ABSTRACT:
Variable selection for structured covariates lying on an underlying known graph is a well-studied problem which has received considerable attention, with the primary focus being on discrete mixture approaches. However, most of the existing approaches are not scalable to high dimensional settings, for example, in genomic studies involving tens of thousands of genes lying on known pathways. We propose a Bayesian shrinkage approach which incorporates prior information through shrinkage parameters, with the coefficients for two connected variables in the graph being encouraged to have a similar degree of shrinkage. We fit our model via a computationally efficient EM algorithm which is scalable to high dimensional settings (p~100000). We establish theoretical properties for fixed as well as increasing dimensions, even when the number of variables increases faster than the sample size. We demonstrate the advantages of our approach via a simulation study, and apply the method to a real data example.

This is a joint work with Changgee Chang and Suprateek Kundu.