# カーネル多変量解析
# 第2章　カーネル多変量解析の仕組み

佐々木研究室
三沢　博章

# 特徴抽出からの導入

基本となる線形モデル　$y = w^T x$

これではxとyの間の直線的な関係しかモデル化できない。

↓　高次の多項式をつかう

$$f(x) = \sum_{m=1}^{d} w_m x^m \tag{2.1}$$

xに何か定まった非線形変換を施して、

高次元空間に写像することを**特徴抽出**とよぶ

非線形関数で特徴抽出されたベクトル（特徴ベクトル）を

$$\phi(x) = (\phi_1(x), \ldots, \phi_d(x))^T$$

式(2,1)の例で言えば、

$$\phi(x) = (x, x^2, x^3, \ldots, x^d)^T$$

というd次元の特徴ベクトルを抽出していることに相当する。

# 特徴抽出からの導入（2）

特徴抽出された空間において線形モデルを考えると、

$$f_w(x) = w^{\mathrm{T}}\phi(x) = \sum_{m=1}^{d} w_m \phi_m(x) \qquad (2.2)$$

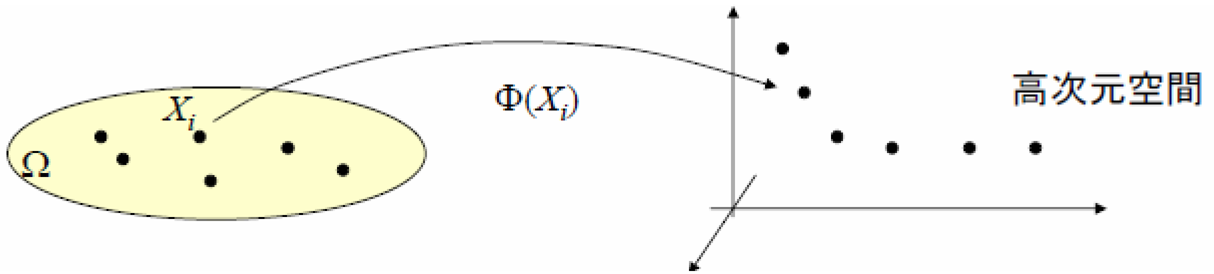
と書ける。（パラメータと特徴ベクトルの内積）

なぜ、線形から非線形へとするのか？

線形な構造だけ見ていても、

複雑なデータに対して、構造が捕らえにくい。

データが数値ベクトルにある必要がある。



# 特徴抽出からの導入（3）

特徴抽出を導入するメリット・・・

特徴抽出しない線形モデルは入力xは実ベクトルだったが、特徴ベクトルが実ベクトルならばxは実ベクトルでなくてもよい

つまり、、、

これにより、xが文字列やグラフ構造といった複雑な対象の場合でも実ベクトルと同じように手法を適用できる。

# 特徴抽出を用いたカーネル関数の定義

変数xの集合Xの２つの要素x、x´に対し、カーネル関数
k（x、x´）は、それぞれの特徴ベクトルの内積

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^{\mathrm{T}}\phi(\mathbf{x}') = \sum_{m=1}^{d} \phi_m(\mathbf{x})\phi_m(\mathbf{x}') \qquad (2.3)$$

