Efficient Support Recovery via 
Weighted Maximum-Contrast Subagging

Jelena Bradic, PhD
Assistant Professor, Department of Mathematics
University of California, San Diego

Wednesday, April 16, 2014
3:30pm - 4:30pm, CHS 43-105
Refreshments served at 3:00 PM in room 51-254 CHS

ABSTRACT: In this paper, we study subagging in the context of non-smooth estimation and model selection in sparse and large-scale regression settings where both the number of parameters and the number of samples can be extremely large. This setup is very different from high-dimensional regression and is such that Lasso estimator might be inappropriate for computational, rather than statistical reasons. We show that subagging of Lasso estimators results in discontinuous estimated support set and is never able to recover sparsity set when at least one of aggregated estimators has probability of support recovery strictly less than 1. Therefore, we propose its randomized and smoothed alternative, which we call weighted maximum-contrast subagging. We develop theory in support of the claim that proposed method has tight error control over both false positives and false negatives, regardless of the size of a dataset. Unlike existing methods, it allows for oracle-like properties, even in cases of non-oracle-like properties of aggregated estimators. Furthermore, we design an adaptive procedure for selecting tuning parameters and appropriate optimal weighting scheme. Finally, we validate our theoretical findings through extensive simulation study and analysis of a part of the million-song-challenge dataset.