混合整数線形最適化によるカーネル SVM の変数選択

05000845　東京農工大学，オクトーバー・スカイ　*田村 隆太　TAMURA Ryuta
01309510　筑波大学　高野 祐一　TAKANO Yuichi
01606760　東京農工大学　宮代 隆平　MIYASHIRO Ryuhei

1.　はじめに

本稿では，カーネル法を用いた 2 値分類非線形サポートベクトルマシン (SVM) に対する変数選択問題を扱う．SVM の分類モデルに含まれる説明変数の数が多いほど，学習データに対する分類精度は向上する．しかし，多数の説明変数を含んだモデルでは，モデルの複雑化や過剰適合，データの取得や保管コストの増加，問題サイズの増大などの問題を引き起こす可能性がある．このため，変数選択を用いて，適切な説明変数集合による分類モデルを構築することが重要となる．

線形 SVM の変数選択問題に対しては，混合整数線形最適化問題 (MILO) として定式化する厳密解法が提案されている．対して非線形 SVM の変数選択問題には厳密解法が構築されておらず，DC アルゴリズムや再帰的特徴減少法 (RFE) などの発見的解法が主流となっている．

非線形 SVM の変数選択に対する指標の一つとして，特徴空間における 2 クラスの重心間距離が知られている [1]．ガウスカーネルを用いてこの距離を最大化する問題は，指数関数を含む目的関数を持つ混合整数非線形最適化問題 (MINLO) として定式化できるが，目的関数が非凸非凹な形であるため，汎用のソルバーなどを用いて解くことが難しかった．

本研究では，ガウスカーネルによる特徴空間上の 2 クラスの重心間距離最大化を目的とする変数選択問題に対して，MILO としての定式化を提案する．

2.　2 クラスの重心間距離

本稿では，連続する正整数の添字集合を $[n] := \{1, 2, \ldots, n\}$ と表記する．

以下では，事例 i ($i \in [n]$) が $(\boldsymbol{x}_i, y_i)$ からなる入力データを扱う．$\boldsymbol{x}_i := (x_{i1}, x_{i2}, \ldots, x_{ip})^\top$ は p 個の説明変数からなるベクトル，$y_i \in \{+1, -1\}$ は被説明変数（クラスラベル）である．さらに，各クラスに属する事例の添字集合を $N(+1)$, $N(-1)$ として，事例 i のクラスラベルをそのクラスの事例数で除した ψ_i を定義する:

$$\boldsymbol{\psi} := (\psi_i)_{i \in [n]} := \left(\frac{y_i}{|N(y_i)|} \right)_{i \in [n]} \in \mathbb{R}^n.$$

ここで，入力データ $\boldsymbol{x}$ を特徴空間へ移す写像 $\boldsymbol{\phi} : \mathbb{R}^p \to \mathcal{X}$ を考えると，$\boldsymbol{\phi}$ による特徴空間上の 2 クラスの重心間距離は，カーネル関数を用いて以下のように表現できることが Neumann ら [1] によって示されている:

$$\left\| \frac{1}{|N(+1)|} \sum_{i \in N(+1)} \boldsymbol{\phi}(\boldsymbol{x}_i) - \frac{1}{|N(-1)|} \sum_{i \in N(-1)} \boldsymbol{\phi}(\boldsymbol{x}_i) \right\|_2^2$$
$$= \sum_{i=1}^{n} \sum_{h=1}^{n} \psi_i \psi_h k(\boldsymbol{x}_i, \boldsymbol{x}_h),$$

ただし $k(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{\phi}(\boldsymbol{x})^\top \boldsymbol{\phi}(\boldsymbol{x}')$ は特徴空間における内積を表すカーネル関数である．

これにより，高次の特徴空間における情報を，元の変数の空間で扱うことが可能となる．また，本稿ではカーネル関数として，次のガウスカーネルを用いる:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_h) = \exp\left(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{hj})^2 \right)$$
$$= \exp\left(\sum_{j=1}^{p} c_{ihj} \right),$$

ここで γ は利用者が指定するスケーリング定数であり，c_{ihj} は簡略化のため定数項をまとめたものである．ガウスカーネルは無限次元の特徴空間における内積を表すことが知られている．

3.　混合整数最適化問題としての定式化

特徴空間上での 2 クラスの重心間距離を最大化するような説明変数集合を選び出す変数選択問題を考える．まず，変数ベクトル $\boldsymbol{z} := (z_1, z_2, \ldots, z_p)^\top$

