# 混合整数非線形計画問題を用いたAIC最小化

木村　圭児

九州大学大学院数理学府数理学専攻

指導教員：脇　隼人 九州大学マス・フォア・インダストリ研究所

## 1. AIC最小化

統計学におけるモデル推定では，赤池情報量規準 (Akaike's information criterion: AIC) という情報量規準を用いて変数選択を行うことがある．AICが最小であるような変数の組合せを求めることで，データとの当てはまりの良さを損なわないように予測精度の良いモデルを推定することができる．AICができるだけ小さい値をとるような変数の組合せを求める一般的な手法として，ステップワイズ法が知られている．ステップワイズ法は，Rなどの既存の統計ソフトウェアに実装されていて，計算が高速なため広く用いられている．しかし，ステップワイズ法は局所探索をしているとみなせるので，得られる変数の組合せに対するAICが最小であるとは限らない．

## 2. 既存手法と提案する手法

ロジスティック回帰における変数選択に対して，線形近似を用いて混合整数線形計画問題を用いる手法が提案されている [1]．この手法によって得られる解は，AICが最小であると保証されないが，ステップワイズ法と比べ質の良い解である．一方，線形回帰における変数選択に対しては，混合整数2次錐計画問題を用いる手法が提案されている [2]．この手法は最適性が保証され，説明変数の候補の数が30個以下であれば現実的な時間で求解できる．

本研究では，混合整数非線形計画問題 (MINLP) を用いて，AICが最小である変数の組合せを求める手法を提案する．求解には，分枝限定法のフレームワークを提供しているSCIP (Solving Constraint Integer Program)[3] を用いる．SCIPは自由度が高いソフトウェアであり，細かな解法を制御するプログラムを実装することができる．ベンチマークデータ [4] に対して，提案する手法がどの程度の規模の問題まで求解できるか紹介する．

## 3. MINLPとして定式化

説明変数の候補の数を $p$ とし，$I_p = \{1, 2, \ldots, p\}$ とする．各 $z_j$ $(j \in I_p)$ に対して，

$$z_j = \begin{cases} 1 & (j\text{ 番目の説明変数が採用されるとき}) \\ 0 & (j\text{ 番目の説明変数が採用されないとき}) \end{cases}$$

とし，$\boldsymbol{z} = (z_1, z_2, \ldots, z_p)^T$ とモデルの係数パラメータ $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ をMINLPの変数とする．$j$ 番目の説明変数を採用しないときは，対応する係数パラメータ $\beta_j = 0$ である．次の形に定式化されるMINLPを本研究では取り扱う：

$$\begin{aligned} (\mathrm{P}) \quad \min_{\boldsymbol{\beta}, \boldsymbol{z}} \quad & f(\boldsymbol{\beta}) + \lambda \sum_{j \in I_p} z_j \\ \text{s.t.} \quad & -Mz_j \le \beta_j \le Mz_j \quad (j \in I_p) \\ & z_j \in \{0, 1\} \quad (j \in I_p) \end{aligned}$$

ただし，$\lambda$ は正の定数，$M$ は十分大きい正の定数とする．目的関数にAICを与えると，$f(\boldsymbol{\beta})$ は観測データとモデルとの誤差の大きさを表し，$\lambda \sum z_j$ が説明変数の数のペナルティとして機能する．制約式より，$z_j = 0$ の場合は $\beta_j = 0$ となり，これは対応する説明変数が採用されないことを意味する．

## 4. 下界値と上界値の計算

分枝限定法では，分枝操作により $z_j$ $(j \in I_p)$ を1あるいは0に固定することで部分問題を生成する．生成された部分問題に対して，$z_j$ $(j \in I_p)$ の添字に関して次の集合を定義する．

$$\begin{aligned} Z_1 &\overset{\text{def}}{=} \{\, j \in I_p \mid z_j \text{ は 1 に固定されている} \,\} \\ Z_0 &\overset{\text{def}}{=} \{\, j \in I_p \mid z_j \text{ は 0 に固定されている} \,\} \\ Z &\overset{\text{def}}{=} \{\, j \in I_p \mid z_j \text{ はまだ固定されていない} \,\} \end{aligned}$$

問題 (P) の部分問題の緩和問題 $(\mathrm{R}_1)$ は，

$$\begin{aligned} (\mathrm{R}_1) \quad \min_{\boldsymbol{\beta}, \boldsymbol{z}} \quad & f(\boldsymbol{\beta}) + \lambda \sum_{j \in Z} z_j + \lambda \,\#(Z_1) \\ \text{s.t.} \quad & \beta_j \in \mathbb{R},\ z_j = 1 \quad (j \in Z_1) \\ & \beta_j = 0,\ z_j = 0 \quad (j \in Z_0) \\ & -Mz_j \le \beta_j \le Mz_j,\ 0 \le z_j \le 1 \quad (j \in Z) \end{aligned}$$

