ATELIER INTERNATIONAL « PERSPECTIVES SUR L'ANALYSE DES DONNÉES EN DIMENSIONS SUPÉRIEURES II »

30 MAI – 1 JUIN, 2012

INTERNATIONAL WORKSHOP ON THE PERSPECTIVES ON HIGH-DIMENSIONAL DATA ANALYSIS II
MAY 30 – JUNE 1, 2012

Tuning parameter selection in high dimensional penalized likelihood

Yingying Fan* fanyingy@marshall.usc.edu

Cheng Yong Tang† statc@nus.edu.sg

How to appropriately select the tuning parameter is essential in regularized likelihood methods for high dimensional data analysis. For this purpose, we first examine the generalized information criterion (GIC) for generalized linear models when the dimensionality of covariates p increases exponentially with the sample size n. To ensure consistently identifying the true model, we find that the penalty on the model complexity in GIC should diverge at the rate of some power of $\log p$ depending on the tail probability behavior of the response variables. This reveals that using the AIC or BIC to select the tuning parameter may not be adequate to consistently identify the true model. We then propose to select the tuning parameter in the regularized likelihood methods by optimizing the GIC with appropriate model complexity penalty. We show that the proposed approach for tuning parameter selection consistently identifies the true model from the candidate models with probability tending to 1. We confirm the performance of the proposed procedure by numerical simulations and a gene expression data analysis.

*Information and Operations Management Department, University of Southern California, Marshall School of Business, Los Angeles, CA 90089 USA.

†Department of Statistics and Applied Probability, National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077, SINGAPORE.