Syllabus

Instructor: Dr. Donald S. Williamson	**Time:** TR 4:55pm – 6:10pm
Email: Piazza (required), williads@indiana.edu	**Place:** Zoom (see Canvas for link)
Office Hours: TBD	**Course Page**: Canvas
AI: Junyi Fan	**Email:** Piazza (required), fanjun@iu.edu
Office Hours: TBD	**Place:** Zoom (see Canvas for link)

Overview: This course will introduce graduate students to the practical aspects of machine learning through the application of machine-learning concepts to real-world problems. This course will cover fewer learning algorithms and spend less time on math and theory, though some theory and math will be covered to ensure understanding. More time will be spent on implementation skills required for algorithms to work on a variety of datasets. The students will gain an understanding of the full machine-learning pipeline, including data pre-processing, feature extraction, algorithm development, and performance evaluation. Students will gain first-hand knowledge of machine learning by implementing various algorithms and concepts in the Python programming language. Students will also be exposed to popular Python machine-learning libraries (e.g., PyTorch, SciKit-Learn, etc.).

Objectives: By the end of the course, students will:

- Become familiar with the full pipeline of applying machine learning to different problems
- Gain skills regrading the collection, processing, and labeling of data
- Understand and implement the most popular machine-learning algorithms
- Improve their theoretical understanding of machine learning
- Enhance their Python programming skills and utilize ML libraries (PyTorch, Scikit, TensorFlow, . . .)
- Design experiments that reasonably assess performance
- Evaluate multiple learning algorithms across several tasks
- Asses potential societal impacts of the machine-learning algorithms

Prerequisites: Applied Machine Learning draws from many fields, so an undergraduate-level understanding of probability, statistics, calculus, linear algebra and programming in Python is assumed. Exposure to these topics is helpful, but you may need to do some background reading to supplement material covered in class. External resources will be linked from time to time to help the students in this regard. While it is not expected that students should be an expert Python programmer, they are expected to have the ability to debug their own code and learn how to use new libraries. Due to the complexity of the material, the size of the class and the limited staff resources, the instructors will prioritize helping students with questions about the class material over questions of how the programming language works.

Recommended Texts: This is a restricted list of various interesting and useful books that will be touched during the course. You need to consult them frequently.

- Geron, A. *Hands-On Machine Learning with Scikit-Learn & TensorFlow: concepts, tools, and techniques to build intelligent systems.* 1st Edition. O'Reilly. 2017. ("HOML")
- Marsland, S. *Machine Learning An Algorithmic Perspective.* 2nd Edition. CRC Press Taylor & Francis Group. 2015. ("MLAP")
- Raschka, S., & Mirjalili, V.. *Python Machine Learning: Machine Learing and Deep Learning with Python, scikit-learn and TensorFlow 2.* 3rd Edition. Packt. 2019. ("PML")

Grading Policy: Four main components will be used to assess student learning, which includes, participation, homework assignments, quizzes and a final project. The grading policy for each of these components are as follows: **Participation(10%), Homework (30%), Quizzes (30%), and Project (30%)**. Grades are based on straight scale 99-100 A+, 93-98 A, 90-92 A-, 88-89 B+, 83-87 B, 80-82 B-, 78-79 C+, etc. Grades may be positively curved upon the discretion of the instructor.

Participation: students are expected to attend class, ask questions during class and answer questions that are posed during class. The questions will either be asked verbally, through Zoom polls or through TopHat (https://uits.iu.edu/tophat).

As this is a graduate course, attendance is expected though not required. Due to the online nature of the course, we understand that students (and instructors) may experience technical difficulties. To circumvent this, all lectures will be recorded and posted to Canvas. In this event, students are responsible for any missed content or assignments.

We found students did significantly poorer on homework when they merely watched recorded lectures as opposed to attending live classes–they also found it very difficult to focus on their own and keep to a schedule. Each lecture may have a learning quiz that will be required to submit by the end of the day–this may be paired programming assignment too!

Homework: Approximately five homework assignments will be given. The assignments will contain programming problems as well as some theoretical questions involving math and intuition. While the focus of the programming questions will be to effectively employ machine learning algorithms to practical problems, they might also contain the implementation of certain algorithms.

