カーネル多変量解析
3.4　判別分析と正準相関分析

06T4007A　伊藤輝将

カーネル判別分析(1)

・(線形)判別分析

-クラスを表す離散の目的変数がある場合の次元圧縮法

判別分析ではクラス識別そのものではなく、クラス識別の前段階として有効と考えられる特別な基準を設定する

まず、カーネル法を導くため、次元圧縮の関数として特徴ベクトルの線形和で定義される1変数関数式(3.57)を考える

$$f(x) = w^t \phi(x) \qquad (3.57)$$

各サンプルx^(i)にはクラスラベルy^(i)がついている。クラス数をcとし、y^(i)は1,2,・・・,cのどれかの値をとるとする

カーネル判別分析(2)

基準は以下のように書ける

$$\max_{f} \frac{\sigma_{B}^{2}}{\sigma_{w}^{2}} \qquad (3.58)$$

判別分析をカーネル法で行う場合、カーネル法の記述能力が高すぎてこの基準そのままではクラス内分散が0となる解が得られる

そこで回帰のときと同様にλ||w||^2という正則化項を考える

すると、リプレゼンター定理から

$$f(x) = \sum_{i=1}^{n} a_i k(x^{(i)}, x) \qquad (3.59)$$

という形を成すことができる

カーネル判別分析(3)

クラス内分散σ_{w}^2は、各クラスlごとの分散をσ_l^2としたとき、そのクラスのメンバー数で重みづけた平均として以下の式で定義できる

$$\sigma_{\mathrm{w}}^2 = \frac{1}{n}\sum_{l=1}^{c} n_l \sigma_l^2 \qquad (3.60)$$

σ_l^2を計算するためにまず、各クラスlに属するサンプルに対するf(x)の平均を求めると以下となる

$$\mu_l = \frac{1}{n_l}\sum_{i=1}^{n}\{\alpha_i \sum_{j:y^{(i)}=l} k(x^{(i)}, x^{(j)})\} = \sum_{i=1}^{n} \alpha_i \overline{k_{i,l}} \qquad (3.61)$$

ただし、

$$\overline{k_{i,l}} = \frac{1}{n_l}\sum_{j:y^{(j)}=l} k(x^{(i)}, x^{(j)}) \qquad (3.62)$$

とおいた

カーネル判別分析(4)

したがって、クラスlのサンプルに対する $f(x^{(i)})$の分散は

$$
\begin{aligned}
\sigma_l^2 &= \frac{1}{n_l}\sum_{j:y^{(j)}=l}(f(x^{(j)})-\mu_l)^2 \\
&= \frac{1}{n_l}\sum_{j:y^{(j)}=l}\left(\sum_{i=1}^{n}\alpha_i(k(x^{(i)},x^{(j)})-\overline{k_{i,l}})\right)^2 \\
&= \frac{1}{n_l}\sum_{i=1}^{n}\sum_{i'=1}^{n}[\alpha_i\alpha_{i'}\sum_{j:y^{(j)}=l}(k(x^{(i)},x^{(j)})-\overline{k_{i,l}})(k(x^{(i')},x^{(j)})-\overline{k_{i',l}})] \\
&= \alpha^T S_l \alpha \qquad (3.63)
\end{aligned}
$$

となる.ただし、行列 S_lのi,i’ 成分を

$$
(S_l)_{ii'} = \frac{1}{n_l}\sum_{j:y^{(j)}=l}(k(x^{(i)},x^{(j)})-\overline{k_{i,l}})(k(x^{(i')},x^{(j)})-\overline{k_{i',l}} \qquad (3.64)
$$

とした

カーネル判別分析(5)

式(3.63)のように σ_l^2 はαについての二次形式で書けるので、

$$\sigma_w^2 = \frac{1}{n}\sum_{l=1}^{c} n_l \sigma_l^2 = \alpha^T V_w \alpha, \qquad (3.65)$$

が得られる. V_wは S_l から

$$V_w = \frac{1}{n}\sum_{l=1}^{c} n_l S_l \qquad (3.66)$$

一方、クラス間の分散は各クラスの平均どうしの分散として

$$\sigma_B^2 = \sum_{l=1}^{c} \frac{n_l}{n}(\mu_l - \mu_T)^2 \qquad (3.67)$$

と定義される

カーネル判別分析(6)

