Stat 610: Statistical Computing

Meeting time: Tuesdays and Thursdays, 11:15-12:30pm Meeting location: BH 342
Website:

Instructor: Prof. Julia Fukuyama jfukuyam at iu dot edu
Office hours: Mondays 12:45-2:45 Informatics East, Room 201

Associate Instructor: Mr. John Koo
Office hours: TBA TBA

Course Overview

As a statistician, you will need to manipulate data, optimize, and simulate. You will also need to know enough about how the methods you use work to diagnose problems when they arise and to be able to implement modified versions when the standard implementations don't suit your purposes.

You also need to write accurate, clean, maintainable, demonstrably correct code. To that end, the first half of the class will be devoted to how to program well, with statistical tasks giving us the computational problems.

Once we have the software engineering down, we will move on to the algorithms used in applied statistics. These can be roughly broken up into optimization methods and stochastic simulation methods.

Textbooks

The primary textbook for the first half of the course with be *The Art of R Programming*, by Norman Matloff. *The R Cookbook*, by Paul Teetor, will also be useful. The primary textbook for the second half of the course will be *Numerical Analysis for Statisticians*, by Kenneth Lange. Additional readings will be posted on the course website.

Class Schedule

Dates and topics subject to change.

Week 1

- Data types and data structures
- Flow control and looping
- Reading: Matloff Chapters 1-6, Chapter 7.1

Week 2

- Text
- Regular expresions
- Reading: Matloff Chapter 11
- Homework 1 out, due September 10

Week 3

- Writing and calling functions
- Code refactoring
- Reading: Matloff Sections 7.6, 7.9, 14.1-14.3
- Homework 2 out, due September 17

Week 4

- Split/apply/combine
- Reading: Matloff Chapter 10, R Cookbook Chapter 6
- No new homework

Week 5

- Shape changing/merging/transformations
- Debugging
- Reading: Matloff Chapter 13
- Homework 3 out, due October 1

Week 6

- Testing/top-down design
- Reading: Matloff Section 7.6
- Homework 4 out, due October 8

Week 7

- Object-oriented programming, classes
- Packages and namespaces
- Reading: Matloff Chapter 9
- No new homework (study for the midterm)

Week 8

- Midterm
- Version control/git
- No new homework

Week 9

- Performance enhancement and code profiling
- Fitting and using statistical models
- Reading: Matloff Chapter 14
- Homework 5 out, due October 29

Week 10

- Gradient methods
- Homework 6 out, due November 5

Week 11

- Newton's Method

Week 12

- More specialized optimization topics: convex optimization and the EM algorithm
- Homework 7 out, due November 19

Week 13

- Random number generation
- Independent Monte Carlo

Week 14

- Monte Carlo methods
- Homework 8 out, Due December 10

Week 15

- Monte Carlo Methods, wrapping up

Assessment

Assessment will be based on a combination of homework, an in-class midterm, and a final project. Final grades will be based on:

- – 40% homework

- – 30% midterm

- – 30% final project

There will be 8 homeworks over the course of the semester, generally graded out of 5 points, with one point for a good-faith effort at all the problems, 5 points for correct answers with clean code, and an intermediate number of points otherwise.

Homeworks will be assigned on Sundays and due the following Tuesday (9 days later). At the time the homework is assigned, we will generally not have covered all the material needed to complete the homework, but we will have covered everything by the Thursday before the due date. The idea is to give you the homework early enough that you can think about it while the material is being covered in lecture. Therefore, it will generally be a good idea to take a look at the homework when it is assigned even if you aren't able to complete all the problems yet.

Course Policies

Late Policy

Late homework will be penalized at one point per 24 hours, and no assignments will be accepted more than three days late (e.g. a homework due on Tuesday at 11:59pm will not be accepted later than the following Friday at 11:59pm). No more than three late days can be used over the course of the semester. Special accommodations may be granted if you ask very early.

Academic Integrity

You are expected to abide by the guidelines of the IU Code of Student Rights, Responsibilities, and Conduct (http://studentcode.iu.edu/responsibilities/academic-misconduct.html) regarding cheating and plagiarism. Any ideas or materials taken from another source must be fully acknowledged and cited.

Disability Accommodation

Please contact me if you require assistance or academic accommodations for a disability. You should establish your eligibility for disability support services through the Office of Disability Services for Students in Wells Library W302, 812-855-7578.