15.S60 --- Computing in Optimization and Statistics

Days: Tue Thu (9am-12pm)
1/8/2018 – 2/1/2018
Place: E51-325
Credits: 3 Units (Credit/Fail or Listener Only)

Instructors: Prof. Dimitris Bertsimas (dbertsim@mit.edu)
Phil Chodrow (pchodrow@mit.edu)
Brad Sturt (bsturt@mit.edu)

TAs: Colin Pawlowski, Andy Zheng, Zachary Blanks, Arthur Delarue, Konstantina Mellou, Chris McCord, Jackie Baek

Course Content and Objectives:

The "big data revolution" has placed added emphasis on computational techniques for decision-making with data. Large-scale optimization, data analysis and visualization are now commonplace among researchers and practitioners alike. More than ever, there is a need not only to develop techniques, but also to implement and use them in computational practice.

15.S60 is a multi-session workshop on software tools for informing decision-making using data, with a focus on contemporary methods in optimization and statistics. We concentrate on teaching elementary principles of computational practice using common software and practical methods. By the end of the course, students will possess a baseline technical knowledge for modern research practice. Class participation and individual hands-on coding are stressed in each session.

The course is divided into 8 self-contained modules. Each module consists of a 3-hour, interactive workshop where participants learn a specific software tool. Class participation, group code-reviews and individual hands-on coding are stressed in each session. At the end of the module, participants will be able to use the software and techniques learned in their own research. Participants will also leave each workshop with code they, themselves, have authored to use for future reference.

Prerequisites:

This course is NOT entry-level. The prerequisites are:

- **Required**: Instructor permission. Please email bsturt@mit.edu to request permission. Briefly describe your interest in the course as well as your academic and research background.
- **Required**: Familiarity with a modern programming language such as C++, Python, Java, R, Matlab, or Julia. Modules will be taught in R and Julia but prior knowledge of these specific languages is not assumed.
- **Helpful**: Familiarity with optimization at the level of 6.255J/15.093J. Familiarity with elementary statistical concepts. Experience working with data in a research or industry setting.

Course Materials:

All software used in this course is either available free for download, under academic license, or through MIT IST. Data sets, software installation instructions, tutorials and reference material will be made available through a class GitHub repository (previous year's found at https://philchodrow.github.io/cos_2017/).

Grading:

Course is only available as Credit/Fail or Listener. To receive credit for the course, attendees must

1. Attend and actively participate in at least 6 of 8 sessions
2. Complete ALL 8 of 8 "Testing your Installation Assignments" (See below)
3. Complete course feedback forms for at least 6 of 8 sessions.

See the "Assignments" section below for more detail.

Module Schedule:

The schedule will include the following topic modules (subject to change):

Introduction

Module 1: Motivation, Terminal, Github
Leaders: Jackie Baek, Brad Sturt

Data Science with R

Module 2: Data Wrangling
Leaders: Andy Zheng

Module 3: Statistical Modeling and Machine Learning in R
Leaders: Colin Pawlowski

Module 4: Advanced Techniques for Data Science in R
Leader: Phil Chodrow

Module 5: Deep Learning in R
Leader: Zachary Blanks

Optimization with Julia

Module 6: Introduction to Julia and JuMP, Linear Optimization
Leaders: Arthur Delarue, Konstantina Mellou

Module 7: Nonlinear and Integer Optimization in JuMP
Leaders: Arthur Delarue, Konstantina Mellou

Distributed and Parallel Computing

Module 8: Distributed and Parallel Computing, Engaging
Leader: Chris McCord

Assignments:

Before each session

All software and datasets required for a session should be installed **prior** to that session. Instructors will not delay class to assist with installation issues. Detailed installation instructions are available on Github.

At the end of each set of instructions, you will see a section entitled “Testing your installation.” This section will typically involve downloading a script from the Stellar site and running that script on your computer.

Your homework--due at 8pm the day before each session--is to copy and paste the output of this script into a text document, and upload this text document to Stellar.

Some installation processes make take several hours to complete. Please plan ahead.

During each session

Participants should bring a laptop to all sessions or make prior arrangements to share with another student. Participants are expected to participate in coding exercises, class discussion and any group-code reviews.

Each session is approximately 3 hours long. Please make sure your laptops are charged before the session and bring a power-adapter as necessary.