Do Simpler Models Exist and How Can We Find Them?

Cynthia Rudin
Duke University
Durham, NC
cynthia@cs.duke.edu

ABSTRACT

While the trend in machine learning has tended towards more complex hypothesis spaces, it is not clear that this extra complexity is always necessary or helpful for many domains. In particular, models and their predictions are often made easier to understand by adding interpretability constraints. These constraints shrink the hypothesis space; that is, they make the model simpler. Statistical learning theory suggests that generalization may be improved as a result as well. However, adding extra constraints can make optimization (exponentially) harder. For instance it is much easier in practice to create an accurate neural network than an accurate and sparse decision tree. We address the following question: Can we show that a simple-but-accurate machine learning model might exist for our problem, before actually finding it? If the answer is promising, it would then be worthwhile to solve the harder constrained optimization problem to find such a model. In this talk, I present an easy calculation to check for the possibility of a simpler model. This calculation indicates that simpler-but-accurate models do exist in practice more often than you might think. I then briefly overview several new methods for interpretable machine learning. These methods are for (i) sparse optimal decision trees, (ii) sparse linear integer models (also called medical scoring systems), and (iii) interpretable case-based reasoning in deep neural networks for computer vision.

This is joint work with Lesia Semenova, Ron Parr, Xiyang Hu, Margo Seltzer, Chaofan Chen, Oscar Li, Alina Barnett, Daniel Tao, and Berk Ustun.

KDD '19, August 4–8, 2019, Anchorage, AK, USA.

ACM ISBN 978-1-4503-6201-6/19/08.
DOI: https://doi.org/10.1145/3292500.3330823

BIOGRAPHY

Cynthia Rudin is a professor of computer science, electrical and computer engineering, and statistical science at Duke University, and directs the Prediction Analysis Lab, whose main focus is in interpretable machine learning. Previously, Prof. Rudin held positions at MIT, Columbia, and NYU. She holds an undergraduate degree from the University at Buffalo, and a PhD from Princeton University. She is a three time winner of the INFORMS Innovative Applications in Analytics Award, was named as one of the "Top 40 Under 40" by Poets and Quants in 2015, and was named by Businessinsider.com as one of the 12 most impressive professors at MIT in 2015.

She is past chair of both the INFORMS Data Mining Section and the Statistical Learning and Data Science section of the American Statistical Association. She has also served on committees for DARPA, the National Institute of Justice, and AAAI. She has served on three committees for the National Academies of Sciences, Engineering and Medicine, including the Committee on Applied and Theoretical Statistics, the Committee on Law and Justice, and the Committee on Analytic Research Foundations for the Next-Generation Electric Grid. She is a fellow of the American Statistical Association and a fellow of the Institute of Mathematical Statistics. She will be the Thomas Langford Lecturer at Duke University during the 2019-2020 academic year.