PS 590 - Intro to Supervised Machine Learning

Fall 2025

Taught by: Fabricio Vasselai **(**vasselai@illinois.edu)
When: Tuesdays, 9:30am-11:50am.
Classoom: 404 David Kinley Hall.
Office Hours: (in person) Tuesday 11:50am-12:50pm at 426 David Kinley Hall.
(remotely) variable (usually either T, Th or Fr, late afternoon), by appointment.

Course Overview

This course will cover the fundamentals of Supervised Machine Learning (SML), with an applied focus and geared towards Political and Social Scientists. Students will be introduced to key prediction techniques like k-nearest neighbors, naive bayes classifier, regularized linear and non-linear regression, decision trees, random forests, SVM and a variety of deep neural networks. Besides, the course will cover parameter optimization and also how to evaluate SML model predictions using different types of cross-validation (Monte Carlo, leave-k-out, block) and with different metrics. The course ends with a discussion of bias, fairness and trustworthiness in SML and how to use interpretable / explanation methods to address such issues. Note that the course will cover some of the crucial mathematical foundations behind each technique, but a lot of the focus will be both on having the students learn the critical concepts and develop deep intuitions about the techniques, and on the application of each method using R programming language.

Prerequisites

Students should only take this course if they:

- Sucessfully took PS 530 and PS 531 (or equivalent);
- Have been introduced to reading and writing R code (including for-loops, while-loops and custom functions).

Basic knowledge of Linear Algebra and Probability is strongly desirable. Basic knowledge of multi-variate calculus would help. A review of key topics in those areas will be offered in the beginning of the course. DataCamp online courses on R programming language will be made freely available to those enrolled, in case students want to review / improve their R skills.

Computers in class

Students should bring a laptop to class or plan to work with someone who has a laptop. ATLAS Share is a university program that provides long term loaner computers to current LAS undergraduate students who may not have access to their own computer. Make sure that you install R statistical software and RStudio IDE in your working laptop.

Course Structure

Class sessions will consist of a mix of lecture and going over R code.

All course content and readings are organized on Canvas under "Modules" on the left-hand side of the screen. Students should read all of the eventual readings listed for a given class session ***ahead of class time*** and be prepared for in-class discussion.

While there is no one actual *textbook* for this course, the following books are recommended:

Murphy (2012). Machine Learning: A Probabilistic Perspective, MIT Press.
Hastie, Tibshirani, and Friedman (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, Second Edition.

Bishop (2006). Pattern Recognition and Machine Learning, Springer.

Mohri, Rostamizadeh, and Talwalkar (2012). Foundations of Machine Learning, MIT Press.

Duda, Hart, and Stork (2001). Pattern Classification, Wiley.

Communications

Strictly all regular course-related communication will happen via Piazza, (which means not via email), a very handy on-line questions-and-answers platform. In Piazza, students can ask questions anonymously, both publicly and privately (although the latter should be saved only for discussing personal matters or when the question being asked contains part of a solution for a homework). This way, we centralize our communication, and it makes it so that all questions and doubts you might have end up reaching all colleagues. Another neat feature of Piazza is that students can also answer to students. I **will** consider your Piazza activity (asking & answering questions) in your participation grade.

In the rare circumstance that you should feel the need to send me a direct email, start the subject of the email with "[PS590]" (not just "PS590"), otherwise there will be a high likelihood that I miss your email.

Always expect 2 business days for me to answer to any form of communication (that is about 48 hours, not including weekends).

Evaluation

- *DataCamp Ungraded Courses* ***(20%)***
- *In-Class Quizzes* ***(30%)***
- *Problem sets* ***(20%)***
- *Final Exam* ***(20%)***
- *Participation* ***(10%)***

DataCamp courses

In most weeks, students will have to complete DataCamp courses / course chapters online, that will be assigned beforehand. These must be completed always before class time. They will not be graded for accuracy, just for completion.

Plagiarism and Academic Honesty

According to the Student Code, 'it is the responsibility of each student to refrain from infractions of academic integrity, from conduct that may lead to suspicion of such infractions, and from conduct that aids others in such infractions.' I expect you to be familiar with and understand the university's policies on academic honesty for this course. Please consult the Student Code for more information (https://studentcode.illinois.edu/article1/part4).

