

Since coming on board at IIIT-Bangalore, I (Professor Lakshminarayan) have started a course stream in statistics and data mining. In spring 2004, I introduced a class in Statistics and Data Mining. The class included a long introduction to probability and statistics which consumed a quarter of the semester, and the last three quarters were focused on statistical methods in data mining.

I tried to maintain a nice balance between theory and practice. As it was a three credit class, we met three times a week, for 1-15mins each. One class period a week was a dedicated lab session, where we worked large data sets from the real world.

All the data analysis and modeling was done using SAS programming language, SAS suite of statistics and mathematical algorithms and the SAS Enterprise Miner. As the class focused on exploratory data analysis, modeling and pattern recognition, we used the univariate procedures, regression (simple, multiple, logistic), and generalized linear model, discriminate analysis (parametric and non-parametric), cluster analysis (Fastclus, Cluster, E-M algorithm), and Neural networks among others. All analysis was conducted using the procs listed above using SAS/STAT, as well as SAS Enterprise Miner.

It was, indeed, valuable that SAS was able to provide the “Text Miner” node within Enterprise miner. I introduced the students to text mining, and the availability of hierarchical and E-M algorithm was certainly helpful. A term project was a part of the course evaluation, and I had a team of students design an artificial neural networks algorithm using back-propagation. Comparing our code against the ANN with within EM was again helpful for benchmarking.

My students are now very conversant with SAS and many are considering taking the SAS certification examination. SAS is ubiquitous, SAS is comprehensive, SAS is user friendly, and SAS technical support is simply outstanding. AS SAS scores highly across all these important dimensions, it’s certainly the gold standard for statistical software.

Choudur K. Lakshminarayan Ph.D
Professor of Statistics and Computer Science
Indian Institute of Information Technology
26/C Electronics City, Hosur Road
Bangalore 560100, India
Tel: +91 80 28527627, Ext. 123.
e-mail: choudur.lakshminarayan@iiitb.ac.in
e-mail: choudur.lakshminarayan@hp.com