = \frac{1}{N} \sum_{i=1}^N \left(X^i - \frac{1}{N} \sum_{j=1}^N X^j \right) \left(X^i - \frac{1}{N} \sum_{j=1}^N X^j \right)^T$$

a : 分散共分散行列 V_{XX} の最大固有値に対応する固有ベクトル
 d 個まで取れば d 次元表現を得る。

例2: 正準相関分析(CCA, canonical correlation analysis)

□ 正準相関分析 – 2つ多次元変数間の相関を見る手法 –



$a^T X$ と $b^T Y$ の相関が最大になるように、ベクトル a, b を決める

$$\rho = \max_{\substack{a \in \mathbf{R}^m \\ b \in \mathbf{R}^n}} \frac{\frac{1}{N} \sum_i \left(a^T X^i - \frac{1}{N} \sum_j a^T X^j \right) \left(b^T Y^i - \frac{1}{N} \sum_j b^T Y^j \right)}{\sqrt{\frac{1}{N} \sum_i \left(a^T X^i - \frac{1}{N} \sum_j a^T X^j \right)^2} \sqrt{\frac{1}{N} \sum_i \left(b^T Y^i - \frac{1}{N} \sum_j b^T Y^j \right)^2}}$$

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

一般化固有値問題として解ける

線形から非線形へ

■ 線形手法の限界

- 非線形なデータの構造は捕らえにくい
複雑なデータに対して線形な構造だけを見ているでは不十分なこともある
- データが数値ベクトルである必要がある
非数値的データには別途考慮が必要

■ カーネル法：非線形データ解析手法

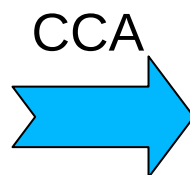
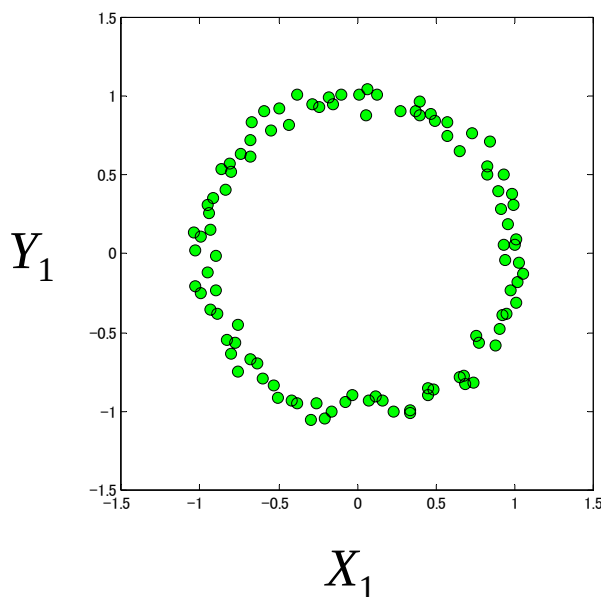
- カーネル法 … 非線形データ解析の一般的枠組み
 - データの非線形な関係を考慮できる
 - 既存の線形手法を容易に非線形化できる
 - 非数値的データにも適用可能

