CSCI 381

Machine Learning

Spring 2025 MFW 11:35am–12:45 pm MEY 129

http://cs.wheaton.edu/~tvandrun/cs381

Thomas VanDrunen

☎630-752-5692 630-639-2255 ✉Thomas.VanDrunen@wheaton.edu

Office: MEY 163 Office hours: Drop-in: MWF 3:30–4:30pm; Or by appointment through Calendly

---

Contents

Catalog description Theory, algorithms, and applications of machine learning. Machine learning techniques including k-nearest neighbors, expectation-maximization, neural nets, support vector machines, and principal component analysis. Ethical considerations for how machine learning applications are used and how they affect society. Prerequisites: CSCI 345 and MATH 245.

Textbook. Han Veiga and Ged. *The Mathematics of Machine Learning.* De Gruyter, 2024.

Purpose of the course. There are several models for what a course in *machine learning* could be. At one extreme are courses that present machine learning as an advanced area of probability and statistics. At the other are courses that teach students to patch together machine learning applications from libraries without needing to understand the mathematics behind them. Both of those models neglect the *algorithms* of machine learning, and neither of them reflect our intention in this course.

Instead, this course is designed to present a balanced approach to machine learning: as we consider a selection of machine-learning techniques, students will competently but not exhaustively explore the *mathematics* of machine learning; they will practice applying machine learning *libraries* to solve real-world problems; but especially they will learn the *algorithms* of machine learning by implementing them from directly.

Goals and objectives. The goals of this course are that students will be able to

1. Articulate the goals and use the terminology of machine learning correctly.
2. Complete the laboratory exercises that introduce widely-used machine learning libraries and apply those libraries to machine-learning problems.
3. Identify the elements of widely-used machine-learning algorithms.

The objective of the course is that students will be able to

1. Explain how machine learning techniques are derived from probability theory and statistics.
2. Implement machine-learning algorithms correctly.
3. Discuss social and ethical issues around machine learning using technical knowledge and Christian ethics.

In addition to these, together we have the general objective of seeing statistical inference and machine learning as a way of knowing God's world and a tool for doing good, to God's glory.

COURSE OUTLINE. The course is organized into a sequence of machine learning techniques. The major assignments and where they fall are indicated in bold.

I. Prolegomena (3 days)

   A. Terms and concepts
   B. Python libraries

II. The nature of data (2 days)

   A. Tables, attributes, and dimensionality
   B. k-nearest neighbors

III. Linear regression (4 days)

   A. Mathematical derivation of simple linear regression
   B. Multiple regression and regularization
   C. Gradient descent

IV. Logistic regression (3 days)

   A. Derivation from regression to classification
   B. Multiclass classification

V. Gaussian mixture models and EM (4 days)

   A. Continuous probability
   B. Distributions and statistics
   C. Gaussian mixture models
   D. Expectation maximization

VI. Support vector machines (5 days)

   A. Linear programming
   B. Concepts of SVMs
   C. The math of SVMs, including the kernel trick
   D. SVM algorithms using quadratic programming

VII. Principal component analysis (4 days)

   A. Concepts of PCAs
   B. Eigenvectors and eigenvalues
   C. PCA algorithms using eigenvalues

VIII. Neural nets (7 days)

   A. Perceptrons and MLPs
   B. Perceptron training
   C. The feed-forward and back-propogation algorithms
   D. Deep learning and convolutional neural nets

IX. Ethics of machine learning (3 days)

Course procedures

How we do this course. The "typical" coverage of one of the machine-learning techniques would follow a four-day pattern of

a. Concepts

b. Applications using libraries (in a lab activity)

c. Mathematical details

d. Algorithm with accompanying

In reality, different topics will take different amount of class time, from two to seven days. But regardless of length, the coverage will be in the spirit of this four-day plan. Concurrently with these class days, students will have readings, programming assignments, and quizzes (administered on Canvas).

Note that for the purposes of this course, I use *labs* to refer activities done during a class meeting (but in the CSCI lab) and *project* to refer to an open-ended, semester-long assignment. Smaller, more frequent sets of exercises that involve programming I simply call *programming assignments*.

Implementation platform. Code examples, labs, and programming projects will be done using Python 3. Students without prior experience in Python are responsible for learning the basics of Python on their own. Resources for learning Python can be found on the course website. We will make extensive use of certain libraries, (especially in lab, less so in projects). The main library we will use is `scikit-learn`; additionally we will use `numpy`, `scipy`, `pandas`, `matplotlib`, `tensorflow`, and `cvxopt`.

