多重共線性を考慮した回帰式の変数選択問題の定式化

田村　隆太，小林　健，高野　祐一，宮代　隆平，中田　和秀，松井　知己

重回帰分析では，複数の説明変数の間に強い線形従属性（多重共線性）が存在すると，回帰式の推定が数値的に不安定になり，分析結果の信頼性が低下する．このような多重共線性を回避するための簡単かつ有効な方法として，説明変数の有用な部分集合を選択して回帰式に使用する変数選択がある．本稿では，多重共線性を回避する制約条件の下で回帰式の残差二乗和を最小化する変数選択問題を，最適化ソルバーで直接求解可能な混合整数最適化問題に帰着する三種類の定式化を紹介する．

キーワード：重回帰分析，変数選択，多重共線性，混合整数最適化，条件数，分散拡大要因

1. はじめに

多重共線性とは重回帰分析においてしばしば直面する問題であり，回帰式の複数の説明変数の間に強い線形従属性が存在することである [1–3]．説明変数間の強い従属関係は，個々の説明変数が目的変数に与える影響を不明瞭にし，偏回帰係数の推定が数値的に不安定になるため，分析結果の信頼性も損なわれる．

多重共線性を回避する方法として，説明変数の有用な部分集合を選択して回帰式に使用する変数選択が有効である．また近年は，変数選択問題の厳密解法として混合整数最適化が注目を集めている．この解法は 1970 年代から提案されてはいたが [4]，当時は計算機が非力だったために実用規模の問題を解くことは難しかった．しかし，計算機や最適化アルゴリズムの性能が飛躍的に向上した現在では [5]，混合整数最適化が実用的な変数選択の手法となりつつある．ステップワイズ法 [6] に代表される多くの発見的解法とは対照的に，混合整数最適化の最大の利点は，回帰式の評価指標に関して最良の部分集合を選択できることである．

本稿では，多重共線性を回避する制約条件の下で，回帰式の残差二乗和を最小化する変数選択問題を対象とする．2 節では関連研究を紹介し，3 節では多重共線性の指標である相関係数行列の条件数と分散拡大要因（variance inflation factor, VIF）について説明する．4 節では，条件数を用いた混合整数半正定値最適化問題による定式化 [7] を提示する．5 節では，VIF を用いた混合整数二次最適化問題による二種類の定式化 [8] を提示する．これらの混合整数最適化問題は，最適化ソルバーで直接求解することが可能である．多重共線性を回避する制約条件は一見扱い難そうに思えるが，上手く変形することで扱いやすい制約条件に帰着することが可能であり，このような定式化の醍醐味が読者に伝われば幸いである．最後に 6 節で本稿のまとめを述べる．

2. 関連研究

本節では，多重共線性を回避する方法について説明し，混合整数最適化による変数選択の既存研究を紹介する．

2.1 多重共線性を回避する方法

多重共線性を回避する方法として，直交変換，罰則付き回帰，変数選択などの方法が知られている [2]．直交変換は主成分回帰 [9, 10] や PLS 回帰 [11, 12] に代表される方法であり，相関の強い説明変数を合成して使用する．直交変換を施すことで，推定の数値的不安定性は解消される．一方で，変数を合成してしまうと元の変数の効果がわかり難くなるため，分析結果の解釈が難しくなり [13]，予測精度の悪化や外れ値の影響があることも指摘されている [14, 15].

たむら　りゅうた
株式会社オクトーバー・スカイ
〒 183–0055 東京都府中市府中町 1–25–12
こばやし　けん
株式会社富士通研究所
〒 211–8588 神奈川県川崎市中原区上小田中 4–1–1
たかの　ゆういち
専修大学ネットワーク情報学部
〒 214–8580 神奈川県川崎市多摩区東三田 2–1–1
みやしろ　りゅうへい
東京農工大学大学院工学研究院
〒 184–8588 東京都小金井市中町 2–24–16
なかた　かずひで，まつい　ともみ
東京工業大学工学院経営工学系
〒 152–8552 東京都目黒区大岡山 2–12–1

罰則付き回帰には ridge 回帰 [16]，lasso 回帰 [17]，elastic net [18] などの方法があり，これらの方法では偏回帰係数の値をゼロに近づける罰則項を導入して推定する．このことにより，モデルの過剰適合が軽減され予測精度の向上が見込めるが，一方で罰則項の存在により偏回帰係数の推定値に偏りが生じ，分析結果の解釈の観点からは望ましいとは言えない [13, 19]．

