Stat 710: Statistical Computing

Meeting time: Tuesdays and Thursdays, 1-2pm — Meeting location: BH 317
Website: jfukuyama.github.io/teaching/stat710

Instructor: Prof. Julia Fukuyama — jfukuyam at iu dot edu
Office hours: Thursdays 2:15-4:15pm — Informatics East, Room 201

Associate Instructor: Mr. John Koo
Office hours: Mondays 10:30-11:30am, Wednesdays 1-2pm — Informatics East, Room 103

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. The class will be primarily in R, with one homework in python and the option of another homework in python for those who would like more experience with the language.

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. For optimization, we will cover gradient descent, stochastic gradient descent, the EM algorithm, and topics in convex optimization. Stochastic algorithms will include rejection sampling, Metropolis-Hastings, and Gibbs sampling.

Textbooks

The primary textbook for the course with be *The Art of R Programming*, by Norman Matloff. *The R Cookbook*, by Paul Teetor, will also be useful. 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 January 22

Week 3

- Writing and calling functions
- Code refactoring
- Reading: Matloff Sections 7.3-7.6, Chapters 3 of *Clean Code: A Handbook of Agile Craftsmanship* by Robert Martin (Chapters 1 and 2 are also good and useful, read them if you have time).
- Homework 2 out, due January 29

Week 4

- Input and output
- Split/apply/combine 1
- Reading: Matloff Chapter 10, R Cookbook Chapter 6
- No new homework

Week 5

- Split/apply/combine 2
- Shape changing/transformations
- Homework 3 out, due February 12

Week 6

- Debugging
- Testing/top-down design

- Reading: Matloff Chapter 13, Section 7.6
- Homework 4 out, due January 19

Week 7

- Testing/top-down design 2
- Object-oriented programming, classes
- Reading: Matloff Section 7.6, Chapter 9
- No new homework (study for the midterm)

Week 8

- Midterm
- Performance enhancement and code profiling
- Reading: Matloff Chapter 14
- No new homework

Week 9

- Fitting and using statistical models
- Homework 5 out, due March 19

Week 10

- Convex theory
- Gradient methods for unconstrained optimization (gradient descent, stochastic gradient descent, coordinate descent, backtracking methods)
- Homework 6 out, due March 26

Week 11

- Constrained optimization, Lagrange multipliers, barrier method
- Statistical applications and cvx

Week 12

- The EM algorithm
- Independent Monte Carlo
- Reading: Lange Chapter 10 for EM, Lange Chapter 21 for Monte Carlo

Week 13

- ABC
- Markov chains

Week 14

- Metropolis-Hastings
- Gibbs sampling

Week 15

- Neural nets: backpropagation
- Neural nets: some examples

Assessment

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

- 40% homework
- 30% midterm
- 30% final exam

There will be 10 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.