ここで μ_T はサンプル全体に関する平均

$$\mu_T = \frac{1}{n}\sum_{i=1}^{n}(\alpha_i \sum_{j=1}^{n} k(x^{(i)}, x^{(j)})) \qquad (3.68)$$

である. μ_l 、μ_T が α の1次式で書けているから、σ_B^2 も2次形式となり以下のような形になる

$$\sigma_B^2 = \alpha^T V_B \alpha \qquad (3.69)$$

2次形式の比の最適化(1)

σ_B^2/σ_W^2の最大化は、分母も分子もパラメータの2次形式になっていることに注意すると、最適化は次のように考えれば簡単になる

↓

・分母、分子とも2次形式であれば、αを定数倍しても値は変化しない

→分母を定数1として仮定しても問題ない

すると、この制約下で分子を最大化する問題に帰着される

2次形式の比の最適化(2)

制約付きの最適化問題はラグランジュの未定乗数法で定式化できる

この場合ラグランジュ関数

$$L(\alpha) = \alpha^T V_B \alpha - \lambda(\alpha^T V_W \alpha - 1) \qquad (3.70)$$

の極値問題として解けるが、このままでは過学習を起こすので、正則化項を加えた

$$\tilde{\mathrm{L}}(\alpha) = \alpha^T V_B \alpha - \lambda\{\alpha^T (V_W + \zeta K)\alpha - 1\} \qquad (3.71)$$

を考える必要がある

2次形式の比の最適化(3)

$\tilde{L}(\alpha)$ をαについて微分すると

$$V_B \alpha = \lambda (V_W + \zeta K)\alpha \qquad (3.72)$$

という一般化固有値問題として定式化できる

・こうして得られたαを式(3.59)に代入して得られた関数が線形判別分析をカーネルにより一般化したカーネル判別の解である

・この形は線形判別分析とほぼ同じで違いは以下

- V_B、V_W がカーネル関数によって定義されている点
- 正則化項に対応するζKが加わっている点

2次形式の比の最適化(4)

したがって、線形判別分析を実装したプログラムを以下のように変更するだけでカーネル判別分析のプログラムが出来上がる

・サンプルを並べた行列Xを入力とする代わりに、グラム行列Kを入力として使ってクラス内、クラス間の分散共分散行列 V_W、V_B を計算する

・次に、V_W にξKを加えたものを改めてクラス内の分散共分散行列とし、一般化固有値問題を解く

判別分析が満たす特徴的な性質(1)

- 線形判別分析は、クラス内分散が(平均的に)単位行列となるように線形変換を施してから、クラス平均ベクトルの主成分分析することと等価である

  →クラス内の分散を偏りがないように補正した上で、できるだけクラス平均ベクトルが散らばるような部分空間を求めている

- カーネル判別分析についても特徴ベクトルの空間で同じことをしていると解釈できる(図3.12)

