ISyE 6416 – Computational Statistics - Spring 2015

Tentative Syllabus

This course covers a set of optics regarding statistical computing and machine learning, including the following:

- Bisection and functional iteration
- Newton's method and application in statistics
- Regression and logistic regression
- Gaussian mixture and Hidden Markov models
- EM algorithm
- Model selection and cross validation
- Principle component analysis (PCA)
- Splines
- Simulation and numerical integration
- Monte-Carlo Markov chain
- Boostrapping
- Other topics (optional): Density estimation and smoothing, tree methods and boosting, change point detection, compressed sensing and matrix completion, clustering and spectral clustering

Instructor: Prof. Yao Xie, Groseclose #339, email: yao.xie@isye.gatech.edu

Class Time and Location: Tues, Thurs, 8:05am - 9:25am, Instr Center 213

Instructor Office Hour: Tues at noon, Groseclose 339

Class TA: Minkyoung Kang, mkk@gatech.edu

Class rescheduling: We will have to reschedule 1 to 2 lectures because of instructor's traveling to conferences. Videotaped lectures will be posted online. Please check with announces and emails for makeup lecture arrangements.

Class Website: main course website http://www2.isye.gatech.edu/~yxie77/isye6416.html

Tsquare website includes

- Homework assignments and solutions
- Exam and solution
- Your course grades on exams and homework

Make sure the scores in T-square are consistent with what you got. We will not make any change in grading for works older than 2 weeks.

Class Mailing List: Registered students are automatically subscribed to the class mailing list.

Textbook: Numerical Analysis for Statisticians, Kenneth Lange.

References:

An introduction to statistical learning: with applications in R, G. James, D. Witten, T. Hastie, R. Tibshirani.

Computational Statistics handbook with MATLAB, W. L. Martinez and A. R. Martinez.

Elements of Computational Statistics, J. E. Gentle.

Numerical linear algebra, Lloyd N. Trefethen and David Bau III.

Numerical Recipes, 3rd edition, William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery.

The elements of Statistical Learning: Data Mining, Inference, and Predictions, 2nd edition, Trevor Hastie, Robert Tibshirani, and Jerome Friedman.

Machine learning: A probabilistic perspective, K. P. Murphy.

Prerequisites: undergraduate level basic probability, linear algebra, and statistics.

Honor Code: For any question involving Academic Honor Code issues, please consult me, the class teaching assistant, or <u>www.honor.gatech.edu</u>

Software: MATLAB. You will use this software for homework assignment.

Grading Policy: Class Participation 5%, Submitting Teaching Evaluation – 5%, Homework - 30%, Midterm - 30%, Final - 30%

Homework: There are roughly biweekly homework assignments. Late homework will not be accepted. You may collaborate with another classmate on homework, but you have to write up and turn in your own solutions, as well as write your team member's name on the homework. The lowest homework grade will be dropped.

Midterm and Final: There will be one in-class midterm exam and one final exam. The exams are close book. But you may bring four pages (two-sided, or equivalently 8 one-sided pages) notes. On the notes, you can write down anything you like. But the notes have to be self-made and you may not copy from others. You are not allowed to use your cell phone. **No make-ups.** Any regarding request has to be submitted within 1 week after the midterm is handed back. Please write the detail of your request and explain the reason on a separate sheet attached to the original answer sheet.

Midterm will be arranged after we finish the topic of model selection and cross validation. Final exam has been scheduled on **Thursday April 30, 8:00-10:50am** in classroom.