[Math 218] Introduction to Statistical Learning

Fall 2022

Section:	MATH 0218A	**Office:**	Warner 214
Class hours:	MWF 9:05-9:55	**E-mail:**	btang@middlebury.edu
Room:	Warner 104	**Office Hours:**	W 15:00-17:00
Prereqs:	MATH 118		F 12:00-13:00 or Su 18:00-20:00?

Professor: Becky Tang

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.

Topics covered will include: linear regression, logistic regression and linear discriminant analysis; cross-validation and bootstrap, model selection and regularization methods (ridge and lasso); tree-based methods, random forests and boosting; principal components and clustering (k-means and hierarchical). 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
- ❏ Have fun learning new methods

Textbooks and Course Materials

Online Resources: Lecture notes and slides will be provided to students throughout the semester.

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 Policy and Rules of Conduct

- **Students are expected to physically show up to class and actively participate**, conditional on classes being in-person. Exceptions include previously-communicated illness or planned absence.
- **Please arrive on time.** I expect students 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 fellow students. 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 between students**, especially for clarification or help! However, please be mindful of volume so that the conversations will not be disruptive to the class.

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). 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.

Disability Resource Center (DRC)

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.

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.
- 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.

Component	Percentage
Homework	35%
Labs	15%
Midterm 1	20%
Final Project	25%
Participation	5%

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.

Option	**Assignment**	**Released**	**Due**
1	Lab	Friday (in-class)	Sunday 23:59
	Homework	Monday 12:00	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 - What is Statistical Learning? W - Review of R and Linear Algebra F - **Lab:** activities covering discussed topics.
2	9/19	M - Linear regression W - Linear regression (cont.) F - **Lab:** activities covering discussed topics.
3	9/26	M - Classification - HW 1 assigned W - Classification (cont.) F - **Lab:** activities covering discussed topics.
4	10/3	M - Resampling - HW 2 assigned W - Resampling (cont.) F - **Lab:** activities covering discussed topics.
5	10/10	M - Review for Midterm I - Introduce final project W - Midterm I F - No class. Please complete mid-semester feedback form.
6	10/17	M - Linear model selection + Shrinkage - HW 3 assigned W - Linear model selection + Shrinkage (cont.) F - **Lab:** activities covering discussed topics.
7	10/24	M - Tree-based methods - HW 4 assigned W - Tree-based methods (cont.) F - **Lab:** activities covering discussed topics.
8	10/31	M - Support vector machines - HW 5 assigned W- SVMs (cont.) F - **Lab:** activities covering discussed topics.
9	11/7	M - Survival analysis - HW 6 assigned

		W - Survival analysis (cont.) F - **Lab:** activities covering discussed topics.
10	11/14	M - Unsupervised learning - HW 7 assigned W - Unsupervised learning (cont.) F - **Lab:** activities covering discussed topics.
-	11/21	Thanksgiving break
11	11/28	M - Project work day W - Additional topic F - Additional topic (cont.)
12	12/5	M + W - Work on final project F - Project presentations
	12/12	M - Project presentations (cont.)

NOTE: The above actual dates 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**.