Foundations of Machine Learning
cs4262-5262

Jonathan Sprinkle

Day/Time: MWF 10:10-11:00 pm
Location: Featheringill 298

https://brightspace.vanderbilt.edu/d2l/home/403139

Instructor: Jonathan Sprinkle
Institute for Software Integrated Systems, 1025 16th Ave S.
Email: jonathan.sprinkle@Vanderbilt.Edu

Teaching Assistant: Quan Liu
FGH 333
Email: quan.liu@Vanderbilt.Edu

Office Hours: Monday, 1:30-2:30
And by appointment.
Instructor may cancel office hours **without** advance Brightspace/email notice.

TA Office Hours: Wed. 1:30-2:30
TA may hold office hours via zoom if requested.
TA may cancel office hours **with** advance Brightspace/email notice.

Textbook(s): There are no plans for *required* textbook purchases. Many references may be provided, which will be made available freely, but which may contain required reading.
Some general supplemental references include:
(1) *An Introduction to Statistical Learning.* G. James, D. Witten, T. Hastie, and R. Tibshirani. Available at https://www.statlearning.com
(2) *Pattern Recognition and Machine Learning* Christopher M. Bishop.

Prerequisites: CS 3251; one of MATH 2810, 2820, or 3640; one of MATH 2410, 2500, 2501, or 2600.

I want you to succeed in this course. Through this syllabus, you can understand the commitment you should make in order to be successful. It is my wish that you will, at the end of this course, be able to interpret how machine learning can solve problems, respect where its shortcomings are, and understand how it can impact society. I hope to make the material interesting and approachable—enough that you can continue to pursue this topic in further courses, or as part of your career in industry or academia.

My commitment in this class is to use the time set aside for meetings in approaching the course content through presented materials, as well as through group exercises or individual reflection. My primary modality will involve handwritten materials that we synchronously edit together, so we can move at the same pace.

Your commitment in this class is to review materials prior to coming to our meetings, and to commit to come to class meetings and to participate in group and individual exercises. Your attendance may not be required in all class meetings—but missing those meetings may mean you are unable to have an equivalent experience.

Course Description:

This course provides theoretical and algorithmic foundations of machine learning and statistical pattern recognition. Material will be presented through lectures, homework assignments (involving both theory and programming), and a final project. The set of topics and areas covered by this course, and upon which you may be tested, include:

- regression;
- support vector machines;
- neural networks;
- clustering and dimensionality reduction;
- bias/variance tradeoffs;
- feature selection;
- data pre-processing.

We will also highlight applications of machine learning, and discuss issues in ethics and data privacy.

In addition to the previous topics, cs5262 students are expected to undertake a more extensive final project, with an expectation that project work is at a level that it could be submitted to a technical conference relevant to the project topic. This may include fewer assumptions on ingested data, more analysis of the scalability of an approach, more detailed comparisons to the state of the art, and myriad other differences.

Students with Disabilities:

Most students learn in this course through direct lecture on the board, and through completion of homework assignments. I may add a limited number of online videos of worked problems, and derivations that will be skipped in class, in favor of examples to be performed in class.

If you have a learning or physical disability, or if you learn best utilizing a particular method, please discuss with me how I can best accommodate your learning needs. I am committed to creating an effective learning environment for all learning styles. However, I can only do this successfully if you discuss your needs with me in advance of the quizzes, papers, and notebooks. I will maintain the confidentiality of your learning needs. If appropriate, you should contact Vanderbilt's Student Access Services to get more information on accommodating disabilities. Visit Student Access for more information on how to get started.

Communications outside of Meeting Times

The instructor and teaching assistant will be responsive on email during office hours, if there are no synchronous appointments during office hours. Email communications outside of this time will incur delay, due to research, service, and work-life-balance. If an office hour has passed since sending an email and you have received no reply, please re-send it to ensure that it is not lost.

Course Outline:

The listing of weekly course lecture topics may be found on Brightspace, and is summarized in the following Table. The topics schedule is subject to change without notice due to class progress—changes in due dates for assignments and quizzes will be rare, and distributed through both Brightspace and email in addition to in-class announcements. In the event of class cancellation, advance notice *via Brightspace and/or email* will be given, but any homework due that day will still be due unless otherwise notified via email.

Week of	**Monday Topic**	**Wednesday Topic**	**Friday Topic**	**Other**
01/09	Overview and Logistics	Linear Regression	Python Tutorial	
01/16	- No class meeting -	Linear Regression	Linear Regression	Homework 0 Due, 01/20
01/23	Probabilistic interpretations	Polynomial regression	Model selection, cross validation	Homework 1 Due, 01/27 Project Teams due 01/27
01/30	Locally weighted regression	Locally weighted regression	Classification: logistic regression	Project Proposals Due, 02/03
02/06	Logistic regression (cont.)	Regularization	Classification: Gaussian discriminant analysis	Homework 2 Due, 02/10
02/13	Gaussian discriminant analysis (cont.)	Classification: Naïve Bayes	Classification: Naïve Bayes	
02/20	Naïve Bayes (cont.)	Classification: Support Vector Machines	SVMs (cont.)	Homework 3 Due, 02/24
02/27	SVMs (cont.)	Recap and midterm review	Midterm Quiz	Midterm Quiz, 03/03
03/06	Feature selection, performance metrics	K-nearest neighbors	Clustering	Homework 4 Due, 03/10 Project Interim Reports due, 03/10
03/13	- Spring -	- Break -	No class meetings	
03/20	Dimensionality Reduction, PCA	Dimensionality Reduction, PCA	Ethics and data privacy	Project Interim Reports due, 03/24
03/27	Intro to neural networks	Neural networks and deep learning	Neural networks and deep learning	Homework 5 Due, 03/31
04/03	Recap and course review	Final Quiz	Final project presentations	Final Quiz, 04/05
04/10	Final project presentations	Final project presentations	Final project presentations	
04/17	Final project presentations	Final project presentations	Final project presentations	
04/24		Final Project Paper Due: W 4/26, 2:00–5:00 pm, no late days		

