Integer Programming for Learning Directed Acyclic Graphs from Continuous Data

Hasan Manzour
Ph.D. Candidate
Industrial and Systems Engineering
University of Washington

Abstract: With the deluge of data, the field of statistical learning has gained rapid prominence and has grown in complexity by the problems that science and industry bring to its door. To cope with the ever increasing complexity of statistical learning problems, the development of new optimization models and efficient algorithms becomes imperative. Traditionally, continuous optimization has received a lot of attention from the statistics community. However, in recent years, a body of work has shed light on the importance of discrete optimization in the field of statistical learning.

In this talk, I will discuss how discrete optimization can successfully solve a central problem in machine learning: learning Bayesian Networks from data. A Bayesian Network (BN) is a rich framework that represents the joint probability distribution and dependency structure among a set of random variables in the form of a Directed Acyclic Graph (DAG). An essential part of this problem entails learning the DAG structure from observational data. While statistical properties of DAG estimation have been extensively studied, the development of efficient computational tools to learn DAG structure from observational data is still an open challenge. We cast this problem in the form of a mathematical programming model which can naturally incorporate a superstructure in order to reduce the set of possible candidate DAGs. We use the penalized negative log-likelihood score function and propose a new mixed-integer quadratic optimization (MIQO) model, referred to as a layered network (LN) formulation. Computational results demonstrate that the proposed formulation outperforms existing mathematical formulations and scales better than available algorithms.

Bio: Hasan Manzour is a Ph.D. candidate in the Department of Industrial and Systems Engineering at the University of Washington working with Prof. Simge Küçükyavuz and Prof. Ali Shojaie. His research interests primarily lie in the intersection of statistical learning and optimization theory. In particular, Hasan is currently focusing on applying discrete optimization and combinatorial algorithms to formulate and solve computationally challenging statistical learning tasks.