CSCI 381

Machine Learning (As CSCI 394. Seminar: Machine Learning)

Spring 2021 MFW 12:55–2:05 pm CHR 223

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

Thomas VanDrunen

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

Office: SCI 163 Office hours: 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. Deisenroth, Faisal, and Ong. *Mathematics for Machine Learning.* Cambridge UP, 2020.

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 scratch.

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, sandwiched between a general introduction at the beginning (including a summary of probability and statistics background) and an exploration of ethical questions at the end. For a schedule, see the course website.

I. Introduction and basic definitions

II. Probability and statistics background

III. Machine learning techniques

   A. Linear regression
   B. Gaussian mixture models and expectation-maximization
   C. Neural nets
      1. Multi-layer perceptrons
      2. Deep learning
   D. Support vector machines
   E. Principal component analysis
   F. Reinforcement learning

IV. Ethics

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 project)

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 and occasionally short conceptual assignments and/or a quiz to enforce the reading and concepts.

Readings. Students are expected to do all the assigned readings. Occasional quizzes will be given and/or short exercises or summaries will be assigned to enforce that students do the readings.

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`.

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.

PROJECTS. Most of students' work in this course outside of class meetings will be in projects. We expect five projects, each implementing one of the following machine learning techniques or algorithms:

1. k-nearest neighbors
2. Expectation-maximization
3. Multilayer perceptron training
4. Support vector machines
5. Principal component analysis

TESTS. There will be two tests, the first to be held during class time about a third of the way through the course, and the other held during the final exam block (Tues, May 4, 10:30am).

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, achieving at least 50% of the points for projects, and having at least a 50% average on the three tests.

For students who have met the minimum requirements, their *semester score* is the geometric mean of their participation score (lab activities and other, smaller assignments), their project score, and their test score, with projects counted twice. That is, your semester score is

$$\sqrt[4]{Participation \cdot Projects^2 \cdot Tests}$$

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

Policies etc

ACADEMIC INTEGRITY. Collaboration among students in the class is permitted on projects and most assignments. All code turned in for projects and similar assignments must 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 projects, 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 project may be turned in after the last day of class (Apr 30), thus if a project 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 project is due, or earlier. Beyond the allowance for late days, projects will not be accepted late. Other assignments will not be excepted late.

Attendance. Students are expected to attend all class periods, in person whenever possible. It is courtesy to inform the instructor when a class must be missed or when you must attend virtually.

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, when Test 2 is held, is Tuesday, May 4, at 10:30 am. I do not allow students to take finals early (which is also the college's policy), so make appropriate travel arrangements.

Gender-inclusive language. The college requires the following statement to be included on all syllabi: *For academic discourse, spoken and written, the faculty expects students to use gender inclusive language for human beings.*

Confidentiality and mandatory reporting. I'm committed to help maintain a safe learning environment on campus. As a faculty member I am required to share with College authorities any information about sexual misconduct 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 the college's policies is available at `www.wheaton.edu/sexualassaltresponse`.

Special needs. *Institutional statement:* Wheaton College is committed to providing reasonable accommodations for students with disabilities. Any student with a documented disability needing academic adjustments is requested to contact the Academic and Disability Services Office as early in the semester as possible. Please call 630.752.5941 or send an e-mail to `jennifer.nicodem@wheaton.edu` for further information.

My own statement: If you have a documented need for accommodations, I will have received a letter on your behalf from the Disability 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.

Office hours. Please schedule office hours through Calendly. I am trying to make myself available as much of the time as possible, but times may vary from week to week. Normal office hour times are almost-all-day Tuesday and Thursday and 8:30-10pm Monday through Thursday. I am not normally available during the day Monday, Wednesday, and Friday.

Electronic devices. Under normal circumstances my intent is for my courses to be electroinic-device-free zones. But these aren't normal circumstances. So instead I ask of you, whether you are joining class in-person or remotely, *please do not use your laptop, table, phone, etc, for anything other than class activities.* "Class activities" means looking at an electronic version of the textbook, looking at your solutions to daily work, taking notes, and using Zoom (if joining class remotely). Please refrain from from all other uses of electronic devices. In particular, **NO TEXTING OR USING SOCIAL MEDIA DURING CLASS MEETINGS.**

All this, the Lord willing.