The Annals of Applied Statistics
2013, Vol. 7, No. 2, 1244–1246
DOI: 10.1214/13-AOAS640

LETTER TO THE EDITOR

BY YUAO HU, YE TIAN AND HENG LIAN

Nanyang Technological University

The paper by Alfons, Croux and Gelper (2013), *Sparse least trimmed squares regression for analyzing high-dimensional large data sets*, considered a combination of least trimmed squares (LTS) and lasso penalty for robust and sparse high-dimensional regression. In a recent paper [She and Owen (2011)], a method for outlier detection based on a sparsity penalty on the mean shift parameter was proposed (designated by "SO" in the following). This work is mentioned in Alfons et al. as being an "entirely different approach." Certainly the problem studied by Alfons et al. is novel and interesting. However, there is actually a connection between the LTS approach and that of She and Owen (2011). This connection can be roughly seen from Theorem 4.1 and Proposition 4.1 of She and Owen (2011), where iterative thresholding was related to penalized regression and also to the M-estimator. In particular, although not explicitly mentioned in She and Owen (2011), from this one can derive the close relationship between hard thresholding, L_0 penalty and LTS [the relationship between hard thresholding and L_0 penalty was mentioned on page 630 of She and Owen (2011)]. Given that LTS regression is not directly posed as an M-estimator, the following proposition can be directly shown via elementary arguments.

PROPOSITION 1. *Using the notation of Alfons et al., if* $(\hat{\boldsymbol\beta}, \hat{\boldsymbol\gamma})$ *is a minimizer of* $\sum_{i=1}^n (y_i - \mathbf{x}_i'\boldsymbol\beta - \gamma_i)^2 + \lambda_1\|\boldsymbol\beta\|_1 + \lambda_2\|\boldsymbol\gamma\|_0$ *and* $\|\boldsymbol\gamma\|_0 = n - h$, *then* $\hat{\boldsymbol\beta}$ *is the minimizer of* $\sum_{i=1}^h (\mathbf{r}^2(\boldsymbol\beta))_{i:n} + \lambda_1\|\boldsymbol\beta\|_1$.

PROOF. Obviously we have $\hat\gamma_i = y_i - \mathbf{x}_i'\hat{\boldsymbol\beta}$ if $(y_i - \mathbf{x}_i'\hat{\boldsymbol\beta})^2 > \lambda_2$ and $\hat\gamma_i = 0$ if $(y_i - \mathbf{x}_i'\hat{\boldsymbol\beta})^2 < \lambda_2$. Thus, we can profile out $\boldsymbol\gamma$ and get exactly the LTS problem. □

The result above says that a solution of SO is a solution of some LTS problem and, thus, the set of solutions that can be obtained by SO (by varying λ_1 and λ_2) is a subset that can be obtained by LTS (by varying λ_1 and h). Obviously, if for any fixed λ_1 and $h \geq n/2$, we can make $\|\boldsymbol\gamma\|_0 = n - h$ by choosing an appropriate value for λ_2, then the two will be the same. Numerically, we do find occasionally some values of $n - h$ cannot be obtained by $\|\boldsymbol\gamma\|_0$. In the numerical example below with sample size $n = 59$, $h = 45$ (25% trimmed) can be achieved in both cases.

Received January 2013; revised February 2013.