図3.12

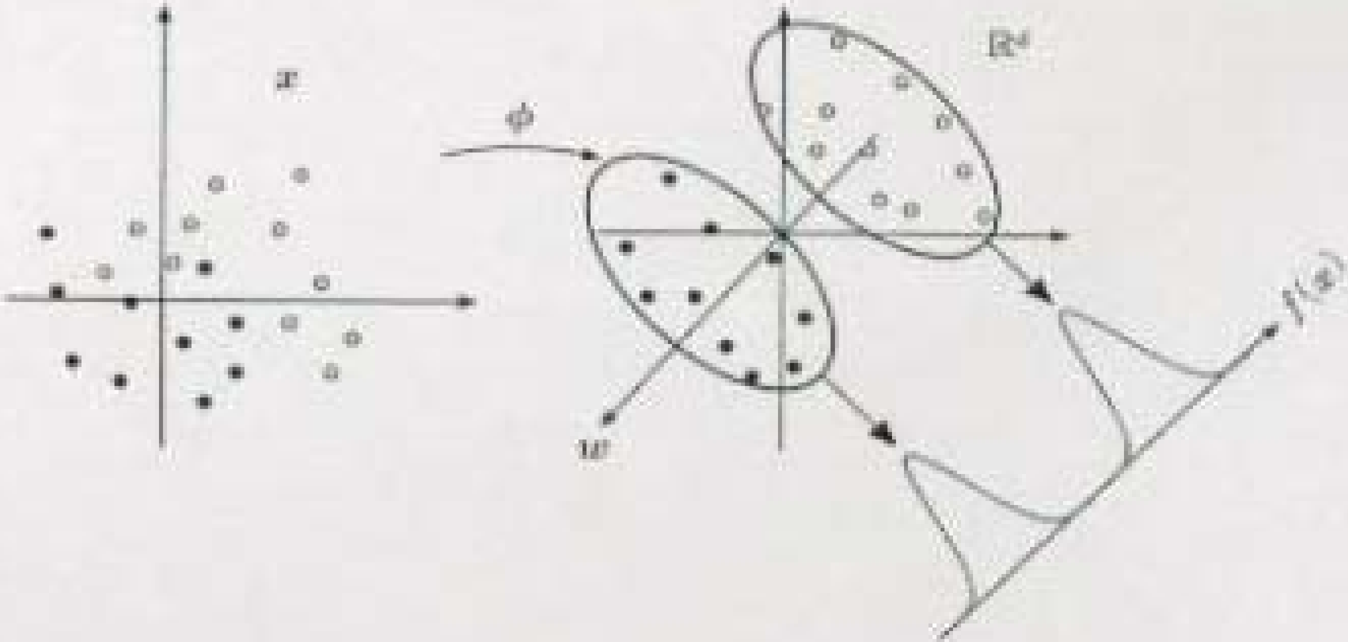

図 3.12　カーネル判別分析の概念図．特徴ベクトルの空間に移した後，クラスとクラスをできるだけ分離するような低次元空間への射影を求める．

判別分析が満たす特徴的な性質(2)

- 判別分析は主成分分析としてみると、クラス平均ベクトルだけしかサンプル数がないと言える
- これらのベクトルは常に「クラス数-1」次元の部分空間の上にあるため、判別分析で抽出される部分空間の次元は最大で「クラス数-1」であることに注意する必要がある(それ以下の固有値は0となる)

複数情報源の統合(1)

・正準相関分析

複数の情報源に共通して含まれる情報だけを抽出することによって、情報を統合するための次元圧縮法

ここでは二つの多変量x,yが同時に観測されるという状況を考える

→xとyの関係を調べるのに、すでに今まで述べた手法を使うことも考えられる

それぞれの問題点を次にまとめる

複数情報源の統合(2)

1.関数の枠組みとしてとらえ、xからyへの関数、あるいはyからxへの関数を学習するというやり方が考えられる

-関数の出力が連続値(回帰)か離散値(クラス識別)に限定され、xやyがより複雑なデータ構造の場合は限界がある

-仮に連続変数どうしだとしても非常に高次元どうしの関数を直接学習するのはパラメータの数が多すぎて次元の呪いを受けやすい

-特にxやyがそれぞれ固有の情報を持っていた場合はそれがノイズとなってしまう

2.xとyを並べて一つのベクトルとして扱い、教師なし学習の枠組みで次元圧縮をするという方法が考えられる

-必ずしもxとyの両方に含まれている情報を取り出してくるとは限らない

複数情報源の統合(3)

- 問題点を解決するために、正準相関分析では、x,yをそれぞれ同じ次元の空間に次元圧縮する

- 以下では簡単のため、1次元の関数f(x)、g(y)という関数を求めることを考える

  その際、f(x)とg(y)がxとyに共通して含まれる情報を担うようにするために、それらの間の相関係数が最大になるように関数fやgを決める

カーネル正準相関分析(1)

従来からある線形正準相関分析ではf,gはxの線形関数に取るが、カーネル正準相関分析では特徴ベクトルの線形関数に拡張する.まず、

$$f(x) = a^T \phi_x(x), \qquad g(y) = b^T \phi_y(y) \qquad (3.73)$$

という形の関数を考える.xとyは形の違うデータなので、特徴抽出も別々の関数 ϕ_xと ϕ_yを使う(図3.13)

