STAT 253: Statistical Machine Learning

Macalester College, Spring 2023

Section 01: TTh 9:40-11:10am, OLRI 241
Section 02: TTh 1:20-2:50 pm, THEATR 002

Your Instructor

James Normington (he/him)

Website:
jamesnormington.github.io
Email:
jnorming@macalester.edu

Office Hours
- TTh: 3-4pm
- Th: 12:15pm - 1:15pm
- Location: Olin-Rice 230 or by Zoom (with prior scheduling)
- By appointment – please don't hesitate to use this option!

Course Description

This class is an introduction to the exciting world of statistical machine learning. Broadly, statistical machine learning consists of *tools and algorithms to learn from data*. Insights from machine learning help us take action by enabling us to *make predictions* and *understand uncertainty*. Applications of machine learning algorithms are used everywhere from *finance, biology, medicine, social sciences, language, and the humanities*. We will build on the linear and logistic regression modeling techniques covered in STAT 155 to understand tools of *regression and classification* more broadly. Where in 155 we fit models to make inferences about the population, in 253 we will train models to be used for prediction. We will also learn about *unsupervised methods* that can help you find underlying structure in data. Along the way, you will *gain computing skills in R* that will allow you to use these tools in any profession.

Objectives

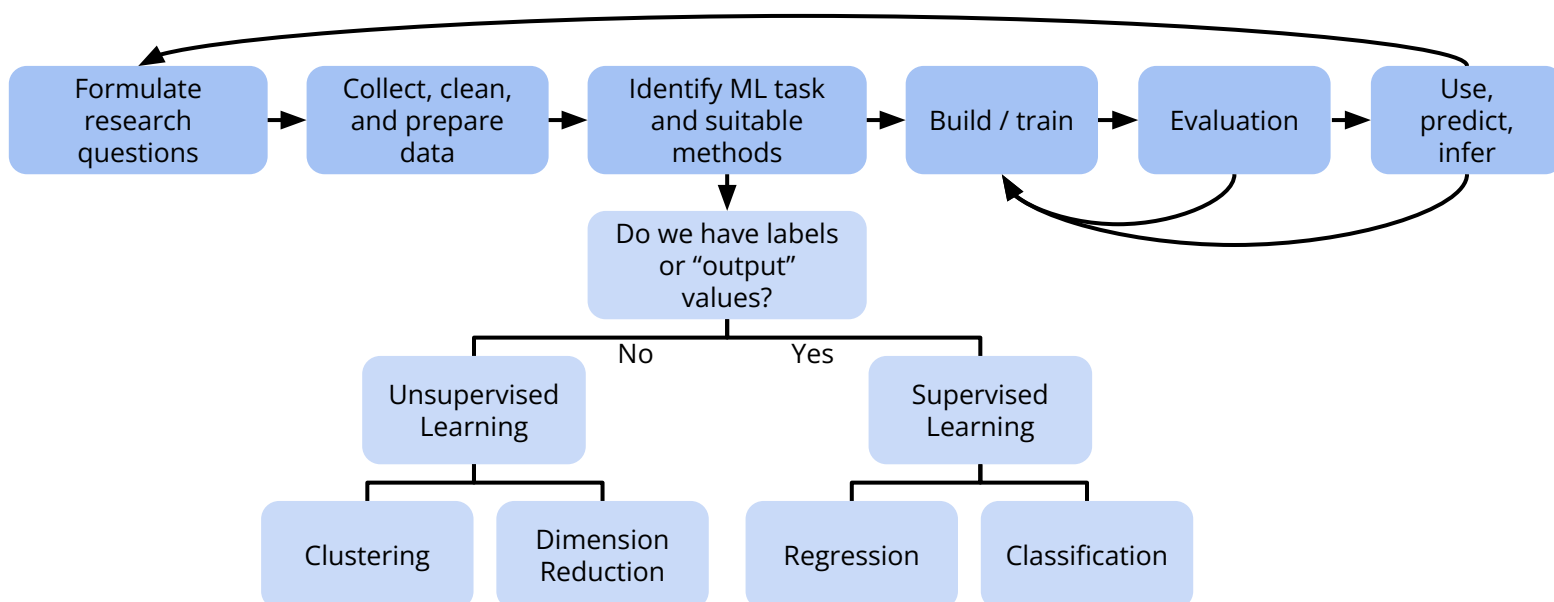**Algorithm Choice and Evaluation**   **R Programming**   **Communication**   **Collaboration**   **Machine Learning and Society**

Assignments and Feedback

The 4 pieces below form the core means of active practice and feedback.