Grade Policy:

cs4262-5262		Grade Demarcation	
Homework:	40%	$[90-100]$	A
Quiz 1:	20%	$[80-90)$	B
Project:	30%	$[70-80)$	C
Quiz 2:	10%	$[60-70)$	D
		$[0-60)$	E

Assignment breakdown is divided equally between assignments. Assignment of grades is done according to a "modified-contract" method. The above scale represents a minimum guarantee. However, the instructor reserves the right to "upward curve" the final grade of the entire class, or of one or more individuals whose objective performance improves as the term progresses.

Homework Assignments:

These will be primarily Python-based, but may include mathematical problem solving/derivation, short answers to theoretical or practical problems (included with your Python notebook), separate writing assignments, etc. All Homework Assignments will be posted and submitted via Brightspace. You are encouraged to form study groups, but you should write up solutions to homework assignments independently. Any submitted homework assignment discussed with your study group should list names of anyone you worked with. Don't google for answers, don't look at previous years' solutions. List any external resources you used.

Project Companion:

The final project is an opportunity to investigate machine learning approaches on a problem of your choice, and to gain more experience working with "real world" datasets. You are encouraged to identify a problem and domain that interests you! Please feel free to contact us anytime for feedback or assistance, and look to the Project Companion to help schedule and scope your project.

Late Days:

You have a total of 5 late days that can be used for the course assignments. Once these late days are exhausted, any assignment turned in late will be penalized 20% *per late day.* If a project milestone is submitted late, and one or more project members have no remaining late days, only those project members will be penalized. These will be counted automatically in Brightspace; you don't have to email the course staff in order to use them. **N.B. You *cannot* use any late days on the final project writeup.**

Participation:

Your active participation in this class will aid in your comprehension of the topics, and set a standard for active learning the rest of your life. You are encouraged to take notes (even if the lecture notes will be available) as you will learn and remember more when you write it down. Please commit to reviewing previous lecture notes, and reading assigned materials, before class.

Academic Integrity:

This material is such that it welcomes collaboration among students across disciplines—and I encourage you to form study groups in order to understand the fundamentals better, and address individual questions prior to coming to office hours.

However, it may not be so easy to know whether your study group is infringing on issues of academic integrity. For that, the following recommendations should be observed:

1. **Please ask if you are uncertain**.
2. Studying together is okay, but write down solutions and code *independently*. Any submitted homework assignment discussed with your study group should list names of anyone you worked with.
3. Don't google for answers, don't look at previous years' solutions. Resist these urges to find the answer fastest, and your reward will be a deeper understanding of the content. *Referring to previous years' solutions, even without transcribing them as your answers, is an honor code violation.*
4. By all means, google for answers to python syntax challenges, and errors in the use of external libraries. But don't cross the line into looking for answers that tie to specific intellectual problems in this course.
5. You may not copy code from online resources unless explicit permission has been given in the assignment description. If you use any online resources in completing the assignment (beyond looking up syntax of a programming language), you must acknowledge the source on your homework submission. Please ask if uncertain.

If you are struggling with the content, or think you need quick answers because of other deadlines, use your late days or come to see me in office hours. Please familiarize yourself with the Honor Code (see https://www.vanderbilt.edu/student_handbook/the-honor-system/). If violations of the Honor Code occur as a matter of this course, I am obligated to refer them to the Honor Council.

Class attendance:

- If you feel sick, or may have been in contact with someone who is infectious, stay home. Except for seeking medical care, avoid contact with others and do not travel.
- Notify me through email (even if through a friend or relative who is writing for you), if you will be missing for an extended period.
- Quizzes may be missed only due to illness, and excusal for a quiz requires a note from campus health. Quizzes may not be made up or rescheduled if missed without an excusable illness. This means that an interview cannot prompt rescheduling of a quiz or other due date. Exceptions require a letter from the Dean of Students or equivalent Dean. The dates of quizzes should be excluded from your availability for interviews.

Life challenges:

If life, graduation, classes, social pressure, whatever is coming at you hard and you aren't sure what to do, the University Counseling Center can provide a consultation. Many successful professionals are successful now, because they had access to counseling when they needed it most. Visit https://www.vanderbilt.edu/ucc/ or call (615) 322-2571 to find out how to take the next step.