The Hong Kong
Polytechnic University
香港理工大學

DEPARTMENT OF APPLIED MATHEMATICS
應 用 數 學 系

Academy of Mathematics
and Systems Science

The AMSS-PolyU Joint Research Institute

Distinguished Lecture on

Impact of Dimensionality and Correlation Learning from Ultra-high Dimensional Data

by Professor Jianqing Fan, Princeton University

ABSTRACT

Model selection and classification using high-dimensional features arise frequently in many contemporary statistical studies such as tumor classification using microarray or other high-throughput data. The impact of dimensionality on classifications is largely poorly understood. We first demonstrate that even for the independence classification rule, classification using all the features can be as bad as the random guessing due to noise accumulation in estimating population centroids in high-dimensional feature space. In fact, we demonstrate further that almost all linear discriminants can perform as bad as the random guessing. Thus, it is paramountly important to select a subset of important features for high-dimensional classification, resulting in Features Annealed Independence Rules (FAIR). The connections with the sure independence screening (SIS) and iterative SIS(ISIS) of Fan and Lv (2008) in model selection will be elucidated and extended. Further extension of the correlation learning results in independence learning for feature selection in general loss functions will also be discussed. The proposed choice of the optimal number of features, or equivalently, the threshold value of the test statistics is based on an upper bound of the classification error. Simulation studies and real data analysis support our theoretical results and demonstrate convincingly the advantage of our new classification procedure.

BIOGRAPHY

Jianqing Fan is Frederick L. Moore'18 Professor of Finance and the past president of the Institute of Mathematical Statistics. He is the co-editor of Econometrical Journal and an associate editor of The Journal of American Statistical Association. Professor Fan has coauthored two highly-regarded books entitled "*Local Polynomial Modeling*" (1996) and "*Nonlinear time series: Parametric and Nonparametric Methods*" (2003), and published over 100 articles on computational biology, financial econometrics, semiparametric and non-parametric modeling, statistical learning, nonlinear time series, and other aspects of theoretical and methodological statistics. He was awarded with the 2000 COPSS President's Award, the Humboldt Research Award in 2006 and the Morningside Gold Medal of Applied Mathematics in 2007 for his outstanding achievement in statistics and applied mathematics. He was also elected as fellow of The American Association for the Advancement of Science, Institute of Mathematical Statistics and American Statistical Association.

Date: 16th December, 2008 (Tuesday)

Time: 11:00a.m. – 12:00p.m.
(Tea reception starts at 10:30a.m.)

Venue: AG710

ALL ARE WELCOME