STAT 890
Selected Topics: Statistical Computing

Spring 2011
Day Course

Students requiring accommodations as a result of disability, must contact the Centre for Students with Disabilities 778-782-3112 or csdo@sfu.ca

Instructor: Dr. Brad McNeney

Textbook:

Numerical Analysis for Statisticians, 2nd ed., Author: Kenneth Lange, Publisher: Springer, Year: 2010.

Course Description:

This course provides an overview of computational statistics methods for frequentist and bayesian inference.

Course Outline:

1. Review of numerical issues: machine precision, random number generation, probability integral transform
2. Review of likelihood inference: maximum likelihood, profile likelihood, constrained likelihoods and Lagrange multipliers
3. Optimization methods: gradient methods, gradient free methods (eg. golden section search), IRLS
4. Other computational stat tools: EM algorithm, data augmentation
5. Basic sampling methods: bootstrap, jackknife
6. Review of Bayesian methods: posterior modes, Bayes factors, credible intervals
7. Basic Monte Carlo methods: MC integration and its errors, importance sampling
8. MCMC: Metropolis Hastings, burn-in, diagnostics, proposal acceptance, effective sample size
9. MCMC extensions: simulated annealing and population MCMC
10. If time/expertise permits: supervised and unsupervised Machine learning, reversible jump MCMC, simulated tempering

Grading Scheme:

Wiki participation
Project presentations
Grading is subject to change.

Students should be aware that they have certain rights to confidentiality concerning the return of course papers and the posting of marks. Please pay careful attention to the options discussed in class at the beginning of the semester. Students are reminded that Academic Honesty is a cornerstone of the acquisition of knowledge. Scholarly integrity is required of all members of the University. Please consult the General Guidelines of the calendar for more details.

Revised October 2010

Teaching details:
Course Objectives

Course objectives include

- an understanding of the limitations of computers to solve problems
- practice at breaking a real problem down into a series of smaller subproblems and tasks.
- practice at researching appropriate methodology for a specific task, assessing the suitability of the methodology and communicating with collaborators about your assessment and the reasons for it.
- experience collaborating with others
- practice at communicating findings on both the numerical methods and the real problems that motivate them.
- build a body of knowledge for future problems
- gain exposure to collaborative tools such as wikis and version control systems

Projects

- Small team of (e.g. 3-4) students solve a real problem in some area of statistics using numerical methods.
- Teams required to program in a high-level statistical language and provide working code with documentation and data to illustrate their projects.
- Teams document the evolution of their projects, including the obstacles encountered and how they were overcome, through a class wiki that highlights the process of development and builds community. It is hoped that learning will become a community focus (e.g. Reggio Emilia approach).
- Part-way through, teams make their draft projects available on the wiki to be viewed and evaluated by both the other teams and the instructor. Teams receive feedback on their drafts from everyone in an open-development process.
- Part of course grade is based on Wiki participation

- End of course: Teams give two presentations
1) a general presentation giving an overview of their project and the results (possibly by poster)
2) a more specific presentation on the methodologic challenges and the numerical methods used to tackle them.

Everyone on the team needs to present some aspect of the work in this second presentation. An important goal is to give everyone some practice speaking in front of an audience.

Both presentations are evaluated by the other teams and the instructor. The team receives a grade.

Example areas of application:
- disease mapping
- resource management
- clinical trials
- medical research
- statistical genomics

Examples of statistical methods that give rise to interesting numerical issues:
- in general, likelihood and Bayesian inference for various complex statistical models such as
    * generalized linear mixed models
    * generalized additive models
    * missing data problems
    * measurement error models
    * functional data analysis
    * differential equations modelling