

混合整数最適化によるロジットモデルの変数選択

高野 祐一

筑波大学 システム情報系

統計的モデルを作成する際に、候補変数の中から有用な説明変数を選択することは変数選択と呼ばれる。変数選択には、過剰適合を軽減し予測性能を向上させる、モデルを単純化し結果の解釈を容易にする、などの利点がある。本講演では、統計的な離散選択モデルである各種のロジットモデルに対して、混合整数最適化による変数選択手法を紹介する。

ロジスティック回帰モデル [1] ロジスティック回帰（2項ロジット）モデルでは2クラス分類を対象とする。ロジスティック損失関数と呼ばれる非線形の1変数関数が対数尤度関数に含まれるため、混合整数最適化による変数選択問題を最適化ソルバーで直接求解することは難しい。そこで本講演では、ロジスティック損失関数に区分線形近似を施し、最適化ソルバーによる直接求解が可能な定式化を提案する。

逐次ロジットモデル [2] 逐次ロジットモデルでは、ロジスティック回帰による2クラス分類を繰り返して多クラス分類を行なう。それゆえロジスティック回帰モデルの場合と同様に、対数尤度関数にロジスティック損失関数が含まれる。既存研究ではテイラー展開に基づく2次近似を用いた定式化が提案されているが、本講演では区分線形近似を用いた定式化を提案する。

順序ロジットモデル [3] 順序ロジットモデルは、順序関係が定められた複数のクラスを分類する手法である。順序ロジットモデルの対数尤度関数には非線形の2変数関数が含まれる。そこで本講演では、区分線形（接線）近似を2変数関数に拡張した接平面近似による定式化を提案する。

参考文献

- [1] Sato, T., Takano, Y., Miyashiro, R., & Yoshise, A. (2016). Feature subset selection for logistic regression via mixed integer optimization. *Computational Optimization and Applications*, **64**(3), 865–880.
- [2] Sato, T., Takano, Y., & Miyashiro, R. (2017). Piecewise-linear approximation for feature subset selection in a sequential logit model. *Journal of the Operations Research Society of Japan*, **60**(1), 1–14.
- [3] Naganuma, M., Takano, Y., & Miyashiro, R. (投稿中). Feature subset selection for ordered logit model via tangent-plane-based approximation.