[Math 218] Introduction to Statistical Learning

Fall 2022

Section:	MATH 0218A	**Professor:**	Becky Tang (she/her)
Class hours:	MWF 9:05-9:55	**Office:**	Warner 214
Room:	Warner 104	**E-mail:**	btang@middlebury.edu
Prereqs:	MATH 118	**Office Hours:**	W 3-5pm + TBD

Course Description

This course is an introduction to modern statistical, machine learning, and computational methods to analyze large and complex data sets that arise in a variety of fields, from biology to economics to astrophysics. The theoretical underpinnings of the most important modeling and predictive methods will be covered, including regression, classification, clustering, resampling, and tree-based methods. Student work will involve implementation of these concepts using open-source computational tools. Additional topics will be taught if time permits. Computing is done in R. The final project will give students an opportunity to apply some of these methods to a dataset of their choosing.

Key Learning Outcomes

- ❑ Introduce fundamental tools for building predictive models, including some "state-of-the-art" methods in data science
- ❑ Understand the role of model assessment, comparison, and selection
- ❑ Learn how to use the vast collection of tools in R to implement the methods learned
- ❑ Learn to effectively communicate results through written assignments and final project presentation
- ❑ Have fun learning new methods

Textbooks and Course Materials

Online Resources: Content (slides, homework, schedule) will be housed mainly on the course website: https://math218-fall2022.github.io/website/. I encourage you to bookmark this website.

textbook: There is no required textbook for this course. However, the following may be useful:

- *An Introduction to Statistical Learning with Applications in R, Second Edition*, J. Gareth, et. al.

Class Expectations

- **You are expected to physically show up to class and actively participate**, conditional on classes being in-person. You are an integral part of the class community! Exceptions include previously-communicated illness or planned absence.
- **Please arrive on time.** I expect everyone, myself included, to arrive on time and dedicate full attention during the class. In turn, I will do my best to always end class at the designated time.
- **Use of laptops will be often be necessary during the lectures and labs.** However, please be respectful of both the professor and your peers. Please ensure that your laptop has sufficient battery for the duration of the class.
- **Cell phones should be turned to silent**. I don't mind cell phones in class, but please silence them so as to not disrupt the class.
- **Please ask questions!**
- **I encourage discussion amongst yourselves**, especially for clarification or help! However, please be mindful of volume so that the conversations will not be disruptive to the class.

Resources

- ❑ **Office hours**. This time is meant for you! Please come by to ask questions, chat with me, or work on homework. You should never worry about disturbing me during this time. The TA's office hours are also a wonderful resource.
- ❑ **One-on-one meetings**. If you would like to meet with me one-on-one, please send me an e-mail or approach me after class so we can schedule a time.
- ❑ **Campuswire**: Online forum where you can ask and answer questions. You may do so anonymously, to the entire class, or just for me to see. Please utilize this resource if you want some clarification regarding anything *except* the midterm. However, I ask that you do not post questions that share solutions or partial solutions. When in doubt, please ask a question privately ("to the instructor").
- ❑ **Your peers**. I encourage students to work together and discuss material! However, unless the assignment explicitly states that it is to be completed as group work, the submitted material must be your own.

College policies and resources

Academic Integrity

As an academic community devoted to the life of the mind, Middlebury requires every student to reflect complete intellectual honesty in the preparation and submission of all academic work. Details of our Academic Honesty, Honor Code, and Related Disciplinary Policies are available in Middlebury's handbook.

Honor Code Pledge

The Honor Code pledge reads as follows: "I have neither given nor received unauthorized aid on this assignment." It is the responsibility of the student to write out in full, adhere to, and sign the Honor Code pledge on all examinations, research papers, and laboratory reports. Faculty members reserve the right to require the signed Honor Code pledge on other kinds of academic work.

Disability Access and Accommodation

Students who have Letters of Accommodation in this class are encouraged to contact me as early in the semester as possible to ensure that such accommodations are implemented in a timely fashion. For those without Letters of Accommodation, assistance is available to eligible students through the Disability Resource Center (DRC). https://www.middlebury.edu/office/disability-resource-center. The DRC provides support for students with disabilities and facilitates the accommodations process by helping students understand the resources and options available and by helping faculty understand how to increase access and full participation in courses. DRC services are free to all students. Please contact ADA Coordinators Jodi Litchfield and Peter Ploegman of the DRC at ada@middlebury.edu for more information. All discussions will remain confidential.

