Tuning parameter selection in LASSO regression

Gianluca Sottile & Vito M. R. Muggeo

Università degli Studi di Palermo
Dipartimento Scienze Economiche, Aziendali e Statistiche

gianluca.sottile@unipa.it

Abstract

We propose a new method to select the tuning parameter in lasso regression. Unlike the previous proposals, the method is iterative and thus it is particularly efficient when multiple tuning parameters have to be selected. The method also applies to more general regression frameworks, such as generalized linear models with non-normal responses. Simulation studies show our proposal performs well, and most of times, better when compared with the traditional Bayesian Information Criterion and Cross validation.

Introduction

In the context of high-dimensional data, typically only a small number of variables are truly informative whilst other are redundant. Selecting the appropriate variables is a crucial step of data analysis process. An undertted model excluding truly informative variables may lead to severe estimation bias in model fit, whereas an overtted model including redundant uninformative variables, increases the estimated variance and hinders model interpretation.

Among dierent variable selection methods discussed in literature, penalized regression models have gained popularity and attractiveness. Selection of variables is controlled by the tuning parameter which encourages model sparsity. Well known procedures include the Least Absolute Shrinkage and Selection Operator (LASSO, Tibshirani, 1996), the Smoothy Clipped Abso- lute Deviation (SCAD, Fan and Li, 2001), the Adaptive LASSO (ALASSO, Zou, 2006), and the Elastic Net (Zou and Hastie, 2005). In penalized regression, the tuning parameters balance the trade-o between model fit and model sparsity, and selecting an appropriate value is the key point.

In literature, traditional criteria to select the tuning parameter include Cross-Validation (CV), Generalized Cross Validation (GCV), Mallows Cp, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) or its extension.

Main Objectives

1. To find a goodness of fit which could correctly select the true active set of variables in the classical Gaussian model framework.
2. To build an iterative algorithm to use the 'new' goodness of fit to select efficiently the tuning parameter λ.
3. To compare the 'new' goodness of fit with the trditional criteria CV and (E)BIC in terms of Mean Squared Error and Degree of freedom.

Materials and Methods

In LASSO regression with sample size n and p covariates, the objective is to find a solution of the following optimization problem:

$$\min_{\beta} \sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p}|\beta_j|. \quad (1)$$

As discussed in the Introduction, typically one fixes the tuning parameter λ that balances the trade-off between sparsity and fitting, and then minimizes objective (1) by means of any of optimization algorithms recently developed, e.g. gradient descent or lars.

To set up an iterative algorithm to find λ, we borrow the Schall algorithm idea successfully employed to estimate the variance components in random effects models. More specifically, starting from an initial guess, $\lambda^{(0)} = .001$, say, the algorithm alternates estimation of lasso regression (at fixed λ) and updating of the tuning parameter via the variance ratio properly modified to account for the L_1 penalty:

