



DISSERTATION PRESENTATION AND DEFENSE

Interpretable Data Science: Applications and Theory in Climate Science and Public Health

WHEN

April 18th, 2022
2:00PM

WHERE

Jones Laboratory, Room 304, 5747 S. Ellis Avenue



Abby Stevens, PhD candidate

New statistical and machine learning methods have led to important advances in image and natural language processing, genetics, digital advertising, and other fields where there is an abundance of high-quality digital data and strong market incentives for automating tasks. This thesis intentionally focuses on areas such as climate science and public health, as they have suffered in this space without the same scale of training data and private investment.

Part 1 of this thesis focuses on applications in climate science, specifically forecasting precipitation at seasonal timescales, which is notoriously difficult due to complex dependence structures, nonstationarities, and a lack of a reliable observational record. To address some of these challenges, we develop a regularization regression scheme using a graph-guided regularizer that simultaneously promotes sparsity and similar coefficients for highly correlated covariates. We propose a novel way of combining climate model simulations and observations by using large ensemble simulations from a climate model to construct this regularizer, highlighting the potential to combine optimally the space–time structure of predictor variables learned from climate models with new graph-based regularizers to improve seasonal prediction. In Part 2, we develop a fast and flexible method for estimating variable importance (VI) measures with large neural networks. Our VI measure of interest analyzes the difference in predictive power between a full model trained on all variables and a reduced model that excludes the variable(s) of interest, which can be expensive to compute. We replace the need for fully retraining a wide neural network to estimate the reduced model by a linearization initialized at the full model parameters. We provide inferential guarantees for our method and verify its performance on simulated and real data. Part 3 of this thesis describes the development of city-scale synthetic populations for use in an agent-based model (CityCOVID) that simulates the endogenous transmission of COVID-19 and measures the impact of public health interventions. We then extend this work to empirically investigate the impacts of social determinants of health (SDOH) on COVID-19 transmission and mortality in order to better target interventions and counter social inequities in healthcare.

stat.uchicago.edu



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS