Marten H. Wegkamp

Sparsity oracle inequalities via ℓ_1 regularization in nonparametric models

Complex models need to be postulated and new computational and theoretical tools need to be developed for the analysis of high dimensional data sets with relatively small sample sizes. When the data is scarce, target recovery is possible if there exists a sparse approximating submodel of small dimension relative to the sample size regardless of the dimension of the initial approximating model. We will (a) define sparsity and target recovery in general high dimensional, small sample size settings and (b) show that empirical risk minimization with an ℓ_1 regularization allows for target recovery, under minimal assumptions. We will focus on three important functional estimation problems: density estimation, regression and classification. They can be placed in the following general framework. The parameter of interest f_0 is the minimum taken over all measurable functions f of some risk function $R(f)$ of interest. The estimators of f_0 studied here will be linear combinations of functions from a given dictionary $\mathcal{F}_M = \{f_1, \ldots, f_M\}$, where M may depend on n, the size of the sample. The main purpose of our proposal is to define an oracle target, which will be an ideal approximation of f_0 with possibly many fewer elements of $\mathcal{F}_M$, and to show that the oracle target can be recovered by computationally tractable techniques, such as empirical risk minimization with an ℓ_1 regularization, even if $M > n$.

The great popularity of ℓ_1 regularized estimation is owed to its computational feasibility and dimension reduction properties. We will discuss a unified framework for the theoretical understanding of ℓ_1 regularized estimation. For each class of problems the oracle target f^* will be the linear combination of functions from $\mathcal{F}_M$ that achieves the theoretical bias-variance trade-off; the coefficients of this ideal combination will be collected in the oracle vector λ^*. Whereas in regression models the meaning of "bias" and "variance" are well agreed upon, these notions are only now beginning to crystalize in classification problems, and have not yet been studied in the less standard problem of classification with a reject option. To determining whether the oracle target can be recovered via an estimate $\widehat{f} = \sum_j \widehat{\lambda}_j f_j$, one typically establishes oracle inequalities for the excess risk $R(\widehat{f}) - R(f_0)$; this, however, gives no information about the behavior of $\widehat{\lambda}$ relative to the oracle vector λ^*. We advance the use of a novel type of oracle inequality, where the aim is to show that the sum of the excess risk and a quantity that is proportional to the ℓ_1 distance between $\widehat{\lambda}$ and λ^* achieves the optimal bias variance trade-off. Via such general inequalities we propose to show that ℓ_1 regularized estimates adapt to the unknown sparsity of f_0 relative to $\mathcal{F}_M$, for a number of measures of adaptation.