を定義し，z_j ($j \in [p]$) を j 番目の説明変数を選択するとき 1, そうでないとき 0 となる 0–1 変数を表すものとする．これを用いると，今考えている変数選択問題は次の式 (1)–(3) により MINLO として記述できる:

$$\begin{aligned}
\underset{\boldsymbol{z}}{\text{maximize}} \quad & \sum_{i=1}^{n}\sum_{h=1}^{n}\psi_i\psi_h \exp\left(\sum_{j=1}^{p} c_{ihj}z_j\right) && (1)\\
\text{subject to} \quad & \sum_{j=1}^{p} z_j \le \theta, && (2)\\
& \boldsymbol{z} \in \{0,1\}^p, && (3)
\end{aligned}$$

ここで θ は利用者が指定する定数であり，選択する説明変数の個数（基数）の上限値である．この最適化問題の目的関数は非凸非凹な形であり，このままでは大域的最適解を求めることは難しい．

本研究では，以下の定理で問題 (1)–(3) が MILO に変形可能であることを示す．

定理． ($\boldsymbol{e}^*, \boldsymbol{z}^*$) を MILO (4)–(9) の最適解とする．このとき，$\boldsymbol{z}^*$ は MINLO (1)–(3) の最適解となる．

$$\begin{aligned}
\underset{\boldsymbol{e},\boldsymbol{z}}{\text{maximize}} \quad & \sum_{i=1}^{n}\sum_{h=1}^{n}\psi_i\psi_h e_{ihp} && (4)\\
\text{subject to} \quad & \sum_{j=1}^{p} z_j \le \theta, && (5)\\
& e_{ih0} = 1 \quad (i \in [n], h \in [n]), && (6)\\
& -Mz_j \le e_{ihj} - e_{ih,j-1} \le Mz_j \\
& \qquad (i \in [n], h \in [n], j \in [p]), && (7)\\
& -M \cdot (1 - z_j) \\
& \quad \le e_{ihj} - e_{ih,j-1} \cdot \exp\left(c_{ihj}\right) \\
& \quad \le M \cdot (1 - z_j) \\
& \qquad (i \in [n], h \in [n], j \in [p]), && (8)\\
& \boldsymbol{e} \in \mathbb{R}_{+}^{n \times n \times (p+1)},\ \boldsymbol{z} \in \{0,1\}^p, && (9)
\end{aligned}$$

ただし e_{ihj} ($i \in [n], h \in [n], j \in \{0\} \cup [p]$) は補助的に導入する非負連続変数，$M$ は十分に大きな正定数である（M の上界として $M = 1$ が導出できる）．

証明． 任意の $\boldsymbol{z} \in \{0,1\}^p$ について，以下の関係が満たされることを示せばよい:

$$e_{ihp} = \exp\left(\sum_{j=1}^{p} c_{ihj}z_j\right).$$

まず，各 $(i,h,j) \in [n] \times [n] \times [p]$ について以下の式が成り立つことに注意する:

$$\exp\left(c_{ihj}z_j\right) = \begin{cases} \exp(0) = 1 & \text{if } z_j = 0 \\ \exp\left(c_{ihj}\right) & \text{if } z_j = 1\,. \end{cases}$$

よって，式 (7), (8) から次の等式が導かれる:

$$e_{ihj} = e_{ih,j-1} \cdot \exp\left(c_{ihj}z_j\right) \quad (j \in [p]).$$

この関係を $e_{ih0} = 1$ から再帰的に適用することで，e_{ihp} は次のように定まる:

$$\begin{aligned}
e_{ihp} &= \prod_{j=1}^{p} \exp\left(c_{ihj}z_j\right) \cdot e_{ih0} \\
&= \exp\left(\sum_{j=1}^{p} c_{ihj}z_j\right). \qquad \text{（証明終）}
\end{aligned}$$

さらに筆者らの論文 [2] では，問題の性質を利用して，問題 (4)–(9) の計算効率の改善が可能であることを示している．本稿では詳細は省略するが，それらの結果を簡単に記す．

- 冗長な変数の削減: 連続変数 e_{ihj} の個数を，$(p+1)n^2$ 個から $(p+1)n(n-1)/2$ 個へ削減
- 冗長な制約の削減: big-M 制約 (7), (8) の本数を，$4pn^2$ 本から $pn(n-1)$ 本へ削減
- タイトな big-M 定数の導出: 各 (i,h,j) について，定数 M の上界を 1 から $1 - \exp(c_{ihj})$ まで減少

4. 計算機実験

実験結果の詳細については，発表会当日に報告する．

参考文献

[1] J. Neumann, C. Schonörr and G. Steidl (2005): Combined SVM-based feature selection and classification. *Machine Learning*, **61**, 129–150.

[2] R. Tamura, Y. Takano and R. Miyashiro (2022): Feature subset selection for kernel SVM classification via mixed-integer optimization. *arXiv*:2205.14325.