From I-Score, a Model-free Feature Selection Statistic, to Explainable AI

In-Chi Hu

Department of Statistics, George Mason University, U.S.A.

Abstract

Recently, I-Score, introduced in Chernoff, Lo and Zheng (2009), has been successfully incorporated into convolutional neural networks (CNN) and recurrent neural networks (RNN) to improve both prediction accuracy and explainability. In this talk, we discuss some new results and further properties of I-score, and explore their implications to explainable AI (XAI). I-score properties discussed include its connections with predictivity of a subset of variables, AUC, multi-class classification problems, and why I-score can increase prediction accuracy without sacrificing explainability. If time allows, we will also discuss some properties of backward dropping algorithm (BDA)- a greedy algorithm to maximize the predictivity of a set of variables.