として定義される。
式(2,1)の1変数多項式の例ではカーネル関数は

$$k(x, x') = \sum_{m=1}^{d} x^m (x')^m \qquad (2.4)$$

となる。
このように定義されたカーネル関数の重要な性質は、
式(2,2)が実はカーネル関数の線形和で表されることである。

# 内積のカーネルによる表現

式(2,2)は、十分多くの$x_1, x_2, \ldots$を適切に選ぶことにより、

$$f(x) = \sum_i \alpha_i k(x_i, x) \quad (2.5)$$

の形でいくらでも近似できる。
式(2,3)によりfは

$$f(x) = \sum_i \alpha_i \phi(x_i)^{\mathrm{T}} \phi(x) \quad (2.6)$$

と書ける。これはパラメータwを

$$w = \sum_i \alpha_i \phi(x_i) \quad (2.7)$$

と限定してよいことを意味している。
この性質は「再生性」という性質から導ける（6章参照）

# リプレゼンター定理

式(2.5)は

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

前節で述べたこの式に似ているが、

この式は「与えられたサンプル」点x^(i)でのカーネル関数の線形和

式(2.5)は「十分多くの」点$x_i$でのカーネル関数の線形和

↓

ある特別な正則化を行うことで可能となる。

↓

**リプセンター定理**

損失関数に正則化を加えて最適化する問題において、正則化項が

λ||w||^2ならば、最適解はx^(i)(i=1,2,…,n）をサンプル点として

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x^{(i)}, x) \qquad (2.8)$$

# リプレゼンター定理の証明

サンプルの特徴ベクトルの線形和を

$$w_0 = \sum_{i=1}^{n} \alpha_i \phi(x^{(i)}) \qquad (2.9)$$

とおく、wは、これにすべての$\phi$(x^(i))に直行する$\xi$を加えた

$w = w_0 + \xi$ (2.10)

という形に書ける。

このwとサンプルx^(j)の特徴ベクトル$\phi$(x^(j))との内積は、

$$f_w(x^{(j)}) = w^T \phi(x^{(j)}) = (w_0 + \xi)^T \phi(x^{(j)}) = w_0^T \phi(x^{(j)}) \quad (2.11)$$

となる。

# 続き

式(2.11)よりサンプル点だけの関数値$f_w$(x^(j))で決まる
損失関数の値はξによらない。
正則化項のほうは、
λ||w||^2=λ(||$w_0$||^2+||ξ||^2)　　　(2.12)
となり、最小値はξ＝0のときであるので、
||w||^2を正則化項にもつ場合に、
損失関数と正則化項の和を最小にする解はw=$w_0$である。
カーネル関数の定義より、式(2.11)は、$\phi(x^{(i)})^{\mathrm{T}}\phi(x)=k(x^{(i)},x)$
となるので、式(2.8)を得る

# 正定値性からの導入

前節でのカーネル関数・・・特徴ベクトルの間の内積
1章でのカーネル関数・・・サンプル間の近さを表す

内積は同じ方向を向いてるときに大きな値をとり、
直行していると0になる。
　　・・・特徴量で見たときのxとx´の類似度を表している
つまり、xとx´の類似度を表すような関数k(x,x´)があるとき
それをカーネル関数として使用できる。
↓
類似度をカーネルを結びつけるためには、
**正定値性**という性質が重要

# 正定値性からの導入(2)

関数k(x,x´)が正定値であるとは、

$$K = \begin{pmatrix} k(x_1,x_1) & k(x_2,x_1) & \cdots & k(x_n,x_1) \\ k(x_1,x_2) & k(x_2,x_2) & \cdots & k(x_n,x_2) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_1,x_n) & k(x_2,x_n) & \cdots & k(x_n,x_n) \end{pmatrix} \qquad (2.15)$$

その2次形式が常に非負

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j K_{ij} \geq 0 \qquad (2.16)$$

が任意のn次元ベクトル $\alpha=(\alpha_1,\ldots,\alpha_n)^T$ について成り立つこと

# カーネルと正定値性

これまでのカーネル関数

特徴ベクトルを計算　→　内積を計算

しかし、

関数（正定値性を満たしてる）　→　内積計算×

正定値性をカーネル関数の定義とすると、計算量が減らせる。

↓

**カーネルトリック**

例：$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left\{-\beta\|\boldsymbol{x} - \boldsymbol{x}'\|^2\right\} \qquad (2.17)$$

この数式は、ガウスカーネルと呼ばれる。

ガウスカーネルは特徴ベクトルの間の内積を表現できる。

また、正定値性も満たす。

# 多項式カーネル

実数ベクトルに対し使われるカーネルで
多項式カーネルというものがある。

$$k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^{\mathrm{T}}\mathbf{x}' + c)^{p} \tag{2.25}$$

（pは自然数、c≧0）