



統計科學研究所

INSTITUTE OF
STATISTICAL SCIENCE



S.E.M.I.N.A.R.

S.T.A.T.I.S.T.

Seminar

Title : Random Integrated Subdata Ensemble (RISE) for
Big Data Model Building

Speaker : Prof. Lih-Yuan Deng (鄧利源教授)
(University of Memphis, U.S.A.)

Time : 10:30 AM~12:00 PM, Thursday, Jan 5, 2023

Place : Auditorium, B1F, Institute of Statistical Science

Abstract

We discuss our newly proposed Random Integrated Subdata Ensemble (RISE) method to build a more efficient big data model such as variable selections and/or model building. Generally speaking, large sample size tends to make some "variables" statistically significant while they may not have real "practical importance". Therefore, it is more likely to overly select such variables simply because of using (relatively) large sample size. For most "big data" applications, the number of observations for the whole data and training data can easily exceed few thousands. Consequently, most "typical" statistical variable selection procedures tends to overly select "less important" variables which are "statistically significant". RISE is a better strategy to choose and analyze various subdata of a "smaller size" that can combine or ensemble various results to build a more efficient and reliable model.

※ Online live streaming through Cisco Webex will be available.

※ The tea reception will be held at 10:10.