Center for Teaching, Learning, and Research (CTLR)

The CTLR provides academic support for students in many specific content areas and in writing across the curriculum through both professional and peer tutors. The Center is also the place where students can find assistance in time management and study skills. These services are free to all students. go.middlebury.edu/connect

Grading

- Unless otherwise noted, assignments should be turned in on Canvas.
- Late work will always be considered within one week of the original due date. Unless otherwise stated, the late policy is as follows: for every 24-hour period the assignment is late, 10% from the maximum possible grade will be deducted.
- I will do my best to return assignments within one week of submission.
- Regrade requests: I do allow regrade requests, which must be submitted within one week of when the assignment is returned. These should be entered as a "Comment" on the corresponding assignment in Canvas, after which you must notify me via e-mail that you have a regrade request. Keep in mind that regrade requests do not guarantee points back, and may result in a lower score if I see that I incorrectly marked something as correct.

Component	Percentage
Homework	30%
Labs	20%
Midterm	20%
Final Project	25%
Participation	5%

Types of assignments

- **Problem**. Assigned weekly and typically completed individually. In homework, you will apply what you have learned during lecture to analyze output from computational analyses, as well as dive a bit more into the theory and nuances behind the methods. While there may be computational exercises in homework, implementation of methods will largely be relegated to labs (see below).
- **Lab**. Assigned mostly weekly, labs are an opportunity for you to implement the methods learned in class in R. Labs are usually started in class, with the expectation that they can be finished within class time or with minimal time outside of lecture. *Lowest score will be dropped.*
- **Midterm exam**. One take-home midterm exam designed as an opportunity to assess the knowledge and skills you've learned. The midterm exam will include analysis and computational tasks related to the content discussed in lectures, labs, and homework assignments.
- **Final project**. The purpose of the project is to apply what you have learned throughout the semester to analyze an interesting data-based research question of your choosing. The final project will include a written component as well as an in-person class presentation. The presentations are scheduled to take place during the last week of classes, though there will be opportunity to revise the written portion up until the end of finals. **You must present the final project to pass the course.**
- **Participation**. I expect you to be present and active learners in this course! Throughout the semester, there will be various small assignments that will count towards participation (e.g. reflection exercises, giving constructive feedback).

Some weeks, both a lab and homework will be assigned. Other weeks, just one of the two. The following table displays the typical combinations of release and due dates. For both homework and lab, your completed work should be turned in on Canvas as a PDF.

Option	Assignment	Released	Due
1	Lab Homework	Friday (in-class) Friday 12:00	Sunday 23:59 Thursday 23:59
2	Lab only	Friday (in-class)	Sunday 23:59
3	Homework only	Friday 12:00	Thursday 23:59

Tentative Course Content

Week	Date	Topic
1	9/12	M - Welcome! W - What is Statistical Learning? F - Review of R and Linear Algebra - **Lab:** activities covering discussed topics. - HW 01 assigned Su - Lab due at 11:59pm
2	9/19	M - Linear regression W - Linear regression (cont.) R - HW 01 due at 11:59pm F - **Lab:** activities covering discussed topics. - HW 02 assigned Su - Lab due at 11:59pm
3	9/26	M - Classification W - Classification (cont.) R - HW 02 due at 11:59pm F - **Lab:** activities covering discussed topics. - HW 03 assigned
4	10/3	M - Resampling W - Resampling (cont.) R - HW 03 due at 11:59pm F - **Lab:** activities covering discussed topics.

5	10/10	M - Linear model selection - Introduce final project W - Linear model selection (cont.) F - No class. Please complete mid-semester feedback form.
6	10/17	M - Midterm I W - Shrinkage F - **Lab:** activities covering discussed topics. - HW 04 assigned
7	10/24	M - Tree-based methods W - Tree-based methods (cont.) R - HW 04 due at 11:59pm F - **Lab:** activities covering discussed topics. - HW 05 assigned
8	10/31	M - Unsupervised learning W - Unsupervised learning (cont.) R - HW 05 due at 11:59pm F - **Lab:** activities covering discussed topics. - HW 06 assigned
9	11/7	M - Survival analysis W - Survival analysis (cont.) R - HW 06 due at 11:59pm F - **Lab:** activities covering discussed topics. - HW 07 assigned
10	11/14	M - Project work day W - Support Vector Machines R - HW 07 due at 11:59pm F - SVMs (cont.) - Project proposals due at 11:59pm
-	11/21	Thanksgiving break
11	11/28	M - Project work day W - Additional topic F - Additional topic
12	12/5	M + W - Case study R - Final project due 12:00pm

		F - Project presentations
	12/12	M - Project presentations (cont.)

NOTE: The above dates and content may be modified due to the requirements of the class. Also, the indicated dates of course content and homework may be moved backward or forward depending on class progress and my conference travel. **Midterm and final project dates will not change**.