

Course Syllabus

Instructor

John Curtin

- Office hours: Thursdays, 1-2 pm or by appointment in room 326
- Email: jjcurtin@wisc.edu (but please use Slack DM or channel posts for all course communications during this semester)

Teaching Assistants

Michelle Marji

- Office hours: INSERT HOURS
- Email: michelle.marji@wisc.edu

Kendra Wyant

- Office hours: INSERT HOURS
- Email: kpaquette2@wisc.edu

Course Website

https://jjcurtin.github.io/book_iaml/

Communications

All course communications will occur in the course's Slack workspace (<https://iaml-2024.slack.com/>). You should have received an invitation to join the workspace. If you have difficulty joining, please contact me by my email above. The TAs and I will respond to all Slack messages within 1 business day (and often much quicker). Please plan accordingly (e.g., weekend messages may not receive a response until Monday). For general questions about class, coding assignments, etc., please post the question to the appropriate public channel. If you have the question, you are probably not alone. For issues relevant only to you (e.g., class absences, accommodations, etc.), you can send a direct message in Slack to me. However, I may share the DM with the TAs unless you request otherwise. In general, we prefer that all course communication occur within Slack rather than by email so that it is centralized in one location.

Meeting Times

The scheduled course meeting times are Tuesdays and Thursdays from 11:00 - 12:15 pm. Tuesdays are generally used by the TAs to discuss application issues from the homework or in the course more generally. Thursdays are generally led by John and used to discuss topics from the video lectures and readings.

All required videos, readings, and application assignments are described on the course website at the beginning of each unit.

Course Description

This course is designed to introduce students to a variety of computational approaches in machine learning. The course is designed with two key foci. First, students will focus on the application of common, “out-of-the-box” statistical learning algorithms that have good performance and are implemented in tidymodels in R. Second, students will focus on the application of these approaches in the context of common questions in behavioral science in academia and industry.

Requisites

Students are required to have completed Psychology 610 with a grade of B or better or a comparable course with my consent.

Learning Outcomes

- Students will develop and refine best practices for data wrangling, general programming, and analysis in R.
- Students will distinguish among a variety of machine learning settings: supervised learning vs. unsupervised learning, regression vs. classification
- Students will be able to implement a broad toolbox of well-supported machine-learning methods: decision trees, nearest neighbor, general and generalized linear models, penalized models including ridge, lasso, and elastic-nets, neural nets, random forests.
- Students will develop expertise with common feature extraction techniques for quantitative and categorical predictors.
- Students will be able to use natural language processing approaches to extract meaningful features from text data.
- Students will know how to characterize how well their regression and classification models perform and they will employ appropriate methodology for evaluating their: cross validation, ROC and PR curves, hypothesis testing.
- Students will learn to apply their skills to common learning problems in psychology and behavioral sciences more generally.

Course Topics

- Overview of Machine Learning Concepts and Uses
- Data wrangling in R using tidyverse and tidymodels
- Iterations, functions, simulations in R
- Regression models
- Classification models
- Model performance metrics
- ROCs
- Cross validation and other resampling methods
- Model selection and regularization
- Approaches to parallel processing

- Feature engineering techniques
- Natural language processing
- Tree based methods
- Bagging and boosting
- Neural networks
- Dimensionality reduction and feature selection
- Explanatory methods including variable importance, partial dependence plots, etc
- Ethics and privacy issues

Schedule

The course is organized around 14 weeks of academic instruction covering the following topics:

1. Introduction to course and machine learning
2. Exploratory data analysis
3. Regression models
4. Classification models
5. Resampling methods for model selection and evaluation
6. Regularization and penalized models
7. Midterm exam/project
8. Model comparisons
9. Advanced performance metrics
10. Advanced models: Random forests and ensemble methods (bagging, boosting, stacking)
11. Advanced models: Neural networks
12. Natural Language Processing I: Text processing and feature engineering
13. Applications
14. Ethics

- The final exam is during exam week on Tuesday May 7th from 11 - 12:15 in our normal classroom.
- The final project is due during exam week on Wednesday May 8th at 8 pm.

