ルジャンドル多項式を用いた回帰モデルにおける正則化とモデル選択

陳　莉香　(指導教員：吉田裕亮)

1　はじめに

本研究では多項式回帰において, 単項式 (冪乗項) による回帰と Legendre 多項式による回帰の比較及び, Lasso とリッジ正則化を併用した従来の次数決定方法で求められた回帰モデルと Elastic Net に対応する L^q 正則化 ($1 \leq q \leq 2$) で求められたモデルの違いの検討を行った.

2　多項式回帰

多項式回帰とは応答変数 Y を予測変数 X の n 次多項式でモデルを作成する回帰分析における手法の一つである. 回帰係数 $\beta_0, \beta_1, \ldots, \beta_n$ とすると

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \cdots + \beta_n X^n + \varepsilon$$

と表される. ただし, $\varepsilon \sim N(0, \sigma^2)$. 最小二乗法を用いて推定し, 予測変数と応答変数との間に非線形な関係があると仮定された場合に用いられる. 極めて複雑な曲線を表現することが出来るが, 次数 n の値を大きくすれば過学習が発生し, 不自然な形状になってしまう.

過学習とは, 推定されたモデルの訓練データへの当てはまりが過剰に良く, 未知のデータに対しての予測が正常に行えない状態である [1].

3　Legendre 多項式

Legendre 多項式は直交多項式の一種であり、以下の式で定義される.

$$L_n(x) = \frac{(-1)^n}{2^n n!} \frac{d^n}{dx^n}(1-x^2)^n \quad (n = 0, 1, 2, \cdots).$$

これらは区間 $[-1, 1]$ の一様分布に関する直交多項式で

$$\int_{-1}^{1} L_k(x) L_m(x) dx = \delta_{k,m}$$

を満たすことに注意する.

4　正則化回帰モデル

正則化回帰モデルとは, 通常の最小二乗法に罰則項を加えて制約を設けることで, 係数の推定値を安定させる手法である. これにより一般に, 過学習を防ぎ, 汎用性を高めることが出来る.

通常の最小二乗法においては, 以下の式で表される残差平方和 (RSS) を最小化することで回帰係数 $\beta_0, \beta_1, \ldots, \beta_n$ が推定される.

$$RSS = \sum_{i=1}^{P} (y_i - \widehat{y}_i)^2$$

ここで i は i 番目のデータ, P はデータの総数, y_i は観測値, $\widehat{y}_i$ は予測値を表している. 正則化回帰ではこれと異なり, チューニングパラメータ $\lambda > 0$ を用いて負荷項を加えた関数を最小化する. 代表的なモデルとして最も良く知られている以下のものがある.

リッジ回帰

リッジ回帰係数 $\beta_0^R, \beta_1^R, \ldots, \beta_n^R$ は, 以下を最小化する値である.

$$RSS + \lambda \sum_{k=0}^{n} \left(\beta_k^R\right)^2$$

リッジ回帰は少ないデータ数でも機能し, 訓練データへの当てはまりが良いことが特徴である. しかし, 不要な回帰係数の絶対値を縮小するであって, 全ての係数が残される傾向にある.

Lasso 回帰

Lasso 回帰係数 $\beta_0^L, \beta_1^L, \ldots, \beta_n^L$ は以下を最小化する値である.

$$RSS + \lambda \sum_{k=0}^{n} |\beta_k^L|$$

リッジ回帰と比べると訓練データへの当てはまりはやや下がるが, パラメータ λ の値が十分大きい場合に不要な係数を 0 にする効果を持つ. つまり, 変数選択を行うことが出来るということが最大の特徴である.

Elastic Net

Elastic Net はリッジ回帰と Lasso 回帰を組み合わせた正則化回帰モデルである.Elastic Net における回帰係数 $\beta_0^E, \beta_1^E, \ldots, \beta_n^E$ は以下を最小化する値である.

$$RSS + \lambda \sum_{k=0}^{n} \left(\alpha \beta_E^2 + (1-\alpha)|\beta_k^E|\right)$$

ここで $0 \leq \alpha \leq 1$ はチューニングパラメータである. α が 0 に近づけば Lasso 回帰の特徴が強くなり, 1 に近づけばリッジ回帰の特徴が強くなる. 両回帰モデルの利点を組み合わせることを目的として用いられる.

本研究では直感的な分かりやすさを追求し, 上記の式に代わり, ほぼ同等な意味を持つ L^q 正則化 ($1 \leq q \leq 2$) を用いる [2].

$$RSS + \lambda \sum_{k=0}^{n} |\beta_k^E|^q$$

5　5 分割交差検証

過学習をどの程度回避しているか, 正則化回帰モデルの精度を評価する方法として 5 分割交差検証がある. データを 5 つのグループに分割し, 4 つを学習データ, 残りの 1 つを評価データとして誤差率となる平均二乗誤差 MSE を算出する. 5 個分すべてのデータが 1 回ずつ評価データになるよう 5 回繰り返し, 誤差率の平均を評価値 $CV_{(5)}$ とする. 数値が小さい程評価が良いとモデルと考える. 本研究ではこの評価方法を用いる.

$$CV_{(5)} = \frac{1}{5} \sum_{i=1}^{5} MSE_i$$

6　実験

6.1　実験 1

目的

多項式回帰において, 単項式 (冪乗項) による回帰と Legendre 多項式による回帰によって推定, および選択されたモデルについてどちらがより良いか調べる.

