MATH/STAT 389: Statistical Learning, Spring 2019
Tuesday, Thursday 09:00-10:15   JPSN G24

Instructor: **Taylor Arnold**
E-mail: tarnold2@richmond.edu
Office: Jepson Hall, Rm 218
Office hours: By appointment (see course website for `calend.ly` link)

Description:

Computational statistics and statistical algorithms for building predictive models from large data sets. Topics include model complexity, hyper-parameter tuning, over- and under-fitting, and the evaluation of predictive performance. Models covered include linear regression, penalized regression, additive models, gradient-boosted trees, and neural networks. Applications are drawn from many areas, with a particular focus on processing unstructured text and image corpora.

Computing:

To facilitate your ability to actually *do* statistics, most class meetings will involve some form of computing. No prior programming experience is assumed or required.

We will use the **R** programming environment throughout the semester. It is freely available for all major operating systems and is pre-installed on many campus computers. You can download it and all supporting files for your own machine via these links:

$$\text{https://cran.r-project.org/}$$
$$\text{https://www.rstudio.com/}$$

You are required to bring a laptop to each class meeting with R installed and running. This requires that you have a computer with an up-to-date version of macOS, Windows, or Linux (iPads and Chromebooks will not suffice). If this is not possible, or becomes a problem during the semester, it is your responsibility to inform me as soon as possible so that we can find an alternative solution.

Course Website:

All of the materials and assignments for the course will be posted on the class website:

$$\text{https://statsmaths.github.io/stat389-s19}$$

The website contains notes, assignment details, and supplemental materials. At the end of the semester, this version of the course will be archived and available for your reference.

Labs:

During most class meetings, you will work on a series of assignments I refer to as 'Labs'. These may be a paper handout with questions or a code file that requires that you fill in answers digitally. In order to succeed in the course you should complete these prior to the next class meeting.
In the first half of the course (pre-Exam II), rather than formally handing them in, you must instead fill out an online questionnaire at some point prior to the next class meeting. I will not accept late submissions. The questionnaire can be found through a link on the course website.

In the second half of the course, you must submit your predictions through GitHub classroom. The link and instructions for doing this are on the course website.

You are excused for forgetting to hand in **two** questionnaires or predictions. Beyond these you will lose one point on your final grade for every missing lab.

Exams:

We will have four exams during the semester. Exams will focus on the material in each section of the course, but due to the cumulative nature of the material each requires understanding previous sections. There may be in-class and take-home components of each exam. The in-class portion will take place on the following days:

- 2019-02-05 (Tue)
- 2019-02-26 (Tue)
- 2019-03-26 (Tue)
- 2019-04-16 (Tue)

Take-home components will typically be due in class the day on which the in-class exam is assigned.

Final Project:

In lieu of a final exam, the course concludes with a final project. The project is due on the second to last class of the semester so that we can accommodate in-class presentations during the final week. More details on the project will be given prior to Spring Break.

Final Grades:

You will receive a numeric score from 0-100 for each of the exams and the final project. Your final numeric grade is determined by taking the average score of your four best grades, rounded to the nearest integer. Finally, subtract one point for every missing lab for which you failed to turn in a form (beyond the grace window of two missing forms).

The mapping from numeric grades to letter grades is given as follows:

A $\Rightarrow$ 90 to 100

B $\Rightarrow$ 80 to 89

C $\Rightarrow$ 70 to 79

I may assign pluses and minuses as needed. When appropriate, I may also modify these cut-off scores to make them more generous (but will not make them more strict).

Attendance:

There is no formal attendance policy for this course. However, if you miss a class it is your responsibility to catch up with the material. E-mail and office hours are not a replacement for attendance. When you are present, I expect you to arrive on time, engage with the material, and give us your full attention.

Office Hours and Email:

There are three easy ways to get help with the course:

- **lab form**: You may ask questions as part of the lab form. This is particularly useful if you have a general misunderstanding about a concept or a very specific question about one of the lab questions.

- **piazza**: I have created a Piazza site for our course; there is a link available on the main course website. This free, private site allows you to ask (optionally anonymous) questions visible to the entire class. I will provide answers to any posted questions as soon as possible. Students may also add their own responses.

- **office hours**: You may also make an appointment to come to my office hours. Please do this using the `calend.ly` link on the course website. Appointments can be made up to a week in advance, but must be booked within 24-hours of the time slot.

If you have personal issues of circumstance, feel free to email me directly. For questions about the course material, however, please instead make a Piazza post so that everyone can benefit from the answer.

Class Policies:

The following class policies address some of the most common questions and concerns that students have. If anything is unclear or not covered below, please feel free to ask for clarification at any point in the semester.

- **Academic honesty:** Cheating and plagiarism are grave scholarly offenses and potential grounds for expulsion; they are also a major barrier to your intellectual development. You are expected to familiarize yourself with the entirety of the University of Richmond's Honor Code. If you are confused or unsure about appropriate citation protocol or any other aspect of the Honor code, please consult me before turning in an assignment.

- **Special approval:** If you have special approval forms for extra time on exams or any other circumstances I should know about, please speak with me as early as possible so that we can best accommodate your needs.

- **Late work:** You are expected to submit all work on time. The final project will be accepted after the due date with a 10-point deduction for each 24 hour period (rounded up) that it is late.

- **Pass/Fail/Withdraw/Incomplete:** If you choose to take this class pass/fail, a passing mark requires you to receive a grade of C- or higher. I do not normally give a final grade of D. I am generally happy to allow students to withdraw after the deadline, with approval of the Dean, without penalty. However, I typically do not give a grade of incomplete (I or Y) to students who have not finished at least three of the exams.

- **Make-up exams:** There are no make-up exams. If you fail to attend an exam without a valid excuse (given to me by email within 24-hours of the exam) you will receive a score of zero. In the event of a valid reason for missing the exam, the missing score will be filled in with the median grade from the three remaining exams.

- **Class conduct:** During class I expect you to refrain from checking email, being on phones, or working on assignments for other classes.

Notice:

I reserve the right to modify this syllabus, with advanced warning, throughout the semester. If necessary, I will email the class list and post an updated version of the document on the course website.