Note that both Python and the libraries mentioned above have been through several recent versions. If you do your assignments using a Python installation on your own computer, make sure you test your solution on the CSCI lab machines before turning it in. I will use the CSCI lab environment to grade your submission, and you are responsible for submitting code that this compatible with the lab installation, which may not be the latest version.

When asking for help with Python codig by email, please either (a) paste the code directly into the body of the email (preferred for smaller pieces of code) or (b) attach the entire file after renaming it so that it doesn't have a `py` extension—for example, change `some_program.py` to `some_program.py.txt`. Please do **not** attach a screenshot or a file with a `py` extension. Wheaton's email service *silently rejects* any email that has a `py` file as an extension.

Laboratory activities. Collaborative in-class lab assignments will constitute a major portion of students' experience in this course. We will use Jupyter notebooks as our programming environment. Students will be penalized for lab activities that are missed and not made up.

Most lab activities will be based on code and ideas from our textbook and the following books:

- Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn & TensorFlow*, O'Reilly, 2017.
- Andreas Müller and Sarah Guido, *Introduction to Machine Learning with Python*, O'Reilly, 2016.
- Sebastian Raschka and Vahid Mirjalili, *Python Machine Learning*, Packt Publishing, 2017.
- Joel Grus, *Data Science from Scratch: First Principles with Python*. O'Reilly 2015.

Programming assignments. Most of students' work in this course outside of class meetings will be in programming assignments. There will be an average of one programming assignment per week, but they will be shorter/easier/more frequent at the beginning of the semester and longer/harder/less frequent at the end.

Tests. There will be a midterm and a final. The midterm is expected to be held in class on Fri, Mar 21 (may be subject to change). The final exam block is Wed, May 7, 10:30 am–12:30 pm. The final exam will be "mostly non-cumulative."

Reflections/responses. Throughout the semester there will be occasional assignments where students will write a reflections on ethical questions, often in response to a reading. See Canvas for details.

Project. Students will complete a semester project with a partner in which they will apply machine learning techniques to a chosen data set to solve a problem or answer a question. The final deliverables of the project will consist in a Jupyter notebook and a short (approximately 10. minute) presentation. See Canvas for details.

Grading. To **pass** this course (receive a grade of D or better), students must perform competently on each goal by completing at least 75% of the the lab activities, completing a semester project, achieving at least 50% of the points on the programming problems, and having at least a 50% average on the midterm and final together.

For students who have met the minimum requirements, their *semester score* is the geometric mean of their scores for the midterm, final, project, programming assignments, and other, where *other* comprises quizzes, lab activities, and reflections. That is, your semester score is

$$\sqrt[5]{Midterm \cdot Final \cdot Project \cdot ProgrammingAssignments \cdot Other}$$

The geometric mean is used because it is self-normalizing: The individual scores will have different scales, but affect the semester score equally.

Letter grades will be determined by score clustering. An estimation of semester grade will be given after the first test and, after that, upon request.

I use the "Gradebook" feature on Canvas only to communicate scores on individual assignments and tests. I do **not** use the Canvas gradebook for my official record-keeping for scores, for calculating semester scores, or determining letter grades. Please **ignore** any grade estimate that Canvas gives you for this course.

Policies etc

Academic integrity. Collaboration among students in the class is permitted on programming assignments. All code turned in for assignments and all written work is to be original Any resources consulted in projects besides the textbook and the official documentation for Python and the libraries used must be cited as in a research paper.

Late assignments. You are allowed three late days on programming and reflection assignments, which may be divided in whole-number units among the projects (for example, one project three days late, or three projects each one day late, etc). *Days* refers to calendar days. No assignment may be turned in after the last day of class (Apr 30), thus if an assignment is due on Apr 30 then no late days may be applied to it. Please inform the instructor that you are using a late day on the day that the assignment is due, or earlier. Beyond the allowance for late days, assignments will not be accepted late. Other assignments will not be excepted late.

Attendance. Students are expected to attend all class periods. It is courtesy to inform the instructor when a class must be missed.

