

## 5.8. 正則化と再生核ヒルベルト空間[Munch]

スプラインを**正則化法**や**再生核ヒルベルト空間**の範囲から考える。

正則化の一般的なクラスは、損失関数 $L(y, f(x))$ 、罰則汎関数 $J(f)$ 、 $J(f)$ の関数空間 $\mathcal{H}$ より、

$$\min_{f \in \mathcal{H}} \left[ \sum_{i=1}^N L(y_i, f(x_i)) + \lambda J(f) \right] \quad (5.42)$$

この内、**罰則汎関数** $J(f)$  についてGirosi et al. は最も一般的な以下の形を考えた。  
ここで、 $\tilde{f}$ は $f$ のフーリエ変換、 $\tilde{G}$ は $\|s\| \rightarrow \infty$ で0に収束する任意の正の関数である。

$$J(f) = \int_{\mathcal{R}^d} \frac{|\tilde{f}(s)|^2}{\tilde{G}(s)} ds \quad (5.43)$$

いくつかの仮定の上で、解は以下のように求められる。

このとき、 $\phi_k$ は罰則汎関数 $J$ の零空間、 $G$ は $\tilde{G}$ のフーリエ逆変換である。

$$f(X) = \sum_{k=1}^K \alpha_k \phi_k(X) + \sum_{i=1}^N \theta_i G(X - x_i) \quad (5.44)$$

評価関数が無限次元空間で定義されているのに、**解は有限次元であることが特徴である。**

## 5.8.1. カーネルにより生成される関数空間

**再生核ヒルベルト空間(RKHS)** は、正定値カーネル  $K(x, y)$  によって得られる関数空間である。この、無限次元の問題が有限次元の最適化問題となる現象は、**カーネル特性** と呼ばれる。

- 正定値カーネル
  1. グラム行列  $K$  が正定値
  2. 各要素がカーネル関数  $k(x_i, x_j)$
  3. カーネル関数  $k$  が再生性を持つ → 計算が楽
- カーネル法 ... データを高次元に射影して線形問題に置き換えると同時に計算量の問題を解決

<https://qiita.com/kilometer/items/82924863daba6795a981>

[https://zenn.dev/akira\\_kashihara/articles/e039331864a498](https://zenn.dev/akira_kashihara/articles/e039331864a498)

## 5.8.2. 再生核ヒルベルト空間の例

### 無限次元の一般化リッジ回帰

$$\min_{\{c_j\}_1^{\infty}} \sum_{i=1}^N \left( y_i - \sum_{j=1}^{\infty} c_j \phi_j(x_j) \right)^2 + \lambda \sum_{j=1}^{\infty} \frac{c_j^2}{\gamma_j} \quad (5.53)$$

$N$ この当てはめ値からなるベクトルは、

1. ガウス確率場の**クリギング推定量**として、(5.57)
2. **平滑化行列** $S_\lambda$  に近い形として、(5.58)

$$\hat{f} = K(K + \lambda I)^{-1}y \quad (5.57)$$

$$= (I + \lambda K^{-1})^{-1}y \quad (5.58)$$

## 罰則付き多項式回帰

$x, y \in \mathbb{R}^d$ に対するカーネル  $K(x, y) = (\langle x, y \rangle + 1)^d$  は、 $\mathbb{R}^d$  上の全次元  $d$  の多項式空間を張る  $M = \binom{p+d}{d}$  個の固有関数を持つ。

$$K(x, y) = \sum_{m=1}^M h_m(x) h_m(y) \quad (5.60)$$

$h(x)$  は、 $M$  個の直交固有関数と  $K$  個の固有値で表せる。

$V$  は  $M \times M$  サイズの直交行列、 $D_\gamma = \text{diag}(\gamma_1, \gamma_2, \dots, \gamma_M)$  である。

$$h(x)^T = V D_\gamma^{1/2} \phi(x) \quad (5.62)$$

- $\text{diag} \dots$  対角行列

次式で表される罰則付き多項式回帰問題を解くのが目標である。

$$\min_{\{\beta_m\}_1^M} \sum_{i=1}^N \left( y_i - \sum_{m=1}^M \beta_m h_m(x_i) \right)^2 + \lambda \sum_{m=1}^M \beta_m \quad (5.63)$$

基底関数の個数  $M$  は極めて大きくなる場合があり、しばしば  $N$  よりも大きくなってしまふ。

しかし、式(5.55)の解関数にカーネル表現を用いると、 $N^2$ 回のカーネル評価と  $O(N^3)$  回の演算で計算可能。

## ガウス動径基底関数

2乗損失誤差に基づくガウスカーネル  $K(x, y) = \exp(-v\|x - x_m\|^2)$

これを使って、ガウス動径基底関数の展開となる回帰モデル、

$$k_m(x) = \exp(-v\|x - x_m\|^2), \quad (m = 1, \dots, N) \quad (5.64)$$

カーネル行列  $K$  の固有値は、

1. 低次元だと滑らか
2. 高次元では波打ちだす

上記の事から、式(5.49)の罰則において **高次元の関数の係数** は低次元の関数の係数より **強く罰則を受ける** ことがわかる。



また、固有関数  $h_l(x) = \sqrt{\hat{\gamma}_m} \hat{\phi}_l(x)$ ,  $(l = 1, \dots, N)$  に対応する特徴空間表現より、

