# High confidence set regularization in sparse high dimensional logistic regression with measurement error

Maorong Rao

Preprint no. 2018-08

**Abstract**

$l_1$ based regularizations such as Lasso and Dantzig selectors succeed in two aspects: the inherent sparsity of $l_1$ accords with the underlying nature of high dimensional data, and the convexity essence paving the way to computational feasibility in high dimension. G. M. James and P. Radchenko extended an algorithm to solve Dantzig Selector for generalized linear models. J. Fan abstracted this framework to the set of convex loss function as High Confidence Set. To fill the gap of theoretical support within this framework, we derive the bound of prediction and parameter errors beyond the scope of logistic loss. We term this classifier as High Confidence Set (HCS) Selector.

An implicit assumption of HCS selection is that the data is collected precisely. However, the data is inevitable to process with measurement error in reality. In response to this challenge, we introduce a new methodology termed High Confidence Set Selector with Measurement Error (MHCS) that can account for measurement error. We further derive the theory and algorithm. Our simulation study provides strong numerical support that is comparable with other popular regularization methods. And due to embedded linearity instinct, HCS and MHCS are versatile to connect with the state-of-art techniques such as word vectors, deep network, transfer learning, etc.

Department of Mathematics, UNC Charlotte, Charlotte, NC 28223