変数選択は，候補となる説明変数の中から有用な部分集合を選択して回帰式に使用する方法であり，部分集合選択，特徴選択，素性選択などと呼ばれることもある．変数選択問題に対しては多くの解法が提案されているが [20–23]，それらの多くは発見的解法やメタ戦略とみなすことができる．多重共線性を除去する場合には，多重共線性がなくなるまで説明変数を一つずつ削除していく方法がよく用いられる [2]．しかながら，このような方法では，回帰式の評価指標に関して最良の部分集合を最終的に得られる保証はない．変数選択問題の厳密解法としては，変数集合の包含関係に基づく枝刈り操作を組み合わせた総当り法 [24] が知られているが，候補変数が多い場合には実用的な方法とは言い難い．そこで本稿では，混合整数最適化による変数選択に着目する．

2.2 混合整数最適化による変数選択

混合整数最適化では，選択する説明変数の数を制限する基数制約の下で，回帰式の残差二乗和や絶対偏差を最小化する問題として，変数選択問題を定式化することが多い [4, 19, 25]．この定式化では選択変数の数を指定する必要があるが，適切な数を事前に指定することは難しい．一方で，回帰式の評価指標に基づいて，選択変数の数も同時に最適化する定式化が提案されており，Mallows の C_p 規準 [26]，自由度調整済み決定係数 [27]，AIC や BIC などの情報量規準 [27]，離散型 Dantzig 選択器 [28] などの評価指標が活用されている．また，ロジスティック回帰 [29, 30]，逐次ロジット回帰 [31]，サポートベクターマシン [32]，クラスター分析 [33] などの分類モデルの変数選択や，分類木 [34] の構築に対しても，混合整数最適化手法が提案されている．変数選択の問題構造を活用して，分枝限定法を高速化する研究も行われている [35, 36]．混合整数最適化による変数選択は，医学 [37, 38]，マーケティング [39, 40]，化学 [41–43]，金融 [44] などの分野の問題に応用され，データ解析環境 R のパッケージも公開されている [45]．

2.3 混合整数最適化による多重共線性の除去

多重共線性を除去するための簡易的な方法として，相関の強い説明変数のペアに対して，少なくとも一方の変数を削除する制約条件を課すことが考えられる [13]．しかしながら，この方法では 3 個以上の変数間に存在する線形従属性を除去することができない．このような線形従属性を除去することができる方法として切除平面法があり，多重共線性を有する変数集合を排除する妥当不等式を逐次的に追加する [13]．変数減少法を利用して妥当不等式を強化し，切除平面法を高速化する方法も提案されている [7]．しかしながら切除平面法を単純に実装すると，混合整数最適化問題を繰り返し解くことになり，計算効率のよい方法とは言い難い．

3. 多重共線性の指標

本節では多重共線性の指標として，相関係数行列の条件数と VIF について説明する．

3.1 線形回帰

n 個の標本 $\{(y_i; x_{i1}, x_{i2}, \ldots, x_{ip}) \mid i = 1, 2, \ldots, n\}$ が与えられ，y_i は標本 i の目的変数の値とし，x_{ij} は標本 i の j 番目の説明変数の値とする．説明変数の添え字集合を $P := \{1, 2, \ldots, p\}$ とし，変数はすべて以下のように正規化されていることを仮定する：

$$\sum_{i=1}^{n} x_{ij} = \sum_{i=1}^{n} y_i = 0, \tag{1}$$

$$\sum_{i=1}^{n} (x_{ij})^2 = \sum_{i=1}^{n} (y_i)^2 = 1 \quad (j \in P). \tag{2}$$

このとき，線形回帰式は以下のように書ける：

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{a} + \boldsymbol{\varepsilon},$$

ただし，$\boldsymbol{y} := (y_1, y_2, \ldots, y_n)^\top$，

$$\boldsymbol{X} := \begin{pmatrix} \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

とし，$\boldsymbol{a} := (a_1, a_2, \ldots, a_p)^\top$ は推定すべき偏回帰係数のベクトル，$\boldsymbol{\varepsilon} := (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$ は各標本に対する残差のベクトルとする．

