

Applied Machine Learning

INFO 4604/5604; Fall 2020

Monday, Wednesday, Friday; 12:40PM–1:30PM

This course will meet remotely

Zoom: <https://cuboulder.zoom.us/meeting/91652413042>

Canvas: <https://canvas.colorado.edu/courses/62561>

Abram Handler

Instructor, Information Science

E-mail: abram.handler@colorado.edu

Office hours: Tuesday, Thursday from 1PM to 2 PM (<https://cuboulder.zoom.us/my/abehander>)

Course Description

This course offers an applied introduction to machine learning. It focuses on using existing software tools to solve practical problems with machine learning methods. The course also presents more theoretical perspectives, with an eye towards selecting, analyzing, debugging and interpreting methods in practice.

This course requires proficiency with Python programming; students should be comfortable learning new Python libraries with minimal support. There is no specific math requirement for the course, however students should feel comfortable with mathematical notation and quantitative reasoning. Prior exposure to concepts from probability and discrete math (e.g. INFO 2301) will help.

The course offers:

- An introduction to foundational concepts in machine learning
- Hands-on practice applying machine learning methods to practical problems
- An introduction to machine learning tools in Python

By the end of this semester you will be able to:

- Identify when machine learning can help solve a problem
- Identify which approaches are appropriate for a particular problem
- Be comfortable using standard machine learning tools in Python
- Understand underlying concepts well enough to use and modify machine learning code

4605 vs 5604

Students in this class are enrolled in either 4604 (undergraduate-level) or 5604 (graduate-level). **Graduate students should be enrolled in 5604.** Students in 5604 students will be assigned additional readings, as well as additional problems on homeworks, quizzes and exams. 4604 students can do the 5604 problems for extra credit.

Remote Course Policies

Class will meet three times per week (Monday, Wednesday, Friday) from 12:40PM–1:30PM on Zoom. There will be lectures and in-class activities during class time. **Attendance is required.** We will often follow a

[flipped classroom](#) model. You will be asked to watch a prerecorded lecture before class. During class, you will work on activities based on the lecture. This will be clearly indicated on Canvas.

In order to fully participate remotely, students will need access to a personal computer with a web camera, a reliable high-speed internet connection, and a minimally disruptive background environment. You will often do coding assignments during class so please join with a computer and not a phone. Please email the instructor immediately if you expect any issues with remote participation.

Books

Required

4604/5604: [Raschka and Mirjalili \(2019\) Python Machine Learning, 3rd Edition](#). Available as an e-book from [Packt](#). Earlier editions are also acceptable. Companion code is available [here](#).

5604 only: [Understanding Machine Learning: From Theory to Algorithms](#). Available as a free e-book.

Optional

[Kevin Murphy, Machine Learning: a Probabilistic Perspective](#). This is a standard textbook for machine learning courses, with comprehensive coverage of the field. It is useful both as an introduction and as a reference book for students who want to do research in machine learning.

[Diez, Cetinkaya-Rundel and Barr, OpenIntro Statistics](#). Available as a free e-book. This book introduces many of the mathematical concepts used in this course, and you may find it to be a helpful supplement. Material on probability may be especially useful.

Quizzes

This class will include regular, announced quizzes. Quizzes will be open book. You can use any materials from the course (or any website, though this will not be necessary) to answer questions on a quiz. Because everyone sometimes has a bad day or needs more time to master a new concept, the instructor will automatically drop the lowest five quiz scores from your final grade.

In-class activities

This class will sometimes include many in-class coding activities, which you will complete in breakout rooms. You should expect to work with the people in the breakout room as you do the assignment. On random days, you will turn in in-class activities to Canvas at the end of class. These activities will be graded loosely based on effort and correctness (check minus, check, check plus), with an understanding that you have limited time to complete the work.

Attendance

Attendance for this class is required. The instructor will take attendance on random days to calculate your attendance score. In-class activities will also be used to calculate your attendance score.

Homeworks

This class will have regular homeworks. You may complete the homeworks alone or with one partner. If you work with a partner, you must list the partner on your assignment. Homeworks will be challenging! It will be important to start early.

Late work

The instructor understands that students are busy with many obligations. Therefore, across the whole semester, you will be allotted a total of five free late days to turn in assignments. If you turn in an assignment

within 24 hours after the deadline you will be deducted 1 late day. If you turn in an assignment within 48 hours after the deadline you will be deducted 2 late days, and so on. If you have used five or fewer late days so far during the semester (including the most recent assignment) you will not be penalized for late work. However, after you have used up your late days, late homework will not count for credit except in special circumstances.

Grading

Final grades are calculated according to the following distribution:

- 20% Homework
- 20% In-class activities
- 15% Midterm
- 20% Final Project
- 20% Quizzes (online, via Canvas)
- 5% Attendance and participation

Letter grades will follow a typical scoring distribution (A if $\geq 93\%$, A- if $\geq 90\%$, B+ if $\geq 87\%$, B if $\geq 83\%$, B- if $\geq 80\%$, C+ if $\geq 77\%$, and so on). Do not expect most grades to be curved, though exam grades may be curved if needed.

Computing Requirements

Students will need to use statistical computing software as well as teleconferencing software to participate in class. [Jupyter notebooks](#) written in Python 3 will be used for all in-class examples and assignments. We will use many Python libraries such as [scikit-learn](#). The [Anaconda distribution](#) of Python 3.5 (or above) is *strongly* recommended. It makes it easy to install all of the supporting libraries we will use in this course. Lectures will include exercises and presentations, with the expectation that students participate with their own computers. If students do not have access to a computer to use for computing or Zoom, they should immediately email the instructor. Students who require technical assistance should email the instructors with the code and data they are working with, a summary of their debugging efforts to date, and attend an instructor's office hours.