正準相関分析ではサンプル係数

$$p_n = Cor_n[f(x), g(y)] = \frac{Cov[f(x), g(y)]}{\sqrt{Var_n[f(x)]}\sqrt{Var_n[g(y)]}} \qquad (3.74)$$

の最大化が目的である

図3.13

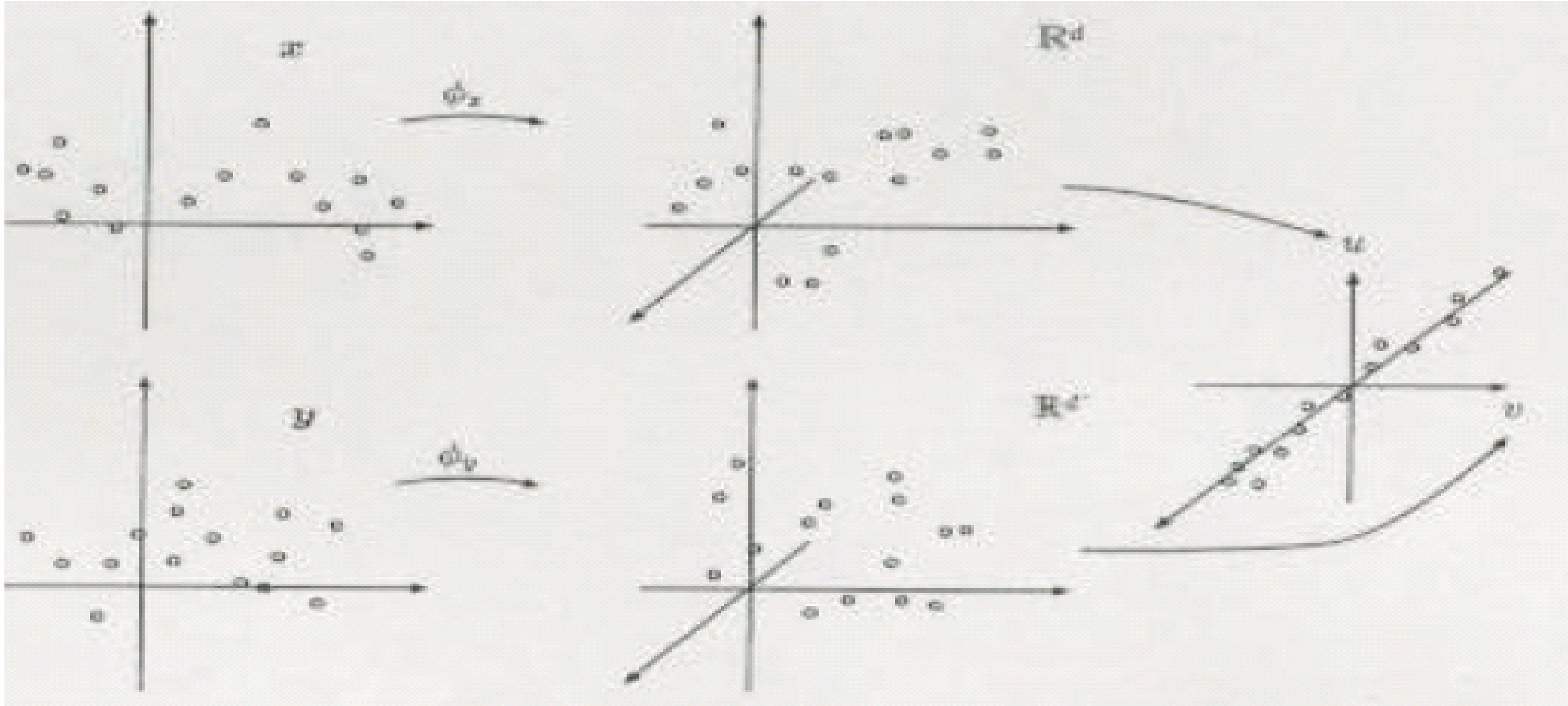

図 3.13　カーネル正準相関分析の概念図．x と y それぞれを特徴ベクトル $\phi_x(x)$, $\phi_y(y)$ に変換してから線形変換により，同じ次元の部分空間(それぞれ図の u 軸と v 軸に対応)に射影し，相関ができるだけ大きくなるようにする．