非線形データ解析の重要性

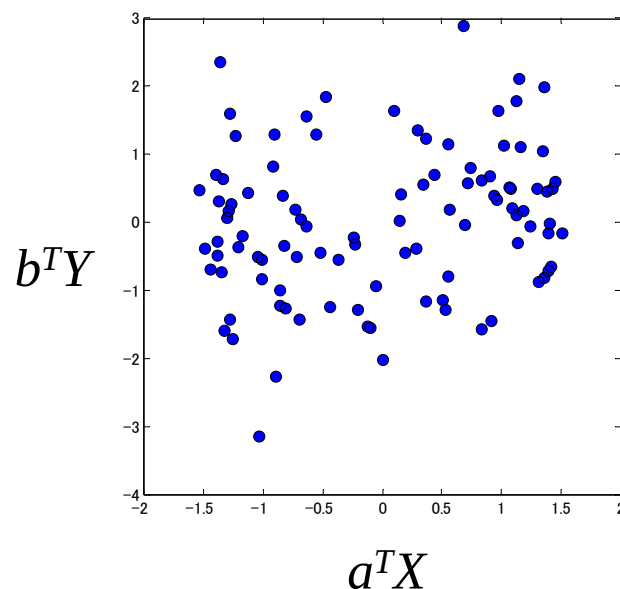
■ 正準相関分析の例

□ $X = (X_1, X_2), Y = (Y_1, Y_2)$

X_1 と Y_1 が強い非線形な
依存関係を持つ



非線形な関係は
発見できない



非線形データ解析の重要性

■ 非線形変換＋CCA

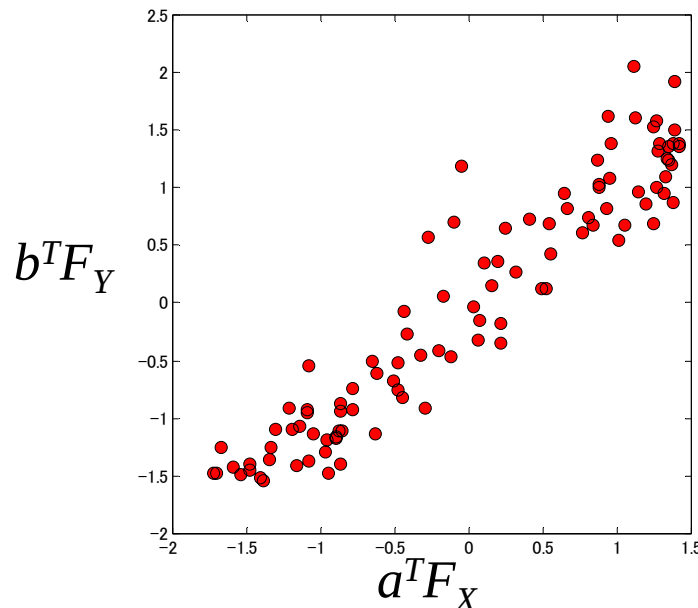
2次式による変換

$$(X_1, X_2) \Rightarrow (X_1, X_2, X_1^2, X_1 X_2, X_2^2) = F_X$$

$$(Y_1, Y_2) \Rightarrow (Y_1, Y_2, Y_1^2, Y_1 Y_2, Y_2^2) = F_Y$$

それぞれ
5次元データに変換

変換後のデータに対してCCA → 非線形相関が考慮できる



相関係数 = 0.963

非線形化の困難

- 高次相関 → 多次元データでは計算量が爆発

もとの変数: m 次元

d 次式の空間 $\cdots$ 高次元

$$\begin{array}{ccc} X_1^1, \dots, X_m^1 & \xrightarrow{\quad \Phi \quad} & (X_1^1)^d, \dots, (X_m^1)^d, (X_1^1)^{d-1}X_2^1, \dots, (X_1^1)^{d-1}X_3^1, \dots \\ \vdots & & \vdots \\ X_1^N, \dots, X_m^N & & (X_1^N)^d, \dots, (X_m^N)^d, (X_1^N)^{d-1}X_2^N, \dots, (X_1^N)^{d-1}X_3^N, \dots \end{array}$$

$$\frac{(m+d-1)!}{d!(m-1)!} \text{ 次元}$$

$$\text{例) } m = 128, d = 3 \Rightarrow 3.6 \times 10^5 \text{ 次元}$$

莫大な計算量



カーネル法

高次元の内積計算

1. イントロ：線形から非線形へ
2. カーネル法：高次元の内積計算
3. カーネル法の具体例：カーネルPCAとカーネルCCA
4. グラフに対するデータ解析
5. まとめ

非線形変換としてのカーネル法

■ カーネル法の原理

もとの変数

X^1

$\vdots$

X^N

Φ



非線形写像
(うまい性質を持つ)

高次元ベクトル空間(無限次元でよい)

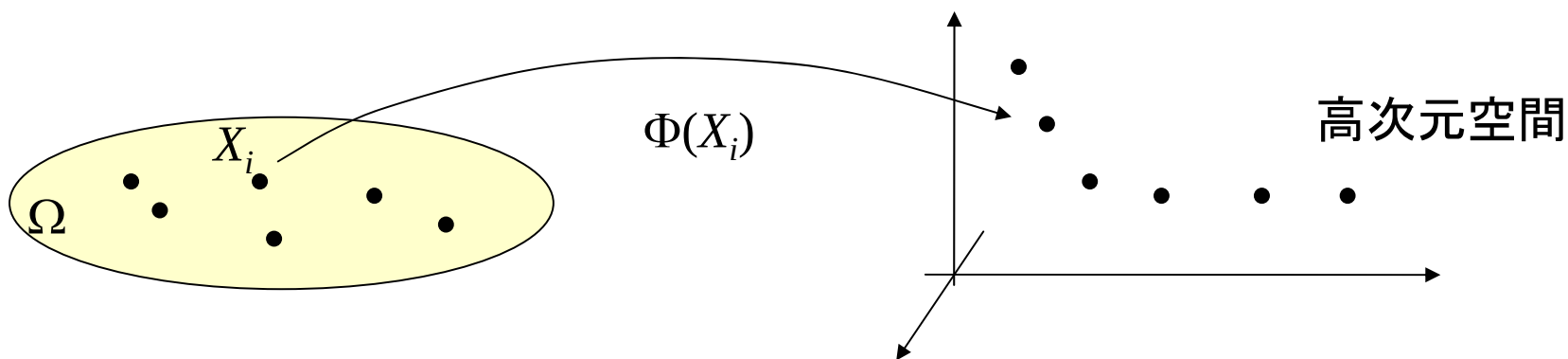
$\Phi(X^1)$

$\vdots$

$\Phi(X^N)$

変換後のデータに
線形の手法を用いる

非線形相関などが
考慮できる



カーネル法における内積計算

■ 内積計算

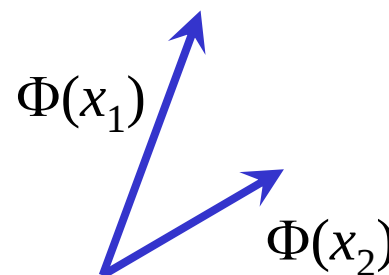
- 線形手法の計算 . . . 内積計算が本質 (相関行列、分散共分散行列)

内積 $\Phi(X_i) \cdot \Phi(X_j)$ が計算できればOK

- カーネル法: 変換 Φ を与える代わりに
内積が

$$\Phi(x_1) \cdot \Phi(x_2) = K(x_1, x_2)$$

と計算できるようなカーネル関数 $K(x_1, x_2)$ を与える。
高次元ベクトル空間と非線形変換は自然と定まる



- カーネル $K(x_1, x_2)$: データ x_1 と x_2 の類似度を定めるようなもの

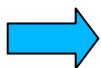
例:カーネル主成分分析

■ 高次元空間内でのPCA

$$\max_{\|F\|=1} \frac{1}{N} \sum_{i=1}^N \left(F \cdot \left\{ \Phi(X^i) - \frac{1}{N} \sum_j \Phi(X^j) \right\} \right)^2$$

高次元空間内のベクトル F は

$$F = \sum_{\ell=1}^N \alpha_{\ell} \left\{ \Phi(X^{\ell}) - \frac{1}{N} \sum_s \Phi(X^s) \right\} \quad \text{の形で探せば十分 (}\because \text{直交性)}$$



$$\max_{\alpha} \frac{1}{N} \sum_{i=1}^N \left(\sum_{\ell=1}^N \alpha_{\ell} \left\{ \Phi(X^{\ell}) - \frac{1}{N} \sum_s \Phi(X^s) \right\} \cdot \left\{ \Phi(X^i) - \frac{1}{N} \sum_j \Phi(X^j) \right\} \right)^2$$

