Estimation and inference for high dimensional, doubly-structured regression models

Parker Knight, Yue Wang, and Tim Randolph

Popular regularized regression models, such as the LASSO and Ridge regression, are unable to incorporate additional structural information into the fitting procedure. This is a disadvantage when variables in the model are highly correlated, as is often the case when analyzing high-throughput biological data. Kernel Penalized Regression (KPR) is an extension of Ridge regression that fits models with respect to variable and subject structure. We have developed an R software package for fitting KPR models and performing inference on the estimated coefficients. To select model tuning parameters, we have implemented both cross validation and Restricted Maximum Likelihood estimation via a mixed model formulation. The package also includes helper functions for data processing and visualization with a supervised biplot. An analysis of microbiome bacterial abundance data with the KPR model showed that the inclusion of structural data can improve model interpretability and statistical power, leading us to conclude that our package can provide great utility to researchers analyzing highly structured data.