Minimax rates of estimation
for high-dimensional linear regression over ℓ_q-balls

Bin Yu
Departments of Statistics and EECS
University of California at Berkeley
www.stat.berkeley.edu/~binyu

Recently there has been much work on sparse linear regression modeling in the $d >> n$ modern theoreical paradiagm where the true model is assumed to be sparse (either hard or weak). In particular, upper bounds have been obtained by various authors for Lasso and variants in terms of ℓ_2 estimation and L_2 prediction errors.

In this talk, we present sharp minimax rate results of convergence for estimation of the parameter vector ℓ_2-norm and L_2- prediction error, among other results. We consider the standard Gaussian linear regression model with where we assume the number of parameters d is greater than the number of parameters n (ie. high-dimensional scaling). Sharp minimax rates are found for estimation of the parameter vector β in ℓ_2-norm and L_2- prediction error, assuming that the true parameter β^* belongs to a weak ℓ_q-ball, with $\|\beta^*\|_q^q \leq R_q$ for some $q \in [0, 1]$ (hard sparsity when $q = 0$). We show that under suitable regularity conditions on the design matrix X, the minimax error in squared ℓ_2-norm and prediction norm scales as $R_q(\frac{\log d}{n})^{1-\frac{q}{2}}$. In addition, we provide lower bounds on rates of convergence for general ℓ_p norm (for all $p \in [1, +\infty], p \neq q$). Our proofs of the lower bounds are information-theoretic in nature, based on Fano's inequality and results on the metric entropy of the ℓ_q-ball ($\|\beta\|_q^q \leq R_q$). Matching upper bounds were derived by direct analysis of the solution to the least-squares algorithm over the ℓ_q-ball ($\|\beta\|_q^q \leq R_q$). We prove that the conditions on X required by optimal algorithms are satisfied with high probability by broad classes of non-i.i.d. Gaussian random matrices, for which RIP or other sparse eigenvalue conditions are violated. For $q = 0$, ℓ_1-based methods (Lasso and Dantzig selector) achieve the minimax optimal rates in ℓ_2 error, but require possibly stronger regularity conditions on the design than the non-convex optimization algorithm used to determine the minimax upper bounds.

(This talk is based on joint work with Garvesh Raskutti and Martin Wainwright at UC Berkeley.)