3.2 相関係数行列の条件数

説明変数の部分集合 $S\ (\subseteq P)$ を選択したと想定し，行列 $\boldsymbol{X}$ の部分行列 $\boldsymbol{X}_S := (\boldsymbol{x}_j)_{j \in S}$ を考える．このとき，回帰式の残差二乗和

$$\|\boldsymbol{y}-\boldsymbol{X}_S\boldsymbol{a}_S\|_2^2 = (\boldsymbol{y}-\boldsymbol{X}_S\boldsymbol{a}_S)^\top(\boldsymbol{y}-\boldsymbol{X}_S\boldsymbol{a}_S)$$

が最小となる偏回帰係数 $\hat{\boldsymbol{a}}_S \in \mathbb{R}^{|S|}$ は最小二乗推定量と呼ばれ，以下の正規方程式を解いて求めることができる：

$$\boldsymbol{X}_S^\top \boldsymbol{X}_S \hat{\boldsymbol{a}}_S = \boldsymbol{X}_S^\top \boldsymbol{y}. \tag{3}$$

集合 S に属する説明変数の相関係数行列は，仮定 (1)–(2) より，以下のように計算できる：

$$\boldsymbol{R}_S := (r_{j\ell})_{(j,\ell)\in S\times S} = \boldsymbol{X}_S^\top \boldsymbol{X}_S.$$

このとき相関係数行列 $\boldsymbol{R}_S$ の条件数は，その最小固有値 $\lambda_{\min}(\boldsymbol{R}_S)$ と最大固有値 $\lambda_{\max}(\boldsymbol{R}_S)$ を用いて，以下のように定義される：

$$\mathrm{cond}(\boldsymbol{R}_S) := \frac{\lambda_{\max}(\boldsymbol{R}_S)}{\lambda_{\min}(\boldsymbol{R}_S)}. \tag{4}$$

条件数が大きいと，正規方程式 (3) の係数行列 $\boldsymbol{R}_S = \boldsymbol{X}_S^\top \boldsymbol{X}_S$ が悪条件となり，方程式の解 $\hat{\boldsymbol{a}}_S$ が数値誤差の影響を強く受けることになる．条件数が 225 を超えると，多重共線性の悪影響が強くなるとされる [2].

3.3　分散拡大要因

VIF は説明変数ごとに定義され，ℓ $(\in S)$ 番目の説明変数の VIF は，相関係数行列 $\boldsymbol{R}_S$ の逆行列の ℓ 番目の対角成分により，以下のように定義される：

$$\mathrm{VIF}(\ell, S) := [\boldsymbol{R}_S^{-1}]_{\ell\ell}. \tag{5}$$

選択された変数の間に強い相関が存在する場合には，相関係数行列 $\boldsymbol{R}_S$ が特異行列に近づき，VIF の値が大きくなる．VIF が 10 を超えると，多重共線性の兆候とされる [2].

VIF には定義 (5) と等価だが異なる形の定義が知られている．この定義を説明するために，ℓ $(\in S)$ 番目の変数を目的変数とし，集合 S に属するほかの変数を説明変数とした以下の線形回帰式を考える：

$$\boldsymbol{x}_\ell = \boldsymbol{X}_{S\setminus\{\ell\}}\boldsymbol{a}^{(\ell,S)} + \boldsymbol{\varepsilon}^{(\ell,S)}, \tag{6}$$

ここで，$\boldsymbol{a}^{(\ell,S)} \in \mathbb{R}^{|S|-1}$ と $\boldsymbol{\varepsilon}^{(\ell,S)} \in \mathbb{R}^n$ はそれぞれ偏回帰係数と残差を表すベクトルである．

式 (3) と同様に，以下の正規方程式を解いて最小二乗推定量 $\hat{\boldsymbol{a}}^{(\ell,S)}$ を求めることができる：

$$\boldsymbol{X}_{S\setminus\{\ell\}}^\top \boldsymbol{X}_{S\setminus\{\ell\}} \hat{\boldsymbol{a}}^{(\ell,S)} = \boldsymbol{X}_{S\setminus\{\ell\}}^\top \boldsymbol{x}_\ell. \tag{7}$$

回帰式 (6) の決定係数は，仮定 (1)–(2) より，最小二乗推定量を用いて以下のように計算できる：

$$R^2(\ell, S) := 1 - \|\boldsymbol{x}_\ell - \boldsymbol{X}_{S\setminus\{\ell\}}\hat{\boldsymbol{a}}^{(\ell,S)}\|_2^2.$$

集合 S に属する説明変数の間に強い線形従属性が存在すると，回帰式 (6) の当てはまりがよくなり，決定係数 $R^2(\ell, S)$ の値が 1 に近づく．

