Welcome to ML4CI

Ashley I Naimi

Spring 2022

Contents

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 fill this gap.

2 Overview of the Course

This short course will be divided into four sections (days):
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" butting, and selecting "Download ZIP".