変数選択法の性能比較：マーケティングデータを例に

川崎　能典　**モデリング研究系　教授**

概要: 預金マーケティングデータを用い，契約の成否を判別・予測するモデルを構成する際の変数選択に関して，様々な手法を比較検討する．用いたデータはMoro et al. (2014) *Decision Support Systems*で分析されたものであり，UC Irvine Machine Learning RepositoryにBank Marketing Data Setとして公開されている．[本報告は植木優夫氏(現・長崎大学情報データ科学学部，執筆当時は久留米大学バイオ統計センター)との共同研究である．より詳細な報告は Kawasaki and Ueki (2015)を参照されたい．]

1. 預金マーケティングデータ

記録された顧客接触記録は41,188件，応答変数は定期預金を契約してくれたか($y = 1$)否か($y = 0$)の2値．説明変数は，連続・カテゴリカル合わせて19個用意されている．カテゴリカル変数としては，職種(job)，配偶者の有無(marital)，学歴(education)，事故歴(default)，住宅ローンの有無(housing)，その他借入(loan)，接触方法(contact)，最終接触月(month)，最終接触曜日(day_of_week)，以前のキャンペーン時の結果(poutcome)がある．連続型変数としては，年齢(age)，今回キャンペーンでの接触回数(campaign)，前回キャンペーン最終接触日からの経過日数(pdays)，今回キャンペーン以前の接触回数(previous)といった顧客属性と，雇用変動率(emp.var.rate)，消費者物価指数(cons.price.idx)，消費者信頼指数(cons.conf.idx)，欧州3ヶ月指標金利(euribor3m)，雇用者数(nr.employed)といったマクロ経済指標が含まれる．

カテゴリカル変数に関しては，当該変数がm個のカテゴリーを有するときは，そこから$m-1$個のダミー変数を生成する．例えば貸出事故履歴(default)では，yes, no, unknownの3つのカテゴリーがあるが，default-yesとdefault-unknownに対してそれぞれダミー変数を用意する．このようにして今回の問題では，最終的に連続・カテゴリー合わせて52個の説明変数を用意した．

2. 円滑閾値型推定方程式

いまd個の説明変数の未知係数を$\underline{\theta} = (\theta_1, \dots, \theta_d)'$と書くことにし，$L(\underline{\theta}) + \sum_{j=1}^{d} w_j \theta_j^2/2$という形の罰則付き損失関数を考えよう．これはすなわち，$\frac{\partial L(\underline{\theta})}{\partial \theta_j} + w_j \theta_j = 0\ (j = 1, \dots, d)$に等しい．

ここで$\delta_j \in [0, 1]$を使って$w_j = \delta_j/(1-\delta_j)$とパラメータを取り直すことができる．このとき，上掲の推定方程式は$(1-\delta_j)\partial L(\underline{\theta})/\partial\theta_j + \delta_j \theta_j = 0$ $(j = 1, \dots, d)$という形になる．$\delta_j = 1$(このとき$w_j = \infty$)ならば$\theta_j = 0$すなわちスパース解となることに注意．この方法を，円滑閾値型推定方程式(Smooth-Threshold Estimating Equation, 略してSTEE)と呼ぶ．

δ_jの決め方は，adaptive lassoにおける議論の類推から，$\sqrt{n}$一致性を持つ何らかの初期推定量$\hat{\theta}_j^{\mathrm{ini}}$を使って，$\hat{\delta}_j = \min\left(1, \lambda/|\hat{\theta}_j^{\mathrm{ini}}|^{1+\gamma}\right)$とする．チューニング・パラメータ$(\lambda, \gamma)$はBIC型規準で選択する．

STEEでは変数選択と同時に変数のグルーピングも考慮できる．更に$\sum_{j=1}^{d}\sum_{k>j}^{d} \hat{w}_{jk}(\theta_j - \theta_k)^2/2$という罰則項を加えればよい．罰則の重み$\hat{w}_{jk}$は，$\hat{\delta}_{jk} = \min\left(1, \lambda/|\hat{\theta}_j^{\mathrm{ini}} - \hat{\theta}_k^{\mathrm{ini}}|^{1+\gamma}\right)$により$\hat{w}_{jk} = \hat{\delta}_{jk}/(1-\hat{\delta}_{jk})$と決めればよい．

