Class  
Meetings: 
Section 3 - MW 1:50-3:10pm 
Section 4 - MW 3:30-4:50pm  
Location: Zoom  

Lab Sessions: 
L01 - F 9:10am-10:30am or  
L02 - F 10:50am-12:10pm  
Location: Zoom  

Professor: Chris Clapp (he/him) 
Email: cclapp@uchicago.edu  
Office Hours: TBD  
or by appointment  
Location: Zoom  

TAs: TBD  
Email: TBD  
Office Hours: TBD  
Location: Zoom  

Course Description  
We live in an information age where many of our decisions and actions are tracked. Digital information is being produced and recorded at a stifling pace. This has led to an explosion of “big data” questions. Coupled with cheap computing power and expanded data storage, new developments across statistics, computer science, and data-driven social sciences like economics have led to new ways to answer those questions. This class will teach you a set of tools to detect patterns in data, then use what you have learned from the data to predict important outcomes or describe the salient relationships among inputs. While these machine learning tools are being used extensively in marketing, finance, and business, we will focus on their public policy applications.  

Learning Objectives: What’s My Incentive for Taking This Course?  
Specifically, the purpose of the course is to introduce students to a wide array of the fundamental methods in modern machine learning. Each week, we will learn about and discuss a different suite of techniques and their applications to public policy. The course objective is for students to be able to use those techniques to carry out basic machine learning analyses to inform better policy and make the world a better place, as well as to become informed and critical consumers of machine learning research.  

Overall Approach  
The world is (extra) crazy right now because of the global pandemic, racial injustice, the associated civil unrest.... We’re all dealing with a lot of new issues, disruptions, and stress. So, as a guiding principle, I’m
going to be (extra) nice to you! The class is designed to be as flexible and accommodating as possible while still maintaining academic standards. Please feel free to give me feedback on what’s working well and what is not. We will adjust the class further as needed.

If you get sick, are caring for a sick relative, have a tough situation at home, or anything else that becomes an obstacle to your coursework, please inform me and your academic advisor as soon as you are able. We will all work together to develop appropriate accommodations.

Finally, please be nice to each other, your TA, and me. This is semester is going to be challenging, but we’re all in this together!

**Prerequisites**

The official prerequisites are:

- PPHA 30535 Data and Programming for Public Policy I and
- PPHA 30536 Data and Programming for Public Policy II.

This course is the third installment of the three-quarter core sequence of the Data Science Certificate at the Harris School of Public Policy. Students at Harris and in the College may enroll, with permission of the instructor, without having taken previous courses in the sequence. However, it is necessary for MPP students to take the full sequence in order to meet the necessary requirements of the Data Science Certificate.

**Evaluation**

Your final grade in this course will be related to performance in several areas. The weight placed on each component will be as follows:

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Sets</td>
<td>55%</td>
</tr>
<tr>
<td>Midterm Quiz</td>
<td>15%</td>
</tr>
<tr>
<td>Final Quiz</td>
<td>20%</td>
</tr>
<tr>
<td>Participation</td>
<td>10%</td>
</tr>
</tbody>
</table>

There are four problem sets in this class. Each problem set will have two sections. The first section presents problems to help students cement their understanding of previously taught material, and the second section asks high-level questions about upcoming material to motivate the lecture. These questions are designed to give students time to ponder the methodological questions which will be answered in the lecture. For full credit, the latter questions need to be answered thoughtfully, but not necessarily correctly.

You are welcome (and encouraged) to form study groups to work on the problem sets together. But you should write your own code and your own solutions.

The midterm quiz will review theoretical material from the first-half of the class. A set of practice questions will be handed to the students a week before the exam.

The final quiz is of the same format as the midterm, and covers material from the whole course. A set of practice questions will be handed to the students a week before the exam.

