

学位論文要旨

Ridge Parameters Optimization based on Minimizing Model Selection Criterion in Multivariate Generalized Ridge Regression

(モデル選択規準最小化に基づく多変量一般化リッジ回帰のリッジパラメータ最適化)

氏名 大石 峰暉

本論文では、複数の目的変数と複数の説明変数の間に線形関係を仮定したモデルである多変量線形回帰モデルを扱う。具体的には、各目的変数を説明変数と回帰係数の線形結合によって記述するモデルで、未知パラメータ行列である回帰係数行列の推定法が問題となる。その推定法としてもっとも基本的な手法の1つが最小二乗推定法で、残差平方和を最小にするパラメータを回帰係数の推定量とする手法である。最小二乗推定法により得られる最小二乗推定量は陽な形で得られるが、不偏性や漸近正規性などの理論的に優れた性質を有するため、最小二乗推定量は簡単でありながらも良い推定量であるといえる。しかしながら、最小二乗推定量は常に良い推定量といえるわけではなく、例えば説明変数間に多重共線性が生じるとき、推定量の分散が大きくなるという意味で最小二乗推定量は良い推定量とはいえない。さらに、一次従属な説明変数の組が存在すれば最小二乗推定量は計算できなくなってしまう。

このような問題に対処するための手法として、多変量一般化リッジ回帰に着目する。多変量一般化リッジ回帰では、リッジパラメータとよばれる複数の正則化パラメータにより回帰係数を縮小推定することで、多重共線性から生じる推定量の分散の増大を抑えることができ、一次従属な説明変数が存在するときでも一意な推定量の存在を保証することができる。このように多変量一般化リッジ回帰は有用であるが、推定量の縮小量を調整するパラメータであるリッジパラメータは未知であるため、より良い推定量を得るためにはリッジパラメータの最適化を考えなければならない。このような最適化問題を解くための手法としては、モデル選択規準を最小にするパラメータを最適なものとするモデル選択規準最小化法がよく知られており、例えば一般化 C_p (GC_p) 規準最小化法を用いれば、最適なリッジパラメータを陽な形で得ることができる。しかしながら、モデル選択規準最小化問題の最適解は一般的には陽な形で得られないことが知られており、その他の規準の最小化に基づく最適なリッジパラメータを得るためには数値探索アルゴリズムを用いた繰り返し計算が必要となる。このような意味で多変量一般化リッジ回帰は使いやすい手法とはいえない。そこで本論文では、多くのモデル選択規準を統一的に表すためのクラスを定義し、その規準の最小化に基づいて多変量一般化リッジ回帰のリッジパラメータを効率的に最適化するためのアルゴリズム

ムを提案する。

本論文で提案するアルゴリズムは Ohishi, Yanagihara & Fujikoshi (2020, *J. Statist. Plann. Inference*) に基づくものである。Ohishi, Yanagihara & Fujikoshi (2020) は多変量一般化リッジ回帰の特別な場合である一般化リッジ回帰に関するリッジパラメータの最適化に関する論文である。一般化リッジ回帰は一変量モデル、すなわち目的変数が 1 つのモデルに対して提案されたものであり、そのリッジパラメータの選択問題において、Ohishi, Yanagihara & Fujikoshi (2020) では残差平方和と一般化自由度で構成されるモデル選択規準をそれらの二変数関数として統一的に表現するためのクラスが定義され、その一般的なモデル選択規準の最小化に基づく最適リッジパラメータを高速に計算するためのアルゴリズムが提案されている。

多変量一般化リッジ回帰において、リッジパラメータ最適化のためのモデル選択規準を定義するためには行列に関する距離の定義が重要となる。多変量モデルにおいては、Lawley-Hotelling トレース規準 (LH 規準)、尤度比規準、Bartlett-Nanda-Pillai トレース規準のような平均構造に基づく距離の規準や、最大尤度関数、一般化最小二乗関数のような共分散構造に基づく距離の規準がよく知られている。本論文では、まずモデル選択規準をこれら 5 つの規準を含むような広いクラスの関数として定義する。さらに、この広いクラスの関数を最小にするような最適リッジパラメータを高速に計算するためのアルゴリズムを提案する。特に、LH 規準に基づくモデル選択規準の最小化は Ohishi, Yanagihara & Fujikoshi (2020) と同様の枠組みで扱うことができ、高速な計算が可能であることを示す。その他の距離に基づくモデル選択規準に最小化には、反復法と座標降下法の 2 つのアルゴリズムを用い、反復法のための反復関数や座標降下法の更新式を導出する。さらに、上記の 5 つの規準は LH 規準とその他の規準の 2 つに大別することができるが、それぞれの利点を掛け合わせた新たなリッジパラメータの最適化法である Plug-in Iteration Method も提案する。

ところで、一般化リッジ回帰は罰則付き推定法の 1 つでもあり、スパース性のない推定法であることが知られている。一方でスパース性を持つ罰則付き推定法として adaptive-Lasso 回帰がよく知られており、チューニングパラメータとよばれる 1 つの正則化パラメータによるスパースな縮小推定量が得られる。これら 2 つの推定量は全く異なるものとして認識されているが、Ohishi, Yanagihara & Kawano (2020, *Ann. Inst. Statist. Math.*) では直交する説明変数を持つ一変量線形回帰モデルにおいて、モデル選択規準最小化法により正則化パラメータを最適化した後の 2 つの推定量が理論的に完全に等しくなることが示された。本論文では Ohishi, Yanagihara & Kawano (2020) と同様に、直交する説明変数を持つ多変量線形回帰モデルにおいて、モデル選択規準最小化法により正則化パラメータを最適化した後の多変量一般化リッジ推定量と多変量 adaptive group Lasso 推定量の関係についても述べ、これらが理論的に完全に等しくなることを示す。