カーネル正準相関分析(2)

式(3.74)は関数の記述能力が高すぎるので回帰分析や判別分析同様過学習となってしまい、サンプル以外のデータに対しては意味のない結果しか得られない

↓そこで、、、

正則化を行う

リプレゼンター定理が適用できるように$\zeta_x \|a\|^2 + \zeta_y \|b\|^2$ を正則化項として選ぶ

するとf(x),g(y)はサンプル点におけるカーネル関数を使って以下のように書ける

$$f(x) = \sum_{i=1}^{n} \alpha_i k_x(x^{(i)}, x), \qquad g(y) = \sum_{i=1}^{n} \beta_i k_y(y^{(i)}, y) \qquad (3.75)$$

カーネル正準相関分析(3)

式(3.75)の k_x, k_y はx,yそれぞれの特徴ベクトルの内積で定義されるカーネル関数

$$k_x(x, x') = \phi_x(x)^T \phi_x(x'), \qquad k_y(y, y') = \phi_y(y)^T \phi_y(y') \qquad (3.76)$$

である

相関係数を計算するためにはまず、f(x)とg(y)の分散や共分散を計算する

分散はカーネル主成分分析で考えた式(3.22)と同じで

$$Var_n[f(x)] = \frac{1}{n}\alpha^T K_x J_n K_x \alpha \qquad (3.77)$$

$$Var_n[g(y)] = \frac{1}{n}\beta^T K_y J_n K_y \beta \qquad (3.78)$$

となる

カーネル正準相関分析(4)

同様にしてf(x)とg(y)の共分散は

$$
\begin{aligned}
&Cov_n[f(x), g(y)] \\
&= \frac{1}{n}\sum_{j=1}^{n}[\sum_{i=1}^{n}\alpha_i(k_x(x^{(i)}, x^{(j)}) - \frac{1}{n}\sum_{l=1}^{n}k_x(x^{(i)}, x^{(l)}))] \times \\
&\qquad [\sum_{i=1}^{n}\beta_i(k_y(y^{(i)}, y^{(j)}) - \frac{1}{n}\sum_{l=1}^{n}k_y(y^{(i)}, y^{(l)}))] \\
&= \frac{1}{n}\alpha^T K_x J_n K_y \beta \qquad (3.79)
\end{aligned}
$$

次に相関係数 p_n を最大にする

相関係数も判別分析の場合と同様に分母や分子に現れるものはやはり2次形式なので、固有値問題に帰着させることができる

カーネル正準相関分析(5)

ラグランジュの未定乗数法を適用すると、正則化項も含めて

$$L(\alpha,\beta)=K_x J_n \beta-\frac{\lambda_x}{2}\alpha^T(K_x J_n K_x+\zeta_x K_x)\alpha$$

$$-\frac{\lambda_y}{2}\beta^T(K_y J_n K_y+\zeta_y K_y)\beta \qquad (3.80)$$

というラグランジュ関数の極値問題として解ける

ここで、正則化項はα、βそれぞれの2次形式なので制約項の部分に入れ込んで書いた

つまり、もともと分数が1という制約だったのは、正則化と合わせて

$$\alpha^T(K_x J_n K_x+\zeta_x K_x)\alpha=\beta^T(K_y J_n K_y+\zeta_y K_y)\beta=1 \qquad (3.81)$$

という制約に変更することに相当する

カーネル正準相関分析(6)

L(α , β)を α , β についてそれぞれ微分して0とおけば

$$K_x J_n K_y \beta = \lambda_x (K_x J_n K_x + \zeta_x K_x) \alpha \qquad (3.82)$$

$$K_y J_n K_x \alpha = \lambda_y (K_y J_n K_y + \zeta_y K_y) \beta \qquad (3.83)$$

となる.ここで上の式のそれぞれの両辺に α ^T、β ^T を左から掛けると、左辺はいずれも

$$\alpha^T K_x J_n K_y \beta \qquad (3.84)$$

になり、これは最大化したい相関の値になっている

そこで右辺どうしを比較すると、式(3.81)の制約条件から $\lambda_x = \lambda_y$ が成り立つことがわかる

カーネル正準相関分析(7)