手法

1) n 次多項式によるモデルを設定する. そのモデルに従うサンプルを 60 個与える.
2) 上で作成したデータをもとに Lasso 回帰を行い, λ の値を変化させ係数の変化を観察.
3) より速く 0 に収束する係数を除き, 最適次数を選択.
4) 最適次数を用いてリッジ回帰でモデルを作成し, 5 分割交差検証で最も評価が良いものを最適な λ とする.

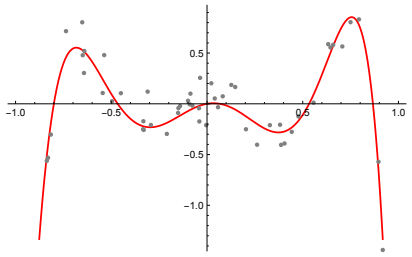

図 1: 7 次関数のモデル (赤) と与えられたデータ (点)

結果

Legendre 多項式を用いた場合のモデルは, 冪乗項を用いた場合のモデルよりも評価値が良かった. 図は 7 次関数で実験を行った結果である. 冪乗項で回帰, 選択を行った最適モデルの評価値 $CV_{(5)}$ は 0.067977 であった. 一方 Legendre 多項式を用いた場合の最適モデルの評価値 $CV_{(5)}$ は 0.031072 であった.

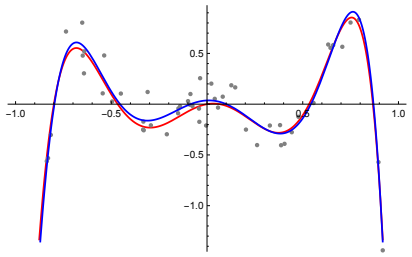

図 2: 冪乗項で回帰, 選択した最適モデル (青)

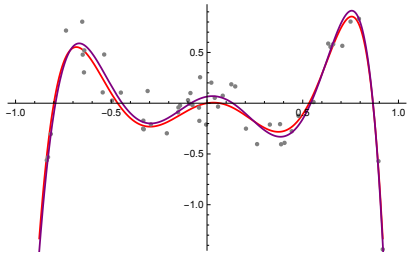

図 3: Legendre 多項式で回帰, 選択した最適モデル (紫)

6 次関数や 8 次関数における実験でも同様の結果となり, Legendre 多項式を用いた場合の方がより良いモデルが得られると考えられる.

6.2　実験 2

目的

Legendre 多項式を用いて多項式回帰を行い, L^q 正則化を用いてモデルを選択する. 実験 1 で得られた, 従来の選択手法で得られたモデルよりも良い評価が得られるかどうかを調べる.

手法

1) 実験 1 と同じデータを用いる.
2) 次数を高く設定し, 過学習を発生する状態で回帰.
3) L^q 正則化で λ, q の値を変化させモデルを作成.
4) 5 分割交差検証で最も評価が良いものを最適とする.

結果

結果として, L^q 正則化を利用して選択したモデルが従来の手法で選択したモデルの精度を上回ることはなかった. 表 1 に比較した評価値 $CV_{(5)}$ を示す. 図 4 の 7 次関数の例にあるように, フィッテングはそれほど悪くはないが, 従来の手法より改善されることはなかった.

設定モデルの次数	従来の手法	L^q 正則化を利用
6 次関数	0.002428	0.037986
7 次関数	0.031072	0.041085
8 次関数	0.122004	0.135726

表 1: 従来の手法, 及び L^q 正則化を利用して得たモデルの評価値 $CV_{(5)}$

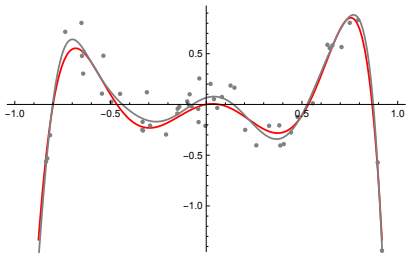

図 4: 7 次関数で L^q 正則化を利用して得たモデル (灰)

7　今後の課題

今回はシミュレーションデータを用いたが, 今後は実データで実験を行い, 同じ結果になるかどうか確かめたい. 加えて実験 2 の通り, Elastic Net をうまく利用することが出来なかったため, 有効に活用出来る事例を探索していきたい.

参考文献

[1] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R*. Springer, 2013.

[2] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2017.