問題 $(\mathrm{R}_1)$ を解くことで下界値を得ることができるが，より効率良く解くために次の性質を利用する．

  

問題 ($\mathrm{R}_1$) の制約式より，問題 ($\mathrm{R}_1$) の最小解では $z_j = |\beta_j|/M$ ($j \in Z$) が満たされる．この性質を用いれば，問題 ($\mathrm{R}_1$) は次の問題 ($\mathrm{R}_2$) に変形することができる．

$$
\begin{aligned}
(\mathrm{R}_2) \quad \min_{\boldsymbol{\beta}} \quad & f(\boldsymbol{\beta}) + \lambda \sum_{j \in Z} \frac{|\beta_j|}{M} + \lambda \,\#(Z_1) \\
\text{s.t.} \quad & \beta_j \in \mathbb{R} \quad (j \in Z \cup Z_1) \\
& \beta_j = 0 \quad (j \in Z_0)
\end{aligned}
$$

次に，問題 ($\mathrm{R}_2$) の目的関数の第二項を取り除いた問題 ($\mathrm{R}_3$) について考える．

$$
\begin{aligned}
(\mathrm{R}_3) \quad \min_{\boldsymbol{\beta}} \quad & f(\boldsymbol{\beta}) + \lambda \,\#(Z_1) \\
\text{s.t.} \quad & \beta_j \in \mathbb{R} \quad (j \in Z \cup Z_1) \\
& \beta_j = 0 \quad (j \in Z_0)
\end{aligned}
$$

これらの緩和問題には，次の関係が成立する．

$$
\text{最小値}\,(\mathrm{R}_1) = \text{最小値}\,(\mathrm{R}_2) \geq \text{最小値}\,(\mathrm{R}_3)
$$

したがって，問題 ($\mathrm{R}_3$) を緩和問題として解くことで，部分問題の最小値の下界値を得る．線形回帰における AIC 最小化の場合，問題 ($\mathrm{R}_3$) は制約無し凸二次計画問題となる．よって，線形方程式を解くことで問題 ($\mathrm{R}_3$) の最小解を求めることができる．ロジスティック回帰における AIC 最小化の場合，問題 ($\mathrm{R}_3$) は制約無し凸計画問題となる．本研究では，ニュートン法を用いて問題 ($\mathrm{R}_3$) の最小解を求めている．

分枝限定法では，問題 (P) の最小値の上界値も必要である．問題 ($\mathrm{R}_3$) の最小解から問題 (P) の実行可能解を生成することができる．生成した解を用いて，問題 (P) の最小値の上界値を得る．

## 5. 効率良く解くための工夫

提案する手法では，次のような工夫を SCIP に実装することで，問題 (P) を効率良く解いている．

(I) 親問題の緩和問題を用いた下界値の更新
(II) データの一次従属性の利用
(III) 分枝変数の決め方
(IV) ニュートン法における初期解の生成
(IV) ステップワイズ法による上界値の更新

## 6. 数値実験

既存の手法（MISOCP [2] とステップワイズ法の一つである変数減少法）と提案する手法 (MINLP) の実験結果を紹介する．数値実験には，[4] で公開されているデータを使用する．計算時間は制限時間を 5,000 秒とする．5,000 秒で解けない場合は，>5000 と記し，AIC は上界値を記す．

| データ | $p$ | 手法 | AIC | time(sec) |
|---|---|---|---|---|
| 回帰分析における AIC 最小化 | | | | |
| sfM | 26 | MINLP | 2926.9 | **4.0** |
| | | MISOCP [2] | 2926.9 | 255.0 |
| bc | 32 | MINLP | **508.4** | **90.2** |
| | | MISOCP [2] | 508.6 | >5000 |
| crime | 100 | MINLP | **3410.3** | >5000 |
| | | MISOCP [2] | 3469.3 | >5000 |
| ロジスティック回帰における AIC 最小化 | | | | |
| ma | 19 | MINLP | **611.9** | 14.3 |
| | | 変数減少法 | 619.6 | 1.7 |
| breast | 34 | MINLP | **147.0** | 297.2 |
| | | 変数減少法 | 152.1 | 1.0 |
| statG | 62 | MINLP | **958.2** | >5000 |
| | | 変数減少法 | 1016.1 | 10.9 |

## 参考文献


[1] T. Sato, Y. Takano, R. Miyashiro and A. Yoshise, "Feature subset selection for logistic regression via mixed integer optimization," *Computational Optimization and Applications*, 2015.
[2] R. Miyashiro and Y. Takano, "Mixed integer second-order cone programming formulations for variable selection," *European Journal Operational Research*, **247**, pp. 721–731, 2015.
[3] T. Berthold, A. Gleixner, S. Heinz, T. Koch and Y. Shinano, "Solving mixed integer linear and nonlinear problems using the SCIP Optimization Suite," 2012.
[4] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/