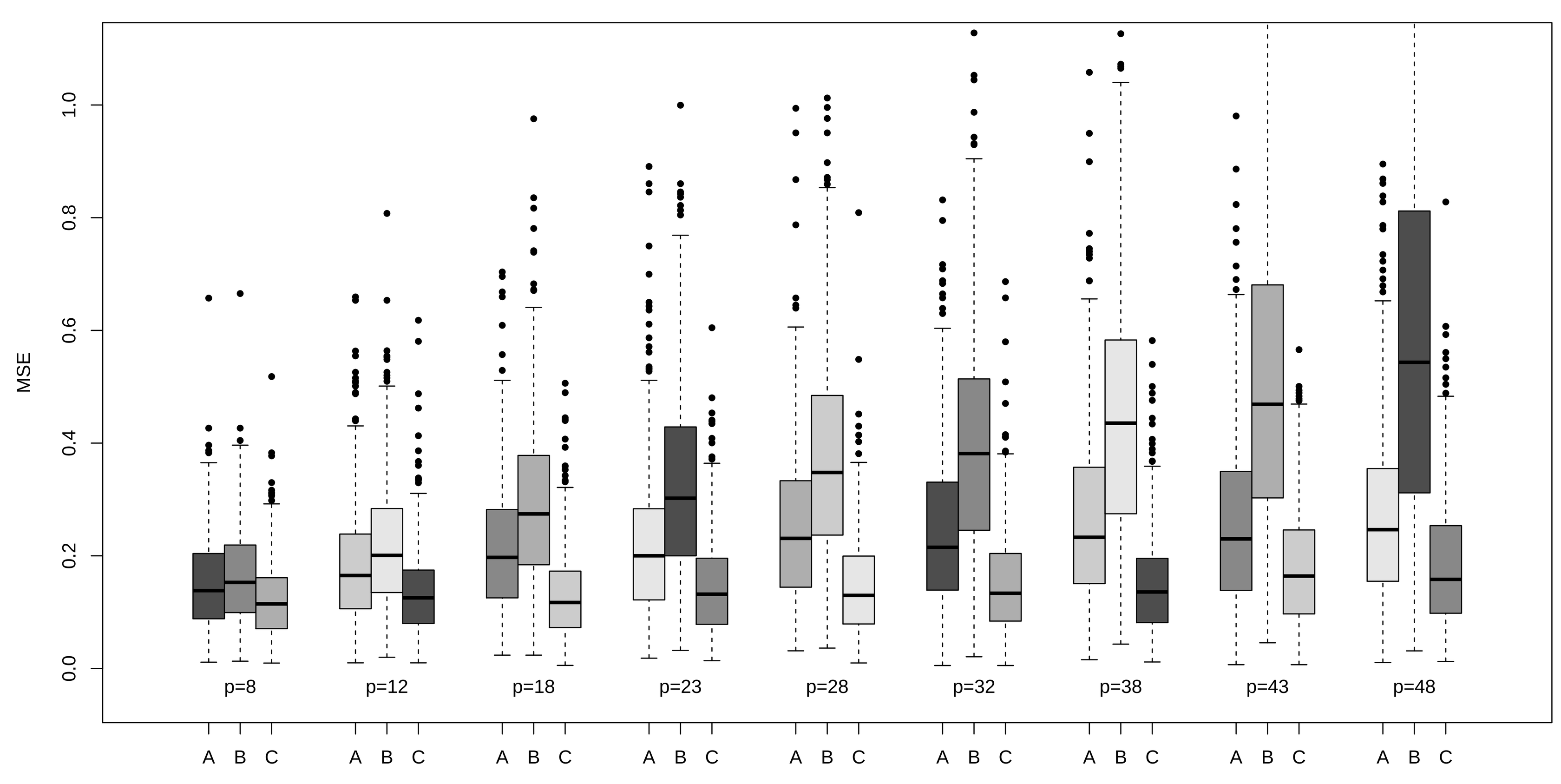
$$\frac{||y - X\beta_\lambda||_2^2/(n-df)}{K * ||\beta||_1/df}. \quad (2)$$

Simulation Studies

Some simulations have been undertaken to compare the traditional selection criteria, BIC, GCV, CV with respect to the proposed algorithm. To assess the performance of each selection criterion, we report degrees of freedom (df, the number of non null coefficients), and the mean squared error (MSE) corresponding to OLS fits including only the informative covariates selected.

The simulated data come from $y = X\beta + \epsilon$, where $X \sim N_p(0_p, \Sigma_p)$, ($\Sigma_{jk} = 0.5^{|j-k|}$) and $\epsilon \sim N(0, 1)$. For two sample sizes (50, 200), two different scenarios have been considered: in the first scenario ($n > p$), $p \in \{.15n, .25n, \dots, .95n\}$ and true coefficients $\beta = (5, 4, 3, 2, 1, 0, \dots, 0)^T$. In the second scenario, $n < p$, $p \in \{1n, 2n, \dots, 64n\}$ and β as in the first scenario.

