

國家理論科學研究中心學術演講

2013 NCTS Statistics Seminar

Speaker: Prof. George C. Tseng 曾建城

(Dept. of Biostatistics, Human Genetics and Computational
Biology, University of Pittsburgh)

Time: AM 10:40-11:30, Mon. May 27, 2013

(10:20-10:40am 茶會於交大統計所 428 室舉行)

Topic: Machine Learning Bias Correction for Minimal-error Classifier and
a Meta-analysis Framework for Sparse K-means in Genomic
Applications

Abstract:

In this talk, I will cover two topics we recently developed for genomic applications. In the first part, we investigated the machine learning bias when one utilizes many classifiers and chooses the best to report. We studied the properties of the bias in relation to the sample size and classifiers used. We proposed an inverse power law method to correct the bias and compared it to conventional nested cross-validation.

Finally we used large-scale empirical gene expression data to recommend a practical guideline for practitioners. In the second part, we extended the sparse K-means algorithm to a meta-analysis framework to combine multiple gene expression profiles for improved disease subtype discovery. The result showed more stable and accurate sample clustering to identify meaningful disease subtypes.

Sponsors: 國家理論科學研究中心

交通大學統計學研究所

清華大學統計學研究所

Place:

交大綜合一館 427 室