全変動正則化を用いた加法モデルによる 効率的な2値分類手法の研究



IBIS2022 2022.11.22 つくば国際会議場 武田優真 伊従寛哉 松島慎

(東京大学教養学部) (東京大学情報理工学系研究科) (東京大学総合文化研究科)

概要

全変動による目的関数の正則化はデータへの局所的適合性が高く、過学習を起こしにくいという利点がある。本研究では全変動正則化を用いた加法モデルによる分類手法の開発を目的とする。学習過程における目的関数の最小化問題は従来の手法ではデータ数nに対しO(n²)の計算量を必要とする。本研究では動的計画法を用いた方法を提案し、O(n)の計算量で最小化問題が解けることを示した。

背景

加法モデル:予測関数や識別関数を各特徴量の寄与の和で表したモデル。各関数 f_j をノンパラメトリックな手法を用いて学習する

$$f(x) = \sum_{j} f_j(x_j)$$

ただし $x = (x_i) \in \mathbb{R}^d, f_i: \mathbb{R} \to \mathbb{R}$ である。

利点:線形モデルなどのようなパラメトリックな手法よりも柔軟に関数を表現することができ、予測性能の向上を期待できる。各変数の寄与を説明可能。

問題点:従来の学習手法では計算量が多く実用的でない。

問題設定

本研究では**二値分類問題**に着目。学習データ $(x_i) \in \mathbb{R}^{n \times d}$, $(y_i) \in \{-1,1\}^n$ に対し以下の目的関数最小化問題を解くことで識別関数 $f: \mathbb{R}^d \to \mathbb{R}$ を得る。

$$\min_{f} \sum_{i=1}^{n} \log \left(1 + \exp\left(-y_i f(x_i)\right)\right) + \lambda R(f)$$

ただしRは正則化項、 λ はハイパーパラメータ。 この識別関数を加法モデルを用いて学習する。

既存手法

加法モデルの学習手法: Backfitting Algorithm [1]

加法モデル中の各関数 f_i について収束するまで以下の問題をiを変えながら解く。

$$f_j^* = \underset{f_j}{\operatorname{argmin}} \sum_{i=1}^n \ell \left(\sum_{j=1}^d f_j(x_{i,j}), y_i \right) + \lambda R(f_j)$$

ただしℓは損失関数。

これはℝ→ℝの関数を求める最適化問題。

ℝ→ℝの関数の学習手法

正則化の手法により異なる形の関数を学習する。

	<u>正則化手法の</u> <u>比較</u>	Smoothing Spline[2]	Total Variation Smoothing[3]	Trend Filtering[4][5]	
	R(f)	$\int_{x_1}^{x_n} \{f''(x)\}^2 dx$	$TV(f^{(k)})^{(\grave{\pm}\mathtt{1})}$	$ D^{(k+1)}eta _1^{(\grave{\pm}2)}$	
	局所的適合性	×	0	0	
-	解の形	区分3次関数	区分k次関数	区分k次関数	
	計算量	$O(n^2)$	$O(n^2)$	O(n)	

注1: $TV(f) = \sup_{z_1 < z_2 < \dots < z_k} \sum_i |f(z_{i+1}) - f(z_i)|$

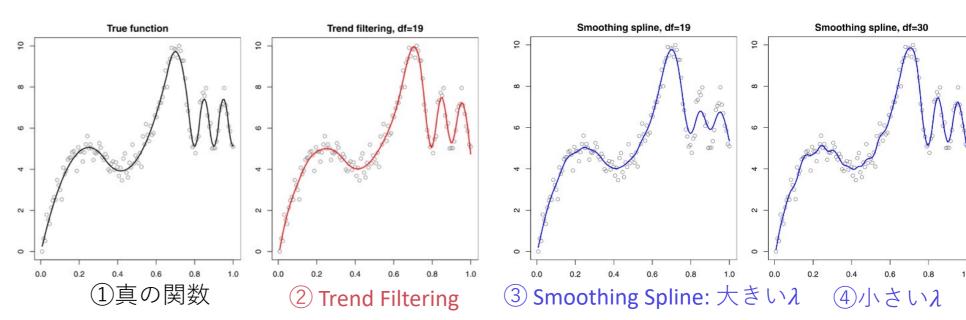
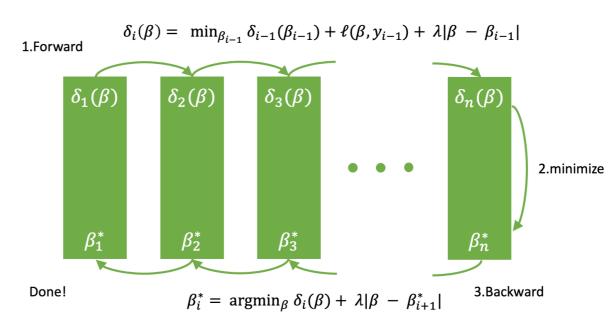