Required Textbooks and Software

All required textbooks are freely available online (though hard copies can also be purchased if desired). There are eight required textbooks for the course. The primary text for which we will read many chapters is:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. [-@ISL]. An Introduction to Statistical Learning: With Applications in R (2021; 2nd Edition). ([website](#))

We will also read sections to chapters in each of the following texts:

- Wickham, H. & Grolemund, G. [-@RDS]. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (1st ed.). Sebastopol, CA: O'Reilly Media, Inc. ([website](#))
- Hvitfeldt, E. & Silge, J.. [-@SMLTAR] Supervised Machine Learning for Text Analysis in R ([website](#))
- Kuhn, M. & Johnson, K. [-@APM]. Applied Predictive Modeling. New York, NY: Springer Science. ([pdf](#))
- Kuhn, M., & Johnson, K. [-@FES]. Feature Engineering and Selection: A Practical Approach for Predictive Models (1 edition). Boca Raton, FL: Chapman and Hall/CRC. ([website](#))

- Kuhn, M. & Silge, J. [-@TID] Tidy Modeling with R. ([website](#))
- Molnar, C. [-@IML] Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. [website](#)
- Silge, J., & Robinson, D. [-@TMR]. Text Mining with R: A Tidy Approach (1st ed.). Beijing; Boston: O'Reilly Media. ([website](#))
- Wickham, H. [-@TSG]. The Tidy Style Guide. ([website](#))

Additional articles will be assigned and provided by pdf through the course website.

All data processing and analysis will be accomplished using R (and we recommend the RStudio IDE). R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.

Grading

- Quizzes (13 anticipated): 15%
- Application assignments (11 anticipated): 25%
- Midterm application exam: 15%
- Midterm conceptual exam: 15%
- Final application exam: 15%
- Final conceptual exam: 15%

Final letter grades may be curved upward, but a minimum guarantee is made of an A for 93 or above, AB for 88 - 92, B for 83 - 87, BC for 78 - 82, C for 70 - 77, D for 60-69, and F for < 60.

Exams, Application Assignments and Quizzes

- The midterm application exam will be due during the 7th week of the course on Wednesday, March 6th at 8 pm.
- The midterm conceptual exam will be administered during class on Thursday, March 7th.
- The final exam is during exam week on Tuesday May 7th from 11 - 12:15 in our normal classroom.
- The final project is due during exam week on Wednesday May 8th at 8 pm.
- Approximately weekly quizzes will be administered through Canvas and due each Wednesday at 8 pm
- Approximately weekly application assignments will be submitted via Canvas and due each Wednesday at 8 pm.

Application Assignments

The approximately weekly application assignments are due on Wednesdays at 8 pm through Canvas. These assignments are to be done individually. Please do not share answers or code. You are also encouraged to make use of online resources (e.g., stack overflow) for assistance. All assignments will be completed using R markdown to provide both the code and documentation as might be provided to your mentor or employer to fully describe your solution. Late assignments are not accepted because problem solutions are provided immediately after the due date. Application assignments are graded on a three-point scale (0 = not completed, 1 = partially completed and/or with many errors, 2 = fully completed and at least mostly correct). Grades for each assignment will be posted by the following Monday at the latest.

ChatGPT

I suspect you have all seen discussions of all that ChatGPT can do by now and its impact on teaching and assessment. I believe that AI like ChatGPT will eventually become an incredible tool for data scientists and programmers. As such, I view these advances with excitement. Of course, I don't plan to assign a grade to ChatGPT so I want to make sure that we are clear on when you can and when you cannot use it. Given that I expect AI like ChatGPT to become a useful tool in our workflow as professionals, now is the time to start to learn how it can help. Therefore, you are free to use it during any of our **application assignments** AND the **application questions on the mid-term and final exams**. Code from ChatGPT is unlikely to be sufficient in either context (and my testing suggests it can be flat out wrong in some instances!) but I suspect that it will still be useful. In contrast, you CAN NOT use ChatGPT to answer the conceptual questions on the two exams or the weekly quizzes. Those questions are designed to assess your working knowledge about concepts and best practices. That information must be in YOUR head. I am not yet certain how I will administer that part of those exams (in class vs. HonorLock), but I wanted to be 100% clear that use of ChatGPT to answer those questions will be considered cheating and handled as such if detected. There will be a zero tolerance policy for such cheating. It will be reported to the Dean of Students on first offense.

Student Ethics

The members of the faculty of the Department of Psychology at UW-Madison uphold the highest ethical standards of teaching and research. They expect their students to uphold the same standards of ethical conduct. By registering for this course, you are implicitly agreeing to conduct yourself with the utmost integrity throughout the semester.

In the Department of Psychology, acts of academic misconduct are taken very seriously. Such acts diminish the educational experience for all involved – students who commit the acts, classmates who would never consider engaging in such behaviors, and instructors. Academic misconduct includes, but is not limited to, cheating on assignments and exams, stealing exams, sabotaging the work of classmates, submitting fraudulent data, plagiarizing the work of classmates or published and/or online sources, acquiring previously written papers and submitting them (altered or unaltered) for course assignments, collaborating with classmates when such collaboration is not authorized, and assisting fellow students in acts of misconduct. Students who have knowledge that classmates have engaged in academic misconduct should report this to the instructor.

