[Download a pdf copy of syllabus](#)

Basic information

DEPARTMENT: Epidemiology

COURSE NUMBER: TBD

COURSE TITLE: Machine Learning for Causal Inference

CREDIT HOURS: 2

SEMESTER: Fall %%current_date_yyyy%%

CLASS HOURS AND LOCATION: TBD

INSTRUCTOR NAME: Ashley I. Naimi

INSTRUCTOR CONTACT INFORMATION

- **EMAIL**: ashley.naimi@emory.edu
- **SCHOOL ADDRESS OR MAILBOX LOCATION**: CNR 4013
- **OFFICE HOURS**: TBD

TEACHING ASSISTANT INFORMATION:

- **NAME**: TBD
- **EMAIL**: TBD
- **OFFICE HOURS**: TBD (location: TBD)

COURSE DESCRIPTION

This doctoral level course is focused on the use of machine learning methods for estimating cause-effect relations in observational and randomized trial data. This topic differs from more traditional introductory or intermediate level machine learning courses, which are often focused specifically on predictive analytics. In this course, we will cover why inference (statistical and causal) is a challenge when machine learning methods are used to estimate causal effects and obtain valid inference statistics.

By the end of this course you should have a solid understanding of:

- The challenges of estimating causal effects with data.
- When/why one should use machine learning to estimate causal effects.
- The curse of dimensionality and how double robust methods can be used to "lift" the curse.
- Different types of double robust estimators (targeted minimum loss based estimation, augmented inverse probability weighted estimators, double debiased machine learning estimator).
- How to estimate causal effects with machine learning methods using R

The literature on ML and causal inference is very expansive and complex. There are some key essential concepts that are not easy to understand for those with little to no formal technical background. The purpose of this course is to help you with your efforts in filling this gap.

This course **is an applied course.** Some of the theoretical challenges will be introduced, but in a very accessible, intuitive format.

PRE-REQUISITES

This course will build on basic and intermediate analytic methods and causal inference concepts covered in EPI 545, EPI 560 and EPI 760.

Necessary skills and concept include: reading data into R, basic data cleaning in R (e.g., subsetting data, finding missing values, merging data), operating on data.frames (e.g., changing column names, row names, summarizing rows/columns of data using simple statistics), basic graphics (e.g., plot or ggplot2), marginal standardization (g computation, parametric g formula), inverse probability weighting, basic causal estimands (average treatment effect, effect of treatment on the treated), and identifiability.

COURSE LEARNING OBJECTIVES

- Understand the basic causal roadmap of picking a suitable estimand and estimator for a given research question.
- Understand the curse of dimensionality and bias-variance tradeoffs.
- Understand why a double-robust estimator mitigates problems introduced by the curse of dimensionality.
- Understand the difference between TMLE and AIPW.
- Understand how stacking (SuperLearner) combines several machine learning algorithms into a single meta-algorithm.
- Deploy stacking in R using the `SuperLearner` and `sl3` packaages.
- Deploy TMLE and AIPW in R using the `tmle`, `tmle3` and `AIPW` packages.

ATTENDANCE POLICY

In person attendance in this short course is expected.

EVALUATION

TBD

Assignments

There will be one final project to be completed at home. There will also be one in class exercise on each day of the short-course.

Grade composition

Your grade will be determined as follows:

- 75% final project
- 25% in-class exercises

Grade scale

- A = 95 -- 100%

- A- = 90 -- 95%
- B+ = 87 – 90%
- B = 83 – 87%
- B- = 80 – 83%
- C = 65 – 80%
- F = <65%

Final grades are not rounded and the lower limit of each letter grade is inclusive, so e.g., 93.0 is an A, while 92.9 is an A-.

COURSE STRUCTURE

This short course will consist of a combination of in class lectures, in class exercises, and at home assignments. Students will be expected to have R and RStudio installed and working on their computers. In addition, the following packages should be installed and in working order:

```
"tidyverse", "here", "sandwich", "lmtest", "boot", "ranger", "ggplot2",
"broom", "SuperLearner", "tmle", "AIPW", "ranger", "xgboost", "e1071",
"nnet", "glmnet", "remotes"
```

You will also have to install the `tlverse` library, which is available only on GitHub. The best way to do this is to use the `install_github()` function in the `remotes` package. However, you will have to address the potential GitHub API limits, which can lead to installation errors. To deal with this problem, you will need your own GitHub account.

The easiest way to address this issue is to use a Github personal access token (PAT). There are a number of ways to do this, and it's important to read the basic information on PATs. Within R and RStudio, one straightforward way to manage PATs is to install and use the `usethis` package, which has a suite of functions available for creating and integrating PATs. Once you've installed `usethis`, you can:

- Use `usethis::browse_github_pat()` to create a GitHub token

- Use `usethis::edit_r_environ()` and add the environment variable by adding the following line to the R environment file: `GITHUB_PAT = 'your_github_token'`.

- Restart R (so that the GITHUB_PAT is read) and try to reinstall the packages that were resulting in the API limit error.

Be aware: **your Github PAT is a password, and should be treated as such.**

Health considerations

At the very first sign of not feeling well, stay at home and reach out for a health consultation. Please consult the campus FAQ for how to get the health consultation. As you know, Emory does contact tracing if someone has been diagnosed with COVID-19. A close contact is defined as someone you spend more than 15 minutes with, at a distance less than 6 feet, not wearing facial coverings. This typically means your roommates, for example.

RSPH POLICIES

Accessibility and Accommodations

Accessibility Services works with students who have disabilities to provide reasonable accommodations. In order to receive consideration for reasonable accommodations, you must contact the Office of Accessibility Services (OAS). It is the responsibility of the student to register with OAS. Please note that accommodations are not retroactive and that disability accommodations are not provided until an accommodation letter has been processed.

Students who registered with OAS and have a letter outlining their academic accommodations are strongly encouraged to coordinate a meeting time with me to discuss a protocol to implement the accommodations as needed throughout the semester. This meeting should occur as early in the semester as possible.

Contact Accessibility Services for more information at (404) 727-9877 or accessibility@emory.edu. Additional information is available at the OAS website at http://equityandinclusion.emory.edu/access/students/index.html

Honor Code

You are bound by Emory University's Student Honor and Conduct Code. RSPH requires that all material submitted by a student fulfilling his or her academic course of study must be the original work of the student. Violations of academic honor include any action by a student indicating dishonesty or a lack of integrity in academic ethics. Academic dishonesty refers to cheating, plagiarizing, assisting other students without authorization, lying, tampering, or stealing in performing any academic work, and will not be tolerated under any circumstances.

The RSPH Honor Code states: "*Plagiarism is the act of presenting as one's own work the expression, words, or ideas of another person whether published or unpublished (including the work of another student). A writer's work should be regarded as his/her own property.*"

Laney Academic Integrity Statement

You are expected to uphold and cooperate in maintaining academic integrity as a member of the Laney Graduate School. By taking this course, you affirm your commitment to the Laney Graduate School Honor Code, which you can find in the Laney Graduate School Handbook. You should ensure that you are familiar with the rights and responsibilities of members of our academic community and with policies that apply to students as members of our academic community. Any individual, when they suspect that an offense of academic misconduct has occurred, shall report this suspected breach to the appropriate Director of Graduate Studies, Program Director, or Dean of the Laney Graduate School. If an allegation is reported to a Director of Graduate Studies or a Program Director, they are in turn required to report the allegation to the Dean of Laney Graduate School.

COURSE CALENDAR AND OUTLINE

TBD