STEEの一致性，漸近正規性，神託性，チューニング・パラメータ選択に関するBIC型規準の一致性等，理論的性質についてはUeki (2009), Ueki and Kawasaki (2011)を参照．

3. 模擬予測実験

41,188ケースはランダムに2 : 1に訓練データと検証用データに分割される．このデータ分割は実験中独立に10回繰り返される．予測力はAUCで比較する．この時，比較される予測モデル構成法(変数選択法)は以下の10類型14通りのやり方である．

1. 多重ロジスティック回帰 (MLR): 全52変数を投入
2. L_2-罰則項付き多重ロジスティック回帰 (L2): Rのパッケージglmnetを使用
3. 自動変数選択型STEE (STEE.AVS): MLRの推定結果を初期値として利用
4. 自動変数選択型STEE (STEE.AVS-L2): L2の推定結果を初期値として利用
5. 自動変数グルーピング型STEE (STEE.AG): 変数選択に加えて変数のグルーピングも実施，初期値はMLRの結果を利用
6. 自動変数グルーピング型STEE (STEE.AG-L2): 変数選択に加えて変数のグルーピングも実施，初期値はL2の結果を利用
7. Lasso: パッケージglmnetを利用．チューニング・パラメータを10分割交差検証法で選ぶ場合 (LASSO-CV10) とBIC型規準を用いる場合 (LASSO-BIC)とを考察
8. Elastic net: パッケージglmnetを利用．チューニング・パラメータを10分割交差検証法で選ぶ場合 (ENET-CV10) とBIC型規準を用いる場合 (ENET-BIC)とを考察．L_1罰則とL_2罰則との混合率は0.1刻みで$\{0, 0.1, 0.2, \dots, 1\}$から選択.
9. Smoothly-clipped absolute deviation (SCAD): Rのパッケージncvregを利用．チューニング・パラメータを10分割交差検証法で選ぶ場合 (SCAD-CV10) とBIC型規準を用いる場合 (SCAD-BIC)とを考察.
10. Minimum concave penalty (MCP): Rのパッケージncvregを利用．チューニング・パラメータを10分割交差検証法で選ぶ場合 (MCP-CV10) とBIC型規準を用いる場合 (MCP-BIC)とを考察.

4. 実験結果

ランダムなデータ分割を10回繰り返したときの，AUCの平均と標準偏差が以下の表に示す．驚くほど結果に差がない．UCIリポジトリにデータを提供する際，原著者たちは入念に変数選択を行って提供したと思われる．精度の改善には交互作用項の導入等，何らかの追加的試みが必要だろう．

方法名	AUC平均	AUC-SD	方法名	AUC平均	AUC-SD
MLR	79.3	0.7	LASSO-BIC	79.1	0.8
L2	79.3	0.7	ENET-CV10	79.4	0.7
STEE.AVS	79.3	0.7	ENET-BIC	79.1	0.8
STEE.AVS-L2	79.3	0.6	SCAD-CV10	79.3	0.7
STEE.AG	79.2	0.8	SCAD-BIC	79.1	0.7
STEE.AG-L2	79.2	0.7	MCP-CV10	79.3	0.7
LASSO-CV10	79.4	0.7	MCP-BIC	79.3	0.7

予測力に大きな差は出なかったが，STEE.AGとSTEE.AG-L2では変数のグルーピングを考えることができ，効果が同じ大きさの変数群ではパラメータを節約することができる．以下に10回反復の実験過程で一定頻度(6回以上)観察されたグルーピングの結果を報告しておく．

頻度	変数1	変数2
7	job-services	job-blue.collar
7	education-basic.9y	jon-entrepreneur
8	euribor3m	job-student
7	day_of_week-fri	education-basic.4y
6	education-high.school	education-basic.4y
6	previous	day_of_week-fri

参考文献

[1] Ueki, M. (2009), A note on automatic variable selection using smooth-threshold estimating equations, *Biometrika*, 96, 1005–1011.

[2] Ueki, M. and Kawasaki, Y. (2011) Automatic grouping using smooth-thresholding estimating equations, *Electronic Journal of Statistics*, Vol. 5, 309-328.

[3] Kawasaki, Y. and Ueki, M. (2015) Sparse predictive modeling for bank telemarketing success using smooth-threshold estimating equations, *J. Jpn. Soc. Comput. Statist.*, 28, 53–66.