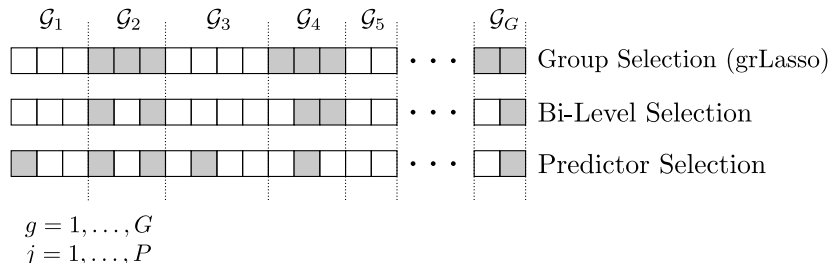


# Incorporating Grouping Information Into Bayesian Decision Tree Ensembles

Junliang Du  
Antonio R. Linero

# Grouping Structures



**Common scenarios:** omics, with groups corresponding to groups of genes or groups of SNPs.

## Additive Models

Assume target  $f(x)$  decomposes additively as

$$f(x) = \sum_{t=1}^m g(x; \mathcal{T}_t, \mathcal{M}_t),$$

for some *adaptively chosen* basis functions  $g(x; \mathcal{T}_t, \mathcal{M}_t)$ .

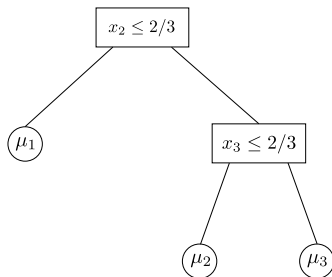
**BART:** basis functions are decision trees; similar in many respects to gradient boosting + decision trees.

# Variable Importance

**Define** the *variable importance*  $s_j$  of predictor  $j$  as

$$\Pr(\text{a given decision rule uses predictor } j).$$

For example, the probability of splitting on  $x_2$  and  $x_3$  in this tree is  $s_2 \cdot s_3$ .



**Near sparse  $s \implies$  small subset of predictors used.**

# Overlapping Group BART

**LDA-like model:** Sampling predictor  $j$  arises by

1. sampling a group according to  $\pi$ ; and
2. sampling a predictor-within-group according to  $w_g$ .

# Overlapping Group BART

**LDA-like model:** Sampling predictor  $j$  arises by

1. sampling a group according to  $\pi$ ; and
2. sampling a predictor-within-group according to  $w_g$ .

Set

$$s = W\pi, \quad \pi \in \mathbb{S}_{G-1}, \quad w_g \in \mathbb{S}_{P-1}.$$

# Overlapping Group BART

**LDA-like model:** Sampling predictor  $j$  arises by

1. sampling a group according to  $\pi$ ; and
2. sampling a predictor-within-group according to  $w_g$ .

Set

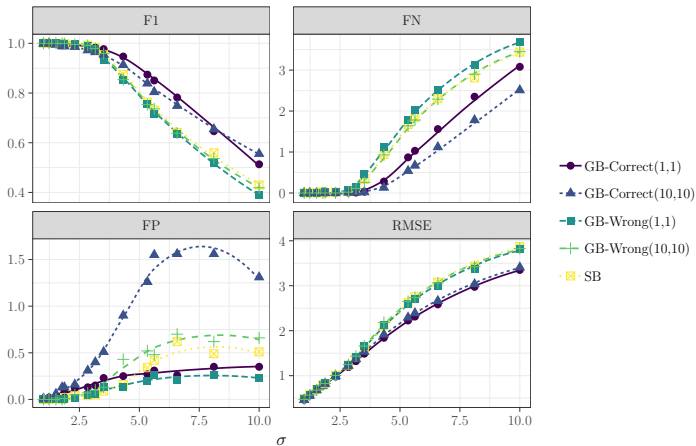
$$s = W\pi, \quad \pi \in \mathbb{S}_{G-1}, \quad w_g \in \mathbb{S}_{P-1}.$$

Incorporate grouping information into sparsity pattern of  $w_g = (w_{g1}, \dots, w_{gP})$ .

Sparsity inducing prior on  $\pi$  and  $w_g \implies$  bi-level selection!

# Simulation Studies

Nonparametric ground truth (one relevant group, 5 relevant predictors, 50 members of group, 500 predictors).





## Breast Cancer Data

Cross validation suggests encouraging performance on breast cancer dataset of Van De Vijver et al. (2002) (classification of metastatic/non-metastatic tumors)

Method	Average Heldout Deviance
OG-BART	620
SBART	646 (0.005)
OG-Lasso	797 (< 0.0001)
cMCP	698 (0.014)

Thanks!

- Bleich, J., Kapelner, A., George, E. I., and Jensen, S. T. (2014). Variable selection for BART: An application to gene regulation. *The Annals of Applied Statistics*, 8(3):1750–1781.
- Van De Vijver, M. J., He, Y. D., Van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., and Marton, M. J. (2002). A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine*, 347(25):1999–2009.