

## Abstract

For a number of years, mutual information has been used as a statistical model selection criterion for choosing the structure of a graphical model. In this work, we outline recent research we have been pursuing along these lines. We begin by reviewing the classic results in the literature, both positive and negative, for learning the best  $k$ -trees representations of a given distribution. Our fortune in this endeavor, it turns out, depends not only on  $k$  but also the learning criterion (e.g., maximum likelihood vs. some other measure). We moreover describe algorithms for exact and approximate partitioning of sets of random variables. We next describe a discriminative model selection heuristic based on the difference between two information theoretic quantities, something that constitutes a valid signed  $I$ -measure for sets of random variables. There is no known way to exactly optimize this measure, but we describe various heuristics for this purpose, some with reasonable local properties and based on the concept of submodularity. We present results on several data sets.