したがってそれを改めてλと書くことにすると、λは正則化付きの相関に等しいので出来るだけ大きなλを選ぶ必要がある.これは、

$$\begin{pmatrix} 0 & K_x J_n K_y \\ K_y J_n K_x & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \lambda \begin{pmatrix} K_x J_n K_x + \zeta_x K_x & 0 \\ 0 & K_y J_n K_y + \zeta_y K_y \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \qquad (3.85)$$

という一般化固有値問題になり、最大固有値に対応する固有ベクトルが解となる

図3.14にカーネル正準相関の実行例を示す

図3.14

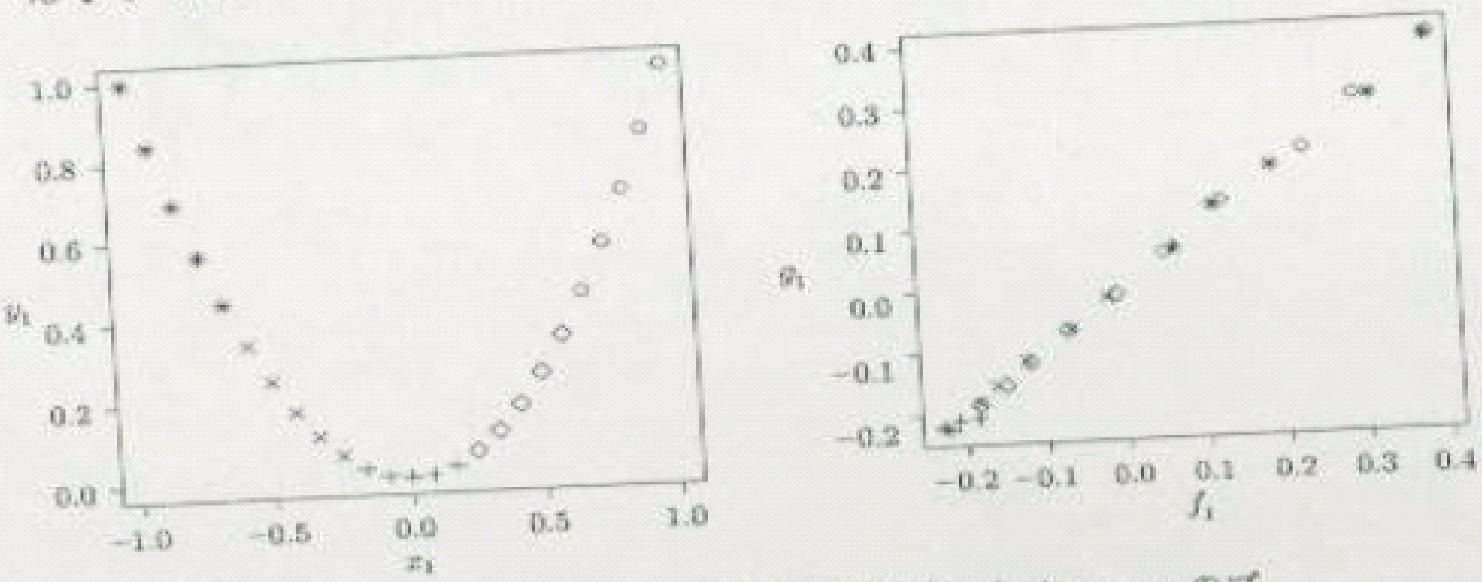

図 3.14　カーネル正準相関分析の実行例．左は x_1, y_1 のプロット（x_2, y_2 は独立な正規乱数($\mu=0$, $\sigma^2=1$)）．右は正準相関分析により相関係数最大の空間に移した $f(x)$, $g(y)$ のプロット．

カーネル正準相関の問題点

・一般にカーネル関数を使って非線形変換に広げてしまうとそれらの性質は一般に成り立たない

-例えば全てのデータを1点に移す、つまりf(x),g(y)を定数関数にするというのは明らかに無意味な変換である

→これを防ぐために分散が1であるという制約条件を入れている

-しかし全てのデータを異なる2点に移すなどはそれでは防げない

・適切なカーネル関数を選び、正則化パラメータをクロスバリデーションなどで適切に選べば比較的望ましい結果が得られることが実験的に知られている