

# **Welcome to ML4CI**

Ashley I Naimi

June 2022

## **Contents**

1	Welcome to Machine Learning for Causal Inference	2
2	Overview of the Course	2
3	Some Logistics	4
4	R Code Logistics	4

# 1 Welcome to Machine Learning for Causal Inference

Welcome to ML4CI!

In this short course, we will spend the next few days learning about the basics behind using machine learning methods for estimating cause-effect relations.

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 challenges of using machine learning methods for estimating causal effects
- how to estimate causal effects with machine learning methods in 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.

## 2 Overview of the Course

Over the next four days, we will cover the following topics:

Day 1 (morning):

- Introduction to the Datasets
- Potential Outcomes, Estimands, Identifiability
- Parametric Regression for Effect Estimation
  - G Computation
  - Inverse Probability Weighting

Day 1 (afternoon):

- Machine Learning for Effect Estimation: The Curse of Dimensionality
- Double Robust Methods: Some Intuition
  - Augmented Inverse Probability Weighting (AIPW)
  - Targeted Minimum Loss-Based Estimation (TMLE)

Day 2 (morning):

- Modeling the Exposure and the Outcome
- Machine Learning Algorithms 1:
  - Neural Networks via nnet package
  - Gradient boosting via xgboost

Day 2 (afternoon):

- Machine Learning Algorithms 2:
  - CARTs and Random Forests via ranger
  - Support Vector Machines via e1071

Day 3 (morning):

- Meta Learners for the Exposure and Outcome Models: Stacking
- SuperLearner and sl3
  - Tuning Parameter Grids
  - Selection Algorithms

Day 3 (afternoon):

- Estimating Effects in Example Datasets 1
  - TMLE3 + sl3 for the ATE, ATT, and ATU
  - AIPW + sl3 for the ATE, ATT, and ATU

Day 4 (morning):

- Estimating Effects in Example Datasets 2
  - TMLE3 + sl3 for the ATE, ATT, and ATU
  - AIPW + sl3 for the ATE, ATT, and ATU

Day 4 (afternoon):

- Machine Learning for Causal Effect Estimation: Wrapping Up
  - Alternative Estimands
  - Time-Dependent Exposure and Confounder Modeling
  - Mediation Analysis
  - Further Reading/Learning Materials

### 3 Some Logistics

All the materials for this short course are available in the following GitHub Repository: GH

If you have git installed on your computer, you can download all the materials via the command line:

```
git clone SOMETHING
```

Or you can use a GUI of your choice to clone the repo.

If you don't have git installed on your computer, you can simply download all the materials by visiting my GH webpage, clicking the "clone" button, and selecting "Download ZIP".

### 4 R Code Logistics

In this short course, we will be using the R programming language throughout. In order to run the code we will be using during the course of the workshop, it will be important to have a few packages pre-installed on your computer.

The easiest way to install these packages would be to paste the following code in the R console<sup>1</sup> and run it:

```
packages <- c("tidyverse", "here", "sandwich",
  "lmttest", "boot", "ranger", "ggplot2",
  "broom", "SuperLearner", "tmle", "AIPW",
  "ranger", "xgboost", "e1071", "nnet",
  "glmnet", "remotes")

for (package in packages) {
  if (!require(package, character.only = T,
    quietly = T)) {
    install.packages(package, repos = "http://lib.stat.cmu.edu/R/CRAN")
  }
}

##
## Attaching package: 'zoo'
```

<sup>1</sup> For the beginner, I highly recommend using RStudio as your IDE of choice, particularly if you are not familiar with R and R programming. There are many excellent resources for installing and setting up R and the RStudio IDE. Here is a good getting started guide, provided by Garrett Golemund: <https://rstudio-education.github.io/hopr/starting.html>

```
## The following objects are masked from 'package:base':  
##  
##   as.Date, as.Date.numeric  
##  
## Attaching package: 'boot'  
  
## The following object is masked from 'package:survival':  
##  
##   aml  
  
## The following object is masked from 'package:lattice':  
##  
##   melanoma  
##  
## Attaching package: 'foreach'  
  
## The following objects are masked from 'package:purrr':  
##  
##   accumulate, when  
  
## Loaded gam 1.20  
  
## Super Learner  
  
## Version: 2.0-28  
  
## Package created on 2021-05-04  
  
##  
## Attaching package: 'Matrix'  
  
## The following objects are masked from 'package:tidyr':  
##  
##   expand, pack, unpack  
  
## Loaded glmnet 4.1-4  
  
## Welcome to the tmle package, version 1.5.0-1.1  
##  
## Major changes since v1.3.x. Use tmleNews() to see details on changes and bug fixes
```

```
##
## Attaching package: 'xgboost'

## The following object is masked from 'package:dplyr':
##
##      slice

##
## Attaching package: 'e1071'

## The following object is masked from 'package:Hmisc':
##
##      impute
```

```
for (package in packages) {
  library(package, character.only = T)
}

remotes::install_github("tlverse/tlverse")
```

```
## Skipping install of 'tlverse' from a github remote, the SHA1 (76c0b732) has not changed since last i
## Use `force = TRUE` to force installation
```

```
library(tlverse)
```

If you run into any problems installing and loading these packages, feel free to message me on the Slack channel for the course.

I look forward to working with you!