Generalized Group Lasso for Patient Subgroup Selection

Wenxuan Deng

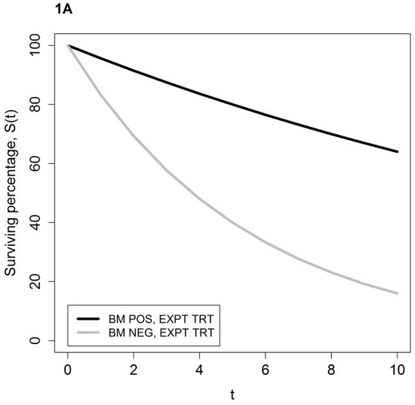
Takeda Pharmaceuticals U.S.A., Inc.

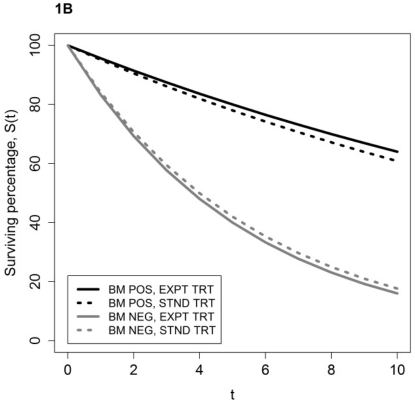
Wenxuan.Deng@takeda.com

July 15, 2018

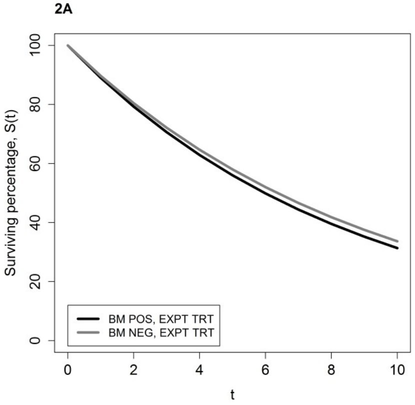
Overview

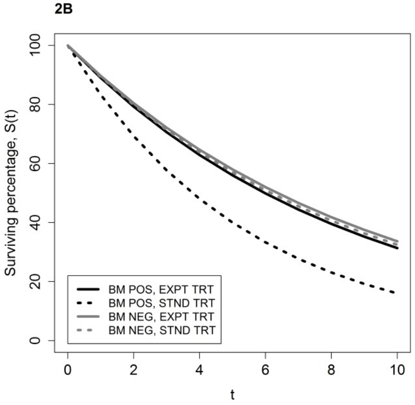
Prognostic Biomarkers





Predictive Biomarkers





Tree-based Methods

Regression trees GUIDE[Loh, 2018]:

- piecewise-linear Model
- examine residual patterns for each treatment level

Cannot repeat even the most naive simulation in GUIDE paper.

Reason: limited sample size. Even two splits will results in small sample size in each branch. The results would be highly unstable. Tree-based method is not appropriate to clinical trial dataset and identify prognostic and predictive biomarkers.

Ordinary Linear Model

$$Y = X\beta + W\tau + G\alpha + W \otimes G\gamma + \epsilon$$

- X: Baseline variables
- W: Treatment variables
- G: Main effects of genes, i.e. expression levels, SNP or mutation
- $W \otimes G$: Interaction effects of genes and treatment
- ϵ: Random errors

Group lasso

We choose group lasso for its ability to

- handle high dimensional data
- allow hierarchical structure

However, the current group lasso based methods

- penalize on all parameters
- have no efficient adaptive penalty weights
- do not specifically target on patients treatment subgroup identification

Loss Function

We assume the hierarchical relationship between prognostic biomarkers and predictive biomarkers, that is a predictive biomarker should be a prognostic biomarker.

The loss function is

$$\min_{\theta} f(\theta|Y, X, W, G) + \lambda \sum_{i} \eta_i^I |\gamma_i| + \lambda \sum_{i} \eta_i^M \sqrt{\alpha_i^2 + \gamma_i^2}$$

where $f(\theta|Y, X, W, G)$ is L-2 loss function, i.e. sum of squared errors for ordinary linear model.

$\theta = (\beta, \tau, \alpha, \gamma)$ is parameter vector.

Loss function for ordinary linear model

$$\min_{\theta} \parallel Y-(X\beta+W\tau+G\alpha+W\otimes G\gamma) \parallel^2 +\lambda\sum_i \eta_i^I|\gamma_i|+\lambda\sum_i \eta_i^M\sqrt{\alpha_i^2+\gamma_i^2}$$

Denote $X_I^{(l)} = W \otimes G_l$, and $X^{(l)} = [G_l, X_I^{(l)}]$ is the lth group of the main and interaction effects of gene l. Thus, due to KKT conditions, we let

$$\eta_i^I = \parallel X^{(i)} \parallel_2$$

$$\eta_i^M = \sqrt{\parallel G_i \parallel_2^2 + 2(1-\sqrt{\frac{2}{\pi}}) \parallel X_I^{(i)}}$$

References

Loh, Wei-Yin, Michael Man, and Shuaicheng Wang.
"Subgroups from regression trees with adjustment for prognostic effects and postselection inference."
Statistics in medicine (2018).

The End