Lasso: Blasso algorithm and a model selection consistency result

Bin Yu
Statistics Department, UC Berkeley
binyu@stat.berkeley.edu, www.stat.berkeley.edu/user/binyu

Information technology advances are making data collection possible in most if not all fields of science and engineering and beyond. Statistics as a scientific discipline is challenged and enriched by the new opportunities resulted from these high-dimensional data sets. Often data reduction or feature selection is the first step towards solving these massive data problems. However, data reduction through model selection or L_0 constrained optimization leads to combinatorial searches which are computationally expensive or infeasible for massive data problems. A computationally more efficient alternative to model selection is L_1 constrained optimization or Lasso optimization.

In this talk, we will describe the Boosted Lasso (BLasso) algorithm that is able to produce an approximation to the complete regularization path for general Lasso problems. BLasso consists of both a forward step and a backward step. The forward step is similar to Boosting and Forward Stagewise Fitting, but the backward step is new and crucial for BLasso to approximate the Lasso path in all situations. For cases with finite number of base learners, when the step size goes to zero, the BLasso path is shown to converge to the Lasso path. Experimental results are also provided to demonstrate the difference between BLasso and Boosting or Forward Stagewise Fitting. We can extend BLasso to the case of a general convex loss penalized by a general convex function and illustrate this extended BLasso with examples.

Since Lasso is used as a computationally more efficient alternative to model selection, it is important to study the model selection property of Lasso. I will also present some (almost) necessary and sufficient conditions for Lasso to be model selection consistent in the classical p ¡¡ n setting.

(This is joint work with Peng Zhao at UC Berkeley.)