ADDITIVE MODELS FOR NON-PARAMETRIC REGRESSION: ADVENTURES IN SPARSE LINEAR ALGEBRA

ROGER KOENKER

Abstract. The development of efficient algorithms for sparse linear algebra has significantly expanded the frontiers of statistical computation. This is particularly true of non-parametric regression where penalty methods for additive models require efficient solution of large, sparse least-squares problems. Sparse methods for estimating additive non-parametric models subject to total variation roughness penalties will be described. The methods are embodied in the R packages `SparseM` and `nprq`.

Models are structured similarly to the `gss` package of Gu and the `mgcv` package of Wood. Formulae like

$$y \sim qss(z_1) + qss(z_2) + X$$

are interpreted as a partially linear model in the covariates of X, with nonparametric components defined as functions of z_1 and z_2. When z_1 is univariate fitting is based on the total variation penalty methods described in Koenker, Ng and Portnoy (1994). When z_2 is bivariate fitting is based on the total variation penalty (triogram) methods described in Koenker and Mizera (2003). There are options to constrain the qss components to be monotone and/or convex/concave for univariate components, and to be convex/concave for bivariate components. Fitting is done by new sparse implementations of the dense interior point (Frisch-Newton) algorithms already available in the R package `quantreg`.

Koenker, R., and I. Mizera, (2004). "Penalized triograms: Total variation regularization for bivariate smoothing", *J. Royal Stat. Soc. (B)*, **66**, 145-163.

Koenker, R., P. Ng, and S. Portnoy (1994). "Quantile Smoothing Splines", *Biometrika*, 81, 673-680.