Portfolio: Over the course of the module, you will build up a personal, organized synthesis of your learning in the course. Weekly homework assignments will prompt you to clearly communicate key concepts from class in an organized way in your portfolio. See this grading spreadsheet for the set of concepts covered.

Each week, you will receive (provisional) grades for the current week's concepts and for any revisions made. You may revise sections as many times as you wish throughout the module to improve your grades, but grading and feedback are only possible for the instructional team once per week.

The goal is that this portfolio serves as a reference for you in the future and/or provides fodder for building up an online presence (e.g., personal website or portfolio).

Project: You will use our course tools to perform a substantive analysis of a dataset of your choice. You will make steady progress on this analysis through weekly homework assignments and receive feedback to help finalize your project in the last ~1.5 weeks of class. You can work individually or in groups.

The goal is for the project to give you practice leading and communicating an applied analysis in a context that you might encounter in your future work.

Ethics in ML: Craft concise, thoughtful responses to all 5 weekly homework prompts on ethical considerations when using machine learning tools in practice.

Quizzes: There will be 5-8 short quizzes designed to test your understanding of the material. Quizzes will be open-note and completed in class.

Final Grade Determination

Your **base grade** will be determined by your portfolio, project, and course engagement:

- A-level: All B's and 90% A's on portfolio concepts; A-level project; Meets Expectations on Ethics in ML
- B-level: All C's and 90% B's on portfolio concepts; B-level project; Meets Expectations on Ethics in ML
- C-level: All D's and 90% C's on portfolio concepts; C-level project; Meets Expectations on Ethics in ML
- D-level: 90% D's on portfolio concepts; A-level project; Meets Expectations on Ethics in ML

Your base grade will be **modified** by your overall **quiz average**, as indicated in the table below:

	Base grade			
Quiz average	**A**	**B**	**C**	**D**
[90, 100]	A	B+	C+	C-
[80, 90)	A-	B	C+	C-
[70, 80)	B+	B-	C	D+
< 70	B	C+	C-	D

Academic Integrity

As a globally- and community-oriented institution, Macalester College expects respectful exchange of ideas. Students are expected to be familiar with the college standards on academic integrity (**website**). I encourage you to work with your classmates to discuss material and ideas for your assignments, but in order for you to receive individualized feedback on your own learning, ***all submitted work (including code!) must be written in your own words.***

When you present someone else's work as your own, you are missing a learning opportunity and cheapening the value of your education. If you are tempted to plagiarize someone's work, please consider requesting an extension instead. See below my policy on extensions.

Extensions and Class Attendance

Extensions. Life happens! There are circumstances outside of the classroom that make it difficult to focus on schoolwork. These circumstances are often entirely personal; thus, I will grant most extensions, without the need to specify a reason, when requested in advance of the corresponding deadline. My intent is that students can focus on their physical, mental, personal, and interpersonal wellness so there are fewer obstacles to learning. However, I reserve the right to deny an extension if I feel the student is abusing this policy.

Final grades are due May. 11th; any unsubmitted work, regardless of circumstances, will count as 0 unless waived by the instructor. If your circumstances make it likely that you can't submit a sizeable portion of classwork by then, you should talk with me about alternative arrangements. For deadlines on adding/dropping courses, grading options, and withdrawal, see the Spring 2023 academic calendar here.

Attendance. Attendance is expected, but not mandatory. If you are to miss class, please try to inform me in advance. Your reason for missing class does not need to be included in your message. I reserve the right to adjust the student's final grade if the student makes a habit of missing class without informing me in advance, or if I suspect a student is abusing this policy.

Getting help in STAT 253

Preceptors

We have four preceptors that will be helping out with STAT 253 this semester! They will be providing feedback on your homeworks and holding office hours.