この事実に基づいて，以下のように VIF を定義することができ，式 (5) と等価な定義であることが知られている [1, 3]：

$$\begin{aligned}\mathrm{VIF}(\ell, S) &:= \frac{1}{1 - R^2(\ell, S)} \\ &= \frac{1}{\|\boldsymbol{x}_\ell - \boldsymbol{X}_{S\setminus\{\ell\}}\hat{\boldsymbol{a}}^{(\ell,S)}\|_2^2}.\end{aligned} \tag{8}$$

4.　条件数に基づく定式化

本節では，条件数に基づく変数選択問題の定式化 [7] について説明する．説明変数の選択を表す 0–1 決定変数のベクトルを $\boldsymbol{z} := (z_1, z_2, \ldots, z_p)^\top$ とし，選択された説明変数の添え字集合を $S(\boldsymbol{z}) := \{j \in P \mid z_j = 1\}$ とする．条件数の上限値を表すパラメータを κ (> 1) とし，回帰式の残差二乗和を最小化する変数選択問題は以下のように書ける：

$$\begin{aligned}
&\text{minimize} && \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}\|_2^2 && (9)\\
&\text{subject to} && z_j = 0 \;\Rightarrow\; a_j = 0 \quad (j \in P), && (10)\\
& && \mathrm{cond}(\boldsymbol{R}_{S(\boldsymbol{z})}) \le \kappa, && (11)\\
& && \boldsymbol{a} \in \mathbb{R}^p,\ \boldsymbol{z} \in \{0,1\}^p. && (12)
\end{aligned}$$

$z_j = 0$ の場合には，論理制約 (10) により偏回帰係数 a_j の値がゼロになり，j 番目の説明変数は回帰式から削除される．このような論理制約は，big-M やタイプ 1 の特殊順序集合を用いて表現可能であり，論理制約を直接記述できる最適化ソルバーも存在する．

条件数の上限制約 (11) は，行列の半正定値制約として記述できる．まず，以下の対角行列 $\mathrm{Diag}(\boldsymbol{1} - \boldsymbol{z})$ と，行列のアダマール積 $\boldsymbol{R} \circ (\boldsymbol{z}\boldsymbol{z}^\top)$ を考える：

$$\mathrm{Diag}(\boldsymbol{1} - \boldsymbol{z}) = \begin{pmatrix} 1 - z_1 & 0 & \cdots & 0 \\ 0 & 1 - z_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 - z_p \end{pmatrix},$$

$$\boldsymbol{R}\circ(\boldsymbol{z}\boldsymbol{z}^\top)=\begin{pmatrix} r_{11}z_1z_1 & r_{12}z_1z_2 & \cdots & r_{1p}z_1z_p \\ r_{21}z_2z_1 & r_{22}z_2z_2 & \cdots & r_{2p}z_2z_p \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1}z_pz_1 & r_{p2}z_pz_2 & \cdots & r_{pp}z_pz_p \end{pmatrix}.$$

非負決定変数 $\lambda \in \mathbb{R}_+$ を導入し，条件数の上限制約 (11) は以下のように記述できる：

$$\lambda \boldsymbol{I} - \mathrm{Diag}(\boldsymbol{1}-\boldsymbol{z}) \preceq \boldsymbol{R}\circ(\boldsymbol{z}\boldsymbol{z}^\top) \preceq \kappa\lambda\boldsymbol{I}, \tag{13}$$

ただし，$\boldsymbol{I}$ は適切なサイズの単位行列とする．

ここでは証明の概略を述べる．まず，$\boldsymbol{R}\circ(\boldsymbol{z}\boldsymbol{z}^\top)$ を計算すると，選択された変数の相関係数行列 $\boldsymbol{R}_{S(\boldsymbol{z})}$ が残り，ほかの成分はゼロになる．このとき制約条件 (13) により，

$$\lambda \le \lambda_{\min}(\boldsymbol{R}_{S(\boldsymbol{z})}), \quad \lambda_{\max}(\boldsymbol{R}_{S(\boldsymbol{z})}) \le \kappa\lambda$$

が保証され，条件数の定義 (4) より上限制約 (11) が満たされる．

さらに制約条件 (13) の双線形項 $\boldsymbol{z}\boldsymbol{z}^\top$ を線形化し [46, 47]，目的関数 (9) を半正定値制約を用いて線形化する [48, 49]．問題 (9)–(12) は混合整数半正定値最適化問題に帰着され，SCIP-SDP[1] [50] などの最適化ソルバーを用いて解くことができる．定式化の詳細については論文 [7] を参照されたい．