$\Phi(X^i) \cdot \Phi(X^j) = K(X^i, X^j)$ を使って展開すると、、、

例:カーネル主成分分析

■ カーネルPCAのアルゴリズム

$$\max_{\alpha} \frac{1}{N} \alpha^T \tilde{K}^2 \alpha$$

ベクトル F のノルムは1なので、
制約 $\alpha^T \tilde{K} \alpha = 1$

$$\text{ただし } \tilde{K}_{ij} = K(X_i, X_j) - \frac{1}{N} \sum_{t=1}^N K(X_i, X_t) - \frac{1}{N} \sum_{s=1}^N K(X_s, X_j) + \frac{1}{N^2} \sum_{s,t=1}^N K(X_s, X_t)$$

$\tilde{K}$: データ数のサイズの行列(データの次元には無関係)

➡ (1) $\alpha^{(1)}, \dots, \alpha^{(d)}$: $\tilde{K}$ の大きい d 個の固有値に対応する
ノルム1の固有ベクトルを求める。

(2) データ点 X^i の第 p 主成分 $F^{(p)}$

$$= \left(\sum_{\ell=1}^N \alpha_{\ell}^{(p)} \left\{ \Phi(X^{\ell}) - \frac{1}{N} \sum_s \Phi(X^s) \right\} \right) \cdot \left\{ \Phi(X^i) - \frac{1}{N} \sum_j \Phi(X^j) \right\}$$

$$= \tilde{K} \alpha^{(p)} \text{ の第 } i \text{ 成分}$$

正定値カーネル

■ どのような $K(x,y)$ がカーネルとして使えるか？

対称性 $K(x, y) = K(y, x)$ と、次の正定値性を満たせばOK

任意の点 $x_1, \dots, x_n$ と実数 $\alpha_1, \dots, \alpha_n$ に対して

$$\sum_{i,j=1}^n \alpha_i \alpha_j K(x_i, x_j) \geq 0 \quad \text{(半)正定値}$$

■ 正定値カーネルの例

□ 線形カーネル $K(x, y) = x^T y$ (普通のユークリッド内積)

□ ガウスカーネル $K(x, y) = \exp\left(-\|x - y\|^2 / \sigma^2\right)$ ($\sigma > 0$)

□ 多項式カーネル $K(x, y) = (c + x^T y)^d$ ($c \geq 0, d \in \mathbf{N}$)



カーネル法の具体例

カーネルPCAとカーネルCCA

1. イントロ：線形から非線形へ
2. カーネル法：高次元の内積計算
3. カーネル法の具体例：カーネルPCAとカーネルCCA
4. グラフに対するデータ解析
5. まとめ

カーネル主成分分析(カーネルPCA)

■ カーネルPCAのアルゴリズム

1. カーネルの選択
2. 中心化グラム行列の計算

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

3. $\tilde{K}$ の固有値分解

$$\tilde{K} = \begin{pmatrix} U \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix} \begin{pmatrix} U^T \end{pmatrix} \quad \begin{array}{l} U: \text{直交行列} \\ \lambda_1 \geq \dots \geq \lambda_N \geq 0 \end{array}$$

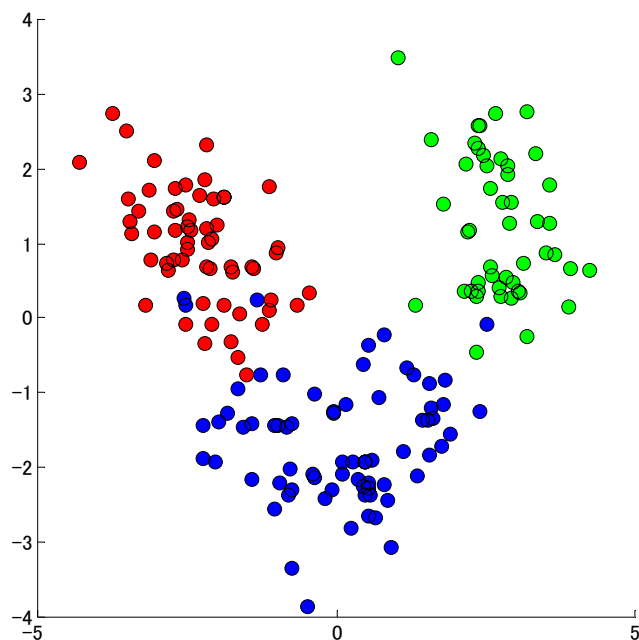
4. 第 i データの第 p 主成分 $= \sqrt{\lambda_p} U_{ip}$

カーネルPCAの例

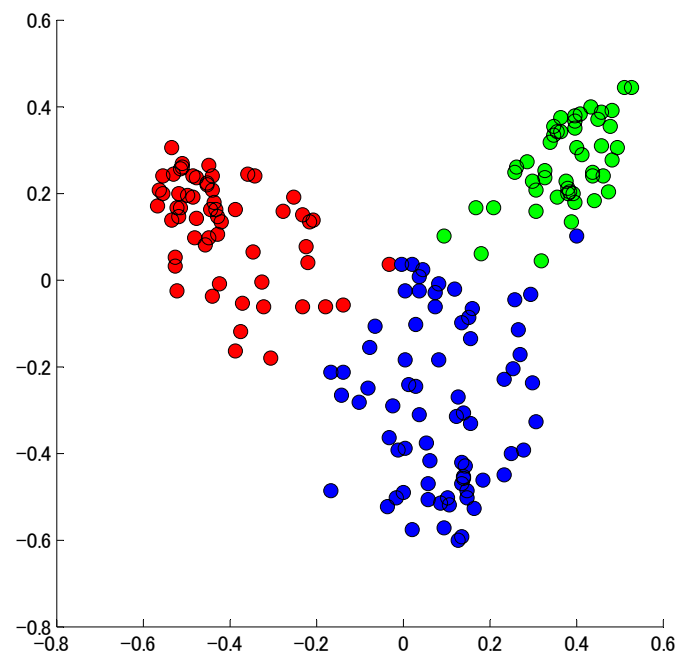
‘Wine’ データ (UCI Machine Learning Repository)