- Ayaa Asoba: aasoba@macalester.edu
- Xiyu (Rita) Li: xli1@macalester.edu
  - Tue 3-5pm, Thurs 3:30-5pm, Library 1st floor
- Kristy Ma (she/her): yma1@macalester.edu
  - Mon & Wed 2:30-4:30pm, Kagin Commons 1st floor
- Wenxuan Zhu: wzhu@macalester.edu
  - Wed & Fri 9-10:30am, Library 1st floor

James' Office Hours
- TTh: 3-4pm
- Th: 12:15pm - 1:15pm
- Location: Olin-Rice 230 or by Zoom (with prior scheduling)
- By appointment – please don't hesitate to use this option!

Additionally, you can always ask me questions over email. Don't be shy!

Course Structure

This class will operate under a "flipped classroom" model. A typical class will proceed as such:

Before Class.

- Complete assigned readings/videos.
- Take notes
- Ask questions!

During Class.

- Short lecture (no more than 30 minutes)
- Group activities
  - **Please bring your laptop to class every day.**
- **Ask questions!!**

After Class.

- Work on HW assignments
- Reflect on in-class learnings
- ***Ask questions!!!!!***

My Teaching Philosophy

Active learning. Statistics is an active discipline. The collection, visualization, analysis, and interpretation of data are all active steps in the statistical modeling process. Because of this, a large proportion of class time will be spent actively doing statistics problems with practical, real-world data.

Statistical thinking over statistical procedure. The rote memorization of manual statistical procedures (i) is unnecessary with modern computing, (ii) requires mathematical prerequisites beyond the scope of this course, and (iii) presents statistical techniques as disparate, non-intuitive procedures. Statistical modeling is best learned through intuitively solving real problems, and leaving the tedium of the procedure up to computers (or, for the inclined, preparation for graduate school exams!).

Diversity, equity, and inclusion. Statistics is, in a word, the study of *variability*. The best statistical models acknowledge and account for all sources of variability, recognizing the diverse and multivariate world we live in. There will be absolutely no discrimination based on race, ethnicity, gender identity, sexual orientation, or any other identities, life experiences, or opinions. If at any point you feel discriminated against, please either reach out to me or find other support at Macalester.

Statistical Software

Computing is an integral part of modern statistics and will thus be a main focus of our course. We will use the R programming language. RStudio, a graphical user interface (GUI) for R, will facilitate our use of R and the creation of R Markdown reports that you will use for submitting assignments. R and RStudio are *free* to use and download.

When you learn a new computing language, you will stumble at the beginning. That is normal! The key is to learn how to overcome those stumbling blocks (or coding errors). Read the error message. Double check the spelling. Consult Google. Post the code and error message on Slack.

Learning machine learning with R/RStudio will take time and effort, but by the end of this semester you'll be leaving this course with a *very marketable and useful skill* that will serve you well in future courses, jobs, and beyond!

Advice for Success in COMP/STAT 112

1. **Actively engage with the course content.** Pay attention in lectures, actively participate during in-class activities, understanding and incorporate my feedback into your Final Project. For Quizzes, do practice problems and understand their solution! If you run out of practice problems, reach out to me.
2. **Ask when you have questions.** Ask during lecture (interrupt me at any time; I'm serious!), come to office hours, visit the preceptors and tutors, ask your classmates, post on Slack. Our best learning comes in the vulnerable moments when we admit we don't understand.

COVID-19 Protocols and Expectations

We all deserve a safe learning environment. Please stay current with the latest COVID-19 guidelines: Shared Community Commitment (Mac Stays Safer 3.0). I will abide with these guidelines very closely.

Some key notes, and my specific expectations:

Vaccination Policy. Macalester requires all employees and students to have received a COVID-19 vaccine and booster, or to have an approved exemption in place. This is not negotiable. If you are eligible for a 2nd booster, I highly recommend it; see more information at the CDC website.

Mask Policy. There will be no requirement to wear a mask in class. However, I will wear a mask in office hours and ask that you do too. During in-class activities and Final Project working sessions, groups should feel free to come up with their own policies (so long as they fit within Macalester's and the State of Minnesota's policies) If there is a disagreement which is harming the group's ability to make progress, please see me. For more information, see the Well-Fitted Mask Policy.

