STAT 4060: Computational Methods for Statistics and Data Science.

Instructor: Ailin Zhang (ailin.zhang@sjtu.edu.cn) TA: Jiayuan Rao

Fall, 2022

Course Description

This course is to introduce the art of computational methods in statistics and data science. With the help of R, Python, and C++, we will write computer code to implement the core algorithms in statistical computing.

We will cover the following topics:

- R and Python Basics
- Least squares regression, sweep operator, QR decomposition
- Eigen computation, Principal Component Analysis
- Logistic regression, Newton-Raphson.
- Feed-forward neural network, back-propagation
- Adaboost, coordinate descent.
- Ridge regression, spline.
- Lasso, stagewise regression, solution path.
- Factor analysis, EM.
- Random number generators, linear congruential, rejection, polar.

When going through the above topics, the focus will be on algorithms and especially programming, instead of theories of learning, inference and computing.

Recommended Reference

The course is **not** built on top of a single textbook. Therefore I would strongly recommend you to participate in the lecture. However, the following references are useful:

1. R Programming for Data Science (2020) by Roger Peng.
   https://bookdown.org/rdpeng/rprogdatascience
2. Introduction to Data Science (2020) by Rafael Irizarry.
   https://rafalab.github.io/dsbook

Grading Policy

The typical JI grading scale will be used. I reserve the right to curve the scale if there are less than 30% of students with grades $\geq$ A. The grade will count the assessments using the following proportions:

- **10%** Attendance and Participation.
  To get the full credit, you need to make > 90% of the attendance check. The rest 10% is your flex-time. Feel free to use it if you have something emergent.

  If you only make > 80% of the attendance check, you will get 6 pt.

  Otherwise, you get 0 pt.
- **40%** Homework
- **20%** Midterm
- **30%** Final Project (15% for presentation, and 15% for the written report)
- **1%*** Extra Credit

Final Project

For your final project, you will analyze some data using the methods we have talked about in class. You will write up your analysis in a written report, and will also make an oral presentation. The presentation will be only 4 minutes each in total.

Office Hour

Wednesday 2-4 PM or by appointment.
You are welcome to chat whatever you like about data science and career planning!

We want you to succeed!

If you are feeling overwhelmed, visit our office hours and talk with us, and we want to help you succeed.