ROBUST FITTING PENALIZED REGRESSION SPLINE MODELS

Shchetinin Eu.Yu., Sariev I.K.

Moscow State Technology University "STANKIN",
Applied Math
Russia, 127994, Moscow, Vadkovsky per., 1, riviera-molto@mail.ru

Penalized regression splines are one of the currently most used methods for smoothing noisy data. The estimation method used for fitting such a penalized regression spline model is mostly based on least squares methods, which are known to be sensitive to outlying observations. In real world applications, outliers are quite commonly observed. There are several robust estimation methods taking outlying observations into account. We define and study S-estimators for penalized regression spline models. Hereby we replace the least squares estimation method for penalized regression splines by a suitable S-estimation method. By keeping the modeling by means of splines and by keeping the penalty term, though using S-estimators instead of least squares estimators, we arrive at an estimation method that is both robust and flexible enough to capture non-linear trends in the data. Simulated data and a real data example are used to illustrate the effectiveness of the procedure.

The main purpose of this paper is to propose robust penalized regression splines that are able to resist the potentially damaging effect of outliers in the sample, and that do not require the separate estimation of the residual scale. To achieve these goals we propose to compute penalized S-regression estimators. In the unpenalized case, these estimators are consistent, asymptotically normal, and have high-breakdown point regardless of the dimension of the vector of regression coeficients. First we show that the solution to the penalized S-regression problem can be written as the solution of a weighted penalized least squares problem. This representation naturally leads to an iterative algorithm to compute these estimators.

We also study how to robustly select the penalty parameter when there may be outliers in the data. We propose a robust penalty parameter selection criteria based on generalized cross-validation that also borrows from the weighted penalized least squares representation of the penalized S-regression estimator. Extensive simulation studies show that our algorithm works well in practice and that the resulting regression function estimator is robust to the presence of outliers in the data.

Литература

1. *Maronna, R. A., M. R. D. and Yohai, V. J.* (2006). Robust Statistics Theory and Methods. John Wiley and Sons,Ltd.
2. *Hall, P. and Opsomer, J.* (2005). Theory for penalized spline regression. *Biometrika*, 92:105-118.
3. *Cantoni, E. and Ronchetti, E.* (2001). Resistant selection of the smoothing parameter for smoothing splines. *Stat. Comput.*, 11(2):141-146.