Schedule

This course will be divided into three units.

Week	Dates	Topic	Method
1	Aug 24 – Aug 28	Intro to machine learning	
2	Aug 31 – Sep 4	Geometry of data, bias and variance	KNN
3	Sep 7 – Sep 18	Parameters and learning	Perceptron
4	Sep 21 – Sep 25	Optimization and loss	Logistic regression
5	Sep 28 – Oct 2	Regularization	Logistic regression
6	Oct 5 – Oct 9	Generative models	Naive Bayes
7	Oct 12 – Oct 16	Train/test split, overfitting	All methods
8	Oct 19 – Oct 23	Non-linear prediction, ensembles	decision trees, random forests

Unit 1: Introduction to ML

Week	Dates	Topic	Details
9	Oct 26 – Oct 30	Data creation	Annotation and agreement
10	Nov 2 – Nov 6	Feature creation	The art of feature engineering
11	Nov 9 – Nov 13	Evaluation and diagnosis	Precision, recall, ROC AUC, confusion matrixes
12	Nov 16 – Nov 20	Review and midterm	

Unit 2: Practical ML

Week	Dates	Topic	Details
13	Nov 23 – Nov 27	FATE	Fairness, accountability, transparency and ethics
14	Nov 2 – Nov 6	Topic models	LDA and Mallet, applications in NLP
15	Nov 9 – Nov 13	Deep learning	Gradients, backpropagation, software packages
16	Nov 16 – Nov 20	Final presentations	

Unit 3: Other topics in ML

Course Policies

In-Class Confidentiality

The success of this class depends on students feeling comfortable sharing questions, ideas, concerns, and confusions about assignments, work-in-progress, and their personal experiences. Students may read, comment, and run on classmates' writing, code, and other class-related content for the sole purpose of use within this class. However, students may not use, run, copy, perform, display, distribute, modify, translate, or create derivative works of another student's work outside of this class without that student's expressed written consent or formal license. Furthermore, students may not create any audio, video, or other records during class time without the instructor's permission nor may students publicly share comments made in class attributable to another person's identity without that person's permission.

Instructor Interaction

The instructor will check e-mail between 8:00 and 18:00 on non-holiday business days and try to respond to emails within 24 hours. E-mailing the instructor or coming to (remote) office hours are the best ways to get help and feedback outside of lecture.

Accommodations for Disabilities

I am committed to providing everyone the support and services needed to participate in this course. If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to the instructor in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the www.colorado.edu/disabilityservices/students. Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition or injury, see Temporary Medical Conditions under the Students tab on the Disability Services website and discuss your needs with the instructors.

Religious Observance

Campus policy regarding [religious observances](#) requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required assignments/attendance. If this applies to you, please e-mail the instructor as soon as possible to make the appropriate accommodations.

Classroom Behavior

Students and instructors each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, color, culture, religion, creed, politics, veteran status, sexual orientation, gender, gender identity and gender expression, age, ability, and nationality. Class rosters are provided to the instructor with the student's legal name. The instructor will honor your request to address you by an alternate name or gender pronoun. Please advise the instructor of this preference early in the semester in order to make appropriate changes. For more information, see the policies on [class behavior](#) and the [student code](#).

Harassment and Discrimination

The University of Colorado Boulder (CU Boulder) is committed to maintaining a positive learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct, discrimination, harassment or related retaliation against or by any employee or student. CU's [Sexual Misconduct Policy](#) prohibits sexual assault, sexual exploitation, sexual harassment, intimate partner abuse (dating or domestic violence), stalking or related retaliation. CU Boulder's [Discrimination and Harassment Policy](#) prohibits discrimination, harassment or related retaliation based on race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. Individuals who believe they have been subject to misconduct under either policy should

contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127. Information about the OIEC, the above referenced policies, and the campus resources available to assist individuals regarding sexual misconduct, discrimination, harassment or related retaliation can be found at the [OIEC website](#).

Honor Code

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the [academic integrity policy](#) of the institution. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, resubmission, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code Council (honor@colorado.edu; 303-735-2273). Students who are found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code Council as well as academic sanctions from the faculty member. Additional information can be found at honorcode.colorado.edu.

Illness

Should a student contract any illness that requires mandatory sequestration, intensive medical treatment, or extended convalescence and disrupts their ability to participate in class and complete assignments, the instructors will try to accommodate their condition without penalty with extensions and incompletes. This also applies if the student has a family member whose diagnosis, treatment, and recovery will affect their ability to participate. Students should notify the instructors as soon as possible of events that will impact their engagement with the class so that we can triage and develop an accommodation plan rather than scrambling at the end of the semester.

Canvas

Once the semester begins, this PDF version of the syllabus will be revised infrequently and any revised requirements will be posted as announcements and updated course schedule to Canvas. The instructor might make changes to the course's schedule, evaluation criteria, policies, *etc.* through announcements in class and on Canvas, so please check Canvas regularly. If you have questions, please email the instructor.

Acknowledgements

This syllabus was typeset in \LaTeX using [Overleaf](#) with the [fbb/Bembo](#) font and is derived from the `memoir` styles adapted by [Kieran Healy](#) and [Benjamin 'Mako' Hill](#). It is modified from a syllabus from [Brian Keegan](#). This course follows curriculum developed by [Michael J. Paul](#).