13次元, 178データ, 3種類のワインの属性データ

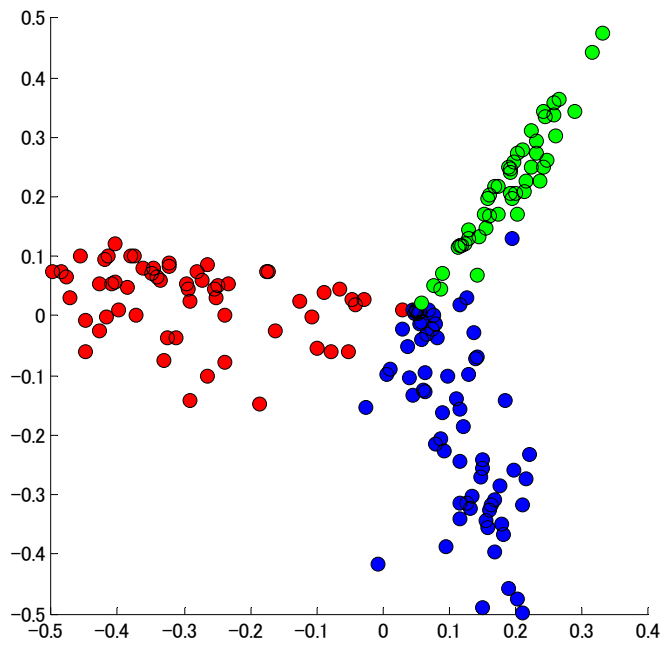
2つの主成分を取った (3クラスの色は参考に付けたもの)



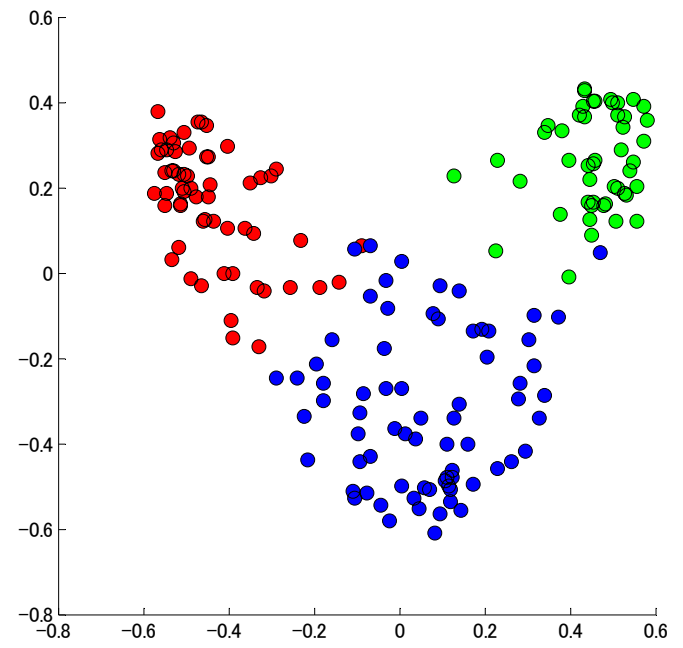
PCA (線形)



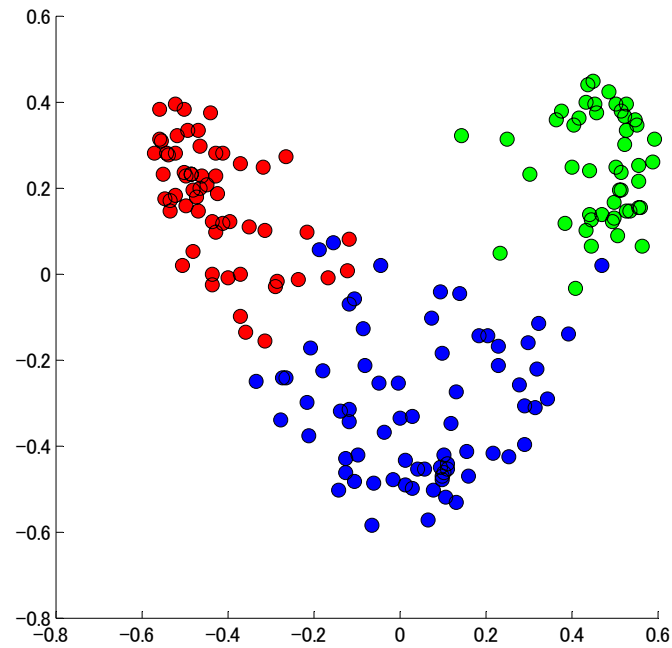
KPCA (Gauss, $\sigma = 3$)



KPCA(Gauss, $\sigma=2$)



KPCA(Gauss, $\sigma=4$)



KPCA(Gauss, $\sigma=5$)

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
(カーネルの種類、カーネル内のパラメータ)

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 (Fisher判別分析、サポートベクターマシン)
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