Class participation grades will be based on your level of active, attentive, inquisitive participation in class discussions and on the discussion board. Note that regular class attendance is generally a necessary (but not sufficient) component of earning a good class participation grade. Given the constraints imposed by remote learning (both the inability to ask questions during asynchronous presentations of the theoretical material and the technological constraints that come with synchronous discussions), we will supplement this participation component by using the Canvas discussion board. Please use the discussion board to post questions and discuss the material covered in the lectures and the readings.
Grades

Grades will be distributed according to the established “Harris” curve in this class (listed in the table that follows).

<table>
<thead>
<tr>
<th>Grade</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1/8</td>
</tr>
<tr>
<td>A-</td>
<td>1/4</td>
</tr>
<tr>
<td>B+</td>
<td>1/4</td>
</tr>
<tr>
<td>B</td>
<td>1/4</td>
</tr>
<tr>
<td>B-</td>
<td>1/8</td>
</tr>
</tbody>
</table>

Pass/Fail (P/F), Withdrawal, and Incomplete grade requests will be handled in accordance with University and Harris policy. Students who wish to take the course pass/fail rather than for a letter grade must use the Harris P/F request form (https://harris.uchicago.edu/form/pass-fail) and must meet the Harris deadline, which is generally 9am on the Monday of the 5th week of courses. To earn a P grade, students taking the course P/F must: complete and submit all assignments; take all exams; and earn a grade that is overall equivalent to at least a C- letter grade.

Materials

Textbooks


  - You can download a free PDF of the book from the author’s website:
    https://faculty.marshall.usc.edu/gareth-james/ISL/.
  
  - Coding examples in the book are written in R, but you can find Python analogs here:

In addition, I will distribute copies of additional readings on Canvas. The material is covered in greater detail in the readings, and they should be considered both an essential supplement and a reference.

Data Analysis and Statistical Software

We will use Python software in this class.

Office Hours

My office hours for this class are listed on the first page of the syllabus. Those hours are for you, so please make use of them (be it with questions about course material, to discuss ideas, or just to chat). You do not need to make an appointment to see me during my office hours; just drop by. I will be on Zoom during those times. If a sufficient number of students attend at the same time and office hours become too crowded to be effective, we will make alternative arrangements.

Please make your best effort to attend during the posted times, but if you have a conflict or want to talk with me one-on-one, you are welcome to make an appointment for another time. I am happy to meet with students outside of office hours. I only ask that you do your absolute best to attend the regularly scheduled office hours since I have many students and there are economies of scale in the production of knowledge. Also, if you know in advance that you cannot make a scheduled appointment, please email me to let me know.
Course Policies

- **Recording**
  
  - The course will be taught synchronously via Zoom. I will record lectures and post them only to Canvas in accordance with University and Family Educational Rights and Privacy Act (FERPA) guidelines. Office hours will not be recorded.
  
  - FERPA is a federal statute that, broadly speaking, guarantees privacy over certain aspects of your educational records. You can view the details of the policy on the registrar’s website (https://registrar.uchicago.edu/records/ferpa/).
  
  - If you record a class, discussion section, office hours, or meeting without permission, or if you share any of the recorded videos without permission, you may be violating eavesdropping laws, copyright laws, or the FERPA statute. So do not post or share any such videos outside of Canvas. This also applies to any manipulated video.

- **General**
  
  - There is no attendance requirement (beyond the first day), but regular attendance is necessary (but not sufficient) to do well in the class.
  
  - The class webpage is available through the Canvas portal. I will use it to post announcements, assignments, and grades. Please check it regularly.
  
  - Email, Canvas postings, and the discussion board are the official means of communication for out-of-class messaging. In other words, you are expected to check your UChicago email account and the Canvas site regularly.
  
  - Email is inefficient. If you have a question about the class or the material, others probably do too! Questions and answers (knowledge) are public goods, so post your question to the discussion board, and feel free to answer questions your classmates ask. I’ll monitor and respond as well.
  
  - If you have a question or concern about something you don’t want to discuss publicly, feel free to email me. I will respond to email within 2 business days (Monday-Friday, 9:00am-5:00pm). I teach multiple classes, so please include “Machine Learning:” as a prefix to your subject.
  
  - Any and all results of in-class and out-of-class assignments and examinations are data sources for research and may be used in published research. All such use will always be anonymous.