All assignments are required to be submitted using Python 3. Why? We often provide starter code for assignments. Due to the size of the class, it is really hard for the AIs to generate template code for multiple programming languages + setup the environments for grading + foresee what problems the students might have using a library so we can provide clear instructions. In previous semesters we have had students from disciplines that mostly use R for their work that were able to pick-up Python quickly and did really well in the class. We have also had students that did portions of their final project using R. The AIs will hold their office hours on multiple days of the week and are also available to answer Piazza posts. Do not get discouraged by the language choice for the assignments, we are more than happy to help if you run into trouble.

To foster a sense of community, all homework will done in pairs. The pairing will rotate based on random selection. This is to help you meet and make friends with students with whom you'll be friends with perhaps your entire life–but certainly during your time at IU and especially in The Luddy School. You will be required to put the name of your partner on your homework–you can conceivably receive different grades if the files differ–but we can discuss those details during the first lecture.

All assignments will be submit through Github: https://github.iu.edu/.

Quizzes: Announced and unannounced quizzes will be given during the semester to assess short-term student learning. These quizzes will be based on recent lectures and/or homework assignments. These quizzes will be administered online, due to the virtual nature of the course. Note that quizzes are individual assignments, and most be completed on your own.

Project: The project constitutes an important part of this class. Students will be expected to work in groups of up to three for the project, throughout the semester. You are expected to work on a machine learning project of considerable size. The project should aim to answer three (3) research questions for which you'll have to design experiments (which models to use, how to evaluate, etc.).

What can count as a research question? For example, using the same insurance dataset:

- Predict whether the person is a low-, moderate- or high-risk individual (classification)
- Given some data about a person and it's risk classification, predict how much that person would cost to the insurance company per year to be able to determine how much of a premium should be charged (regression)
- Are there clusters of claims that might indicate fraud? (clustering)

What is NOT a research question?

- Performing data preprocessing
- Deciding which neural network architecture to use
- Tuning a model's hyperparameters
- Doing anything that falls outside the scope of machine learning. For example, building a website to showcase your project. If you are unsure, ask the AIs.

We are not expecting you to achieve state-of-the-art performance or to create Artificial General Intelligence (AGI) in your work. However, it is required to demonstrate that if an approach did not work originally, that you were able to diagnose it and solve the problem or you tried other approaches that may or may have not worked.

The final group members and the topic should be submitted and approved by the mid-point of the semester, an explicit deadline will soon be announced. You are encouraged to start thinking about the topic as soon as possible. You are strongly encouraged to choose a topic close to your area of interest and something you are passionate about. No project proposal should be based on re-using data from a Kaggle competition unless previously approved by the AIs. The project proposal and the progress report will be graded. We expect meaningful progress to be made by the time that the progress report is due.

There will be final presentations during the final two weeks of classes and students will be graded on the content, work, presentation, report, and teamwork.

- TBD: declare team and topic
- Up to 3 team members. Sign-up on Google Spreadsheet:
- TBD: proposal
- We will provide a LaTeX template
- TBD: meet with AI's for proposal feedback
- Required minor modifications or rejected proposals will have to be re-submitted within one week
- TBD: submit a progress report (and code)
- TBD: final submission (report + code + presentation)
- TBD: project presentations

We recommend that you write a draft of your report as progress is made. Often, students write their reports in the last minute and submit documents that are full of errors. This is a graduate course, so we at least expect that you'll use a spellchecker before submitting.

Trying to do the bulk of the project during the last few weeks is the surest way to fail the course.

Computing Infrastructure: The following will be used to facilitate this course.

- Canvas: Course page
- Piazza: Communication and announcements
- Github.iu.edu: Homework submissions
- Top Hat: In- and post-class quizzes
- Zoom for lectures and office hours
- Python: https://www.python.org/

Late homework submission policy: We expect all work to be finished on time. However, we do understand that there might be times when something unexpected comes up which delays you. Hence, as a late submission policy, we will allow late submissions up to 3 days late, each day carrying a 5 point penalty. If you submit 70 hours after the original deadline, the highest possible score for that assignment will be 85 points, out of 100. Late assignments submitted at 72:01 hours after the original deadline will not be accepted. Please, make the arrangements to start and finish your assignments early. This means you should push to Github well before the time stated on the homework. This means you must start the homework as soon as you can and continually push!

NOTE THAT THIS POLICY ONLY APPLIES TO HOMEWORK ASSIGNMENTS. THE QUIZZES AND PROJECTS MUST BE SUBMITTED ON TIME.