図1: Trend FilteringとSmoothing Splineの局所的適合性の比較。②は①に非常に近い形であることが伺える一方、Smoothing Splineは③に対して未学習,④に対して過学習を起こしている。[4,Fig4]

Trend Filteringにおいてk=0として2乗損失を用いた以下の最小化問題

$$\min_{\beta \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^n (\beta_i - y_i)^2 + \lambda \sum_{i=1}^{n-1} |\beta_{i+1} - \beta_i|$$

は**動的計画法**を用いて解くことができる。[6]



 ℓ は2乗損失とする。 $\delta_1 = 0$ と 定義し、 δ の更新式を用いて順 に関数 δ_i を得る。 δ_n まで計算 をしたら、これを最小化する β_n^* を求め、 β_i^* の更新式によっ て逐次的に β_1^* までを求める。 求めた β_1^* … β_n^* が最小化問題の 最適解。

提案手法

正則化にTrend Filteringを用いた以下の問題

(1)
$$\min_{\beta \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \log \left(1 + \exp\left(-y_i \sum_{j=1}^{d} \beta_{i,j} \right) \right) + \lambda \sum_{j=1}^{d} \sum_{\tau_j=1}^{n} |\beta_{\tau_j+1,j} - \beta_{\tau_j,j}|$$

を解き、全てのi,jに対し $\beta_{i,j} = f_j(x_{i,j})$ となる加法モデルの関数f(x)を識別関数として得る。ただし τ_j は $x_{1,j} \cdots x_{n,j}$ を昇順に並べた時のインデックス。これをBackfitting Algorithmによって解く際各 β_j についての最小化問題を解くが、式(1)の第1項を β_j のみの関数とみなしてこれを $F(\beta_j)$ とおくと、Backfitting Algorithmにおける最小化問題の解は以下の更新式を収束するまで繰り返し用いることで解くことができる。

$$\beta_{j} \leftarrow \underset{\beta'}{\operatorname{argmin}} \frac{1}{2} ||\beta'_{j} - (\beta_{j} - \frac{1}{L} \nabla F(\beta_{j}))||_{2}^{2} + \frac{\lambda}{L} \sum_{\tau_{j}=1}^{n-1} |\beta'_{\tau_{j},j} - \beta'_{\tau_{j}+1,j}|$$

これはk=0としたTrend Filteringの問題であり、動的計画法を用いた手法によって計算量O(n)で解くことができる。

実験

		1. fourclass n = 862, d = 2	2. Employee n = 4653, d = 8	3. skin_nonskin n = 15722, d = 3	4. smoking n = 55692, d = 21	5. smart_grid_stability n = 60000, d = 21	6. smoke_detection n = 62630, d = 12
Accuracy	Total Variation Classifier	0.938	0.801	0.952	0.759	0.972	0.997
	Logistic Regression	0.755	0.681	0.925	0.746	0.812	0.893
Fitting Time (ms)	Total Variation Classifier	1.37	15.4	24.4	708	502	484
	Logistic Regression	0.657	2.47	9.13	134	45.6	533

全データの20%をテストデータとし、学習時には5分割交差検証を行った。
Total Variation Classification: 動的計画法アルゴリズムはC++、それ以外の部分はPythonで実装。Backfitting Algorithmは関数が収束するか、全関数を4回学習するまで実行。
Logistic Regression: Scikit-learnライブラリを使用。penaltyはI1、solverはliblinearとした。
CPU: Intel(R) Core(TM) i5-7360U CPU @ 2.30GHz データ出典

1, 3: https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/

2,4,5,6: https://www.kaggle.com

参考文献

[1]Sadhanala, V. & Tibshirani, R. J., Additive Models with Trend Filtering, 2019 [2]Wood, S. N., Generalized Additive Models: an introduction with R, 2006 [3]Mammen, E. & van de Geer, S., Locally Adaptive Regression Splines, 1997 [4]Kim, S. J., Koh, K., Boyd, S. & Gorinevsky, D., I_1 Trend Filtering, 2009 [5]Tibshirani, R, J., Adaptive Piecewise Polynomial Estimation via Trend Filtering, 2014 [6]Johnson, N. A., A Dynamic Programming Algorithm for the Fused Lasso and L_0 -Segmentation, 2013