

Data Science Seminar
Hosted by Department of Mathematical Sciences

- Date: Tuesday, Oct 26, 2021
- Time: 12:00pm - 1:00pm
- Room: Zoom
- Speaker: Dr. Giles Hooker (UC Berkeley)
- Title: There is No Free Variable Importance: Traps in Interpreting Black Box Functions

Abstract

The field of machine learning - loosely defined as nonparametric statistical modeling - has become enormously successful over the past fifty years, partly by forgoing the parametric models familiar to statisticians. A consequence of this philosophy has been that these methods result in algebraically complex models that provide little human-accessible insight into the workings of the model, or what it might say about the underlying processes generating the data. As these methods have been taken up in high-stakes decision making, demands to “x-ray the black box” have become more prevalent, resulting in a wide variety of approaches to understand what signal the model is capturing or to provide explanations of individual predictions. Unfortunately, many of these methods produce results that can lead to mistaken conclusions about the model, or the underlying processes, or both. This talk reviews two sources of error: distorting the covariate distribution beyond the range where the model performs well and estimating structured surrogates using insufficient data. We show that many popular interpretation/explanation methods suffer from these, potentially resulting in mistaken conclusions or advice, and review the properties necessary to generate reliable explanation or interpretation.

Biography of the speaker: Dr. Hooker is a Professor of Statistics at the University of California, Berkeley. His work has focussed on statistical methods using dynamical systems models, inference with machine learning models, functional data analysis, and robust statistics. He is the author of “Dynamic Data Analysis: Modeling Data with Differential Equations” and “Functional Data Analysis in R and Matlab”. Much of his work has been inspired by collaborations particularly in ecology, human movement, and citizen science data.

This talk is endorsed by the Data Science Transdisciplinary Area of Excellence at Binghamton University.

From:
<https://www2.math.binghamton.edu/> - **Department of Mathematics and Statistics, Binghamton University**

Permanent link:
<https://www2.math.binghamton.edu/p/seminars/datasci/102621>

Last update: **2021/10/20 20:24**

