カーネル多変量解析
第1章
現代の多変量解析とは（1）

06T4007A　伊藤輝将

データマイニング

- データマイニング

  -大量に蓄積されるデータを解析し、その中に潜む項目間の相関関係やパターンなどを探し出す技術

- カーネル法

  -データマイニングに使われる強力な道具の1つ

カーネル法

- 複雑なデータA、Bがあったとき、それらの間の関係をk(A,B)という実数値関数(カーネル関数)によって要約し、全てを数値の世界に持ち込んで処理する

- カーネル関数は関数解析に出てくる「核関数」がその名の由来(詳しくは6章を参照)

関数の推定

・数値データの間の関数関係を推定することを考える

-d個の数値を並べた変数 $\boldsymbol{x}=(x_1,x_2,\ldots,x_d)^{\mathrm{T}}$ とyという変数の組を考え、入力xからyを出力する関数を推定する問題

-サンプルデータとしてx,yの組をn個持っている

-そのうちi番目のサンプルのxを $\boldsymbol{x}^{(i)}=(x_1^{(i)},x_2^{(i)},\ldots,x_d^{(i)})^{\mathrm{T}}$

とし、それに対応するyを $y^{(i)}$ と書く(i=1,2,・・・,n)

-xが1次元なら、2次元平面上の点としてそれぞれのサンプルを描くことができる(図1.1)

図1.1

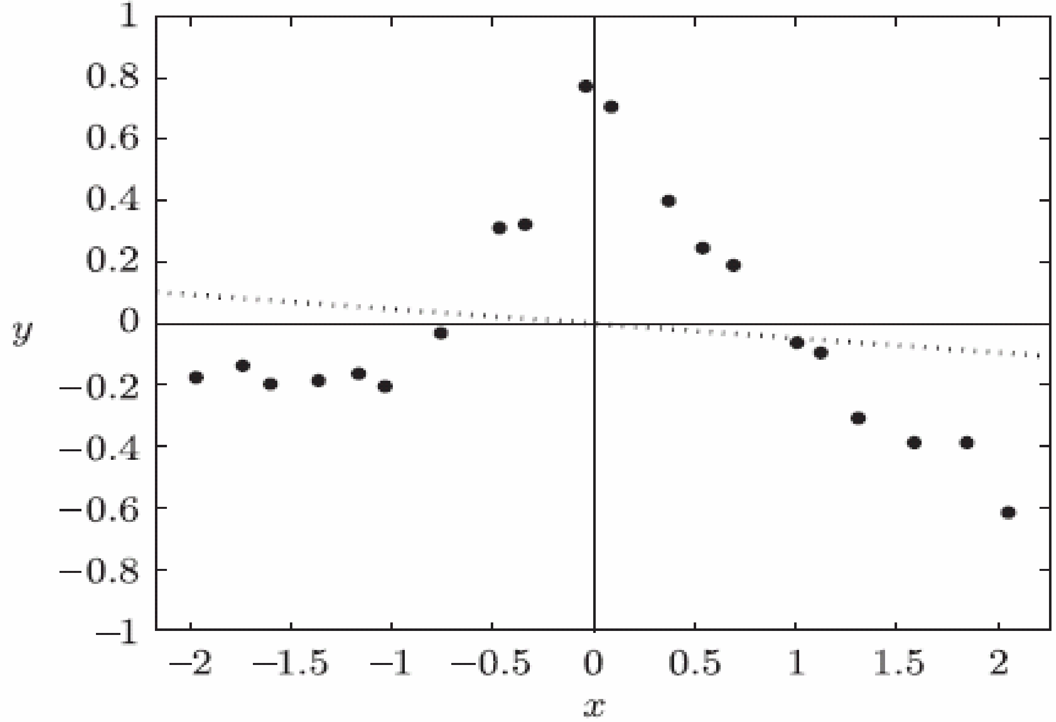

図 1.1　関数推定の例．x から y への関数を与えられた点だけを使って推定する．図の点線は，この点集合にもっともよくあてはまる原点を通る直線，すなわち線形モデルである．

基本の線形モデル(1)

・基本の線形モデル

$$y = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} = \sum_{m=1}^{d} w_m x_m \tag{1.1}$$