Tentative Course Schedule (subject to change):

Week	**Topic**
INTRODUCTION	
Week 1	Course introduction, logistics and Machine Learning basics
Week 2	Data collection, pre-processing and labeling
Week 3	Evaluation and methodology
SUPERVISED LEARNING	
Week 4	Probability review and Naive Bayes
Week 5	Linear and Nonlinear Regression
Week 6	Neural Networks and Deep Learning I
Week 7	Neural Networks and Deep Learning II
Week 8	Neural Networks and Deep Learning III
Week 9	Support Vector Machines
Week 10	Decision Trees, Ensemble Learning and Random Forests
UNSUPERVISED LEARNING	
Week 11	Clustering, KMeans and Nearest Neighbor
Week 12	Gaussian Mixture Models and Nonnegative Matrix Factorization
Week 13	Dimensionality Reduction
Week 14	Reinforcement Learning (time permitting)

Project	
	Project Presentations I (if needed)
Week 15	Project Presentations II
Week 16 (Finals Week)	Project Presentations III (if needed)

Cheating: A significant amount of pre-existing code is obviously available through the internet. All work in this class, however, unless explicitly stated, must be the student's own. Cheating in CS is dumb–you can only learn by doing; therefore, copying someone else's work does not, in any way, help you with your skills. The consequences for cheating are severe.

This site provides information about cheating: https://studentaffairs.indiana.edu/office-student-ethics/misconduct-charges/academic-misconduct.shtml. Quoting the relavent passage:

- ... Coursework performed while misconduct proceedings are underway, however, shall be considered conditional. Conditional work may be affected or eliminated based on a final finding of misconduct or sanction imposed. This may result in loss of course credit, a delay in the awarding of a degree, or revocation of a degree that was awarded prior to a final decision in the misconduct proceedings. If either academic or personal misconduct is discovered that may impact degree conferral or graduation, the Dean of Students may notify the student's academic dean, who may withhold conferral of the degree pending completion of misconduct proceedings. If, after a degree has been conferred, the University determines that the student committed academic misconduct prior to the conferral, the University may revoke the degree . . .

If you're tempted to cheat, speak with the instructor or AIs. The risk is too great and the reward not worth it. Here is more material about cheating.

Communication: This is critical! You must use Piazza to communicate with the instructors. With over 50 students, email isn't feasible. Email will be summarily ignored–that's harsh, probably, but we have to make the class fair for everyone and sensible in response time. Please refer to the following paragraphs about using Piazza. If you have an urgent matter that you would prefer not to share with the whole AI staff, you are allowed to email the instructor privately. If you know you will have issues or something has come up, please contact us as soon as possible. The instructors and AI staff will work with you, but having information sooner will be easier to help dealing with issues.

Piazza is a Question / Answer forum. This will be our main form of communication with instructors and students. We will make all announcements through Piazza. You can access Piazza without signing into Canvas. You can go to https://piazza.com and access it (only after initially looking at it through Canvas). Please follow the Module 0 with more details, specifically Piazza -¿ Logging Into Piazza. Please ensure you follow what Your Name field needs to be. You are responsible to check if we ask questions to your posts.

Class Comportment: Everyone will treat each other with respect and civility. Students who engage in disruptive or rude behavior will be asked to leave the class and, if continually exercised, the course. We require that students connect to Zoom with their laptops, desktops or tablets (e.g. cannot use cell phones) and that they cannot use other devices during the lecture. We also ask that you please complete the guide on Zoom etiquette. Such as, not talking when someone else is talking, being appropriately dressed if video is on, keeping their microphone muted when

they are not talking, using full name when they connect to Zoom, possibly using the chat and 'Raise Hand' features when they have a question. Students coming late or leaving early should contact the instructor beforehand.

Collaboration Policies: Students may freely use any resource that is provided by the instructor for an assignment. Most work is to be completed as paired programming. In general, the work that is submitted for an assignment must be the pair's own. Students may not submit work under their own name that is done by, or in collaboration with, someone else (another pair, for example). Copying solutions from any source, including the web or students in previous offerings of the course, is not allowed. Students should not read or possess copies in any form–physical or electronic–of another student's work. There is no legitimate reason for a student to possess a copy of another student's assignment, to send a copy of student work from one computer account to another, or to be logged-on to another student's account. Providing one's own work to another student is also a violation of these policies. We routinely use software and other tools to detect similarities between submissions. Identical, or nearly identical, submissions will be considered conclusive evidence of plagiarism (aside from paired programming). For programming assignments, students may normally discuss general approaches to assignments, and they may give or receive"consulting" help for specific problems with software or computer programs. A student may look at another student's work only when help is requested. In that situation, the student takes on the role of mentor, and the interaction must be limited to the immediate problem.

