36-602 Perspectives in Data Science II Spring 2019

Course Policies and Syllabus

Instructors: Howard Seltman Baker Hall 232H 412-268-3938 hseltman@stat.cmu.edu
Christopher Peter Makris Baker Hall 232L 412-268-2602 cmakris@stat.cmu.edu

Schedule: Mondays and Wednesdays at 10:30-11:50 in BH 235B

Office Hours: Whenever we are in our offices or by appointment

Course Webpage: https://cmu.instructure.com/ or http://www.stat.cmu.edu/~hseltman/602/

Overview:

This course is a continuation of 36-601, Perspectives on Data Science I. Recall, we have two major goals for the MSP program. One is to advance those skills that will help make you an outstanding statistical scientist. That is the goal of this course. (The other goal is to help you develop general professional skills that will be of value in the workplace, such as, communication, career development and leadership skills.) In this course, you will learn and become proficient in Python, web scraping, SAS macros, and Hadoop. These skills are directly applicable in many workplaces. Equally important, our broader objective is to develop an appreciation for the general principles of data science that transcend specific languages and to learn how to learn a new programming language.

Learning Objectives:

At the end of this course, you should be able:

1. use Python to collect, clean, store, retrieve, and analyze data
2. write R or Python code to automatically extract data from web pages
3. automate tasks in SAS
4. use several different interfaces to implement a map / reduce strategy to collect data distributed across multiple computers in a Hadoop distributed file system
5. ask the appropriate questions to begin to learn a new computer programming language

Course Organization:

Most of the classes will be a mix of lecture and individual or small group programming tasks.

Evaluation:

It is essential to the success of this class that you participate actively. This means that you must do all the assignments including any assigned readings and you must take part in discussions. Attendance is mandatory. Since we are aiming for mastery of material, it will not be unusual to be asked to redo an assignment. The final course grade is based on 20% class participation and 80% homework.

Preliminary Syllabus:

1/14	Presentation for 36-726 & Review of regression interpretation
1/16	Presentation for 36-726 & Object classes in R
1/23	Statistical methods in Python
1/28	dplyr in R
1/30-2/11	Data scraping in Python and R

2/13-3/6	Hadoop and Spark
3/11-3/13	Spring break
3/18-3/25	Writing Python classes
3/27-4/17	SAS Macros
4/22-5/1	(Presentations for 36-726)

Getting in touch with us:

The easiest and most reliable way to get in touch with us is by email. Feel free to send mail at any time to hseltman@stat.cmu.edu or cmakris@stat.cmu.edu. We will respond as soon as we can.

You have unlimited office hours for this class: stop by our offices any time to discuss class material. Please understand that we may not be free to talk to you at that time, and if not, we can make an appointment for a later time.

Academic Integrity:

As a graduate student at Carnegie Mellon University, you are expected to uphold high standards of academic integrity. Please read the official policy at http://www.cmu.edu/policies/student-and-student-life/academic-integrity.html.

In this course, the default policy for individual assignments is that you may discuss them with other students and share short snippets of code, but all writing and the majority of the coding must be down on your own. Similarly, for group assignments, groups may discuss with other groups, but all writing and coding must be done only by the team members, and no team member may turn in an assignment unless he or she has made a substantial contribution and is prepared to answer questions about all parts of the assignment. Exceptions to this default may be listed on particular assignments. For all types of assignments, material taken from any external source must be quoted and appropriately cited. Feel free to talk to me if you have any questions or comments about what constitutes plagiarism or inappropriate collaboration.

Take care of yourself:

Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at http://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

Physically disabled and learning disabled students:

The Office of Equal Opportunity Services provides support services for both physically disabled and learning disabled students. For individualized academic adjustment based on a documented disability, contact Equal Opportunity Services at eos@andrew.cmu.edu or (412) 268-2012.