-実際のデータは直線でうまく当てはめられるとは限らない

↓ そこで、直線からのずれに損失を設定し、損失として二乗誤差

$$r_{\text{square}}(y, \boldsymbol{x}; \boldsymbol{w}) = (y - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x})^2 \tag{1.2}$$

を取れば、ωを求める計算が簡単になる

すべてのサンプルに対する二乗誤差の総和は以下となる

$$R(\boldsymbol{w}) = \sum_{j=1}^{n} r_{\text{square}}(y^{(j)}, \boldsymbol{x}^{(j)}; \boldsymbol{w}) = \sum_{j=1}^{n} (y^{(j)} - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}^{(j)})^2 \tag{1.3}$$

基本の線形モデル(2)

- 続き
  - 式(1.3)をシンプルに書き表すためにサンプルを並べたものをまとめて

$$\boldsymbol{y} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{pmatrix}, \qquad X = \begin{pmatrix} \boldsymbol{x}^{(1)\mathrm{T}} \\ \boldsymbol{x}^{(2)\mathrm{T}} \\ \vdots \\ \boldsymbol{x}^{(n)\mathrm{T}} \end{pmatrix} = \begin{pmatrix} x_1^{(1)} & \cdots & x_d^{(1)} \\ x_1^{(2)} & \cdots & x_d^{(2)} \\ \vdots & \ddots & \vdots \\ x_1^{(n)} & \cdots & x_d^{(n)} \end{pmatrix} \qquad (1.4)$$

のようにベクトルや行列を用いた記号を導入すれば、二乗誤差の和は以下となる

$$R(\boldsymbol{w}) = (\boldsymbol{y} - X\boldsymbol{w})^{\mathrm{T}}(\boldsymbol{y} - X\boldsymbol{w}) \qquad (1.5)$$

基本の線形モデル(3)

- 続き
  - 式(1.5)を最小にするωを求めるには、ωで微分して0となる点を見つければよい

  R(ω)はωの2次式なので、微分すれば

$$\frac{\partial R(\boldsymbol{w})}{\partial \boldsymbol{w}} = 2X^{\mathrm{T}}(\boldsymbol{y} - X\boldsymbol{w}) = \boldsymbol{0} \qquad (1.6)$$

となり、ωについて解けば以下が得られる

$$\boldsymbol{w} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\boldsymbol{y} \qquad (1.7)$$

-ただし、$X^{\mathrm{T}}X$ は逆行列を持つと仮定した

これを図1.1のサンプルに適用して引いたのが図の中に点線で表した直線であるが、誤差が大きく、このサンプルを生成した関数を適切に表現するモデルとは言い難いことがわかる

カーネル関数

・二つの入力 $\boldsymbol{x}=(x_1,\ldots,x_d)^{\mathrm{T}},\ \boldsymbol{x}'=(x'_1,\ldots,x'_d)^{\mathrm{T}}$ から計算される関数k(x,x’)

-由来や満たすべき条件などは後の章で述べる

-ここでは非常によく使われる以下の関数を使う

$$k(\boldsymbol{x},\boldsymbol{x}') = \exp\left(-\beta\|\boldsymbol{x}-\boldsymbol{x}'\|^2\right) \tag{1.8}$$

-カーネル法ではxに対して

$$y = \sum_{j=1}^{n} \alpha_j k(\boldsymbol{x}^{(j)},\boldsymbol{x}) \tag{1.9}$$

という関数を当てはめる

カーネル関数(2)

- 続き
  - 線形の場合と同じように二乗誤差

$$r_k(y,\boldsymbol{x};\boldsymbol{\alpha}) = (y - \sum_{j=1}^{n} \alpha_j k(\boldsymbol{x}^{(j)},\boldsymbol{x}))^2 \tag{1.10}$$

を最小にするように $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_n)^{\mathrm{T}}$ を決める