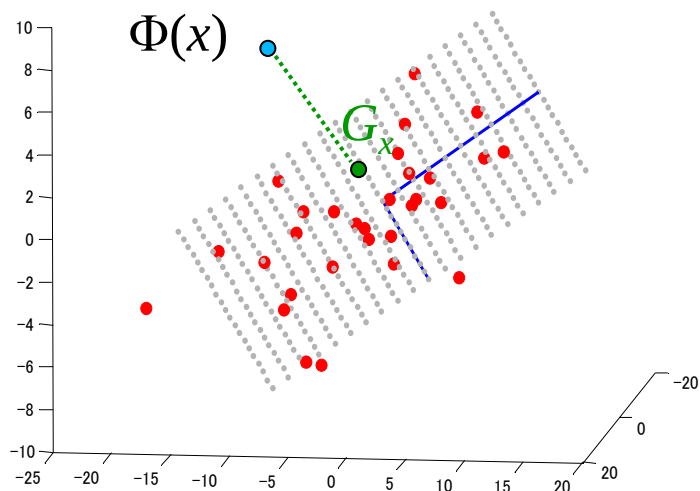
$$F_1, F_2, \dots, F_d$$

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。

Φ の像の中で G_x に最も近いものを $\hat{x}$ とする。

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

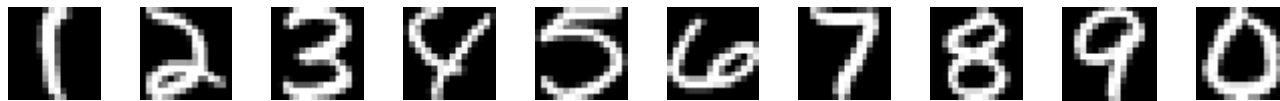
カーネル $K(x_1, x_2)$ を使って表せる



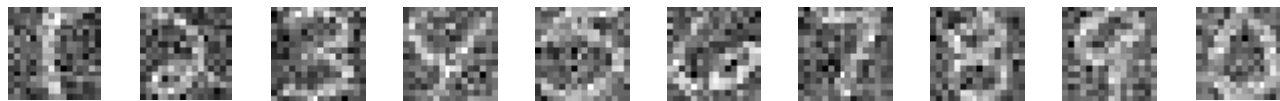
カーネルPCAの雑音除去への応用

■ 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)

Matlab stprtool (V. Franc作) により作成

カーネルCCA

■ 高次元空間内のCCA

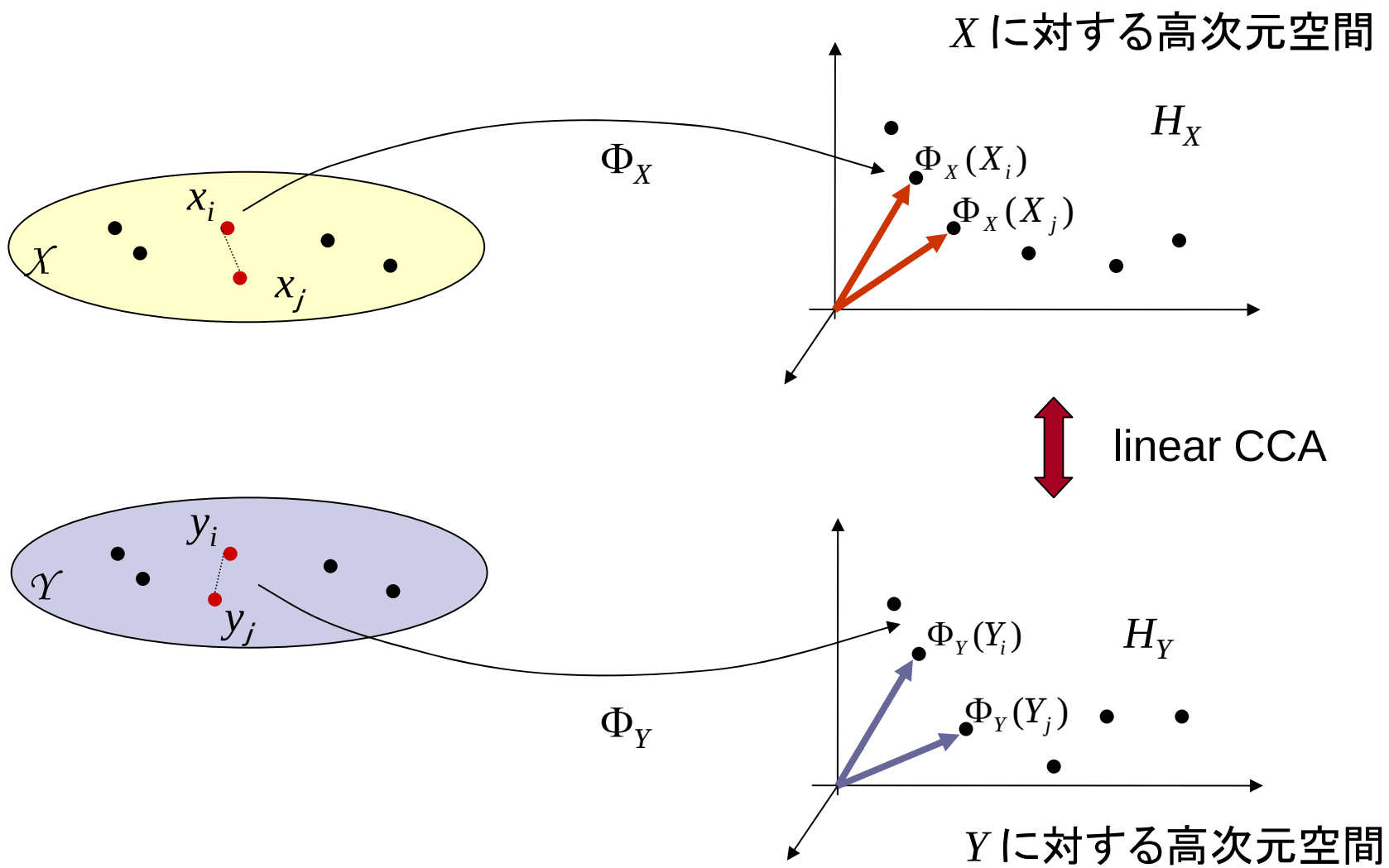
□ 線形のCCA

$$\max_{\substack{a \in \mathbf{R}^m \\ b \in \mathbf{R}^n}} \frac{\text{Cov}[a^T X, b^T Y]}{\left(\text{Var}[a^T X] \text{Var}[b^T Y] \right)^{1/2}}$$

□ カーネルCCA: 非線形変換後のデータのCCA

$$\max_{F, G} \frac{\text{Cov}[F \bullet \Phi_X(X), G \bullet \Phi_Y(Y)]}{\left(\text{Var}[F \bullet \Phi_X(X)] \text{Var}[G \bullet \Phi_Y(Y)] \right)^{1/2}}$$