- **Assignments and Exams**

  - No late assignments will be accepted for any reason, valid or otherwise. Not turning in an assignment, handing it in late, or failing to take an on-line assessment before the link expires will result in a grade of zero. I understand that students sometimes have legitimate reasons for being unable to complete problem sets on time or give their full effort (and that is especially true this quarter), so your lowest assignment grade will be dropped.

  Due to the pandemic, to ensure that students who have medical issues or need to care for sick family members for an extended period of time do not automatically fail the class, I will allow students to write a paper of no more than 10 pages on an application of the topic covered on a missed assignment as a grade replacement. The details of these papers will be shared should they become

---

1. The problem sets constitute the assignments, and the quizzes constitute the exams in this class.
2. Reasons include, but are not limited to: illnesses, athletic competitions, work trips, job fairs, job interviews, travel reservations, relative illnesses, relative funerals, out-of-town weddings, car accidents, car trouble, scooter trouble, tickets to see Billy Joel in concert, and emergency visits to the veterinarian with your dog.
necessary. Following the design of many of our social insurance programs, these papers will be
designed to be optimal (relative to the standard assignment) only for students who truly need to
make use of this option.

– No make-up exams will be given, except in rare cases of serious health problems, family emer-
gency, or other extenuating circumstances in accordance with Harris policy. Doing so would create
concerns about uneven treatment, and I can’t be sure that classmates won’t share information about
what was on the exam. In such a case, notification and/or documentation is required in a timely
manner. Whenever possible, you should contact me before the exam regarding your absence.

Academic Integrity

As a member of the Student Government Judicial Branch as an undergraduate and a graduate student at a uni-
versity where any non-trivial act of lying, cheating or stealing results in expulsion, I take UChicago’s Academic
Honesty & Plagiarism Policy very seriously. All students suspected of academic dishonesty will be reported to
the Harris Dean of Students for investigation and adjudication. The disciplinary process can result in sanctions
up to and including suspension or expulsion from the University. In addition, if in my judgment, the preponder-
ance of the evidence indicates that a student has committed an honor violation on an assignment, that student
will receive an immediate grade of zero for that assignment and cannot earn higher than a B- in the course,
regardless of their performance on other assignments. This is regardless of the outcome of the disciplinary
process. I trust every student in this course to fully comply with all of the provisions of UChicago and Harris’
integrity policies. Here are specific expectations:

• On exams, it is expected that you will neither receive nor give aid, nor access any material other than
  items explicitly outlined in the exam instructions.

• For other assignments, you may (and should!) work with other students, but it is expected that you will
  collaborate on all parts of the assignment (as opposed to the “divide and conquer” method).

• During the entire semester, it is expected that you will not access old problem sets, exams, answer keys,
  student presentations, or any other class material at any time. This includes websites that post solutions
  under the guise of tutoring. (These sites both facilitate cheating and steal the intellectual property of the
  author.)

• During the entire semester and thereafter, it is expected that you will neither post any class material on
  the internet nor share any class materials with other students through any other means. Furthermore, if
  you become aware that this has occurred, you are obligated to let me know immediately.

Americans With Disabilities Act

Students with disabilities needing an academic accommodation should contact UChicago’s Student Disability
Services (SDS). Please see their webpage for contact information (https://disabilities.uchicago.edu). If SDS
determines a disability accommodation is appropriate, you should inform the Harris Dean of Students office by
the end of the first week of class. The Harris Dean of Students office will work with the student and instructor
to coordinate the students’ accommodations implementation. Harris students are not required to submit their
accommodations letter to the instructor, but please feel free to come talk to me if I you are comfortable doing
so. I’m happy to help.

---

3I apologize for the heavy handed tone of this section. It is intended to protect the many honest students who take my class and
academic integrity as a whole.
Mental Health Services

Students differ in how much they know about mental health services. Your use of UChicago’s Student Health and Counseling Services (SHCS) is free, confidential and not linked to your academic file. There are no gains from suffering in silence, so please do not hesitate to make use of the services provided by SHCS if you need them. Please see SHCS’ mental health webpage for services and contact information (https://wellness.uchicago.edu/mental-health/). And if you are having serious mental, physical, or other problems, immediately contact the urgent medical care line at (773) 834-WELL.