ここで、$K_{ij} = k(\boldsymbol{x}^{(j)},\boldsymbol{x}^{(i)})$ を(i,j)成分とするような行列を

$$K = \begin{pmatrix} k(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(1)}) & k(\boldsymbol{x}^{(2)},\boldsymbol{x}^{(1)}) & \ldots & k(\boldsymbol{x}^{(n)},\boldsymbol{x}^{(1)}) \\ k(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}) & k(\boldsymbol{x}^{(2)},\boldsymbol{x}^{(2)}) & \ldots & k(\boldsymbol{x}^{(n)},\boldsymbol{x}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ k(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(n)}) & k(\boldsymbol{x}^{(2)},\boldsymbol{x}^{(2)}) & \ldots & k(\boldsymbol{x}^{(n)},\boldsymbol{x}^{(n)}) \end{pmatrix} \tag{1.11}$$

とおくと、

カーネル関数(3)

- 続き

-二乗誤差の総和は

$$R_k(\boldsymbol{\alpha}) = \sum_{i=1}^{n} r_k(y^{(i)}, \boldsymbol{x}^{(i)}; \boldsymbol{\alpha}) = (\boldsymbol{y} - K\boldsymbol{\alpha})^{\mathrm{T}}(\boldsymbol{y} - K\boldsymbol{\alpha}) \qquad (1.12)$$

と書ける

これは線形モデルの場合のXとωをそれぞれKとαに置き換えただけ

だから、解は(Kが正方行列なら)

$$\boldsymbol{\alpha} = (K^{\mathrm{T}}K)^{-1}K^{\mathrm{T}}\boldsymbol{y} \qquad (1.13)$$

