Sparse Modeling with Uncorrelated Variables

Masaaki Takada[1,2], Taiji Suzuki[3,4], Hironori Fujisawa[1,5]

[1]The Graduate University for Advanced Studies, [2]Toshiba Corporation,
[3]The University of Tokyo, [4]Center for Advanced Intelligence Project, RIKEN,
[5]The Institute of Statistical Mathematics

1 Introduction

Sparse regularization such as ℓ_1 regularization is a quite powerful and widely used strategy for high dimensional learning problems. However, one of the biggest issues in sparse regularization is that its performance is quite sensitive to correlations between features. Lasso can select variables correlated with each other since Lasso has estimate bias, so it results in deterioration of not only its generalization error but also interpretability. We propose a new regularization method, "Independently Interpretable Lasso" (IILasso).

2 Proposed Method

Consider the problem of predicting $y \in \mathbb{R}^n$, given a design matrix $X \in \mathbb{R}^{n\times p}$, assuming a linear model $y = X\beta + \epsilon$, where $\epsilon \in \mathbb{R}^n$ is a noise and $\beta \in \mathbb{R}^p$ is a regression coefficient. We assume without loss of generality that the features are standardized such that $\Sigma_{i=1}^n X_{ij} = 0$, $\Sigma_{i=1}^n X_{ij}^2/n = 1$ and $\Sigma_{i=1}^n y_i = 0$. We propose a new regularization formulation as follows:

$$\min_{\beta} \frac{1}{2n}\|y - X\beta\|_2^2 + \lambda\left(\|\beta\|_1 + \frac{\alpha}{2}|\beta|^\top R|\beta|\right), \tag{1}$$

where $\alpha > 0$ is a regularization parameter for the new regularization term, and $R \in \mathbb{R}^{p\times p}$ is a symmetric matrix whose component $R_{jk} \geq 0$ represents the similarity between X_j and X_k. The last term of (1) is also written as $\frac{\lambda\alpha}{2}\sum_{j=1}^p\sum_{k=1}^p R_{jk}|\beta_j||\beta_k|$. We define R_{jk} for $j \neq k$ as a monotonically increasing function of the absolute correlation $r_{jk} = \frac{1}{n}|X_j^\top X_k|$, so that correlated variables are hard to be selected simultaneously. Our proposed regularizer suppresses selecting correlated variables, and thus each active variable independently affects the objective variable in the model. Hence, we can interpret regression coefficients intuitively and also improve the performance by avoiding overfitting.

3 Theoretical and Numerical Results

We analyze theoretical property of IILasso and show that the proposed method is much advantageous for its sign recovery and achieves almost minimax optimal convergence rate. Synthetic and real data analyses also indicate the effectiveness of IILasso. We provide implementations of IILasso through the publicly available `R` package, `iilasso`.