Software Tools for Operations Research

Time:	9:00 am – 12:00 pm Tuesday and Thursday 7 Jan 2013 – 30 Jan 2013
Place:	7 Jan 2013 - E51-335 All other dates: E62-250
Credits:	3 Units (Credit/Fail or Listener Only)

Instructors:	Prof. Dimitris Berstimas (dberstim@mit.edu) Iain Dunning (idunning@mit.edu) Vishal Gupta (vgupta1@mit.edu)
TAs:	Jerry Kung, John Silberholz, Angie King, Velibor Misic, Miles Lubin, Ross Anderson, Joey Huchette

Course Content and Objectives:

The ``big data revolution" has placed added emphasis on computational techniques in Operations Research. Large-scale optimization, data analysis and visualization are now commonplace among researchers and practitioners alike. Now more than ever, there is a need not only to develop new techniques, but also to implement and use them.

This course is a multi-session workshop focusing on software tools specific to the practice of Operations Research. While other courses focus upon theory and background, our course focuses on the mechanics of using software to apply specific methodologies. A full-list of the covered methodologies and software tools can be seen below in the *Module Synopsis.*

The course is divided into 8 inter-connected 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. A final project that links together many concepts is constructed over the final two modules. At the end of the module, participants should 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. Familiarity with optimization at the level of 15.093J or permission from instructor is required. In this class we will use the R and Julia programming languages, and expect that students will be familiar with programming concepts from the use of either one or both of these languages, or another language such as C, C++, Python, Java, MATLAB, etc.

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 the class GitHub page at https://github.com/IainNZ/ORSoftwareTools2014

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 also the "Assignments" section below for more detail.

Module Synopsis:

Module 1: Introduction to [R]
Leader: Jerry Kung
Date: 7 Jan 2013

Introduces the statistical programming environment [R]. Participants will learn basic functionality, including importing and storing data and performing common analyses. Topics include linear/logistic regression, CART, random forests and clustering. We assume participants are comfortable with basic statistics.

Module 2: Intermediate [R]
Leaders: John Silberholz
Date: 9 Jan 2013

Learn how to manipulate your datasets into the forms you need for further analysis using the apply family of functions in [R]

Module 3: Visualizations in [R]
Leader: Angie King
Date: 14 Jan 2013

Introduces how OR practitioners can use visualization to explore data sets, evaluate models, and present results. Participants will learn to create visualizations with the ggplot2 [R] package. Taking Module 2 before this class is recommended.

Module 4: Modeling Optimization Problems in Julia
Leader: Velibor Misic
Date: 16 Jan 2013

Introduces the Julia-based linear/integer optimization problem modeling language JuMP. Learn how to create data-driven models and best-practices. Class will include an introduction to Julia basics.

Module 5: Big Data
Leader: Iain Dunning
Date: 21 Jan 2013

Learn what Big Data really is, and how to work with large datasets. Covers data representation, algorithms for big data, relevant libraries, and an introduction to computation with MapReduce.

Module 6: Distributed Computing
Leader: Miles Lubin
Date: 23 Jan 2013

Learn about the concepts of distributed computing, including applications and limitations. We will use Julia to solve an optimization in parallel.

Module 7: Project Part 1 (Internet and Databases)
Leader: Iain Dunning
Date: 28 Jan 2013

Introduces the final class project, an internet service that solves travelling salesman problems (TSP) on demand. Discusses how the internet is structured and how clients, servers, clouds, etc. all interact. Build a simple service that is available over the internet with Julia.

Next, databases are introduced, including relational and "NoSQL"-style databases. Learn basics of interacting with a database using SQLite. Build a server that that performs calculations on data from a database.

Module 8: Project Part 2 (MIP Callbacks)
Leaders: Iain Dunning and Ross Anderson
Date: 30 Jan 2013

Learn how to solve hard problems using solver callbacks, including lazy constraint generation and heuristic solutions. We will apply this to solving a TSP.

Finally we will complete the project by connecting our TSP solver to our database and web server. Will cover issues around deployment.

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 Stellar.

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 the relevant scripts into a text document, and upload this text document to Stellar.

See the below timeline for details:

Module	Assignments	Due at 8pm on
1: Introduction. to [R]	Testing R installation	6 Jan
2: Intermediate [R]	Testing shaping packages	8 Jan
3: Visualization in [R]	Testing visualization packages	13 Jan
4: Modeling with Julia	Testing Julia + JuMP	15 Jan
5: Big Data	Testing R packages and Julia	20 Jan
6: Distributed Computing	Testing parallel computing	22 Jan
7: Project Part 1	Testing SQLite, HTTPServer	27 Jan
8: Project Part 2	Testing Gurobi callbacks	29 Jan

Please be aware, some installation processes may take some time 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.

Please note that sessions are approximately 3 hours long. Please make sure your laptops are well-charged before the session and bring a power-adapter as necessary.