CSCI 394

Seminar: Machine Learning

Spring 2019 MFW 12:55am–2:05pm Meyer 131

http://cs.wheaton.edu/~tvandrun/cs394

Thomas VanDrunen

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

Office: SCI 163 Office hours: MWF 3:30–4:30pm;
Th 9:00–10:30am, 11-11:30 and 1:15–3:15pm.

Contents

CATALOG DESCRIPTION (OFFICIAL) Selected topics in Computer Science at each offering, including such subjects as object-oriented design, e-commerce, human computer interface, networking services. May be taken again when a different topic is offered. Prerequisite: Departmental approval.

CATALOG DESCRIPTION (FOR THIS OFFERING) Theory, algorithms, and applications of machine learning, including the statistical background. Machine learning methods including k nearest neighbors, linear regression, neural networks, support vector machines, and expectation maximization. Ethical considerations for how machine learning applications are used and how they affect society. Prerequisites: CSCI 345 and MATH 245.

TEXTBOOK.
Christopher M Bishop, *Pattern Recognition and Machine Learning,* Springer, 2011.

In addition to the required, adopted textbook, I am consulting the following books in my preparation of the course content:

- Kevin Murphy and Francis Bach, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.
- David Barber, *Bayesian Reasoning and Machine Learning*, Cambridge UP, 2012.
- Ian Goodfellow et al, *Deep Learning*, MIT Press, 2016.
- 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.
- Stephen Marsland, *Machine Learning: An Algorithmic Perspective*, Chapman and Hall, 2014.
- Joel Grus, *Data Science from Scratch: First Principles with Python.* O'Reilly 2015.

PURPOSE OF THE COURSE. This an experimental run of what presumably will become a new computer science elective. We currently lack anything related to AI or data science in our elective line-up, so this topic is an obvious one to add. It also was a popular choice in a straw poll conducted among computer science majors in Nov 2017 about possible elective topics. Thank you for coming along on the journey.

That said, this offering will be a *very* rough cut. I'm learning this stuff myself as we go along, and I don't have all the schedule and other details worked out. Be *agile*.

There are several directions a machine learning course can go. It could be an applied probability and statistics course that examines the statistical basis for machine learning. Statisticians and

data scientists with a deep understanding of the statistical models can use standard packages (in Matlab, R, Python, etc) effectively for applications. But that is not this course, since we assume no background in probability and statistics.

At the other extreme one could teach a course that hides as much of the math as possible. Students would learn a catalog of methods by what they do, with only the most abstract understanding of how or why they work. This allows students to patch together machine learning applications quickly.

Both of those approaches sidestep the *algorithms* of machine learning, relying instead on off-the-shelf implementations. My aspiration for this course is a balanced approach (or maybe "full stack," to use a buzzword) that provides both serious-though-introductory coverage of the probability and statistics background and opportunity to try out applications, but moreover emphasizes the algorithms for training models.

COURSE OUTLINE. (For a schedule, see the course website.)

I. Introduction and basic definitions

II. Probability and statistics background

III. Basic techniques

   A. k-nearest neighbors
   B. Linear regression
   C. Naïve Bayes

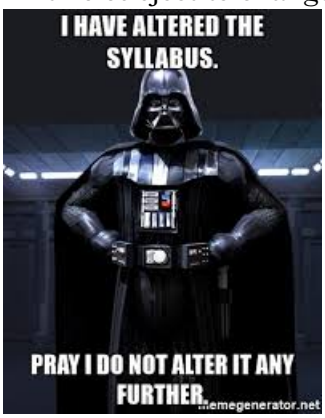
IV. Neural nets

V. Support vector machines

VI. Expectation maximization

VII. Ethics

All this subject to change. . .



LEARNING OUTCOMES. In light of the course's purpose and and the outline of topics, at the end of the course, students will be able to

1. Articulate and use the basic concepts and terminology of machine learning.
2. Derive formulas for statistical models used in machine learning.
3. Identify the uses and liabilities of various machine learning techniques.
4. Use machine learning algorithms—both their own implementations and those from available libraries—to develop machine learning applications.
5. Identify ethical concerns raised by machine learning applications and engage them from a biblical perspective.

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 doing good, to God's glory.

Darth Vader meme from https://memegenerator.net/instance/62411794/darth-vader-i-have-altered-the-syllabus-pray-i-do-not-alter-it-any-further

Course procedures

How we do this course. As an experimental run of a course, it is important for students to be active participants. The instructor will be not as much of a content deliverer as a more senior member of a team as we learn together. The general rhythm of the course is we will meet in the lab most mondays for a hands-on experience; students will learn material from reading the textbook; in non-lab class days the instructor will provide background and clarification to the readings, and students will ask questions on what they read; students will apply the ideas in multi-week programming projects; students will demonstrate their retention and comprehension of the material on the exams.

Readings. Students are expected to do all the assigned readings. The instructor will classify the readings into those that need to be read before the material is addressed in class, those that need to be read after the material is introduced in class, and those that should be skimmed before and read carefully after.

Participation. Various activities will be used to enforce and measure students' participation in class such as quizzes, short assginments, and completion of lab activities.

Projects. Besides readings, most of the work in this course will be in projects. The projects themselves are still to be determined, but at this point I estimate about five.

Projects, labs, and short assignments that involve code will be in Python. 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.

Tests. There will be two tests. The first will be held around one-third of the way through the course (scheduled for Wed, Feb 20 as of the writing of this syllabus) and will cover that first third of the material (basic machine learning concepts, probability and statistics background, and simple techniques). The second test will be held during our final exam block, Thurs May 9, 1:30-3:30 pm and will cover the second two-thirds of the course material—it will not be explicitly cummulative, though it will implicitly, naturally rely on material from earlier in the course.

Grading. Your grade will be based on three things: your *participation score* (sum of points acquired through quizzes, short assignments, labs, etc), your *project score* (sum of points acquired through projects), and your *test score* (cummulative score on the two tests, with the second test weighted twice as much as the first test).

In order to pass the course (that is, to receive a D grade or better), a student must achieve at least 50% of available points on **each** of these individual scores.

For students who have met the minimum requirements, their *semester score* is the geometric mean of these three scores. That is, your semester score is

$$\sqrt[3]{Participation \cdot Projects \cdot Tests}$$

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.

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 (May 3), thus if a project is due on May 3 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. 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, when Tests 3 and 4 are held, is Thursday, May 9, at 1:30 pm. 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. I try to keep a balance: Stop by anytime, but prefer my scheduled office hours. Any time my door is closed, it means I'm doing something uninterruptible, such as making an important phone call. Do not bother knocking; instead, come back in a few minutes or send me an email.

Dress and deportment. Please dress in a way that shows you take class seriously—more like a job than a slumber party. (If you need to wear athletic clothes because of activities before or after class, that's ok, but try to make yourself as professional-looking as possible.) If you must eat during class (for schedule or health reasons), please let the instructor know ahead of time; we will talk about how to minimize the distraction.

Electronic devices. My intent is for class to be an electronic-device-free zone. Please keep all laptops, tablets, phones, etc, silenced and put away. If you absolutely need to check your phone for something, please discreetly step out in to the hall. **NO TEXTING IN CLASS.**