Diversity and Inclusion

Institutional statement on diversity: “Diversity is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals.

The University of Wisconsin-Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background – people who as students, faculty, and staff serve Wisconsin and the world.” <https://diversity.wisc.edu/>

Academic Integrity

By enrolling in this course, each student assumes the responsibilities of an active participant in UW-Madison’s community of scholars in which everyone’s academic work and behavior are held to the highest academic integrity standards. Academic misconduct compromises the integrity of the university. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these acts are examples of academic misconduct, which can result in disciplinary action. This includes but is not limited to failure on the assignment/course, disciplinary probation, or suspension. Substantial or repeated cases of misconduct will be forwarded to the Office of Student Conduct & Community Standards for additional review. For more information, refer to <http://studentconduct.wiscweb.wisc.edu/academic-integrity>

Accommodations Policies

McBurney Disability Resource Center syllabus statement: “The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform faculty [me] of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. Faculty [I], will work either directly with the student [you] or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student’s educational record, is confidential and protected under FERPA.”

<http://mcburney.wisc.edu/facstaffother/faculty/syllabus.php>

UW-Madison students who have experienced sexual misconduct (which can include sexual harassment, sexual assault, dating violence and/or stalking) also have the right to request academic accommodations. This right is afforded them under Federal legislation (Title IX). Information about services and resources (including information about how to request accommodations) is available through Survivor Services, a part of University Health Services: <https://www.uhs.wisc.edu/survivor-services/>

Complaints

Occasionally, a student may have a complaint about a TA or course instructor. If that happens, you should feel free to discuss the matter directly with the TA or instructor. If the complaint is about the TA and you do not feel comfortable discussing it with the individual, you should discuss it with the course instructor. Complaints about

mistakes in grading should be resolved with the TA and/or instructor in the great majority of cases. If the complaint is about the instructor (other than ordinary grading questions) and you do not feel comfortable discussing it with the individual, make an appointment to speak to the Associate Chair for Graduate Studies, Professor Shawn Green, cshawngreen@wisc.edu.

If you have concerns about climate or bias in this class, or if you wish to report an incident of bias or hate that has occurred in class, you may contact the Chair of the Department, Professor Allyson Bennett (allyson.j.bennett@wisc.edu) or the Chair of the Psychology Department Climate & Diversity Committee, Martha Alibali (martha.alibali@wisc.edu). You may also use the [University's bias incident reporting system](#)

Privacy of Student Information & Digital Tools

The privacy and security of faculty, staff and students' personal information is a top priority for UW-Madison. The university carefully reviews and vets all campus-supported digital tools used to support teaching and learning, to help support success through learning analytics, and to enable proctoring capabilities. UW-Madison takes necessary steps to ensure that the providers of such tools prioritize proper handling of sensitive data in alignment with FERPA, industry standards and best practices. Under the Family Educational Rights and Privacy Act (FERPA which protects the privacy of student education records), student consent is not required for the university to share with school officials those student education records necessary for carrying out those university functions in which they have legitimate educational interest. 34 CFR 99.31(a)(1)(i)(B). FERPA specifically allows universities to designate vendors such as digital tool providers as school officials, and accordingly to share with them personally identifiable information from student education records if they perform appropriate services for the university and are subject to all applicable requirements governing the use, disclosure and protection of student data.

Privacy of Student Records & the Use of Audio Recorded Lectures

See information about privacy of student records and the usage of audio-recorded lectures.

Lecture materials and recordings for this course are protected intellectual property at UW-Madison. Students in this course may use the materials and recordings for their personal use related to participation in this class. Students may also take notes solely for their personal use. If a lecture is not already recorded, you are not authorized to record my lectures without my permission unless you are considered by the university to be a qualified student with a disability requiring accommodation. [Regent Policy Document 4-1] Students may not copy or have lecture materials and recordings outside of class, including posting on internet sites or selling to commercial entities. Students are also prohibited from providing or selling their personal notes to anyone else or being paid for taking notes by any person or commercial firm without the instructor's express written permission. Unauthorized use of these copyrighted lecture materials and recordings constitutes copyright infringement and may be addressed under the university's policies, UWS Chapters 14 and 17, governing student academic and non-academic misconduct.

Academic Calendar & Religious Observances

Students who wish to inquire about religious observance accommodations for exams or assignments should contact the instructor within the first two weeks of class, following the [university's policy on religious observance conflicts](#)