Within Group Variable Selection through the Exclusive Lasso

Frederick Campbell[1⋆], **Genevera Allen**[1,2,3,4]

1. Department of Statistics, Rice University
2. Department of Electrical Engineering, Rice University A
3. Department of Pediatrics-Neurology, Baylor College of Medicine
3. an and Dan Duncan Neurological Research Institute, Texas Childrens Hospital
⋆Contact author: fc10@rice.edu

Keywords: Structured Variable Selection, Composite Penalty, Exclusive Lasso

Many data sets consist of variables with an inherent group structure. The problem of group selection has been well studied, but in this paper, we seek to do the opposite: our goal is to select at least one variable from each group in the context of predictive regression modeling. This problem is NP-hard, but we study the tightest convex relaxation: a composite penalty that is a combination of the ℓ_1 and ℓ_2 norms. Our so-called Exclusive Lasso method performs structured variable selection by ensuring that at least one variable is selected from each group. We study our method's statistical properties and develop computationally scalable algorithms for fitting the Exclusive Lasso. We study the effectiveness of our method via simulations as well as using NMR spectroscopy data. Here, we use the Exclusive Lasso to select the appropriate chemical shift from a dictionary of possible chemical shifts for each molecule in the biological sample.