CME594 Introduction to Data Science

Instructor: Professor S. Derrible, 2071 ERF, derrible@uic.edu
Office hours: TBD

Hours: Thursday: 5:00 – 7:30

Location: 104 LH (Lincoln Hall)

Summary: This course introduces students to techniques of complexity science and machine learning with a focus on data analysis. One new technique is covered every week, and will include: scaling laws, network science, cellular automata, agent-based modeling, decision tree learning, neural networks, hierarchical clustering, and principal component analysis. The main assessment is a final paper where the students are asked to pick any data set (preferably from their own research) and apply one or multiple techniques from the course. No coding experience is required, but the course will include some python programming.

Objectives: This course aims to provide students with introductory knowledge of several data science techniques that can be used for data analysis. The material learned should then be useful in the student's own research. More specifically, at the end of this course, students should be able to:
1. explain the main concepts behind all the techniques covered
2. identify the type of technique preferable to use depending on the type of data to analyze
3. use the various python libraries learned to be able to apply these techniques
4. apply rigorously one or multiple of these techniques learned in their own research

Textbook: No textbook is required, but the following books may be useful:

Batty, M., 2013, "The New Science of Cities", MIT Press, Cambridge, MA.
Batty, M., 2007, "Cities and Complexity", MIT Press, Cambridge, MA.
Han, J., Kamber, M., and Pei, J., 2011, "Data Mining: Concepts and Techniques", Elsevier Science.
Murphy, K., 2012, "Machine Learning: A Probabilistic Perspective", MIT Press, Cambridge, MA.
Barabási, A-L., 2014, "Network Science", Creative Commons: CC BY-NC-SA 2.0. PDF V26, 05.09.2014
Epstein, J., and Axtell, R., 1996, "Growing Artificial Societies: Social Science from the Bottom Up", MIT Press, Cambridge, MA.

Software: # NetLogo (sometimes the 32bit version works better even for 64bit computers): https://ccl.northwestern.edu/netlogo/

Python 2.7.xx: https://www.python.org/downloads/
Libraries: NumPy, SciPy, Pandas, igraph, SciKit (for Windows, see: http://www.lfd.uci.edu/~gohlke/pythonlibs/)

or

Anaconda – python 2.7 (recommended package that includes python and most recommended libraries; sometimes the 32bit version works better even for 64bit computers): https://www.continuum.io/downloads

Tentative Grading Policy: Attendance, participation, behavior (15%)
Homework (20%)
In Class Technique Presentation and Application (15%)
Abstract (5%)
Presentation (5%)
Visualization (10%)
Final Paper (30%)

Work submitted late may receive a penalty.

Plagiarism: Plagiarism is a serious offense and it will not be tolerated; see university policy. All reviews, papers and any other submitted material will be run through a plagiarism tool.

Attendance Policy: All students are required to attend the lectures and be on time. If at any moment a student is to be absent, he/she should have discussed it prior with the instructor.

Professional Conduct: Students are always expected to conduct themselves with the utmost respect towards the instructor and their fellow students. Cellphones are to be turned off.

Class Schedule and Readings

TBD