USC Department of Mathematics

Probability & Statistics Seminar

3:30 PM, Friday 24.Apr.09
146 Kaprielian Hall
(Tea beforehand in 116)

Jinchi Lv
Department of Information & Operations Management
USC Marshall School of Business

Sure independence screening for ultra-high dimensional feature space

Variable selection plays an important role in high dimensional statistical modeling which nowadays appears in many areas and is key to various scientific discoveries. For problems of large scale or dimensionality p, estimation accuracy and computational cost are two top concerns. In a recent paper, Candes and Tao (2007) propose the Dantzig selector using L_1 regularization and show that it achieves the ideal risk up to a logarithmic factor $\log p$. Their innovative procedure and remarkable result are challenged when the dimensionality is ultra high as the factor $\log p$ can be large and their uniform uncertainty principle can fail.

Motivated by these concerns, we introduce the concept of sure screening and propose a sure screening method based on a correlation learning, called the Sure Independence Screening (SIS), to reduce dimensionality from high to a moderate scale that is below sample size. In a fairly general asymptotic framework, the correlation learning is shown to have the sure screening property for even exponentially growing dimensionality. As a methodological extension, an iterative SIS (ISIS) is also proposed to enhance its finite sample performance. With dimension reduced accurately from high to below sample size, variable selection can be improved on both speed and accuracy, and can then be accomplished by a well-developed method such as the SCAD, Dantzig selector, Lasso, or adaptive Lasso. The connections of these penalized least-squares methods are also elucidated.