- Any plagiarism (of existing work, of others or even of one's own past work, will result in a zero grade and will be reported to the student's home department and to the University authorities.
- Any copying of others' work will result in a zero grade and will be reported to the student's home department and to the University authorities.
- Unless otherwise explicitly stated in my assignment instructions, **absolutely any usage of ChatGPT or other AI / Generative AI / LLM tools is strictly prohibited**. Again, using any of those will result in a zero grade and will be reported.
- When in doubt, please ask! It is far better to check with us prior to submitting an assignment than waiting.

Please note: it is my work, ethical and even legal responsibility as an instructor to uphold the academic integrity policy of the University, and I fully intend to do so. This is really for real. I have zero flexibility with these things and will go above and beyond to make sure students that are eventually caught infringing the Student Code do face all consequences admissible by University, state and federal statutes.

Note: if you've read this far, go to our Canvas website and submit your all-time favorite meme (or link to a song/video you love, in case you don't have a favorite meme) in the Assignment called "Testing, 123!". Do not tell your colleagues you did this. Let's see how many students catch this!

Submitting Late Assignments

This course has a **very** strict lateness policy for assignment submission:

- Considering their online nature, their short length and their purpose of having students practice code before class, DataCamp courses can never be completed late, under any circumstances.
- In the case of the First and Second graded exercises, assignments submitted:
    - up to 6 hours late will have no penalty;
    - more than 6 but less than 24 hours late will have a -5 points penalty (out of 100);
    - more than 24 but less than 48 hours late will have a -10 points penalty(out of 100);
    - more than 48 but less than 72 hours late will have a -15 points penalty (out of 100);
    - more than 72 but less than 96 hours late will have a -20 points penalty (out of 100);
    - assignments submitted more than 96 hours late **will not be accepted***.

- Given the strict university deadlines for submitting students' final grades, the Final project **cannot be late at all***.

* of course, unless a student has academic accommodations previously approved by the Division of Disability Resources and Educational Services. In that case, I will follow the guidelines from the specific accommodations the student was approved for. See below.

Statement on diversity, inclusion, and disability

The University of Illinois is committed to diversity and rigorous inquiry from multiple perspectives. The Political Science department shares this commitment and seeks to foster productive learning environments based upon inclusion, open communication, and mutual respect for a diverse range of identities, experiences, and positions.

The University of Illinois is committed to ensuring equitable access to our academic programs and services. Students with disabilities who have been approved for the use of academic accommodations by Division of Disability Resources and Educational Services (DRES) and need a reasonable accommodation(s) to participate fully in this course should follow the procedures established by DRES for using accommodations. Timely notifications are required in order to ensure that accommodations can be implemented. Please meet with the instructor to discuss access needs in this class after completing the DRES procedures for requesting accommodations.

- Email: disability@illinois.edu
- Phone: (217) 333-1970

Mental Health

Significant stress, mood changes, excessive worry, substance/alcohol misuse or interferences in eating or sleep can have an impact on academic performance, social development, and emotional wellbeing. The University of Illinois offers a variety of confidential services including individual and group counseling, crisis intervention, psychiatric services, and specialized screenings which are covered through the Student Health Fee. If you or someone you know experiences any of the above mental health concerns, it is strongly encouraged to contact or visit any of the University's resources provided below. Getting help is a smart and courageous thing to do for yourself and for those who care about you.

- Counseling Center (217) 333-3704
- McKinley Health Center (217) 333-2700
- National Suicide Prevention Lifeline (800) 273-8255
- Rosecrance Crisis Line (217) 359-4141 (available 24/7, 365 days a year)

Tentative Course Schedule

Note: *Schedule is subject to change. Check on Canvas for updates as the course progresses.*

Class	Date	Topics
1	**Aug 26**	What is Machine Learning
2	**Sep 02**	Math & Stats review
3	**Sep 09**	k-Nearest Neighbors, Naïve Bayes Classifier, Linear Discriminant Analysis
4	**Sep 16**	Decision Trees, Random Forests, Tree Boosting methods
5	**Sep 23**	Linear, Logistic and Non-linear regressors
6	**Sep 30**	Regularization and Ridge Regression
7	**Oct 07**	Separating Hyperplanes + Kernels
8	**Oct 14**	Constrained Optimization + Support Vector Machines
9	**Oct 21**	Deep Neural Networks (DNNs)
10	**Oct 28**	DNNs - Convolutional Neural Networks
11	**Nov 04**	DNNs - Natural Language Processing
12	**Nov 11**	Generative DNNs
13	**Nov 18**	Model selection, Cross-Validation & Hyperparameter optimization
Nov 27 (no class)		*Thanksgiving*
14	**Dec 02**	Interpretable / Explainable SML
Due Dec 09		*In Person Final Exam*