

# Model Selection and Active Learning for High-Dimensional Data Analysis

梅津, 佑太

<https://doi.org/10.15017/1654664>

---

出版情報 : 九州大学, 2015, 博士 (機能数理学), 課程博士  
バージョン :  
権利関係 : 全文ファイル公表済

(様式 6 - 2)

|        |  |
|--------|--|
| 氏 名    | 梅津 佑太  |
| 論 文 名  | Model Selection and Active Learning for High-Dimensional Data Analysis<br>(高次元データ解析のためのモデル選択と能動学習) |
| 論文調査委員 | 主査 九州大学 准教授 二宮 嘉行<br>副査 九州大学 教授 西井 龍映<br>副査 九州大学 教授 前園 宜彦<br>副査 九州大学 教授 増田 弘毅                      |

### 論 文 審 査 の 結 果 の 要 旨

近年、計算機の発展と測定技術の進歩により、科学や産業界の諸分野において高次元データが蓄積されるようになっており、その解析手法の需要が高まっている。高次元データにおける各変数がなんの制約も受けていなければ解析はほとんど不可能だが、実は情報をもつ変数は少ない（スパースに存在する）ことが多々ある。そして、そのケースに対処できるスパース推定法が現在急速に進歩している。本論文の主テーマである「非凹罰則付き最尤法に対する AIC の開発」では、このスパース推定法というホットトピックにおいても不可欠なモデル選択を、解析的におこなうことを可能とする。一方、非線形の統計解析手法として近年完全に定着しているカーネル法も、高次元データ解析に適した手法である。なぜなら、データを無限次元空間に写像し、写像前のデータが高次元であるか否かに手法がある意味影響を受けないからである。本論文の副テーマである「シンプルな能動学習に基づく欠陥率推定」では、カーネル法の代表といえるサポートベクターマシンにおいて、精度の高い判別を効率よく与えるためのサンプリングの実現を可能とする。

「非凹罰則付き最尤法に対する AIC の開発」は、基本的には「Umezū and Ninomiya (2015), AIC for non-concave penalized method, arXiv preprint arXiv:1509.10688」としてまとめられた内容であり、現在これは内容をより強化した形で国際誌に投稿された状態となっている。このテーマでは、一般化線形回帰モデルに対し、スパース推定法の走りといえる LASSO (least absolute shrinkage and selection operator) 法を改良した Bridge 法や SCAD (smoothly clipped absolute deviation) 法、MCP (minimax concave penalty) 法といった推定法を考えている。これらは、対数尤度関数から未知の回帰係数ベクトル  $\beta = (\beta_1, \dots, \beta_p)$  についての非凹罰則項  $p_\lambda(\beta)$  を引いたものを最大にする  $\beta$  を、回帰係数ベクトルの推定値として与える推定法である。ここで、 $p_\lambda(\cdot)$  は原点でのみ微分不可能かつ各  $|\beta_j|$  に関して単調非減少な関数である。また、 $\lambda$  はチューニングパラメータと呼ばれる正の定数であり、罰則の強さを調整する役目を果たす。これらの推定法は、罰則項の原点での微分不可能性により  $\beta_j$  のいくつかの推定値を厳密に 0 にし、つまり回帰係数ベクトルをスパースに推定する。ここで重要な問題となるのが、チューニングパラメータの値の選択である。というのも、 $\lambda$  を変えらるとどのくらいの数の  $\beta_j$  を 0 にするのが変わるので、この値を恣意的に定めると統計的に妥当な解析結果を与えることができないからである。最も単純なチューニングパラメータの選択法はクロスバリデーションを利用することであるが、これの計算コストは高い。一方、計算機的手法ではない解析

的なアプローチ，例えば情報量規準の開発が近年急速に進んでいる．しかし，それらは基本的にモデル選択の一致性などの良い性質を満たす  $\lambda$  のクラスを与えるアプローチであり，その中から適当なものを選択するルールがない．つまり，チューニングパラメータの選択に任意性が残るという問題が依然として解消されないわけである．本テーマの結果はこの問題を解消する．具体的には，AIC 元来の定義に基づき，Kullback-Leibler ダイバージェンスの漸近不偏推定量として情報量規準を導出している．また，この導出のためには推定量の漸近的性質を求める必要があり，そのための理論の開発がなされている．非凹罰則付き最尤法に対する情報量規準の中で，これは古典的な情報量規準と同じルートをもつ唯一のものといえる．そして古典的な導出に基づくが故の一つの量として与えられるため，恣意性のないモデル選択を可能にしている．この研究において認められるべき点は，競争の厳しいテーマにおいて，解決可能かつ意義のある課題を見出したことである．スパース推定において，AIC 元来の定義に基づいた情報量規準の導出に意義を見出す研究者や推定量の漸近的性質を発展させる研究者は多く存在すると思われるが，それらが同時に実現されていることは，申請者が研究に対する俯瞰力と数理的能力を併せもっていることを示唆する．

「シンプルな能動学習に基づく欠陥率推定」は，「Umezū, Matsuoka, Ikeda and Ninomiya (2015), Defect rate evaluation via simple active learning, Pacific Journal of Mathematics for Industry, 7:8, 1-8」としてまとめられた内容である．研究動機は，富士通研究所との共同研究において課された，製品製造の準備段階における欠陥率推定という問題にある．欠陥リスクに関わる可能性のある変数群の値を変えた複数回の実験的な製品製造をおこなって欠陥が起こるか否かのデータを集め，判別分析を通してリスクを評価する問題において，少ない製造回数で精度の高い評価をおこなうことはコスト上重要である．そこで，判別分析の枠組みで，意図をもって変数群の値を設計する能動学習法を本テーマでは考えている．この研究の価値は，欠陥が起こるか否かの境界が非線形でも良く働くとされる最新手法の能動学習サポートベクターマシンであっても，あまりに非線形性が強い上に欠陥率が低いと適さないことを発見したことにある．そして，その場合でも対処できる簡便な能動学習法を提案し，数値実験と実データ解析を通して提案手法が能動学習サポートベクターマシンよりうまく働くことを確認している．産業界における問題に対し，単に既存の最新手法をあてはめるだけでなく，その欠点を克服して問題解決を遂行した成果といえる．さらに申請者は，やはり富士通研究所との共同研究において要請されたアンケートデータ解析において，高次元かつ特殊なデータ形式に対処できるリッジ型の推定法の提案とその実現に必要な情報量規準の導出をおこない，要請に応じている．これらのことについては，数学分野と産業界を結びつける繋ぎ役としての役割を期待される博士（機能数理学）としての資質を，申請者が有することを示唆する．

以上より，本論文の結果は，統計科学・機械学習の分野，特に高次元データ解析におけるスパース推定法とカーネル法という方法論において，価値ある業績である．よって申請者を博士（機能数理学）の学位を受ける資格があるものと認める．