5. VIF に基づく定式化

本節では，VIF に基づく変数選択問題の二種類の定式化 [8] について説明する．VIF の上限値を表すパラメータを $\alpha\ (>1)$ とし，選択された説明変数に対して VIF の上限制約を課す．このとき，回帰式の残差二乗和を最小化する変数選択問題は以下のように書ける：

$$\begin{aligned} &\text{minimize} && \|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{a}\|_2^2 && (14)\\ &\text{subject to} && z_j=0 \ \Rightarrow\ a_j=0 \quad (j\in P), && (15)\\ & && z_\ell=1 \ \Rightarrow\ \mathrm{VIF}(\ell,S(\boldsymbol{z}))\le\alpha \quad (\ell\in P), && (16)\\ & && \boldsymbol{a}\in\mathbb{R}^p,\ \boldsymbol{z}\in\{0,1\}^p. && (17) \end{aligned}$$

以降では VIF の二種類の定義に基づいて，変数選択問題 (14)–(17) を混合整数二次最適化問題に帰着する．混合整数二次最適化問題は，Gurobi[2] や CPLEX[3] などの最適化ソルバーを用いて解くことができる．

5.1 相関係数行列の逆行列に基づく定式化

本節では，相関係数行列の逆行列に基づく VIF の定義 (5) を利用する．決定変数の正方行列 $\boldsymbol{Q} := (q_{\ell j})_{(\ell,j)\in P\times P}$，$\boldsymbol{U} := (u_{\ell j})_{(\ell,j)\in P\times P}$ を導入する．このとき VIF の上限制約 (16) は，以下のように論理制約と線形制約で記述できる：

$$\begin{aligned} &q_{\ell\ell}\le\alpha \quad (\ell\in P), && (18)\\ &\boldsymbol{Q}\boldsymbol{R}_P+\boldsymbol{U}=\boldsymbol{I}, && (19)\\ &z_j=1 \ \Rightarrow\ u_{\ell j}=0 \quad (\ell\in P,\ j\in P), && (20)\\ &z_j=0 \ \Rightarrow\ q_{\ell j}=q_{j\ell}=0 \quad (\ell\in P,\ j\in P). && (21) \end{aligned}$$

ここでは証明の概略を述べる．制約条件 (20)–(21) を考慮すると，$\boldsymbol{Q}$ の主小行列 $\boldsymbol{Q}_1$ に関する等式条件 $\boldsymbol{Q}_1\boldsymbol{R}_{S(\boldsymbol{z})}=\boldsymbol{I}$ が制約条件 (19) から導かれる．したがって $\boldsymbol{Q}_1=\boldsymbol{R}_{S(\boldsymbol{z})}^{-1}$ となり，制約条件 (18) が定義 (5) に基づく VIF の上限制約 (16) を表す．定式化の詳細については論文 [8] を参照されたい．

5.2 正規方程式に基づく定式化

本節では，正規方程式 (7) に基づく VIF の定義 (8) を利用する．決定変数のベクトル $\boldsymbol{a}^{(\ell)} := (a_j^{(\ell)})_{j\in P\setminus\{\ell\}}$ $(\ell\in P)$ を導入し，以下の制約条件を考える：

$$\begin{aligned} &z_j=1 \ \Rightarrow\ \boldsymbol{x}_j^\top\boldsymbol{X}_{P\setminus\{\ell\}}\boldsymbol{a}^{(\ell)}=\boldsymbol{x}_j^\top\boldsymbol{x}_\ell \quad (j\in P\setminus\{\ell\}), && (22)\\ &z_j=0 \ \Rightarrow\ a_j^{(\ell)}=0 \quad (j\in P\setminus\{\ell\}). && (23) \end{aligned}$$

ここでは p 個の説明変数のうち最初の s 個が選択される，すなわち $S(\boldsymbol{z})=\{1,2,\ldots,s\}$ を仮定し，$\ell\in S(\boldsymbol{z})$ とする．制約条件 (23) より，ベクトル $\boldsymbol{a}^{(\ell)}$ は以下のように分割できる：

$$\boldsymbol{a}^{(\ell)}=\begin{pmatrix}\boldsymbol{a}_1^{(\ell)}\\ \boldsymbol{0}\end{pmatrix}.$$

