

46–927 Machine Learning II

Alex Reinhart

Spring 2023

Professor Alex Reinhart, areinhar@stat.cmu.edu. Please ask questions on Piazza instead of by email if possible.

Lecture MW 3:35–5:05pm Eastern; TCS 250 (Pittsburgh), BRD 507 (New York). When I visit New York, we will move the Wednesday class to Thursday 3:15–4:45pm ET. Dates will be announced in advance.

Office hours Time TBA.

Website <https://canvas.cmu.edu/>

Final exam Saturday, March 4, 10am–1pm ET

TAs Akshay Prasad and Wanshan Li

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.

1. TEXTBOOKS

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 via the CMU Libraries.

- *Elements of Statistical Learning*, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Springer. ISBN: 978-0-387-84857-0, <https://hastie.su.domains/ElemStatLearn/>
- *Introduction to Statistical Learning*, by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Springer. ISBN: 978-1-4614-7137-0, <https://statlearning.com/>

- *Applied Predictive Modeling* by Max Kuhn and Kjell Johnson. Springer. ISBN: 978-1-4614-6848-6

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

2. COMPUTING

This course will rely primarily on Python for computation. The lectures and homework will assume the general proficiency at programming, simulation, and Python that you have developed in your previous courses. As in learning any programming language, some exploration and Internet searching on your part is assumed.

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

3. COURSEWORK

3.1. Course Grade

Homework, a final exam, and a project will contribute to your final course grade. To calculate final grades, your lowest homework grade will be dropped, and then the homework average will count for 40%, the exam for 35%, and the final project for 20%. Attendance will count for the remaining 5% TODO.

3.2. Homework

Homework will generally be due at the beginning of Thursday classes. Homework submitted after the start of class, but within three hours of the deadline, will receive a 30% late penalty. Homework submitted after that point will **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). All homework will be submitted on Gradescope as PDF files.

3.3. Final Exam

The final exam will be on the date set by the MSCF program.

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.** Exams will be proctored according to MSCF policy as found in the Student Handbook at <https://bit.ly/3bUhjBK>.

If you understand everything that was covered in the lectures, and everything that appeared on the homework, 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.

3.4. Project

Students will complete a project applying machine learning to financial data. Projects will be completed in groups of 3-4 students, and will culminate in a written report submitted during the last week of class. During the second half of the mini, homework assignments will become shorter to allow time for the project.

Further details of the project will be discussed in class.

3.5. Regrade Requests

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

If you believe your assignment was graded incorrectly, first check the solutions posted on Canvas. If you believe the grade is in error, you can make a regrade request on Gradescope. Include a detailed explanation of what specific part of the grading you believe is in error, referring to the solutions as needed. Regrade requests must be made **within one week** of when grades are posted on Gradescope.

4. EMAIL

Please do not email TAs with concerns over grading, or any other issues with the course. Instead, make a private post on Piazza or 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). Canvas will be used for important announcements, so make sure that you receive notifications for Canvas announcements.

Please post any **short** questions regarding homework or the lecture to Piazza. (Longer or more detailed questions should be asked in office hours.) *However*, you should be reasonable about expectations of responses to questions on the due dates of homeworks. For quick responses, office hours are preferred to Piazza.

5. LECTURES

I expect you to attend all lectures. Attendance will be taken and is part of your grade (see Course Grade section above). 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 a tablet. These will be posted as a PDF file 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 or download 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.

6. ACADEMIC 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.

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 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.

Similarly, on projects and reports, reading papers and online resources can be valuable to aid your understanding. However, you are responsible for drafting your report, which must demonstrate your understanding and the work you completed. **I expect your reports to be in your own words**, except for quotes given in quotation marks, and all sources of text, ideas, figures, and results must be clearly cited. **It should always be clear what work is yours** and what work was done by others. Using material from outside sources without **clear** attribution will be considered plagiarism. You should ask us for help if you are unsure what attribution would be appropriate.

Cheating on a homework assignment or quiz will, at a minimum, result in an automatic o grade for that assignment. Cheating or plagiarism on a project is typically grounds for course failure. All academic integrity violations will be reported to the Office of Community Standards and Integrity and may be reviewed by the MSCF Steering Committee.

7. 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.

8. 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.

9. DIVERSITY AND INCLUSION

We must treat every individual with respect. We are diverse in many ways, and this diversity is fundamental to building and maintaining an equitable and inclusive campus community. Diversity can refer to multiple ways that we identify ourselves, including but not limited to race, color, national origin, language, sex, disability, age, sexual orientation, gender identity, religion, creed, ancestry, belief, veteran status, or genetic information. Each of these diverse identities, along with many others not mentioned here, shape the perspectives our students, faculty, and staff bring to our campus. We, at CMU, will work to promote diversity, equity and inclusion not only because diversity fuels excellence and innovation, but because we want to pursue justice. We acknowledge our imperfections while we also fully commit to the work, inside and outside of our classrooms, of building and sustaining a campus community that increasingly embraces these core values.

Each of us is responsible for creating a safer, more inclusive environment.

Unfortunately, incidents of bias or discrimination do occur, whether intentional or unintentional. They contribute to creating an unwelcoming environment for individuals and groups at the university. Therefore, the university encourages anyone who experiences or observes unfair or hostile treatment on the basis of identity to speak out for justice and

support, within the moment of the incident or after the incident has passed. Anyone can share these experiences using the following resources:

- Center for Student Diversity and Inclusion: csdi@andrew.cmu.edu, (412) 268-2150
- Report-It online anonymous reporting platform: reportit.net username: tartans password: plaid

All reports will be documented and deliberated to determine if there should be any following actions. Regardless of incident type, the university will use all shared experiences to transform our campus climate to be more equitable and just.