Diversity and Inclusion

UChicago is committed to diversity and rigorous inquiry that arises from multiple perspectives, and Harris encourages thought-provoking discourse that involves not only speaking freely about all issues but also listening carefully and respectfully to the views of others. I concur with this commitment and view the diversity that students bring to my class as a valuable resource and a benefit to learning. I expect to maintain a productive learning environment based on open communication, mutual respect, and non-discrimination. I strive to present materials in a way that is respectful of diverse student backgrounds. As there can always be a gap between intent and execution, suggestions for promoting a positive and open environment are welcomed. Please feel free to correct me on your preferred name and gender pronouns if necessary.

Responsible Employees (Title IX)

All University of Chicago faculty and TAs are classified as “Responsible Employees.” As such, they are required to report any discussions of sexual misconduct, dating violence, domestic violence or stalking to the Title IX Coordinator for the University. This includes the identities of the student making the complaint and alleged perpetrator. You will receive an email once a report is filed, but you are not obligated to meet with anyone or engage in the process. Alternatively, there are “Confidential Resource” employees at the University who do not have an obligation to share identifying information. For more information, including phone numbers, see the UChicago UMatter website (https://umatter.uchicago.edu/find-support/).

Syllabus Change Policy

Except for changes that substantially affect implementation of the evaluation (grading) statement (aside from those I have already noted), this syllabus is a guide for the course and is subject to change with advance notice.

Tentative Course Outline

The weekly coverage might change as it depends on the progress of the class. The “ISL” in the “Reading” column that follows indicates the chapter in the “An Introduction to Statistical Learning” textbook that corresponds to the topic we’re covering in class that day. “PS” is an abbreviation for “Problem Set.”
<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Day</th>
<th>Topic/Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01/11</td>
<td>Mon</td>
<td>Introduction</td>
</tr>
<tr>
<td></td>
<td>01/13</td>
<td>Wed</td>
<td>Review of Basic Material</td>
</tr>
<tr>
<td>2</td>
<td>01/18</td>
<td>Mon</td>
<td>NO CLASS – MLK, Jr. Day!</td>
</tr>
<tr>
<td></td>
<td>01/20</td>
<td>Wed</td>
<td>Multivariate Linear Regression</td>
</tr>
<tr>
<td>3</td>
<td>01/25</td>
<td>Mon</td>
<td>Model Selection: Penalty Function &amp; Resampling Methods</td>
</tr>
<tr>
<td></td>
<td>01/27</td>
<td>Wed</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>02/01</td>
<td>Mon</td>
<td>Model Selection: The Bootstrap, Cross-Validation &amp; Permutation Tests</td>
</tr>
<tr>
<td></td>
<td>02/03</td>
<td>Wed</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>02/08</td>
<td>Mon</td>
<td>Priors, Shrinkage, and Regularization</td>
</tr>
<tr>
<td></td>
<td>02/10</td>
<td>Wed</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>02/15</td>
<td>Mon</td>
<td>Trees &amp; Natural Language Processing</td>
</tr>
<tr>
<td></td>
<td>02/17</td>
<td>Wed</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>02/22</td>
<td>Mon</td>
<td>Midterm</td>
</tr>
<tr>
<td></td>
<td>02/24</td>
<td>Wed</td>
<td>Trees and Random Forest</td>
</tr>
<tr>
<td>8</td>
<td>03/01</td>
<td>Mon</td>
<td>Support Vector Machines &amp; Other Classifiers</td>
</tr>
<tr>
<td></td>
<td>03/03</td>
<td>Wed</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>03/08</td>
<td>Mon</td>
<td>Unsupervised Learning &amp; High-Dimensional Causal Inference</td>
</tr>
<tr>
<td></td>
<td>03/10</td>
<td>Wed</td>
<td></td>
</tr>
</tbody>
</table>