Inference with Mondrian Random Forests

Matias D. Cattaneo[1] Jason M. Klusowski[1] William G. Underwood[1*]

October 13, 2023

Abstract

Random forests are popular methods for classification and regression, and many different variants have been proposed in recent years. One interesting example is the Mondrian random forest, in which the underlying trees are constructed according to a Mondrian process. In this paper we give a central limit theorem for the estimates made by a Mondrian random forest in the regression setting. When combined with a bias characterization and a consistent variance estimator, this allows one to perform asymptotically valid statistical inference, such as constructing confidence intervals, on the unknown regression function. We also provide a debiasing procedure for Mondrian random forests which allows them to achieve minimax-optimal estimation rates with β-Hölder regression functions, for all β and in arbitrary dimension, assuming appropriate parameter tuning.

Keywords: Random forests, regression trees, central limit theorem, bias correction, statistical inference, minimax rates, nonparametric estimation.

[1]Department of Operations Research and Financial Engineering, Princeton University
[*]Corresponding author: wgu2@princeton.edu

Contents