Φ_X と Φ_Y (あるいは K_X と K_Y) は異なってよい

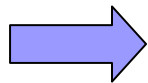


■ カーネルCCAの定式化

$$\max_{F,G} \frac{\frac{1}{N} \sum_{i=1}^N \left(F \cdot \left\{ \Phi_X(X_i) - \frac{1}{N} \sum_j \Phi_X(X_j) \right\} \right) \left(G \cdot \left\{ \Phi_Y(Y_i) - \frac{1}{N} \sum_j \Phi_Y(Y_j) \right\} \right)}{\left(\frac{1}{N} \sum_{i=1}^N \left(F \cdot \left\{ \Phi_X(X_i) - \frac{1}{N} \sum_j \Phi_X(X_j) \right\} \right)^2 \right)^{1/2} \left(\frac{1}{N} \sum_{i=1}^N \left(G \cdot \left\{ \Phi_Y(Y_i) - \frac{1}{N} \sum_j \Phi_Y(Y_j) \right\} \right)^2 \right)^{1/2}}$$

$$F = \sum_{i=1}^N a_i \left(\Phi_X(X_i) - \frac{1}{N} \sum_{j=1}^N \Phi_X(X_j) \right) \quad G = \sum_{i=1}^N b_i \left(\Phi_Y(Y_i) - \frac{1}{N} \sum_{j=1}^N \Phi_Y(Y_j) \right)$$

の形で十分



Kernel CCA

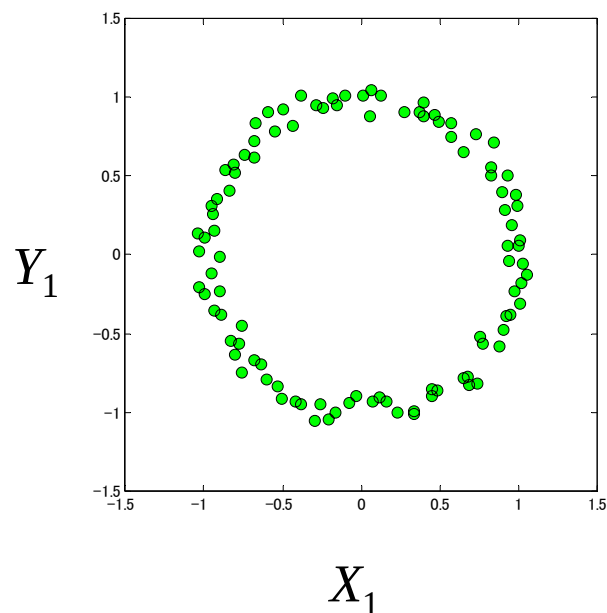
$$\max_{\substack{a \in \mathbf{R}^N \\ b \in \mathbf{R}^N}} \frac{a^T \tilde{K}_X \tilde{K}_Y b}{\sqrt{a^T \left(\tilde{K}_X^2 + N \varepsilon_N \tilde{K}_X \right) a} \sqrt{b^T \left(\tilde{K}_Y^2 + N \varepsilon_N \tilde{K}_Y \right) b}}$$

ε_N : 正則化係数

最適な a と b は一般化固有値問題の解として解ける。

カーネルCCAの適用例

$$X = (X_1, X_2), Y = (Y_1, Y_2)$$



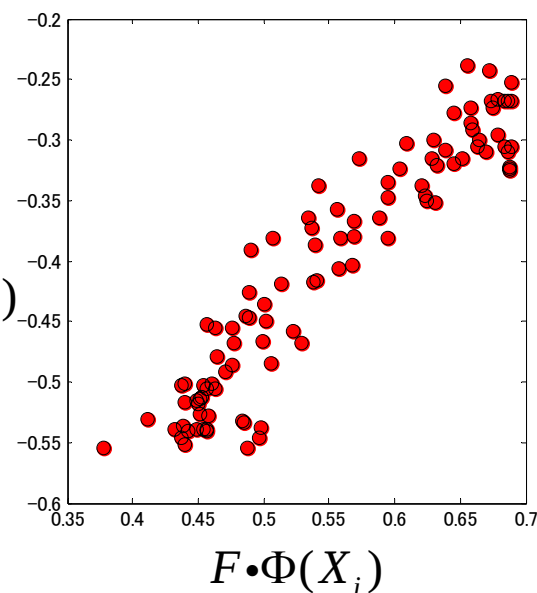
X_2, Y_2 は独立なノイズ

カーネルCCA



ガウスクーネル
 $\exp(-\|z - \tilde{z}\|^2 / 9)$

$G \cdot \Phi(Y_i)$



相関係数 = 0.976

(c.f. 2次式でのCCA,
相関係数 = 0.963)

カーネル法の特徴

- 非線形写像 $\Phi(x)$ を決めるのではなく、内積を表すカーネル $K(x,y)$ を決める。

- 相関行列(分散共分散行列)のかわりに、
グラム行列

$$K(X_i, X_j)$$

が活躍する。

- グラム行列のサイズ = データのサイズ
- カーネル $K(x,y)$ は、データ x, y の類似度を定めている



グラフに対するデータ解析

1. イントロ：線形から非線形へ
2. カーネル法：高次元の内積計算
3. カーネル法の具体例：カーネルPCAとカーネルCCA
4. グラフに対するデータ解析
5. まとめ

非ベクトルデータに対するカーネル

■ カーネルは任意の集合上に定義可能

- 正定値カーネル $K(x, y)$: x, y はベクトルでなくともよい
- 非ベクトルデータ上のカーネルの定義
→ 非ベクトルデータに対するデータ解析が可能
主成分分析、正準相関分析、判別分析 etc
- 複雑な構造を持つデータ(グラフ、ネットワーク、ツリー etc.)のデータ解析に有用
- カーネルをどう定義するか？(類似度をどう決めるか？)
→ 個別の工夫が必要。モデリングの問題。

