

# NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

## C S S E M I N A R

**Title:**           **Visualization of Big Data**

**Speaker:**       Professor KUNG Sun-Yuan  
                  Department of Electrical Engineering  
                  Princeton University

**Date/Time:**    10 June 2015, Wednesday, 04:00 PM to 05:30 PM

**Venue:**           Executive Classroom, COM2-04-02

**Chaired by:**    Dr Chua Tat Seng, KITHCT Chair Professor, School of Computing  
                  (chuats@comp.nus.edu.sg)

### Abstract:

Big data has been widely characterized by 3V's: high volume, high velocity, and high variety. Due to its quantitative (volume and velocity) and qualitative (variety) challenges, big data to the users resembles something like "the elephant to the blind men". It is imperative to enact a major paradigm shift in data mining and learning tools so that information from diversified sources must be integrated together to unravel information hidden in the massive and messy big data and, metaphorically speaking, that the blind men may "see" the elephant.

This talk addresses yet another important "V"-paradigm: "Visualization". Visualization tool is meant to supplement (instead of replace) the domain expertise (e.g. a cardiologist). It provides a big picture to help users formulate critical questions and subsequently postulate heuristic and insightful answers.

Quantitatively, it is vital to address is the curse of high feature dimensionality causing concerns on computational complexity and over-training. In this talk, we shall explore various projection methods for dimension reduction - a prelude to visualization of vectorial and non-vectorial data. For unsupervised learning data, Principal Component Analysis (PCA) has been the primary projection tool for visualization. PCA may be extended to "Multiple Discriminant Analysis" (MDA), amenable for supervised learning applications. Furthermore, we shall develop a "Discriminant Component Analysis" (DCA), which may be viewed as the PCA in the newly introduced Canonical Vector Space (CVS). Based on our experimental, DCA far outperform PCA, both numerically and visually. (If time permits, we shall also show why is DCA is also promising for privacy preserving for cloud and internet processing of personal data.)

Qualitatively speaking, "variety" refers to the fact that big data are often messy and,

moreover, they are often imprecise and incomplete. In order to unravel information hidden in the massive and messy big data, we shall explore kernel learning machine for non-vectorial or incomplete data. More specifically, we shall extend PCA/DCA to kernel PCA/DCA, by introducing a partial correlation kernel function meant for incomplete data analysis. Based on the experiments we have conducted, the partial correlation kernel seems to perform favorably when compared with the (imputed) Gaussian RBF kernel function.

Biodata:

S.Y. Kung received his PhD Degree in Electrical Engineering from Stanford University in 1977. Since 1987, he has been a Professor of Electrical Engineering at the Princeton University.

Kung is a Fellow of IEEE since 1988. He served as a Member of the Board of Governors of the IEEE Signal Processing Society (1989-1991). He was a founding member of several Technical Committees (TC) of the IEEE Signal Processing Society, including VLSI Signal Processing TC (1984), Neural Networks for Signal Processing TC (1991) and Multimedia Signal Processing TC (1998), and was appointed as the first Associate Editor in VLSI Area (1984) and later the first Associate Editor in Neural Network (1991) for the IEEE Transactions on Signal Processing. He presently serves on Technical Committees on Multimedia Signal Processing. Since 1990, he has been the Editor-In-Chief of the Journal of VLSI Signal Processing Systems.

Kung was a recipient of IEEE Signal Processing Society's Technical Achievement Award for his contributions on "parallel processing and neural network algorithms for signal processing" (1992); a Distinguished Lecturer of IEEE Signal Processing Society (1994); a recipient of IEEE Signal Processing Society's Best Paper Award for his publication on principal component neural networks (1996); and a recipient of the IEEE Third Millennium Medal (2000).

On publication side, Kung has authored more than 400 technical papers and numerous textbooks, including "VLSI and Modern Signal Processing" with Russian translation, Prentice-Hall (1985), "VLSI Array Processors", with Russian and Chinese translations, Prentice-Hall (1988); "Digital Neural Networks", Prentice-Hall (1993); "Principal Component Neural Networks", John-Wiley (1996); and "Biometric Authentication: A Machine Learning Approach", Prentice-Hall (2004).

His research interests include VLSI array processors, system modelling and identification, neural networks, wireless communication, sensor array processing, multimedia signal processing, bio- informatics data mining and biometric authentication.