THE UNIVERSITY OF CHICAGO

Department of Statistics
STATISTICS COLLOQUIUM

Jason M. Klusowski

Department of Operations Research & Financial Engineering
Princeton University

Consistency of Decision Trees for High Dimensional Data

MONDAY, November 29, 2021 at 4:30 PM
Via Zoom

ABSTRACT

Decision trees are one of the most elemental methods for predictive modeling. They are often deployed in domains where many explanatory variables are observed and where a high importance is placed on the simplicity and interpretability of the fitted model, such as business and medicine. In this talk, I will show that decision trees are universally consistent in an additive model context, even when the number of predictor variables grows exponentially with the sample size, under natural 1-norm and 0-norm sparsity assumptions. The consistency is universal in the sense that there are no a priori assumptions on the distribution of the explanatory variables, thereby accommodating continuous, discrete, and/or dependent data.