1. 固有値により、ほとんどの関数が急速に0に近づく
2. 特徴空間は無限次元だが、各基底関数に働く縮小の相対料を考えるため、有効な次元が劇的に減る
3. カーネルスケールパラメータ  $v$  は、局所的な関数  $k_m$  を意味し、特徴空間の有効な次元を増加させる

## サポートベクトル分類機

2クラス分類のための**サポートベクトルマシン(SVM)**

$$f(x) = \alpha_0 + \sum_{i=1}^N \alpha_i K(x, x_i)$$

次式を最小化することで、パラメータを求められる。

$$\min_{\alpha_0, \alpha} \left\{ \sum_{i=1}^N [1 - y_i f(x_i)]_+ + \frac{\lambda}{2} \alpha^T K \alpha \right\} \quad (5.67)$$

**サポートベクトル**という名は、 $\hat{\alpha}_i$ の多くが $\hat{\alpha}_i = 0$ となることに由来する。

これは、式(5.67)中の損失関数が**区分的に0**になっているため、0を取る値が増加している。

## 5.9. ウェーブレット平滑化

ウェーブレット平滑化は、任意のウェーブレットを使ってデータや信号の平滑化を行う処理である。

高周波成分を弱めることで、ノイズや細かな変化を取り除き、データや信号の傾向や周期を明らかにできる。

例として、

- ハールウェーブレットは、単純で理解しやすいが、滑らかさは不足している。
  - シムレットウェーブレットは、同様の正規直行基底を持つが、より滑らかである。
- しかし、トレードオフが存在する。

### 5.9.1. ウェーブレット基底とウェーブレット変換[Munch]

ウェーブレットの基底は、単一のスケーリング関数 $\phi(x)$ の**移動と拡大**によって生成される。

この関数 $\phi(x)$ は**ファザー関数**とも言う。

もし、 $\phi(x) = I(x \in [0, 1])$ なら、整数 $k$ を使って、

$$\phi_{0,k}(x) = \phi(x - k)$$

は整数上で**飛んだ値**を取る関数の**正規直行基底**を生成する。

これを**参照空間** $V_0$ と呼ぶ。

$V_{j+1}$ 内の関数は、 $V_j$ の要素と、 $V_j$ の $V_{j+1}$ に対する直行補空間 $W_j$ 内の要素の和で表せる。

これを、 $V_{j+1} = V_j \oplus W_j$ と書く。

$W_j$ 内の要素は**詳細成分**を表し、要素のいくつかの成分を0にしたい場合もある。

**マザーウェーブレット**  $\psi(x) = \phi(2x) - \phi(2x-1)$  により生成された関数  $\psi(x-k)$  は、ハール族  $W_0$  に関する**正規直行基底**を形成する。

よって、 $V_J = V_0 \oplus W_0 \oplus W_1 \oplus \dots \oplus W_{J-1}$

特徴として、

1. **全ての基底**は正規直交
2. **多重解像度解析**が可能

## 5.9.2. 適応的ウェーブレットフィルタリング[Munch]

応答ベクトル $y$ 、 $N$ 個の均等に配置された観測点上で評価された $N \times N$ の正規直交ウェーブレット基底行列 $W$ とする。

このとき、 $y^* = W^T y$ は、 $y$ のウェーブレット変換である。

適応的ウェーブレットでは、SURE(Stein Unbiased Risk Estimation)縮小法という手法を利用する。

最小2乗係数は、0に向かっていき、0になると打ち切られる。

当てはめられた関数(行列)は、逆ウェーブレット変換 $\hat{f} = W\hat{\theta}$ で求められる。

$$\min_{\theta} \|y - W\theta\|_2^2 + 2\lambda\|\theta\|_1 \quad (5.68)$$

$$\hat{\theta} = \text{sign}(y_j^*)(|y_j^*| - \lambda)_+ \quad (5.69)$$

最適な $\lambda$ を選ぶ方法は、 $\sigma$ を雑音の標準偏差推定量として、 $\lambda = \sigma\sqrt{2\log N}$ を求める方法等がある。

空間 $W$ は、多項式や自然スプライン等、任意の正規直交関数を基底から生成できる。この際、ウェーブレットが特別なのは、**基底関数が特別な形**をしているからである。これにより、**時間と周波数に関して、局在化された**表現が可能になる。

SURE基準と平滑化スプライン基準の類似性は、

1. 荒いものから滑らかなものまで、**階層的に構造化**されている。  
ウェーブレットは更に、各分解レベルで**時間的にも局在化**されている。
2. **スプライン $L_2$ 罰則**は純粹に**縮小を誘導**するが、**SUREの $L_1$ 罰則**は**縮小と選択**を行う。

相違点としては、

- 平滑化スプラインが**滑らかさを課す**ことで元の信号の圧縮を達成  
ウェーブレットは**疎性を要求**



実際に、SURE縮小によるウェーブレット当てはめと交差確認を用いた平滑化スプライン当てはめを比較すると、

- 平滑化スプラインでは、分離された**スパイク**を詳細に捉えようと、**詳細成分**が得られている。
- ウェーブレットでは、**スパイクの位置**を適切に抽出している。  
一定の小刻みは動きは余分に残ってしまっている。