正則化手法のフーリエ展開への応用

小林 飛鳥　(指導教員：吉田 裕亮)

1　はじめに

一般に, 回帰分析とは, 幾つかの説明変数からなるモデル式に基づいて, その応答とされる目的変数の振る舞いを予測する方法である. 回帰分析を行う際に説明変数の数が増すと, モデル式のデータへのフィティング精度は上がると同時に過剰なフィッティング現象も発生する. これは統計的機械学習では過学習 (オーバ・フィティング) とよばれ, 最適な予測がされない. 回帰分析において過学習を抑える手法の一つとしては正則化が知られている.

時系列データの離散 Fourier 変換は線形回帰モデルと見做すことも可能である. 本研究では離散 Fourier 展開に 2 次導関数の L^2 ノルムを罰則化項にもつ正則化により, 過学習を抑えデータの局所変動の特徴も捉えた平滑化が可能となることを具体的な実データを用いて検証する. なお最適化には 5-fold クロス・バリデーションを用いることにする.

2　関数近似

回帰モデルによる関数近似の代表例として, 多項式回帰モデルがある. 多項式回帰とは, ひとつの説明変数の冪たちからなる多項式をモデル式とする回帰手法であり, 説明変数として単項式 X^j を用いた線形回帰モデルである.

本研究では, 多項式回帰モデルを一般化し, 単項式に替えて適当な基底関数 $\phi_j(x)$ を用いた

$$f(t) = \sum_{j=0}^{m} b_j \, \phi_j(x)$$

による関数近似を考える. 一般に, どのような基底関数を用いるかによって過学習の制御手法に工夫が必要となってくる.

3　正則化

正則化とは過学習を抑えるために罰則項を加え, 最適化することである. 回帰分析における正則化には Ridge 正則化と Lasso 正則化が良く知られている.

一般に, 回帰分析における過学習は連立 1 次方程式である正規方程式の係数行列が特異行列となり, そのため行列式が 0 に近くなり, 逆行列を経由して回帰係数を求める際の数値計算において, 計算誤差のため回帰係数が安定しないため発生する. このため特に, Ridge 正則化は行列の正則化に相当する操作となる. すなわち, 係数行列は Gram 型であるため, 固有値は正領域に存在するが, 行列式が 0 に近い, つまり固有値が 0 の近傍にあるので, 正のスカラ行列により正値方向に僅かな摂動を加えることにより, 行列が正則化され解である回帰係数を安定させる.

本研究においては、近似関数の凹凸を抑え平滑化を目的として, 近似関数 f の 2 次導関数の 2 乗積分値 (L^2) を罰則項にもちいた正則化を用いる.

すなわち, 与えられたデータ点を $\{(x_i, y_i)\}_{i=1}^{n}$ とするとき

$$\min_f \left\{ \sum_{i=1}^{n} \big(y_i - f(x_i)\big)^2 + \lambda \int \{f''(t)\}^2 \mathrm{d}t \right\}$$

で与えられる. ただし, $\lambda > 0$ は罰則項の重みとなる正則化パラメータであり, この場合は平滑化パラメタ—とも呼ばれる.

4　クロスバリデーション

クロスバリデーション (CV) は予測誤差を推定する方法で最もよく知られている. 標本データを分割し, 一部をテストデータとして取り置き, 残りのトレーニングデータにより学習されたモデルの検証を行手法である.

本研究では, 5-fold クロスバリデーションを用いる. 5-fold CV は以下のように行われる.

1. データを 5 つにランダムに分割し, 1 つをテストデータ, 残り 4 つをトレーニングデータとする.
2. テストデータによる平均 2 乗誤差 MSE を求める.
3. 1. 2. のテストデータを変えながら 5 回繰り返す.

これにより, 平均 2 乗誤差の 5 つの推定値 (MSE) が得られ, 5-fold クロスバリデーションの推定値はこれらを平均することにより得られる.

