Summary and Future Works

Two-Step SP-FSR yields a superior performance among all wrapper methods. We opine that its excellent results are attributed by SP-FSR which is superior in selecting optimal sets of main effect features. Although its performance trails behind `glinternet` most of the empirical problems, we conjecture that Two-Step SP-FSR's performances are consistent under the presence of strong hierarchy. `glinternet` determines its key parameter λ via a grid search of at least 50 different values; hence it can learn better than Two-Step SP-FSR. Ideally, we could apply the grid search on p_0 and p_1 for Two-Step SP-FSR, but it would cause higher computational costs which is a typical tradeoff among wrapper methods.

However, Two-Step SP-FSR is more flexible than `glinternet` since it can incorporate any learning model, such as multinomial, Poisson, and ordinal regression models. We will try different models with Two-Step SP-FSR. Besides, `glinternet` does not have a mechanism to address overidentification issue, i.e. there exist no true interaction features. An alternative is run LASSO to benchmark against `glinternet`, with no guarantee if both will select the same set of optimal main-effect features. Two-Step SP-FSR has a competitive edge over `glinternet` as the former keeps the same set of main-effect features in searching for the pairwise interactions. Two-Step SP-FSR is a wrapper method and hence non-parametric. Unlike `glinternet`, it can be easily modified to include quadratic and higher-order interactions terms.

Though our experiments show that interactions are likely to exist, we suggest running simulation studies for other hierarchical scenarios. The simulation studies will include true strong, weak, anti-hierarchical, pure-interaction, and no-interaction types. So far our experiments in this study focus on real datasets with $n > p$ and all explanatory features are numeric. Therefore, we will also test on datasets where $p \gg n$ and comprise mixed feature types.