このとき制約条件 (22) より，式 (7) と同様の正規方程式

$$\boldsymbol{X}_{S(\boldsymbol{z})\setminus\{\ell\}}^\top\boldsymbol{X}_{S(\boldsymbol{z})\setminus\{\ell\}}\boldsymbol{a}_1^{(\ell)}=\boldsymbol{X}_{S(\boldsymbol{z})\setminus\{\ell\}}^\top\boldsymbol{x}_\ell \tag{24}$$

が導かれ，$\boldsymbol{a}_1^{(\ell)}$ が最小二乗推定量になる．したがって，定義 (8) より VIF は以下のように書ける：

$$\begin{aligned}\mathrm{VIF}(\ell,S(\boldsymbol{z}))&=\frac{1}{\|\boldsymbol{x}_\ell-\boldsymbol{X}_{S(\boldsymbol{z})\setminus\{\ell\}}\boldsymbol{a}_1^{(\ell)}\|_2^2}\\ &=\frac{1}{\|\boldsymbol{x}_\ell-\boldsymbol{X}_{P\setminus\{\ell\}}\boldsymbol{a}^{(\ell)}\|_2^2}.\end{aligned}$$

よって，VIF の上限制約 (16) は以下のように書くこ

[1] http://www.opt.tu-darmstadt.de/scipsdp/
[2] https://www.octobersky.jp/products/gurobi.html
[3] http://www-03.ibm.com/software/products/ja/ibmilogcple

とができ，$z_\ell = 0$ の場合は冗長な制約となり，$z_\ell = 1$ の場合に上限制約を表す：

$$z_\ell \leq \alpha \|\boldsymbol{x}_\ell - \boldsymbol{X}_{P\setminus\{\ell\}} \boldsymbol{a}^{(\ell)}\|_2^2 \quad (\ell \in P). \tag{25}$$

ただし，この制約条件は逆凸制約となっており，混合整数最適化の枠組みで取り扱うことは難しい．しかしながら，正規方程式 (24) を利用して，制約条件 (25) の右辺は以下のように線形化することができる：

$$\begin{aligned} & \|\boldsymbol{x}_\ell - \boldsymbol{X}_{P\setminus\{\ell\}} \boldsymbol{a}^{(\ell)}\|_2^2 \\ = \ & \boldsymbol{x}_\ell^\top \boldsymbol{x}_\ell - 2\boldsymbol{x}_\ell^\top \boldsymbol{X}_{P\setminus\{\ell\}} \boldsymbol{a}^{(\ell)} \\ & + (\boldsymbol{a}_1^{(\ell)})^\top \underbrace{\boldsymbol{X}_{S(\boldsymbol{z})\setminus\{\ell\}}^\top \boldsymbol{X}_{S(\boldsymbol{z})\setminus\{\ell\}} \boldsymbol{a}_1^{(\ell)}}_{\text{式 (24) を代入}} \\ = \ & \boldsymbol{x}_\ell^\top \boldsymbol{x}_\ell - 2\boldsymbol{x}_\ell^\top \boldsymbol{X}_{P\setminus\{\ell\}} \boldsymbol{a}^{(\ell)} + (\boldsymbol{a}_1^{(\ell)})^\top \boldsymbol{X}_{S(\boldsymbol{z})\setminus\{\ell\}}^\top \boldsymbol{x}_\ell \\ = \ & \boldsymbol{x}_\ell^\top \boldsymbol{x}_\ell - \boldsymbol{x}_\ell^\top \boldsymbol{X}_{P\setminus\{\ell\}} \boldsymbol{a}^{(\ell)}. \end{aligned}$$

したがって制約条件 (22)–(23) の下で，制約条件 (25) は

$$z_\ell \leq \alpha(\boldsymbol{x}_\ell^\top \boldsymbol{x}_\ell - \boldsymbol{x}_\ell^\top \boldsymbol{X}_{P\setminus\{\ell\}} \boldsymbol{a}^{(\ell)}) \quad (\ell \in P)$$

となり，VIF の上限制約 (16) は論理制約と線形制約で記述できる．

6. おわりに

本稿では，多重共線性を考慮した変数選択問題に対して，条件数に基づく混合整数半正定値最適化問題による定式化と，VIF に基づく二種類の混合整数二次最適化問題による定式化を提示した．これらの問題は最適化ソルバーで直接求解することができ，求解が終了すれば残差二乗和に関して最良の説明変数の部分集合が得られる．また単一の問題として定式化されているため，計算を途中で打ち切ったとしても，良質な暫定解を得られる可能性が高い．計算機実験の結果については，論文 [7, 8] を参照されたい．