5　離散 Fourier 変換と回帰モデル

本研究では, 離散 Fourier 変換を線形回帰モデルとみて, データへの関数近似を行うことで平滑化を行う. これに, 近似関数の 2 次導関数の L^2 値を罰則項とした正則化を用いる.

n 点からなる時系列データベクトル $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ に対して離散 Fourier 変換は以下のように与えられる.

n 次元ベクトルの族 $\{\mathbf{1}, \boldsymbol{c}_k, \boldsymbol{s}_k : k = 1, 2, \ldots, n\}$ を考える. ただし

$$\mathbf{1} = \frac{1}{\sqrt{n}} \big(1, 1, \ldots, 1\big)^{\mathrm{T}},$$

$$\boldsymbol{c}_k = \frac{1}{\sqrt{n}} \Big(\cos\big(2\pi k \tfrac{j}{n}\big), \cos\big(2\pi k \tfrac{2}{n}\big), \ldots, \cos\big(2\pi k \tfrac{n}{n}\big) \Big)^{\mathrm{T}},$$

$$\boldsymbol{s}_k = \frac{1}{\sqrt{n}} \Big(\sin\big(2\pi k \tfrac{1}{n}\big), \sin\big(2\pi k \tfrac{2}{n}\big), \ldots, \sin\big(2\pi k \tfrac{n}{n}\big) \Big)^{\mathrm{T}},$$

であり, このときベクトルの族 $\{\mathbf{1}, \boldsymbol{c}_k, \boldsymbol{s}_k\}$ の線形結合でベクトル $\boldsymbol{x}$ は, 以下のように展開される. すなわち,

$$\boldsymbol{x} = a_0 \, \mathbf{1} + \sum_{k=1}^{n} \big(a_k \, \boldsymbol{c}_k + b_k \, \boldsymbol{s}_k\big)$$

であり, Fourier 係数 $\{a_k, b_k\}$ は

$$a_k = \langle \boldsymbol{x} \,|\, \boldsymbol{c_k} \rangle, \qquad b_k = \langle \boldsymbol{x} \,|\, \boldsymbol{s_k} \rangle$$

となる.

6　本研究での提案手法

本研究では、まずパワースペクトルによるフィルタリングで振動成分の選択をオーバフィッティングの状態におきながら, 続いて凹凸を抑える 2 次導関数を罰則項とする正則化を施すことによりデータの局所的な特徴を保ちながら平滑化を行うこと提案する.

離散フーリエ変換では, データベクトル $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ に対して (t, x) 平面のデータ点群 $\left\{\frac{2\pi k}{n}, x_k\right\}_{k=1}^{n}$ を考えたとき区間 $(0, 2\pi]$ で, 以下の基底関数展開を行うことに対応する.

$$f(t) = \sum_{k=0}^{n} \left(a_k \cos(kt) + b_k \sin(kt)\right)$$

今、 パワースペクトル p_k に基づき m 個の振動成分 $k_1, k_2, \ldots, k_m$ が選択されたとする. このとき関数展開は,

$$f(t) = \sum_{j=1}^{m} \left(\alpha_j \cos(k_j t) + \beta_j \sin(k_j t)\right) \quad t \in (0, 2\pi]$$

と与えられる.

この場合の正則化項付きの線形回帰モデルは,

$$\min_f \left\{ \sum_{i=1}^{n} \left(x_i - f\left(\frac{2\pi i}{n}\right)\right)^2 + \lambda \int \{f''(t)\}^2 \mathrm{d}t \right\}$$

であるので, 最適化すべきパラメータを用いて最小 2 乗化部は

$$\min_f \left\{ \sum_{i=1}^{n} \left(x_i - \sum_{j=1}^{m} \left(\alpha_j \cos\left(\frac{2\pi k_j i}{n}\right) + \beta_j \sin\left(\frac{2\pi k_j i}{n}\right)\right)\right)^2 + \lambda \int \{f''(t)\}^2 \mathrm{d}t \right\}$$

となる.

近似関数 $f(t)$ の 2 次導関数は

$$f''(t) = -\sum_{j=1}^{m} \left(\alpha_j k_j^2 \cos(k_j t) + \beta_j k_j^2 \sin(k_j t)\right)$$

となるので, 直交性を用いて

$$\int_0^{2\pi} \{f''(t)\}^2 \mathrm{d}t = \sum_{j=1}^{m} \left(\alpha_j^2 k_j^4 + \beta_j^2 k_j^4\right)\pi$$

と非常に簡便化される.

