Promises and Pitfalls of Kernel Interaction Detection

Keli Liu[1]; Feng Ruan[2]

[1]   Voleon Group
[2]   Department of EECS, University of California Berkeley

Abstract:
We tackle the problem of variable selection with a focus on discovering interactions between variables. With p variables, there are $O(p^k)$ possible interactions of order k making exhaustive search infeasible. We show that it is nonetheless possible to identify the variables involved in interactions (of any order) with only linear computation cost, $O(p)$, and in a nonparametric fashion. Our algorithm is based on minimizing a nonconvex objective, carefully designed to have a favorable landscape. We provide finite sample guarantees on both false positives (we show all stationary points of the objective exclude noise variables) and false negatives (we characterize the sample sizes needed for gradient descent to converge to a "good" stationary point).