

Applied Machine Learning in Psychology
PSYCHOLOGY 559
FALL 2021 – Remote at the Start

Location: Zoom 12:30-1:45pm

Times: Most Tuesdays – synchronous, Most Thursdays – asynchronous

Link: <https://unc.zoom.us/j/98547749878>

Instructor: Oscar Gonzalez (he/him/his)

Office Hours: By appointment, over Zoom

Email: ogonza13@unc.edu

Website: Sakai class site

Class Prerequisites: PSYC 210 or equivalent

Reference Text (FREE)

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R* (1st ed.). New York, NY: Springer. (listed as ISL in the tentative schedule)

Download: <https://tinyurl.com/yuedyvm9>

**Most materials needed to succeed in the course will be covered in class. The book provides an additional resource, so students should focus on the bigger points discussed in the book and less on equations. Supplemental readings will also be assigned and posted.

Statistical Software (FREE)

In this class we will use the R statistical language for data analysis. R is a popular, script-based statistical language that you can download for free in your own laptop.

Download: <https://www.r-project.org/>

Overview and Learning Goals for the Course

In this day and age, we interact with many data collection tools. From swiping loyalty cards in the supermarket, movie recommendations by Netflix, or taking driving directions from a GPS, we are leaving a data footprint almost every day. Machine learning algorithms could help us go from raw datasets to valuable information. Machine learning has recently emerged as a major area of statistical research and is making its way into psychology. This course is an introductory seminar on the theory and application of machine learning to data analysis. A lot of research in psychology has focused on hypothesis-driven, explanatory approaches to data analysis. Machine learning could supplement a researcher's analytic toolbox to explore patterns in datasets and assess the predictive value of various combinations of variables on several outcomes.

The overall goals of the course are to...

1. Conceptualize the differences between explanatory and predictive approaches to data analysis
2. Learn the basics of statistical computing in the R statistical environment
3. Be competent in using statistical software to carry out the machine learning algorithms
4. Communicate effectively (written and verbally) what the machine learning algorithms are doing and when to use them.

One of the major aims of this course is to lower the barrier of entry to the field of data science, so the target audience for this course is psychology undergraduates who have a basic statistics foundation.

Structure of the course

Before deeming in-person interactions safe, the course will be synchronous on Tuesdays and asynchronous on Thursdays. On Tuesdays, I will lecture on a particular machine learning algorithm, discuss its theoretical underpinnings, and demonstrate some code on how to carry it out. After we finish covering an algorithm, there will be dedicated days where we have a lab session where I illustrate how to carry out the algorithm in the R statistical environment. Students will then carry out the analysis by themselves, write a short-report, and turn it in for a grade. Towards the end of the semester, I will lecture on special topics in machine learning while students work on their final projects.

All the Tuesday synchronous lectures will be recorded and uploaded to Panopto (available through Sakai). The Thursday asynchronous lectures will be uploaded to Panopto prior to the Thursday class time. Although we will NOT meet as a class on Thursdays, students are **expected** to be available during the Thursday class time to watch the asynchronous lectures, have student group meetings, professor-student check-ins, etc.

Classroom Environment, Course Requirements, Grading, and Honor Code

I value the perspectives of individuals from all backgrounds reflecting the diversity of our students. I broadly define diversity to include race, gender identity, national origin, ethnicity, religion, social class, age, sexual orientation, political background, and physical and learning ability. I strive to make this classroom an inclusive space for all students. Please let me know if there is anything I can do to improve, I appreciate suggestions.

I do not tolerate cheating. The honor code is in effect for this course. Students shall not misrepresent others' work as their own and will give full credit for others' contributions to the extent that these are allowed within the parameters of the assignment. Individual assignments are to be conducted individually. Course grades will be based on three major components:

- **Participation (20%):** There will be several student presentations, discussions on current data topics, and short reflection papers assigned throughout the semester. Late papers will not be accepted. As you may infer, regular class attendance (5% of participation) is expected, and asynchronous video viewing will be audited throughout the semester.
- **Lab write-ups (40%).** There will be five lab write-ups where you report your findings from the lab session. Length of the report will vary but will be at least 2 pages per report in order to meet the General Education requirement of 10 pages of written work. Lab reports are due, roughly, a week after they are assigned. Late reports are penalized with 10% off each day that they are late. After three days, the late report will not be accepted.
- **Final project (40%).** In collaboration with other students, you will take on a comprehensive analysis of a dataset using the algorithms discussed in the class and carried out during the lab sessions. The project will consist of submitting an analytic plan and a final presentation of the results. The final presentation will constitute the "final exam" for the class and will take place during the final exam period. The UNC final exam is given in compliance with the UNC final exam regulations.

The grade scale follows:

A	95-100%	highest level of attainment
A-	91-94%	
B+	88-90%	
B	85-87%	high level of attainment
B-	81-84%	
C+	78-80%	
C	75-77%	adequate level of attainment
C-	71-74%	
D	61-70%	minimal passing level of attainment
F	60% failed	unacceptable performance

Important Deadlines (Tentative)

	Assigned	Due
Reflection paper #1	08/19	08/24
Lab report #1	08/26	09/02
Reflection paper #2	09/09	09/14
Lab report #2	09/14	09/21
Lab report #3	09/23	09/30
Lab report #4	10/12	10/19
Lab report #5	10/28	11/04
Analytic Plan (in class)	11/02	11/02
Ethics Presentation	11/02	11/16
Final (synchronous)	12/03	12/03

University Resources

- **Accessibility Resources and Services:** The University of North Carolina at Chapel Hill facilitates the implementation of reasonable accommodations, including resources and services, for students with disabilities, chronic medical conditions, a temporary disability or pregnancy complications resulting in barriers to fully accessing University courses, programs and activities. Accommodations are determined through the Office of Accessibility Resources and Service (ARS) for individuals with documented qualifying disabilities in accordance with applicable state and federal laws. See the ARS Website for contact information: <https://ars.unc.edu> or email ars@unc.edu.
- **Counseling and Psychological Services:** CAPS is strongly committed to addressing the mental health needs of a diverse student body through timely access to consultation and connection to clinically appropriate services, whether for short or long-term needs. Go to their website: <https://caps.unc.edu/> or visit their facilities on the third floor of the Campus Health Services building for a walk-in evaluation to learn more.
- **Title IX Resources:** Any student who is impacted by discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitation, or stalking is encouraged to seek resources on campus or in the community. Please contact the Director of Title IX Compliance (Adrienne Allison – Adrienne.allison@unc.edu), Report and Response Coordinators in the Equal Opportunity and Compliance Office (reportandresponse@unc.edu), Counseling and Psychological Services (confidential), or the Gender Violence Services Coordinators (gvsc@unc.edu; confidential) to discuss your specific needs. Additional resources are available at safe.unc.edu.

Tentative Schedule*

Date	Week	Lecture	Reference Reading
08/19	1 – on	Why Machine Learning?	Witten et al., (2016) Ch.1
08/24	2 – on	On Learning and Classifying	Turing (1950); ISL p.127-130
08/26	2 – off	Lab: Introduction to R	ISL p.42-51
08/31	3 – on	Classification Trees with CART	ISL p.303-311
09/02	3 – off	CART and Machine Learning Concepts	ISL p.312-316
09/07	4 – on	Random Forest and Ensemble Methods	ISL p.316-321
09/09	4 – off	Evaluating Classification Accuracy	Zou, et al. (2007)
09/14	5 – on	Lab: CART and Ensemble Methods	ISL p.324-332
09/16	5 – off	K-Nearest Neighbors (KNN)	ISL p.39-42; Wu et al. (2008)
09/21	6 – on	Naïve Bayes	Wu et al., (2008)
09/23	6 – off	Lab: KNN and Naïve Bayes	Kuhn & Johnson (2013), Ch. 13
09/28	7 – on	Multiple Regression	ISL p.59-86
09/30	7 – off	Regression with Discrete Predictors	ISL p.86-90
10/05	8 – on	Assumptions of Regression	ISL p.92-102
10/07	8 – off	Regression with Interactions	ISL p.90-92
10/12	9 – on	Lab: Regression	ISL p.109-119
10/14	9 – off	Forward Selection in Regression	ISL p.214-228
10/19	10 – on	Regularization	ISL p.203-214
10/21	10 – off	FALL BREAK	
10/26	11 – on	Logistic Regression for Categorical Outcomes	ISL p.130-138
10/28	11 – off	Lab: Regularization and Logistic Regression	ISL p.154-161; p.251-256
11/02	12 – on	Special topic: Explanation and Prediction	Yarkoni & Westfall (2017)
11/04	12 – off	Special topic: Unsupervised Learning	ISL p.373-401
11/09	13 – on	Special topic: Ethics Presentation Workday	TBD
11/11	13 – off	Special topic: Ethics Presentation View	TBD
11/16	14 – on	Special topic: Machine Learning Research	
11/18	14 – off	FINAL PREP: Stacking Algorithms	Witten et al. (2016), Ch.12
11/23	15 – on	Wrap-up	
11/25	15 – off	THANKSGIVING BREAK	
11/30	16 – on	FINAL PREP: Final project workday	
12/03	16 – on	FINAL PRESENTATIONS	

White – classes that are synchronous or in-person

Gray – classes that are asynchronous

* Instructor reserves the right to make changes to the syllabus, including lab dates and due dates. These changes will be announced as early as possible.

Reference Readings

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. New York, NY: Springer

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). New York: Springer.

Turing, A. M. (1950). Computer machinery and intelligence. *Mind*, 59(236), 433-460.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Zhou, Z. H. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, 14, 1-37.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12, 1100-1122.

Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115, 654-657.

For our in-person interactions:

This semester, while we are in the midst of a global pandemic, all enrolled students are required to wear a mask covering your mouth and nose at all times in our classroom if we have in-person interactions. This requirement is to protect our educational community — your classmates and me — as we learn together. If you choose not to wear a mask, or wear it improperly, I will ask you to leave immediately, and I will submit a report to the Office of Student Conduct. At that point you will be disenrolled from this course for the protection of our educational community. Students who have an authorized accommodation from Accessibility Resources and Service have an exception. For additional information, see Carolina Together.