数理最適化の国際会議 ISMP 2015 の講演で，Dimitris Bertsimas は「私の統計学の授業では混合整数最適化を教えている．その理由は，統計学のさまざまな問題に対して混合整数最適化が利用される時代が将来必ず訪れるからだ」と述べていた．計算機と最適化アルゴリズムの性能は現在も向上を続けている．現状としては，候補変数が多い場合には混合整数最適化による変数選択は膨大な計算時間を要するが，このような状況も徐々に改善されていくだろう．統計的モデル選択の標準的な手法として，混合整数最適化が利用される時代が訪れることを期待して，本稿を終えたい．

謝辞　本研究の一部は JSPS 科研費 JP26560165，JP17K12983 の助成を受けたものです．

参考文献

[1] D. A. Belsley, E. Kuh and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, 2005.
[2] S. Chatterjee and A. S. Hadi, *Regression Analysis by Example*, 5th Edition, John Wiley & Sons, 2012.
[3] C. F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, A. K. Skidmore, D. Zurell and S. Lautenbach, “Collinearity: A review of methods to deal with it and a simulation study evaluating their performance,” *Ecography*, **36**, pp. 27–46, 2013.
[4] T. S. Arthanari and Y. Dodge, *Mathematical Programming in Statistics*, John Wiley & Sons, 1981.
[5] R. E. Bixby, “A brief history of linear and mixed-integer programming computation,” *Documenta Mathematica, Extra Volume: Optimization Stories*, pp. 107–121, 2012.
[6] M. A. Efroymson, “Multiple regression analysis,” *Mathematical Methods for Digital Computers*, A. Ralston and H. S. Wilf (eds.), John Wiley & Sons, pp. 191–203, 1960.
[7] R. Tamura, K. Kobayashi, Y. Takano, R. Miyashiro, K. Nakata and T. Matsui, “Best subset selection for eliminating multicollinearity,” *Journal of the Operations Research Society of Japan*, **60**, pp. 321–336, 2017.
[8] R. Tamura, K. Kobayashi, Y. Takano, R. Miyashiro, K. Nakata and T. Matsui, “Mixed integer quadratic optimization formulations for eliminating multicollinearity based on variance inflation factor,” Optimization Online, 2016, http://www.optimization-online.org/DB_HTML/2016/09/5655.html
[9] I. T. Jolliffe, “A note on the use of principal components in regression,” *Applied Statistics*, **31**, pp. 300–303, 1982.
[10] W. F. Massy, “Principal components regression in exploratory statistical research,” *Journal of the American Statistical Association*, **60**, pp. 234–256, 1965.
[11] H. Wold, “Estimation of principal components and related models by iterative least squares,” *Multivariate Analysis*, P. R. Krishnaiah (ed.), Academic Press, pp. 391–420, 1966.
[12] S. Wold, A. Ruhe, H. Wold and W. J. Dunn III, “The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses,” *SIAM Journal on Scientific and Statistical Computing*, **5**, pp. 735–743, 1984.
[13] D. Bertsimas and A. King, “OR forum—An algorithmic approach to linear regression,” *Operations Research*, **64**, pp. 2–16, 2016.
[14] L. E. Frank and J. H. Friedman, “A statistical view of some chemometrics regression tools,” *Technometrics*, **35**, pp. 109–135, 1993.
[15] A. S. Hadi and R. F. Ling, “Some cautionary notes on the use of principal components regression,” *The*

American Statistician, **52**, pp. 15–19, 1998.

[16] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, **12**, pp. 55–67, 1970.

[17] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B (Methodological)*, **58**, pp. 267–288, 1996.

[18] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **67**, pp. 301–320, 2005.

[19] D. Bertsimas, A. King and R. Mazumder, "Best subset selection via a modern optimization lens," *The Annals of Statistics*, **44**, pp. 813–852, 2016.

[20] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, **97**, pp. 245–271, 1997.

[21] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, **3**, pp. 1157–1182, 2003.

[22] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, **97**, pp. 273–324, 1997.

[23] H. Liu and H. Motoda (eds.), *Computational Methods of Feature Selection*, CRC Press, 2007.

[24] G. M. Furnival and R. W. Wilson, "Regressions by leaps and bounds," *Technometrics*, **42**, pp. 69–79, 2000.

[25] H. Konno and R. Yamamoto, "Choosing the best set of variables in regression analysis using integer programming," *Journal of Global Optimization*, **44**, pp. 273–282, 2009.

