The Annals of Statistics
2007, Vol. 35, No. 6, 2373–2384
DOI: 10.1214/009053607000000460
Main article DOI: 10.1214/009053606000001523

DISCUSSION: A TALE OF THREE COUSINS: LASSO, L2BOOSTING AND DANTZIG

BY N. MEINSHAUSEN,[1] G. ROCHA[2] AND B. YU[3]

University of California, Berkeley

We would like to congratulate the authors for their thought-provoking and interesting paper. The Dantzig paper is on the timely topic of high-dimensional data modeling that has been the center of much research lately and where many exciting results have been obtained. It also falls in the very hot area at the interface of statistics and optimization: ℓ_1-constrained minimization in linear models for computationally efficient model selection, or sparse model estimation (Chen, Donoho and Saunders [5] and Tibshirani [17]). The sparsity consideration indicates a trend in high-dimensional data modeling advancing from prediction, the hallmark of machine learning, to sparsity—a proxy for interpretability. This trend has been greatly fueled by the participation of statisticians in machine learning research. In particular, Lasso (Tibshirani [17]) is the focus of many sparsity studies in terms of both theoretical analysis (Knight and Fu [10], Greenshtein and Ritov [9], van de Geer [19], Bunea, Tsybakov and Wegkamp [3], Meinshausen and Bühlmann [13], Zhao and Yu [23] and Wainwright [20]) and fast algorithm development (Osborne, Presnell and Turlach [15] and Efron et al. [8]).

Given n units of data $Z_i = (X_i, Y_i)$ with $Y_i \in \mathbb{R}$ and $X_i^T \in \mathbb{R}^p$ for $i = 1, \dots, n$, let $Y = (Y_1, \dots, Y_n)^T \in \mathbb{R}^n$ be the continuous vector response variable and $X = (X_1, \dots, X_n)^T$ the $n \times p$ design matrix and let the columns of X be normalized to have ℓ_2-norm 1. It is often useful to assume a linear regression model,

$$Y = X\beta + \varepsilon, \tag{1}$$

where ε is an i.i.d. $N(0, \sigma^2)$ vector of size n.

Lasso minimizes the ℓ_1-norm of the parameters subject to a constraint on squared error loss. That is, $\beta^{\text{lasso}}(t)$ solves the ℓ_1-constrained minimization problem

$$\min_{\beta} \|\beta\|_1 \quad \text{subject to} \quad \tfrac{1}{2}\|Y - X\beta\|_2^2 \le t. \tag{2}$$

Received February 2007.
[1] Supported in part by a scholarship from DFG (Deutsche Forschungsgemeinschaft).
[2] Supported in part by ARO Grant W911NF-05-1-0104.
[3] Supported in part by a Guggenheim Fellowship, NSF Grants DMS-06-05165 and DMS-03-036508, and ARO Grant W911NF-05-1-0104.