SYS 6018: Applied Data Mining

Course Description

Data mining (or knowledge discovery) methods seek to turn data into information, where information consists of relationships between variables (or attributes/predictors). This course describes and investigates methods for mining information from data. We consider both supervised and unsupervised methods for information and knowledge discovery and show their applicability to a range of problems.

Course Information

Lectures: T/Th 8:00am-9:15am in Thornton Hall E-316
Prof. Matthew S. Gerber, Ph.D. (msg8u@virginia.edu); Office Hours: T 1:00pm-2:00pm in Olsson 102-D
Teaching Assistant(s):

- Hunter Li (xl2gr@virginia.edu); Office House: TBD

Textbook / Readings

This course will draw primarily from the following textbook:

> James, Witten, Hastie, and Tibshirani. An Introduction to Statistical Learning. Springer, 2013.

Other readings will be drawn from the research literature and are listed in the course schedule (described below).

Schedule / Components

The schedule can be accessed at the following URL: `https://goo.gl/k4RsA8`. The course contains the following major components:

1. Lectures: I will use the lecture time to present the technical content of the course via a combination of coding demonstrations (in R), chalkboard explanations, and (rarely) projected slides. I will not take attendance, but I will assess your level of participation and factor it into your final grade.
2. Readings: The schedule contains two columns listing required readings. These readings provide necessary background for the associated lectures and must be completed *before the associated lectures*. If you do not finish the readings before class, you will not be able to fully participate in the lecture discussions (a non-negligible part of your final grade).
3. Case Studies: The case studies will provide you with an opportunity to apply the techniques presented in the lectures to non-trivial problems and datasets. All case studies and associated deadlines are described in the schedule.
4. Mid-Term Examination: There will be an in-class mid-term examination. The mid-term examination will consist of closed-notes and open-notes sections. The closed-notes section will test your understanding of concepts. The open-notes section will test your problem solving skills.
5. Final Project: You will work in teams to propose, complete, and present a final project.

Grading and Late Submission Policy

Final grades will be calculated as follows:

- Case Studies: 50%
- Final Project: 25%

- Mid-Term Examination: 20%
- In-Class Participation: 5%

A grade of 'A' indicates superior achievement in the course. A grade of 'B' indicates that the basic requirements were met. Grades of C+ and below indicate that basic requirements were unmet. Half-grades (A+ through C-) are assigned at the instructor's discretion. You must obtain a B- or higher to receive credit for this course.

All assignment submissions will be made through Collab. All submissions will be due by 11:59:59pm on the due date. Late submissions (even 1 second late) will not be accepted and will receive a score of zero. The time stamps produced by Collab will be the authoritative reference for all such decisions. If you have special circumstances (e.g., a documented physical condition) that prevent you from adhering to the posted deadlines, please inform me well in advance of the deadline (at least 1 week) so that I can make arrangement to accommodate you.

Honor Code Except for the final project, all work in this course must be done individually, without help from your peers. You may talk informally about the assignments, but you cannot discuss solutions or show/share code with your peers. The Honor Pledge on each case study report reaffirms this. **DO NOT COPY or PLAGIARIZE!** Anyone caught cheating in this course will receive a failing grade and may be referred to the UVa Honor Committee for further investigation. It is not always easy to tell what qualifies as copying/plagiarism, so do not be afraid to talk to me about it. Such discussions do not imply guilt of any kind.

Honor Offices: 434-924-7602 (Mon-Fri, 9am-5pm); http://www.virginia.edu/honor

Software We will use R, the free environment for statistical computing and graphics. You can obtain R from `http://r-project.org`. Additional instructions for installing R on a Mac can be found on the course schedule. Other languages like S+ and SAS can be used, but I strongly recommend using R since I will be using it during class and in example code files. All example R code and data can be found at the following URL (ignore security warning):

`https://ptl.sys.virginia.edu/msg8u/Personal/Teaching/2015/SYS6018/Examples`

You are encouraged to modify and use this code in your assignments.

Your Well Being The School of Engineering proudly serves as a safe space for its students and aims to promote their well being. If you are feeling overwhelmed, stressed, or isolated, there are many individuals here who are ready and wanting to help. If you wish, you can make an appointment with me and come to my office to talk in private. The staff members in the Student Services Office, located on the third floor of Rouss & Robertson Hall, are also readily accessible to talk during walk-in hours or by setting up an appointment. Alternatively, there are also other University of Virginia resources available. The Student Health Center offers Counseling and Psychological Services (CAPS) for its students. Call 434-243-5150 (or 434-972-7004 for after hours and weekend crisis assistance) to get started and schedule an appointment. If you prefer to speak anonymously and confidentially over the phone, call Madison House's HELP Line at any hour of any day: 434-295-8255. If you or someone you know is struggling with gender, sexual, or domestic violence, there are many community and University of Virginia resources available. The Office of the Dean of Students, Sexual Assault Resource Agency (SARA), Shelter for Help in Emergency (SHE), and UVa Women's Center are ready and eager to help. Contact the Director of Sexual and Domestic Violence Services at 434-982-2774.