[26] R. Miyashiro and Y. Takano, "Subset selection by Mallows' C_p: A mixed integer programming approach," *Expert Systems with Applications*, **42**, pp. 325–331, 2015.

[27] R. Miyashiro and Y. Takano, "Mixed integer second-order cone programming formulations for variable selection in linear regression," *European Journal of Operational Research*, **247**, pp. 721–731, 2015.

[28] R. Mazumder and P. Radchenko, "The discrete Dantzig selector: Estimating sparse linear models via mixed integer linear optimization," *IEEE Transactions on Information Theory*, **63**, pp. 3053–3075, 2017.

[29] D. Bertsimas and A. King, "Logistic regression: From art to science," *Statistical Science*, **32**, pp. 367–384, 2017.

[30] T. Sato, Y. Takano, R. Miyashiro and A. Yoshise, "Feature subset selection for logistic regression via mixed integer optimization," *Computational Optimization and Applications*, **64**, pp. 865–880, 2016.

[31] T. Sato, Y. Takano and R. Miyashiro, "Piecewise-linear approximation for feature subset selection in a sequential logit model," *Journal of the Operations Research Society of Japan*, **60**, pp. 1–14, 2017.

[32] S. Maldonado, J. Pérez, R. Weber and M. Labbé, "Feature selection for support vector machines via mixed integer linear programming," *Information Sciences*, **279**, pp. 163–175, 2014.

[33] S. Benati and S. García, "A mixed integer linear model for clustering with variable selection," *Computers & Operations Research*, **43**, pp. 280–285, 2014.

[34] D. Bertsimas and J. Dunn, "Optimal classification trees," *Machine Learning*, **106**, pp. 1039–1082, 2017.

[35] D. Bertsimas and R. Shioda, "Algorithm for cardinality-constrained quadratic optimization," *Computational Optimization and Applications*, **43**, pp. 1–22, 2009.

[36] K. Kimura and H. Waki, "Minimization of Akaike's information criterion in linear regression analysis via mixed integer nonlinear program," *Optimization Methods and Software*, in press.

[37] B. Ustun and C. Rudin, "Supersparse linear integer models for optimized medical scoring systems," *Machine Learning*, **102**, pp. 349–391, 2015.

[38] B. Ustun and C. Rudin, "Optimized risk scores," In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17)*, pp. 1125–1134, 2017.

[39] T. Sato, Y. Takano and T. Nakahara, "Using mixed integer optimisation to select variables for a store choice model," *International Journal of Knowledge Engineering and Soft Data Paradigms*, **5**, pp. 123–134, 2016.

[40] T. Sato, Y. Takano and T. Nakahara, "Investigating consumers' store-choice behavior via hierarchical variable selection," arXiv preprint, arXiv:1704.00665, 2017.

[41] W. Jian, L. Zhu, Z. Xu and X. Chen, "A variable selection method for soft sensor development through mixed integer quadratic programming," *Chemometrics and Intelligent Laboratory Systems*, **167**, pp. 85–95, 2017.

[42] M. J. Willis and M. von Stosch, "L0-constrained regression using mixed integer linear programming," *Chemometrics and Intelligent Laboratory Systems*, **165**, pp. 29–37, 2017.

[43] Z. T. Wilson and N. V. Sahinidis, "The ALAMO approach to machine learning," *Computers & Chemical Engineering*, **106**, pp. 785–795, 2017.

[44] S. Maldonado, J. Pérez and C. Bravo, "Cost-based feature selection for support vector machines: An application in credit scoring," *European Journal of Operational Research*, **261**, pp. 656–665, 2017.

[45] T. Hastie, R. Tibshirani and R. J. Tibshirani, "Extended comparisons of best subset selection, forward stepwise selection, and the lasso," arXiv preprint, arXiv:1707.08692, 2017.

[46] F. A. Al-Khayyal and J. E. Falk, "Jointly constrained biconvex programming," *Mathematics of Operations Research*, **8**, pp. 273–286, 1983.

[47] G. P. McCormick, "Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems," *Mathematical Programming*, **10**, pp. 147–175, 1976.

[48] M. J. Todd, "Semidefinite optimization," *Acta Numerica*, **10**, pp. 515–560, 2001.

[49] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, **38**, pp. 49–95, 1996.

[50] T. Gally, M. E. Pfetsch and S. Ulbrich, "A framework for solving mixed-integer semidefinite programs," *Optimization Methods and Software*, in press.

