MICHIGAN STATE UNIVERSITY
Department of Statistics and Probability

COLLOQUIUM

Matthew Pratola
Ohio State University

Influential Observations in Bayesian Regression Tree Models

Tuesday, November 30, 2021
10:20 AM - 11:10 AM Eastern Time
Zoom

Abstract

BCART (Bayesian Classification and Regression Trees) and BART (Bayesian Additive Regression Trees) are popular Bayesian regression models widely applicable in modern regression problems. Their popularity is intimately tied to the ability to flexibly model complex responses depending on high-dimensional inputs while simultaneously being able to quantify uncertainties. This ability to quantify uncertainties is key, as it allows researchers to perform appropriate inferential analyses in settings that have generally been too difficult to handle using the Bayesian approach. However, surprisingly little work has been done to evaluate the sensitivity of these modern regression models to violations of modeling assumptions. In particular, we will consider influential observations, which one reasonably would imagine to be common – or at least a concern – in the big-data setting. In this paper, we consider both the problem of detecting influential observations and adjusting predictions to not be unduly affected by such potentially problematic data. We consider two detection diagnostics for Bayesian tree models, one an analogue of Cook's distance and the other taking the form of a divergence measure, and then propose an importance sampling algorithm to re-weight previously sampled posterior draws so as to remove the effects of influential data in a computationally efficient manner. Finally, our methods are demonstrated on real-world data where blind application of the models can lead to poor predictions and inference.

Zoom details can be found at: https://stt.natsci.msu.edu/stt-colloquium-zoom-info/

To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.