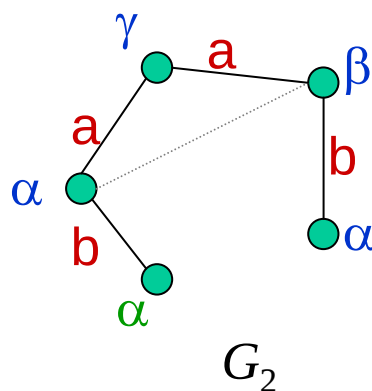
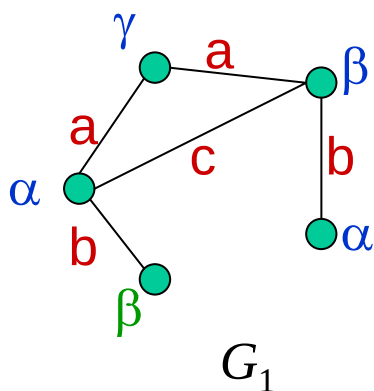
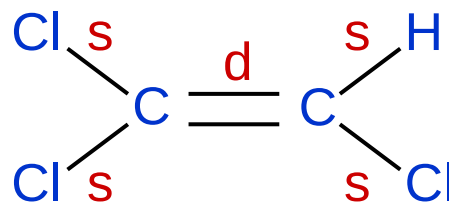
グラフの類似度をはかるカーネル

■ グラフ間の類似度を定める方法

- グラフカーネル (Kashima et al. 2002)

$$\Phi: \text{グラフ } G \mapsto \Phi(G) \in H \text{ ベクトル空間}$$

化合物の毒性予測
に応用



生成可能なパス(ラベル列)
の類似度を $K(G_1, G_2)$ に
用いる

グラフの類似度をはかるカーネル

□ 自然言語処理への応用

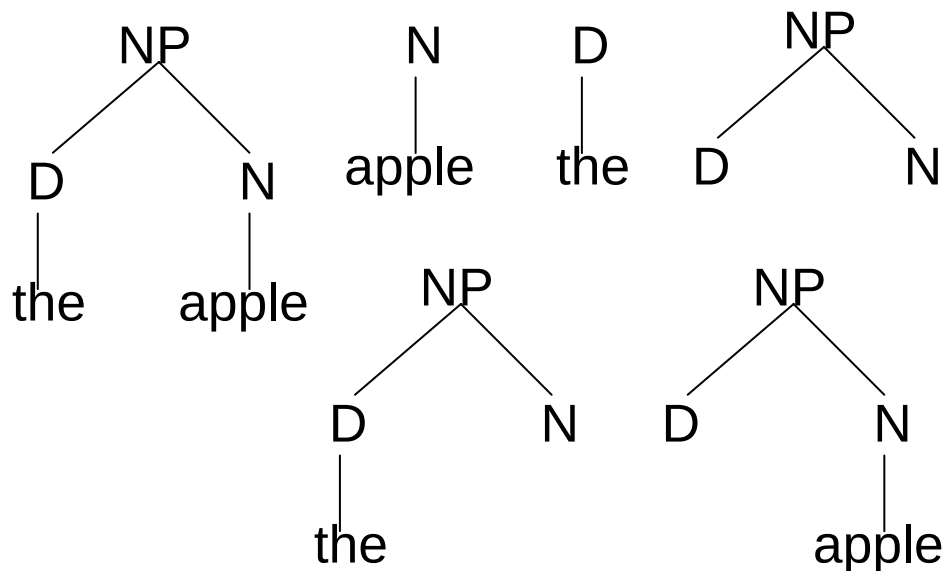
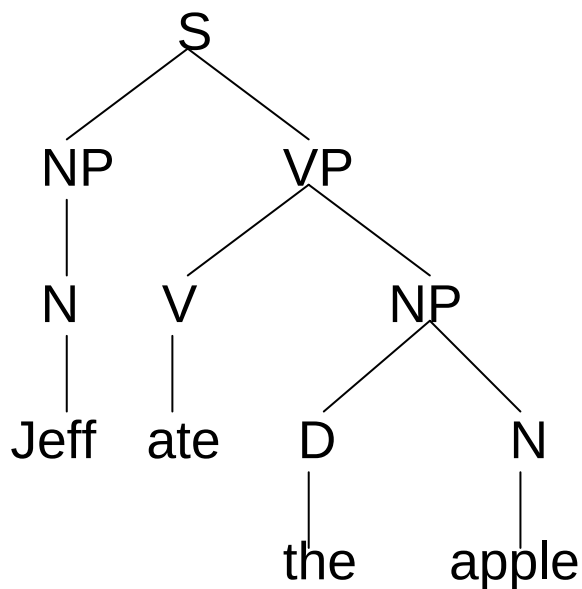
Φ : ツリー $T \mapsto \Phi(T) \in H$ ベクトル空間

サブツリーの一致により類似度を定義

構文解析 (Collins & Duffy 2002)

Jeff ate the apple.

サブツリーの例

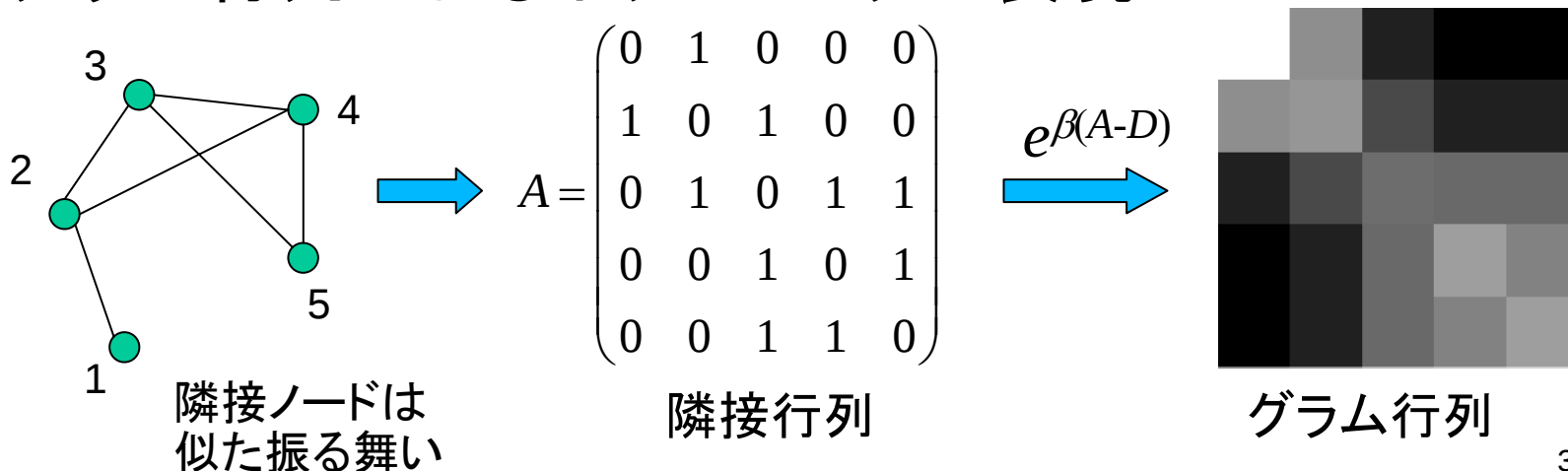


グラム行列によるネットワークデータの処理

■ カーネル手法・・・グラム行列に対する操作

- もとのデータを知らなくても、グラム行列が与えられれば適用可能なものが多い。
- グラム行列 $\simeq$ 類似度行列
正定値な類似度行列が与えられれば、カーネル法を適用可能な場合がある。

■ グラム行列によるネットワークの表現



カーネルCCAによるネットワーク推定

■ タンパク質ネットワークの推定 (Yamanishi et al. 2004)

- 複数の遺伝情報からグラム行列を構成し、逆にネットワークを推定する yeast (*Saccharomyces cerevisiae*)

Gene expression data $\rightarrow K_{X,1}$

Two-hybrid systems $\rightarrow K_{X,2}$

Protein localization data $\rightarrow K_{X,3}$

Phylogenetic profile $\rightarrow K_{X,4}$

$K_Y \leftarrow$ Protein network
(既知の相互作用)

カーネルCCAによるネットワーク推定

- 相互作用ネットワークに関連した特徴をカーネルCCAで抽出

$$K_X = K_{X,1} + K_{X,2} + K_{X,3} + K_{X,4}$$

4種の遺伝情報の統合

$$K_Y$$

ネットワークデータ

カーネルCCA

$$F^{(1)} \cdot \Phi(X_i), \dots, F^{(d)} \cdot \Phi(X_i)$$

遺伝情報に含まれる、ネットワーク推定と関連が強い情報

$$G^{(1)} \cdot \Phi(Y_i), \dots, G^{(d)} \cdot \Phi(Y_i)$$

$$\tilde{K}_{ij} = \sum_{p=1}^d \left(\Phi_X(X_i) \cdot F^{(p)} \right) \left(F^{(p)} \cdot \Phi_X(X_j) \right)$$

ネットワーク推定に有効な類似度

閾値処理によるネットワーク推定

まとめ

■ カーネルを用いた非線形データ解析

- カーネル法 ... カーネルで定まる非線形変換 + 線形データ解析手法
- カーネル ... 内積(類似度)を定義する
- 従来の線形手法の拡張が容易
カーネルPCA、カーネルCCA、カーネル判別分析 etc.
- カーネルの選択は、個別に考える必要がある
- データの分析などへの応用の際は、結果の合理性を考慮すべし
(非線形手法は、局所的な情報のみを強調することもある)
- 複雑な構造を持つデータの処理も可能である
非ベクトルデータに対しても内積定義 → 線形手法の適用可能

References

ホームページ

2004年度統数研公開講座資料「カーネル法 -- 正定値カーネルを用いたデータ解析 --」(福水)

http://www.ism.ac.jp/~fukumizu/ISM_lecture_2004/

カーネル法に関するポータルサイト <http://www.kernel-machines.org/>

ソフトウェア

The Spider (Matlab ツールボックス)

<http://www.kyb.tuebingen.mpg.de/bs/people/spider/index.html>

Statistical Pattern Recognition Toolbox (Matlab ツールボックス)

<http://cmp.felk.cvut.cz/~xfrancv/stprtool/>

Gist (C言語, Support Vector Machine) <http://microarray.cpmc.columbia.edu/gist/>

References

書籍

- 赤穂昭太郎. 「カーネルマシン」 学習システムの理論と実現 3章. 森北出版 (2005)
- Schölkopf, B. and A. Smola. *Learning with Kernels*. MIT Press. 2002.
- Shawe-Taylor, J. and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press (2004)
- Schölkopf, B., K. Tsuda, and J.-P. Vert (eds.). *Kernel Methods in Computational Biology*. MIT Press (2004).

言及した論文

- Kashima, H., K. Tsuda and A. Inokuchi. (2003) Marginalized Kernels Between Labeled Graphs. *Proc. 20th Intern. Conf. Machine Learning (ICML2003)*.
- Collins, M. & N. Duffy. (2002) Convolution Kernels for Natural Language. *Advances in Neural Information Processing Systems 14*.
- Yamanishi, Y., J.-P. Vert and M. Kanehisa. (2004) Protein network inference from multiple genomic data: a supervised approach. *Bioinformatics*. Vol. 20 Suppl. 1, pages i363–i370.

宣伝です

さらに詳しく知りたい方は、、、

統計数理研究所 2006年度公開講座

「カーネル法の最前線 -- SVM, 非線形データ解析, 構造化データ --」

2006年7月上旬 2日間(10時間)

講師: 福水健次 (統数研)

赤穂昭太郎 (産総研)

松井知子 (統数研)

詳細は <http://www.ism.ac.jp/>