7　Fourier 展開に基づく平滑化作用素

これらのことより, の最適化の式を 2 次形式の変分問題に定式化することが可能であり, 最終的に以下の形式まで整理される.

行列 $\mathbf{N}$ を $2m$ 個の基底ベクトルを並べて構成される $n \times 2m$ 行列とする.

$$\mathbf{N} = \left[\boldsymbol{c}_{k_1}\ \boldsymbol{s}_{k_1}\ \boldsymbol{c}_{k_2}\ \boldsymbol{s}_{k_2} \cdots \boldsymbol{c}_{k_m}\ \boldsymbol{s}_{k_m}\right]$$

推定される $2m$ 次の係数ベクトル $\boldsymbol{\mu}$ を

$$\boldsymbol{\mu} = \left(\alpha_1\ \beta_1\ \alpha_1\ \beta_1 \cdots \alpha_m\ \beta_m\right)^T$$

とする. また $2m$ 対角行列 $\boldsymbol{\Omega}$ を

$$\boldsymbol{\Omega} = \mathrm{Diag}\left(k_1^4, k_1^4, k_2^4, k_2^4, \ldots k_m^4, k_m^4\right)$$

とする. このとき 最適化の 2 次形式は

$$\min_{\boldsymbol{\mu}} \left\{ (\boldsymbol{x} - \mathbf{N}\boldsymbol{\mu})^t (\boldsymbol{x} - \mathbf{N}\boldsymbol{\mu}) + \lambda\, \boldsymbol{\mu}^T \boldsymbol{\Omega} \boldsymbol{\mu} \right\}$$

で表現される. この変分より

$$\boldsymbol{\mu} = \left(\mathbf{N}^T\mathbf{N} + \lambda\boldsymbol{\Omega}\right)^{-1}\mathbf{N}^T\boldsymbol{x}$$

で与えられることがわかる.

8　実データへの応用

以下は, 2022 年 7 月 1 日から 2024 年 6 月 30 日までの 2 年間の東京の日最高気温のデータにおいて, オーバフィッティング状態からの, 本研究での提案手法による平滑化結果の実践例である. (青はデータ, 赤は平滑化)

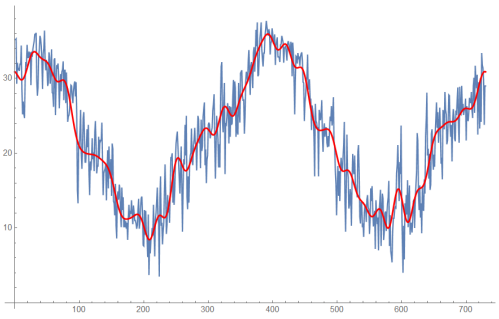

図 1: オーバフィティング状態

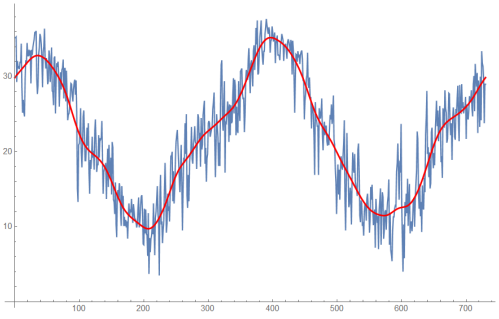

図 2: 平滑化結果

余りに平滑化され過ぎすに, 局所的な変動も適度に捉えた平滑化が実現されていることが分かる. 実際に, 離散 Fourier 変換での選択振動数は $\{1, 2, 4, 6\}$ のみとなり, かなり平滑化されてしまう.

9　まとめ

本研究では離散 Fourier 変換に 2 次導関数を正則化項にもちいる平滑化手法の提案を行った. 実データからも十分に局所的特徴を捉える平滑化が可能であるとわかった.

参考文献

[1] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer, ().