Abstract

In recent times Statisticians are encountering a new variety of datasets with many more variables than observations that we call "Megavariate" data. The methodology for analyzing Megavariate remains a work in progress but new methodology has been developed that are worth of consideration. The central issue of these analyses is how to avoid overfitting the data, or how to separate the true signal from the spurious signal that is generated by the abundance of variables.

I will present a pipeline of methods for analyzing Megavariate data that focuses on the following issues:

1. Individual variable analysis: to identify significant variables borrowing strength from all the variables.
2. Group analysis: procedures to identify significant groups of variables.
3. Enriched methods for supervised and unsupervised classification: procedures that improve the performance of classifiers with the use of enrichment methodology.

In the biological sciences Megavariate datasets are common in experiments with DNA microarrays, DNA sequencing, mass spectroscopy and molecular
imaging, technologies that are now gradually beginning to be more and more widely used in laboratory experiments and in clinical settings. I will illustrate these issues with some data examples.

This is joint work with Dhammika Amaratunga, and my students Lan Yi, Zhenya Cherkas and Volha Tryputsen.

**About the Speaker**

**Javier Cabrera** Ph.D. is the director of Institute of Biostatistics and professor of the department of Statistics and Biostatistics at Rutgers University. He got his Ph.D. from Princeton University and has lectured in Statistics at Rutgers University, National University of Singapore and Hong Kong University of Science & Tech. He is author and coauthor of many publications in the areas of data mining and functional Genomics, including a book in Exploration and Analysis of DNA microarray and Protein array data. He is in the executive board of the Data Analysis working group for the DNA barcode of life. Research funded by NSF, Fulbright Fellow, 2010 SPAIG award for the Pfizer/Rutgers partnership.

Enquiry: 3442 8420

*All are welcome!*

SEEM Seminar 2011-2012/024