

# Introduction to Machine Learning (PhD)

Spring 2020, CMU 10701

**Lectures:** MW, 1:30-2:50pm, Wean Hall

**Recitations:** F, 1:30-2:50pm, Wean Hall

**Instructors:**

Tom Mitchell

Leila Wehbe

**Assistant Instructor:** Brynn Edmunds

**Teaching Assistants:**

Emily Black

Michael Kronovet

Giulio Zhou

Xutong Ren

Dinc Basar

Himanshi Yadav

Oscar Dadfar

**Communication:** [Piazza](#) will be used for discussion about the course and assignments.

**Office Hours:**

- Tom: TBA
- Leila: TBA
- Emily Black: TBA
- Michael Kronovet: TBA
- Giulio Zhou: TBA
- Xutong Ren: TBA
- Dinc Basar: TBA
- Himanshi Yadav: TBA
- Oscar Dadfar: TBA

# Course Description

Machine learning studies the question "How can we build computer programs that automatically improve their performance through experience?" This includes learning to perform many types of tasks based on many types of experience. For example, it includes robots learning to better navigate based on experience gained by roaming their environments, medical decision aids that learn to predict which therapies work best for which diseases based on data mining of historical health records, and speech recognition systems that learn to better understand your speech based on experience listening to you.

This course is designed to give PhD students a thorough grounding in the methods, mathematics and algorithms needed to do research and applications in machine learning. Students entering the class with a pre-existing working knowledge of probability, statistics and algorithms will be at an advantage, but the class has been designed so that anyone with a strong numerate background can catch up and fully participate.

If you are interested in this topic, but are not a PhD student, or are a PhD student not specializing in machine learning, you might consider the master's level course on Machine Learning, 10-601. 10-601 may be appropriate for MS and undergrad students who are interested in the theory and algorithms behind ML. You can see an ML course comparison [here](#).

## Prerequisites

Students entering the class are expected to have a pre-existing working knowledge of probability, linear algebra, statistics and algorithms, though the class has been designed to allow students with a strong numerate background to catch up and fully participate. In addition, recitation sessions will be held to revise some basic concepts.

## Resources

Supplementary readings:

[TM] *Machine Learning*, Tom Mitchell.

[KM] *Machine Learning: A probabilistic perspective*, Kevin Murphy.

[CB] *Pattern Recognition and Machine Learning*, Christopher Bishop.

[HTF] *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Trevor Hastie, Robert Tibshirani, Jerome Friedman.

## Schedule

Tentative schedule is found on [the class website](#), might change according to class progress and interest. Every Friday class is intended to be a recitation.

## Grading

The final grade will be determined as follows:

Exam 1	25%
Exam 2	25%
Homework	25%
Project	25%

## Course Policies

The following policies are adapted from Matt Gormley's [10-601 Fall 2019 Course Policies](#).

### Homework

There will be 5-7 homework assignments. Answers will be submitted on Gradescope and code portions through Autolab.

### Projects

There will be three deliverables (midway report, final report and a review of projects from other students). Each team member's contribution should be highlighted. You should use the project as an opportunity to "learn by doing".

We will have a set of 5 project topics that you can choose from. Alternatively, you can pick your own topic. To be able to do so, you must write a 1-2 page proposal that includes mid-way milestones, schedule a meeting with a TA and convince them of your project.

You may work in teams of 2-3. A student can do a project on their own only in the following situation: (1) you want to design a project around a data set associated with your ongoing research, and (2) you absolutely cannot find another student in our class who would be interested in your problem, and (3) your faculty advisor associated with that ongoing research writes us an email stating that she will be happy to serve as mentor for your class project.

For assistance with the written reports, visit the Global Communication Center (GCC). GCC tutors can provide instruction on a range of communication topics and can help you improve your papers, presentations, and job application documents. The GCC is a free service, open to all students, and located in Hunt Library. You can make tutoring appointments directly on the GCC website: <http://www.cmu.edu/gcc>. You may also browse the GCC website to find out about communication workshops offered throughout the academic year.

## Programming questions

You will submit your code for **programming questions** on the homework to Autolab or Gradescope. After uploading your code, our grading scripts will autograde your assignment by running your program on a VM. This provides you with immediate feedback on the performance of your submission.

**Note: Autolab/Gradescope automatically checks your code against code from other students in the class to identify copying.**

## Gradescope

We use Gradescope to collect PDF submissions of **open-ended questions** on the homework (e.g. mathematical derivations, plots, short answers). The course staff will manually grade your submission, and you'll receive personalized feedback explaining your final marks.

**Regrade Requests:** If you believe an error was made during manual grading, you'll be able to submit a regrade request on Gradescope. For each homework, regrade requests will be open for a maximum of 1 week after the grades have been published. This is to encourage you to check the feedback you've received early!

## Late homework policy

You receive 6 total grace days **for use on any homework assignment except HW1**. We will automatically keep a tally of these grace days for you; they will be applied greedily. No assignment will be accepted more than 4 days after the deadline. This has

the important implication that you may not use more than 4 grace days on any single assignment.

All homework submissions are electronic (see Technologies section below). As such, lateness will be determined by the latest timestamp of any part of your submission. For example, suppose you submit the code part of the homework on time but the written part one hour late, you would have used one of your late days.