Be Safe, Be Respectful. My primary concern is your well-being. If I or a classmate respectfully asks you to cover your sneeze, wash your hands, etc. – please do so. Your classmates may be immunocompromised, or live/work/relate with someone who is. They may also just not want to get sick! Every student has the right to feel safe attending my lectures in person. We each have varying levels of "COVID cautiousness", so please use both common sense and courtesy for others in my classroom and on campus.

Test Frequently. If you feel symptoms associated with COVID-19, **please** get tested. Please see Shared Community Commitment (Mac Stays Safer 3.0) for information on free, accessible testing.

Stay Home when Sick. If you have COVID-19, or any other contagious disease, **please** don't come to class. Email me and we will work out alternate arrangements.

Health and Wellbeing

Your health and well-being should be prioritized over my course. A student cannot effectively learn statistics if they are under physical, mental, emotional, societal, or financial stress. During class time, eat when you are hungry, drink water when you are thirsty, use the restroom, and step away/out if you are upset or need some air. Please do what is necessary so long as it does not impede your or others' ability to be mentally and emotionally present in the course.

Physical Health. Physical exercise is associated with better grades, better physical & mental health, and an improved quality of life. I highly recommend using the Leonard Center to stay physically active. Participation in intramural sports is another great way to exercise, reduce stress, and develop a sense of community on campus.

Mental Health. You can find a list of mental health resources here.

Financial Health. If you have had an unanticipated financial emergency, please consider using the Emergency Aid Program.

Beyond these resources, if you are having difficulties maintaining your well-being, please contact me. I am invested in your full learning journey, which cannot thrive unless you are OK.

The Environment You Deserve

I am committed to helping you learn and succeed in this course and in your time at Macalester. Enabling this means facilitating environments that support you in the ways that you need.

Respect. There will be *no tolerance for discrimination* based on race, ethnicity, gender identity, religion, sexual orientation, disability, and other identities or life experiences, in and out of the classroom. If you see any discrimination whatsoever, please contact me or find other support at Macalester, and be ensured the matter will be dealt with seriously.

Sensitive Topics. Statistics is used to inform policy, research, and understanding across all aspects of our society. The examples and data we use in class will reflect this. Thus, at times our work might touch on sensitive topics. I will try to give warnings for topics that I perceive may be sensitive. Please let me know if, for any reason, you feel that you are unable to participate in the planned activities due to the topic and we can discuss alternatives.

Accommodations. I am committed to creating an accessible and inclusive classroom for all students, including those with disabilities. If you are in need of any accommodations, please contact Disability Services (visit their [website](#), email [disabilityservices@macalester.edu](#), or call 651-696-6275) to schedule an appointment and discuss your needs. Once you've met with Disability Services, please then set a time to meet with me to discuss your accommodation plan for this course. It is important to *arrange this meeting as early in the semester as possible* (ideally within the first week), in order to ensure that your accommodations can be implemented early on. It is your responsibility to make sure you are registered with Disability Services. If you wait until later in the course, I may not be able to accommodate you retroactively.

Religious Observance. Students may wish to take part in religious observances that occur during this semester. I've done my best to schedule deadlines around major holidays, but if you have a religious observance/practice that conflicts with a scheduled class meeting or assignment deadline, *please contact me at least one week in advance* so we can discuss appropriate accommodations.

Title IX. If you or anyone you know has experienced harassment or discrimination on the basis of sex or gender, know that you are not alone. Macalester provides staff and resources to help and support you. More information is available on the [Title IX website](#).

Please be aware that *all Macalester faculty and preceptors are mandatory reporters*, which means that if we become aware of incidents or allegations of sexual misconduct, we are required to share the matter with the Title IX Coordinator. Although we have to make that notification, you control how your case is handled, including whether or not you wish to pursue a formal complaint. If you would like to speak to someone *confidentially*, contact the Hamre Center (651-696-6275), Chaplain staff (651-696-6298), or other local and national resources listed [here](#).