If you are ever in doubt... If you believe you may have received unauthorized help on an assignment, CITE YOUR SOURCE. You can never be accused of academic dishonesty if you are clear (in your assignment) about any help you received. The worst that can happen in this event is that you do not receive full credit for the assignment.

Accommodations for Students with Disabilities: Every attempt will be made to accommodate qualified students with disabilities (e.g. mental health, learning, chronic health, physical hearing, vision neurological, etc.). You must have established your eligibility for support services through the appropriate office that services students with disabilities. Note that services are confidential, may take time to put into place and are not retroactive; captions and alternate media for print materials may take three or more weeks to get produced. Please contact your campus office as soon as possible if accommodations are needed. Find your office at: http://ada.iu.edu/students/index.shtml

Academic Honesty: We take academic integrity very seriously. You are required to abide by the Indiana University policy on academic integrity, as described in the Code of Student Rights, Responsibilities, and Conduct, as well as the Computer Science Statement on Academic Integrity (https://cs.indiana.edu/doc/graduate/cs-graduate-student-handbook-2019-20.pdf). It is your responsibility to understand these policies. Briefly summarized, the work you submit for course assignments, projects, quizzes, and exams must be entirely your own (or entirely that of your group if group work is permitted). If you use the ideas (including text, source code, algorithms, concepts, diagrams, slides, etc.) of others, you must give proper credit with a prominent citation and an explicit indication of which idea(s) or material(s) you borrowed so that another person (e.g. a grader) can easily separate your contribution from the work of others. You may discuss assignments with other students (or students in other groups) at a high level, by for example discussing general methods or strategies to solve a problem, but you must cite the other student in your submission. Looking at someone else's code related to an assignment, whether online or from another student, will almost certainly lead to academic dishonesty. Sharing your assignment code with another student also almost certainly constitutes academic dishonesty.

The consequences of academic dishonesty are extremely serious. We will respond to acts

of plagiarism and academic misconduct according to university policy. In assigning sanctions, we will follow CS Program policy: "The ordinary departmental level penalty for cheating is failure in the course" but "in all cases, the penalty will be more severe than not turning in the assignment." In addition, "the student will no longer be eligible for the guaranteed financial aid provided by the CS program." Moreover, University policy requires us to report the incident to the Dean of Students, who may apply additional sanctions, including expulsion from the university.

Students agree that by taking this course, papers and source code submitted to us may be subject to textual similarity review, for example by Turnitin.com. These submissions may be included as source documents in reference databases solely for the purpose of detecting plagiarism of such papers or codes.

The Computer Science Department seeks to create a friendly and supportive learning environment. We encourage students to work in groups to review material from the lectures and readings, to work practice problems, to study for exams, and to discuss the general ideas and approaches to assignments. However, work submitted for a course must be done independently, unless collaboration on a particular assignment is explicitly permitted. Effective learning is compromised when this principle is violated. This means that the work you turn in must represent only your own work. It must not be based on help from others or information obtained from sources other than those approved by the instructors.

Sexual Misconduct: Indiana University prohibits discrimination on the basis of sex or gender in its educational programs and activities. Discrimination on the basis of sex or gender is also prohibited by federal laws, including Title VII and Title IX. This policy governs the University's response to discrimination based on sex or gender, and all forms of sexual misconduct (which includes sexual harassment, sexual assault, other forms of sexual violence, dating violence, domestic violence, sexual exploitation and stalking. Such behaviors are against the law and are unacceptable behaviors under Indiana University policy. These unacceptable behaviors are hereafter referred to as 'Sexual Misconduct.' The University does not tolerate sexual misconduct and it will take action to prevent and address such misconduct. The University has jurisdiction over all Title IX and related complaints. Questions about Title IX may be directed to Indiana University's Title IX Coordinator, or the Office of Civil Rights. Please visit https://policies.iu.edu/policies/ua-03-sexual-misconduct/index.html