## Extensions

In general, we do not grant extensions on assignments. There are several exceptions:

- **Medical Emergencies:** If you are sick and unable to complete an assignment or attend class, please go to University Health Services. For minor illnesses, we expect grace days or our late penalties to provide sufficient accommodation. For medical emergencies (e.g. prolonged hospitalization), students may request an extension afterwards and should include a note from University Health Services.
- **Family/Personal Emergencies:** If you have a family emergency (e.g. death in the family) or a personal emergency (e.g. mental health crisis), please contact your academic adviser or Counseling and Psychological Services (CaPS). In addition to offering support, they will reach out to the instructors for all your courses on your behalf to request an extension.
- **University-Approved Absences:** If you are attending an out-of-town university approved event (e.g. multi-day athletic/academic trip organized by the university), you may request an extension for the duration of the trip. You must provide confirmation of your attendance, usually from a faculty or staff organizer of the event.

For any of the above situations, you may request an extension **by emailing the assistant instructor(s) at [bedmunds+10601@andrew.cmu.edu](mailto:bedmunds+10601@andrew.cmu.edu) – do not email the instructor or TAs**. The email should be sent as soon as you are aware of the conflict and at least **5 days prior to the deadline**. In the case of an emergency, no notice is needed.

## Audit Policy

Official auditing of the course (i.e. taking the course for an “Audit” grade) is not permitted this semester.

Unofficial auditing of the course (i.e. watching the lectures online or attending them in person) is welcome and permitted without prior approval. We give priority to students taking the course for a letter grade, so auditors may only take a seat in the classroom is

there is one available *10 minutes after the start of class*. Unofficial auditors will not be given access to course materials such as homework assignments and exams.

## **Pass/Fail Policy**

We allow you take the course as Pass/Fail. Instructor permission is not required. What grade is the cutoff for Pass will depend on your program. Be sure to check with your program / department as to whether you can count a Pass/Fail course towards your degree requirements.

## **Accommodations for Students with Disabilities**

If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at [access@andrew.cmu.edu](mailto:access@andrew.cmu.edu).

## **Important Note!**

We want to help you learn the material, and understand unexpected issues arise throughout the semester. If something unexpected comes up, or your are having difficulties that are impacting your learning of the material, come see us! Please, please come. You should also reach out to your academic advisor so they are aware of the situation.

# **7. Academic Integrity Policies**

## **Read this carefully!**

(Adapted from Roni Rosenfeld's [10-601 Spring 2016](#) Course Policies.)

## **Collaboration among Students**

- The purpose of student collaboration is to facilitate learning, not to circumvent it. Studying the material in groups is strongly encouraged. It is also allowed to seek help from other students in understanding the material needed to solve a particular homework problem, provided no written notes (including code) are

shared, or are taken at that time, and provided learning is facilitated, not circumvented. The actual solution must be done by each student alone.

- The presence or absence of any form of help or collaboration, whether given or received, must be explicitly stated and disclosed in full by all involved. Specifically, each assignment solution must include answers to the following questions:

Did you receive any help whatsoever from anyone in solving this assignment? Yes / No.

- If you answered 'yes', give full details: \_\_\_\_\_
- (e.g. "Jane Doe explained to me what is asked in Question 3.4")

Did you give any help whatsoever to anyone in solving this assignment? Yes / No.

- If you answered 'yes', give full details: \_\_\_\_\_
- (e.g. "I pointed Joe Smith to section 2.3 since he didn't know how to proceed with Question 2")

Did you find or come across code that implements any part of this assignment? Yes / No. (See below policy on "found code")

- If you answered 'yes', give full details: \_\_\_\_\_
- (book & page, URL & location within the page, etc.).

- If you gave help after turning in your own assignment and/or after answering the questions above, you must update your answers before the assignment's deadline, if necessary by emailing the course staff.
- Collaboration without full disclosure will be handled severely, in compliance with [CMU's Policy on Academic Integrity](#).

## Previously Used Assignments

Some of the homework assignments used in this class may have been used in prior versions of this class, or in classes at other institutions, or elsewhere. Solutions to them may be, or may have been, available online, or from other people or sources. It is explicitly forbidden to use any such sources, or to consult people who have solved these problems before. It is explicitly forbidden to search for these problems or their solutions on the internet. You must solve the homework assignments completely on your own. We will be actively monitoring your compliance. Collaboration with other students who are currently taking the class is allowed, but only under the conditions stated above.

## Policy Regarding "Found Code":

You are encouraged to read books and other instructional materials, both online and offline, to help you understand the concepts and algorithms taught in class. These

materials may contain example code or pseudo code, which may help you better understand an algorithm or an implementation detail. However, when you implement your own solution to an assignment, you must put all materials aside, and write your code completely on your own, starting “from scratch”. Specifically, you may not use any code you found or came across. If you find or come across code that implements any part of your assignment, you must disclose this fact in your collaboration statement.

## **Duty to Protect One’s Work**

Students are responsible for pro-actively protecting their work from copying and misuse by other students. If a student’s work is copied by another student, the original author is also considered to be at fault and in gross violation of the course policies. It does not matter whether the author allowed the work to be copied or was merely negligent in preventing it from being copied. When overlapping work is submitted by different students, both students will be punished.

To protect future students, do not post your solutions publicly, neither during the course nor afterwards.

## **Penalties for Violations of Course Policies**

All violations (even first one) of course policies will always be reported to the university authorities (your Department Head, Associate Dean, Dean of Student Affairs, etc.) as an official Academic Integrity Violation and **will result in failure in the course.**

## **Take care of yourself**

Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is almost always helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at <http://www.cmu.edu/counseling/>. Consider reaching out to a

friend, faculty or family member you trust for help getting connected to the support that can help.

If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

- CaPS: 412-268-2922
- Re:solve Crisis Network: 888-796-8226

If the situation is life threatening, call the police

- On campus: CMU Police: 412-268-2323
- Off campus: 911