Syllabus for 46-927: Statistical Machine Learning II

Spring 2019

Description: This is the second in a two-part sequence on statistical machine learning. The first course covered tools and approaches for prediction, including both regression and classification, with a focus on understanding the foundations of the methods so that they can be both applied and modified. Topics included foundations of supervised learning, regularized and nonparametric regression, bias-variance tradeoffs, model validation and assessment, classification, and tree-based methods. This second course will expand into advanced topics, with topics drawn from boosting and ensemble methods, support vector machines, mixture models and topic modeling, natural language processing, markov decision processes and reinforcement learning, and neural networks.

Professor: Alex Reinhart
Email: areinhar@stat.cmu.edu. Please ask questions on Piazza instead of by email if possible.
Office: Baker Hall 232K
Office hours: Fridays on Canvas, exact time TBA.

Calendar: Mondays 5:30–8:30 PM.

Lecture 1: March 18 – Pittsburgh
Lecture 2: March 25 – Pittsburgh
Lecture 3: April 1 – NYC
Lecture 4: April 8 – Pittsburgh
Lecture 5: April 15 – NYC
Lecture 6: April 22 – Pittsburgh
Lecture 7: April 29 – Pittsburgh

Final Exam: Friday, May 10th at TBA.

Text: The following three texts are good references for the course. Note that the first two are available for free online from the authors, and the third is available for free download from SpringerLink when you are on the CMU network.

Elements of Statistical Learning, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Springer. ISBN: 978-0-387-84857-0, `https://web.stanford.edu/~hastie/ElemStatLearn/`

Introduction to Statistical Learning, by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Springer. ISBN: 978-1-4614-7137-0, `http://www-bcf.usc.edu/~gareth/ISL/`

Applied Predictive Modeling by Max KuhnKjell Johnson. Springer. ISBN: 978-1-4614-6848-6

As we move into advanced topics, other references will be given in class.

TAs: David Zhao
Kwhangho Kim
Yufei Yi

TA office hours will be held on Canvas on Sundays at 3pm.

If ten minutes pass without any questions, then the TA session may end.

Course Grade: You are responsible for checking your scores regularly on Canvas, and reporting any errors or discrepancies to us.

Homework, and a final exam, and a project will contribute to your final course grade. Your overall homework average will count for 35%, the final exam for 45%, and the final project for 20%.

Computing: This course will rely primarily on Python for computation. We will fall back on R for topics which are not well-developed in current Python packages. The lectures and homework will assume the general proficiency at programming, simulation, and R that you have developed in your previous courses, but will assume no familiarity with Python. As in learning any new programming language, some exploration and internet searching on your part is assumed.

I would highly recommend using Jupyter notebooks for working with Python in this course, including for writing up homework assignments.

Homework: Homework will generally be due at 5:20 PM on Canvas the days of lecture. Homework will usually be available the Tuesday evening after lecture. Because we often discuss homework in detail at the beginning of lecture, late homeworks will generally **not** be accepted without prior arrangement.

The homeworks are a key component of this course, and a large amount of learning will occur during the completion of the exercises. There will be instances where the homework is used to extend material that was taught during the lecture. There will be cases where important results will be presented in the homework. These new results should be considered as being equally important to results presented during lecture.

Further, the difficulty level of the homework exercises is, on average, greater than that of the examples done during lecture. It is not a wise use of our limited lecture time to work through many extended examples, and watching me work an exercise is not a replacement for doing the work yourself.

It is your responsibility to ensure that your homework submissions are readable by the TAs and instructor. Follow the posted guidelines for formatting submissions. The TAs are not obliged to give credit for work they cannot understand, cannot open, or cannot find (e.g. if a solution is buried among pages of computer output).

Email: Please do not email TAs with concerns over grading, or any other issues with the course. Instead, email the instructor.

I assume that you check, at least daily, the email account which is linked up to Canvas (in other words, your `@andrew.cmu.edu` address). Email can be used for important announcements, so make sure that you are regularly reading mail sent this account.

Please post any questions regarding homework or the lecture to Piazza. We will do our best to be helpful and respond promptly. *However,* you should be reasonable about expectations of responses to questions on the due dates of homeworks. The time leading up a lecture tends to be the busiest time of the week for me. This is especially true if I am traveling to New York that day.

Final Exam: The final exam will be on Friday, May 10th at TBA

The final exam will be three hours in length. You will be allowed to bring one sheet of **handwritten** notes, which you have written yourself. (Both sides of an 8.5 by 11 inch sheet of paper.) **You will not be allowed to use a calculator or computer.**

If you understand everything that was covered in the lectures, and everything that appeared on the homeworks, then you are completely ready for the final exam. There is material in the texts that will not be covered in lecture or the homework; you are not responsible for this material on the final. There will be material that is covered in lecture or in the homework, but does not appear in any of the texts; you are responsible for this material on the final.

Exams will be proctored according to MSCF policy as found in the Student Handbook at `https://bit.ly/2o7Yy0W`.

Project: Details of the final project will be discussed in class.

Lectures: I expect you to attend all lectures. Also, I assume you hear everything I say during lecture. In other words, if I have to make an important announcement, and I do so during normal lecture time, I am not obligated to deliver the announcement via any other means. **In particular, I often make important announcements at the beginning of lecture, and you are responsible for being aware of these. These announcements are often on a slide titled "Announcements."**

The lecture notes will be delivered via the document camera or tablet. These will be scanned into a .pdf file and posted on the Canvas site the day following the lecture. If needed, I will clarify, expand, and/or correct the notes prior to posting them.

Prior to lecture, I will post a version of the lecture slides with blanks to be filled in. You can print a copy of this prior to lecture to follow along. **These may not be the entirety of what we will cover during lecture.**

You are encouraged to ask questions at any time during the lectures. If I feel that your question ventures into technical details which are not pertinent to the discussion, I may ask you to either talk after class, come to office hours, or send me an email. You should not interpret this as me discouraging these (or any other) questions.

Integrity: Please read CMU's policy on academic integrity:
`https://www.cmu.edu/policies/student-and-student-life/academic-integrity.html`

You are not allowed to use any materials from past versions of this course. Any such use will be considered cheating, and treated as such.

Collaboration is a tricky issue. The support and assistance of your classmates can go a long way towards helping you to understand the material. Ultimately, however, you are responsible for preparing yourself for the final exam, later courses, and your future. I encourage you to collaborate on understanding the material and your homework. However, everything you submit must be your own work. Your final writeup should be done independently and you must write your own code for computational problems. You are responsible for understanding everything that you submit. **Please ask me if you have any confusion.** Instances where students copy the work of another student will be treated as cheating.

Cheating on a homework assignment will, at a minimum, result in an automatic 0 grade for that assignment. Cheating on an exam is typically grounds for course failure. Academic integrity violations may also be reviewed by the MSCF Steering Committee.

Accommodations: If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.

Wellness: 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. 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 `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.