Predictive and Causal Machine Learning in R
(taught by Prof. Martin Huber)

This course provides an introduction to predictive and causal machine learning based on the software "R".

Predictive machine learning aims at forecasting the value of an outcome of interest, e.g. sales or turnover, based on observing specific patterns of potentially relevant factors (or "predictors") like price, quality, weather, advertisement campaigns etc. That is, predictive machine learning "learns" from patterns among predictors in (past) data to forecast the value of the outcome in the future.

Causal machine learning aims at assessing the causal effect of some action or intervention, like offering or not offering a discount to customers, on an outcome of interest, like customers' buying decision. The assessment of a causal effect requires that customer groups receiving and not receiving a discount are comparable in background characteristics that also affect their buying behaviour (e.g. previous buying behavior, income, education etc.). Causal machine learning can be used to generate such comparable groups in a data-driven way by estimating two separate models for how the characteristics affect the intervention and the outcome. Such approaches also permit detecting customer groups for whom the effect of the intervention is particularly large as a function of their characteristics.

This course first discusses the underlying assumptions, intuition and usefulness of machine learning for forecasting and causal analysis. It then introduces various machine learning algorithms and discusses their application for forecasting and causal analysis. Using the statistical software "R" and its interface "R Studio", these methods are applied to various real-world data sets.

Objectives
- To understand the ideas and goals of machine learning for prediction and causal analysis
- To understand the intuition, advantages, and disadvantages of alternative methods
- To be able to apply predictive and causal machine learning to real world data using the software "R" and its interface "R Studio"

Content
- Introduction to the concepts and purposes of predictive and causal machine learning
- Forecasting based on penalized regression (lasso and ridge regression)
- Forecasting using tree-based approaches (trees, bagging, random forests)
- Model tuning (cross-validation)
- Performance evaluation (out-of-sample testing)
- Causal analysis based on penalized regression (lasso and ridge regression)
- Causal analysis using tree-based approaches (causal trees and causal forests)
- Assessing effect heterogeneity (to find groups for whom the intervention is most effective)
- Application of all methods to real world data using the statistical software "R" and its interface "R Studio"; course participants may bring along their own data sets

Prerequisites
- Basic knowledge of the statistical software "R" and introductory statistics (linear regression), as e.g. provided in the course "Introduction to R"

Duration
- 2 days (roughly 7*45 minutes each day)

Evaluation
- take home exam: project work to be solved in R

ECTS
- 1.0