Selection of tuning parameter in sparse linear regression

Alkomiet Belal
Institute of Information Theory and Automation,
Prague, Czech Republic

Supervisor: Doc. Ing. Václav Šmídl, Ph.D

Researchers have put a decent effort into investigation of penalized regression methods for simultaneous variable selection and coefficients estimation. The aim of variable selection is to identify the best subset among many variables to include in a model in order to improve its prediction performance, provide faster and more cost-effective predictors, and obtain better understanding of the underlying process that generated the data.

The LASSO (Least Absolute Shrinkage and Selection Operator) introduced by Tibshirani in 1996 [4], is a regression method that involves penalizing the absolute size of the regression coefficients by constraining the sum of the absolute values of the estimates. The given maximum for the sum of the estimates is a tuning parameter. The problem can be transformed to optimization of a L2 norm with additional L1 regularization term on the unknown. The tuning parameter is then the multiplier of the L1 norm. Since the tuning parameter affects the coefficient estimation and variable selection, we want to find the optimal value for the tuning parameter to get the most accurate coefficient estimation and best subset of predictors. There are many methods available to select the optimal value of the tuning parameter that results in a the best subset of predictors, we have taested: cross validation (CV) [4], L-curve criterion [1], and Variable Selection stability method [3].

The objective of this work is to evaluate and compare these three methods for selecting the optimal value of tuning parameter in terms of coefficient estimation accuracy and correct variable selection through simulation studies. The practical usefulness of our contribution is demonstrated on an application to real data from the European Tracer Experiment (ETEX) [2].

References

[1] Per Christian Hansen. *The L-curve and its use in the numerical treatment of inverse problems*. IMM, Department of Mathematical Modelling, Technical Universityof Denmark, 1999.

[2] K Nodop, R Connolly, and F Girardi. The field campaigns of the european tracer experiment (etex): Overview and results. *Atmospheric Environment*, 32(24):4095–4108, 1998.

[3] Wei Sun, Junhui Wang, and Yixin Fang. Consistent selection of tuning parameters via variable selection stability. *Journal of Machine Learning Research*, 14(1):3419–3440, 2013.

[4] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.