となる

さらに任意のx,x' に対してk(x,x')=k(x' ,x)が成り立ち、Kは対象行列となる

すなわち $K^{\mathrm{T}} = K$ であるから

$$\boldsymbol{\alpha} = K^{-1}\boldsymbol{y} \qquad (1.14)$$

と書ける

図1.3

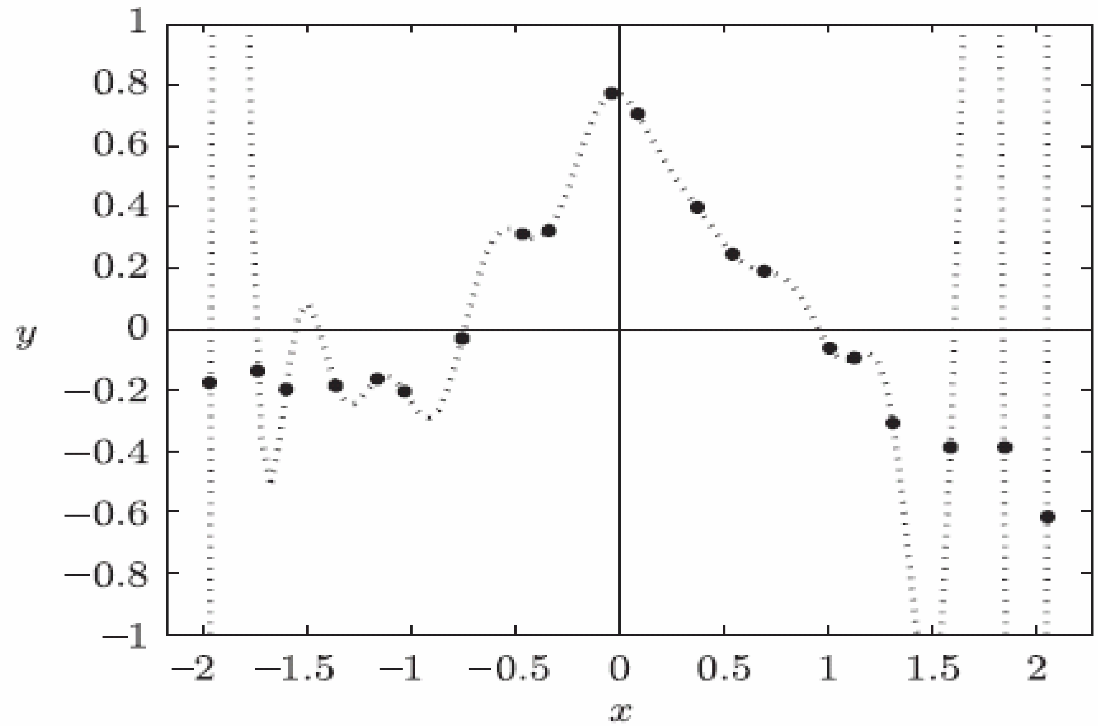

図 1.3　カーネル関数を使った近似(誤差なし)．あてはめた関数(点線で示したもの)は図のはるか外にまで発散している．

正則化

・サンプルに対する誤差のほかに余分な項を付け加えたものを最小化することによって、カーネル関数の表現能力を落とす

-以下の関数の最小化を行う

$$R_{k,\lambda}(\boldsymbol{\alpha}) = (\boldsymbol{y} - K\boldsymbol{\alpha})^{\mathrm{T}}(\boldsymbol{y} - K\boldsymbol{\alpha}) + \lambda\boldsymbol{\alpha}^{\mathrm{T}}K\boldsymbol{\alpha}, \quad \lambda > 0 \tag{1.15}$$

式(1.15)はαの2次関数だから、微分して0とおけば

$$-K(\boldsymbol{y} - K\boldsymbol{\alpha}) + \lambda K\boldsymbol{\alpha} = 0 \tag{1.16}$$

となり、ここでもKが正則だと仮定すれば以下の解が得られる

$$\boldsymbol{\alpha} = (K + \lambda I_n)^{-1}\boldsymbol{y} \tag{1.17}$$

図1.4

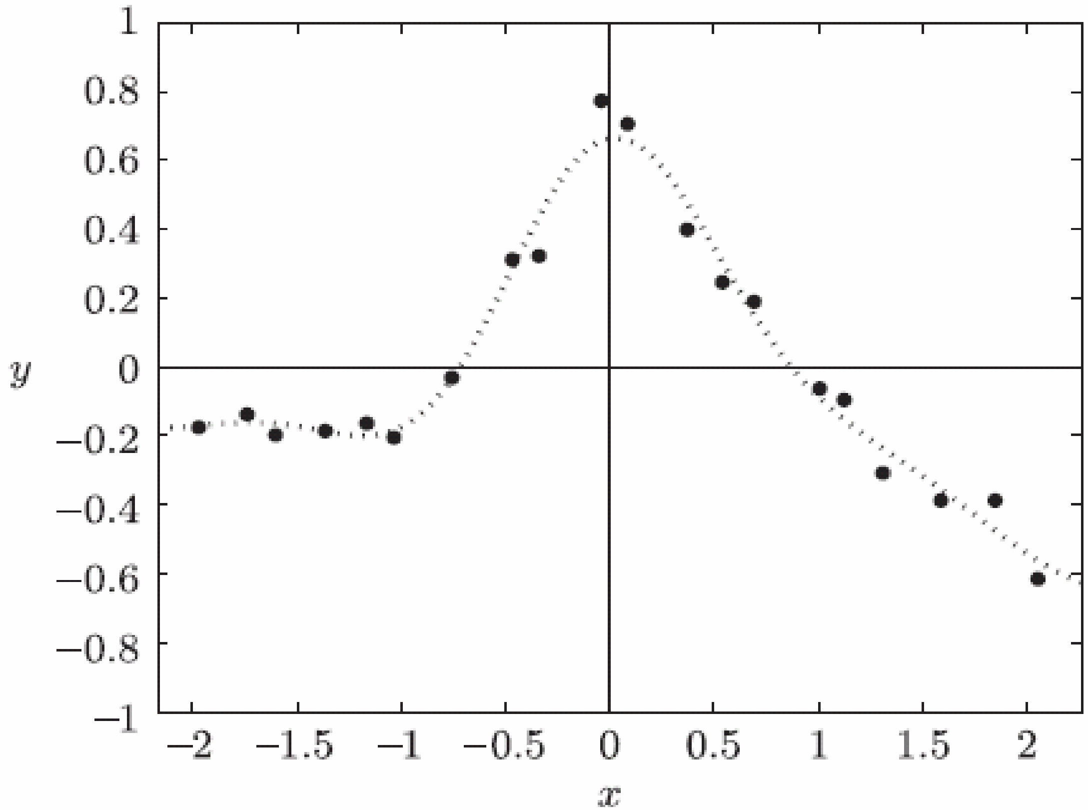

図 1.4　カーネル関数を使った近似(正則化)