Marloes Maathuis, ETH Zürich, Switzerland

FALSE DISCOVERY RATE CONTROL FOR GAUSSIAN GRAPHICAL MODELS

Abstract: We propose a method to control the finite sample false discovery rate (FDR) when learning the structure of a Gaussian graphical model. Our method builds on the recently proposed knockoff idea of Barber and CandÃÍs for linear models. We extend their approach to the graphical model setting by using a local (node-based) and a global (graph-based) step: we construct knockoffs and feature statistics for each node locally, and then solve a global optimization problem to determine the threshold for each node. We then estimate the neighborhood of each node, by comparing its feature statistics to its threshold, resulting in our graph estimate. We establish finite sample FDR control of this procedure. Our proposed method is very flexible, in the sense that one has a lot freedom in the choice of the feature statistics, the optimization problem and the way in which the final graph estimate is obtained. For any given data set, it is not clear a priori what choices of these hyperparameters are optimal. We therefore use a sample-splitting-recycling procedure that first uses half of the sample to select the hyperparameters, and then learns the graph using all samples, in such a way that the finite sample FDR control still holds. Finally, we compare our method to the state-of-the-art in simulations and on a real data set.

This talk is based on joint work with Jinzhou Li.