On Using Predictive Models for Decisions*

Cynthia Rudin

Abstract: We often use predictive models to make a decision afterwards. For instance, we might estimate the number of patients at a medical clinic and then designate resources to serve those patients. The class of accurate predictive models might be quite large (called the "Rashomon effect" by Breiman), which leads to two observations, (i) there may be very accurate, yet very sparse logical models that are naturally useful for decision making, (ii) if the decision problem is coupled over a set of unlabeled points (like scheduling), there may be a large range of decisions resulting from the set of good predictive models.

Considering point (i), I will present Falling Rule Lists (Wang and Rudin, AISTATS 2015). This method is a competitor for CART (Classification and Regression Trees). It produces sparse logical models, which are ordered lists of IF-THEN rules, where the risks monotonically decrease as we go down the list (e.g., IF diabetes THEN risk=50%, ELSE IF hypertension THEN risk=40%, ELSE risk=30%). For medical applications, Falling Rule Lists predicts risk at the same time as it stratifies patients into decreasing risk categories - this makes it very natural to use for decision making, like a medical calculator.

To address (ii) I will present work on Machine Learning with Operational Costs (Tulabandhula and Rudin, Journal of Machine Learning Research, 2013). This paper considers decision problems that are coupled over a small set of unlabeled points. This work is based on the idea that managers have very practical knowledge about the cost of solving their decision problems. This prior knowledge leads to a reduction in the size of the hypothesis space, and better learning theoretic guarantees on the quality of the predictions.

Thus in the first part of the talk I will discuss how predictive modeling can be made to help with decision-making, and in the second part, I will discuss how prior knowledge about decisions can help with prediction. The first part uses Bayesian analysis, and the second part is a statistical learning theory result.

Bio: Cynthia Rudin is an associate professor of statistics at the Massachusetts Institute of Technology associated with the Computer Science and Artificial Intelligence Laboratory and the Sloan School of Management, and directs the Prediction Analysis Lab. Her interests are in machine learning, data mining, applied statistics, and knowledge discovery (Big Data). Her application areas are in energy grid reliability, healthcare, and computational criminology. Previously, Prof. Rudin was an associate research scientist at the Center for Computational Learning Systems at Columbia University, and prior to that, an NSF postdoctoral research fellow at NYU. She holds an undergraduate degree from the University at Buffalo where she received the College of Arts and Sciences Outstanding Senior Award in Sciences and Mathematics, and three separate outstanding senior awards from the departments of physics, music, and mathematics. She received a PhD in applied and computational mathematics from Princeton University. She is the recipient of the 2013 INFORMS Innovative Applications in Analytics Award, an NSF CAREER award, and was named as one of the "Top 40 Under 40" by Poets and Quants in 2015. Her work has been featured in Businessweek, The Wall Street Journal, the New York Times, the Boston Globe, the Times of London, Fox News (Fox & Friends), the Toronto Star, WIRED Science, U.S. News and World Report, Slashdot, CIO magazine, Boston Public Radio, and on the cover of IEEE Computer. She is presently the chair-elect for the INFORMS Data Mining Section, and currently serves on committees for DARPA, the National Academy of Sciences, the US Department of Justice, and the American Statistical Association.

*Machine Learning External Seminar, Gatsby Unit, September 16, 2015.

Papers: Falling Rule Lists. This is joint work with my student Fulton Wang. Proceedings of Artificial Intelligence and Statistics (AISTATS), 2015; http://arxiv.org/abs/1411.5899. This paper won a best paper award from the Statistical Learning and Data Mining Section of the American Statistical Association in 2015.

Machine Learning with Operational Costs. This is joint work with my former student Theja Tulabandhula Journal of Machine Learning Research, 2013. http://web.mit.edu/rudin/www/TulabandhulaRu13.pdf