Examinations. Students are expected to take all tests, quizzes, and exams as scheduled. In the case where a test must be missed because of legitimate travel or other activities, a student should notify the instructor no later than one week ahead of time and request an alternate time to take the test. In the case of illness or other emergencies preventing a student from taking a test as scheduled, the student should notify the instructor as soon as possible, and the instructor will make a reasonable accommodation for the student. The instructor is under no obligation to give any credit to students for tests to which they fail to show up without prior arrangement or notification in non-emergency situations. The final exam block is Wed, May 7, 10:30 am–12:30

pm. I do not allow students to take finals early (which is also the college's policy), so make appropriate travel arrangements.

Accommodations. If you have a documented need for accommodations, I will have received a letter on your behalf from the Learning and Accessibility Services Office. But *please talk to me* about what accommodations are most useful to you. In particular, if you desire accommodations for test-taking, talk to me a reasonable amount time in advance (say, at least two class periods) so arrangements can be made. (See also the College's statement below.)

Office hours. My *drop-in* office hours this semester are MWF 3:30–4:30pm. You can make an appointment through Calendly; I'm available most of the day on Thursday and sometimes on other days.

Electronic devices. My intent is for my courses to be electroinic-device-free zones. Unless you have made special arrangements with me, please keep all laptops, tablets, phones, etc silenced and put away—do **not** have your sitting on your table. If you need to check your phone for something, please discreetly step out into the hall. This policy applies to days with lab activities; only CSCI lab machines may be used, and they may be used only to do the lab activity. In particular, **NO TEXTING OR USING SOCIAL MEDIA DURING CLASS MEETINGS.**

All this, the Lord willing.

College syllabus statements

THE COLLEGE REQUIRES THAT THE FOLLOWING STATEMENTS BE INCLUDED IN ALL SYLLABI.

The "Academic Information" website referred to below is found and https://catalog.wheaton.edu/undergraduate/academic-policies-information/academic-information/

ACADEMIC INTEGRITY. (See "Integrity of Scholarship" on "Academic Information" website.)
The Wheaton College Community Covenant, which all members of our academic community affirm, states that, "According to the Scriptures, followers of Jesus Christ will. . . be people of integrity whose word can be fully trusted (Psalm 15:4; Matt. 5:33-37)." It is expected that Wheaton College students, faculty and staff understand and subscribe to the ideal of academic integrity and take full personal responsibility and accountability for their work. Wheaton College considers violations of academic integrity a serious offense against the basic meaning of an academic community and against the standards of excellence, integrity, and behavior expected of members of our academic community. Violations of academic integrity break the trust that exists among members of the learning community at Wheaton and degrade the College's educational and research mission.

ACCOMMODATIONS. (See "Learning and Accessibility Services" on the "Academic Information" website).
Wheaton College is committed to providing access and inclusion for all persons with disabilities, inside and outside the classroom. Students are encouraged to discuss with their professors if they foresee any disability-related barriers in a course. Students who need accommodations in order to fully access this course's content or any part of the learning experience should connect with Learning and Accessibility Services (LAS) as soon as possible to request accommodations http://wheaton.edu/las (Student Services Building -Suite209, las@wheaton.edu, phone 630.752.5615). The accommodations process is dynamic, interactive, and completely free and confidential. Do not hesitate to reach out or ask any questions.

BEHAVIOR POLICY. (See "Classroom Demeanor" on the "Academic Information" website).

GENDER-INCLUSIVE LANGUAGE. (See "Gender Inclusive Language" on the "Academic Information" website).
Please be aware of Wheaton Colleges policy on inclusive language, "For academic discourse, spoken and written, the faculty expects students to use gender inclusive language for human being."

TITLE IX AND MANDATORY REPORTING. Wheaton College instructors help create a safe learning environment on our campus. Each instructor in the college has a mandatory reporting responsibility related to their role as a faculty member. Faculty members are required to share information with the College when they learn of conduct that violates our Nondiscrimination Policy or information about a crime that may have occurred on Wheaton College's campus. Confidential resources available to students include Confidential Advisors, the Counseling Center, Student Health Services, and the Chaplain's Office. More information on these resources and College Policies is available at http://www.wheaton.edu/equityandtitleIX.

WRITING CENTER. The Writing Center is a free resource that equips undergraduate and graduate students across the disciplines to develop effective writing skills and processes. This academic year, the Writing Center is offering in-person consultations in our Center in Buswell Library, as well as synchronous video consultations onlin Make a one-on-one appointment with a writing consultant here [https://wheaton.mywconline.com/].