Results



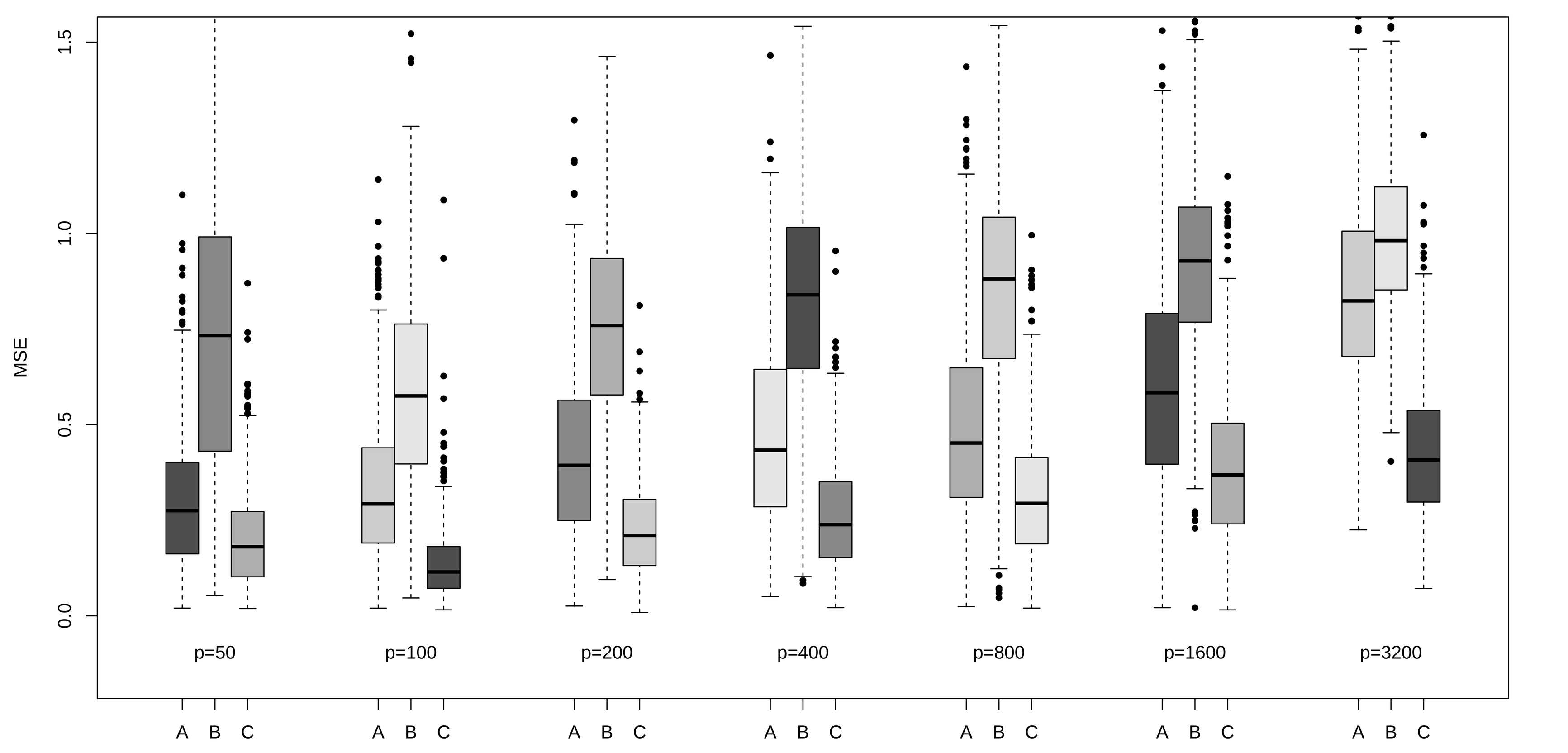


Figure 1: Performance of Tuning Parameter Selector criteria in LASSO Regression ((E)BIC=A, CV=B and the proposed algorithm, C): MSE for $n = 50$ sample size and several p/n ratios. Results based on 500 simulation runs.

	$n = 50$			$n = 200$		
p/n	(E)BIC	CV	New	(E)BIC	CV	New
			$n > p$			
0.15	5.7	6.4	5.1	6.3	10.8	5.0
0.25	6.2	7.6	5.2	6.1	11.6	5.0
0.35	6.5	8.8	5.1	6.4	13.1	5.0
0.45	6.6	9.2	5.4	6.2	13.5	5.0
0.55	7.2	10.5	5.5	6.4	15.0	5.0
0.65	6.6	10.4	5.3	6.0	14.4	5.0
0.75	7.0	11.0	5.3	6.3	15.1	5.0
0.85	6.7	10.9	5.7	6.1	16.0	5.0
0.95	7.1	11.1	5.9	6.2	15.5	5.0
			$n \le p$			
1	7.1	11.1	5.8	6.4	16.0	5.0
2	7.5	13.7	5.1	6.3	17.6	5.0
4	8.6	17.0	5.9	6.2	20.0	5.0
8	9.2	18.2	6.4	6.3	23.9	5.0
16	9.0	19.3	6.6	6.4	27.9	5.0
32	11.0	18.1	7.3	6.6	32.8	5.0
64	20.2	19.2	9.6	6.5	31.1	5.0

Table 1: Performance of Tuning Parameter Selector criteria in LASSO Regression ((E)BIC, CV and the proposed algorithm, New). Average degrees of freedom for $n = 50, 200$ sample sizes and several p/n ratios. Results based on 500 simulation runs.

Table 1 report average degrees of freedom (df), the number of correctly selected parameters are for all criteria $5/5$. Figure 1 report average mean squared errors (MSE). The scenario $n = 200$ has been omitted because the results are similar to $n = 50$ sample size as shown in Table 1.

Results show that the proposed method performs better than the others in all the scenarios, not only in terms of model fit but also in terms of degrees of freedom. Also, The proposed iterative algorithm always exhibits the lowest MSE, but when $n < p$, particularly with small samples ($n = 50$), the new methods performs largely better than the other competitors.

Conclusions

- A 'new' approach/criteria to select iteratively the tuning parameter λ of lasso regression has been proposed.
- The 'new' method attains comparatively better performance in all considered settings.

Forthcoming Research

Results have been presented for the classical Gaussian model, but the proposed approach is favored to be employed in generalized linear models with binary or count responses. Application in very high-dimensional settings ($n \ll p$) that are today one of the most challenging concerns, represents a noteworthy point to be investigated.

References

[1] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

[2] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[3] Hansheng Wang, Runze Li, and Chih-Ling Tsai. Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, 94(3):553–568, 2007.

Acknowledgements

...