Scalable smoothing with BART

Ryan Yee

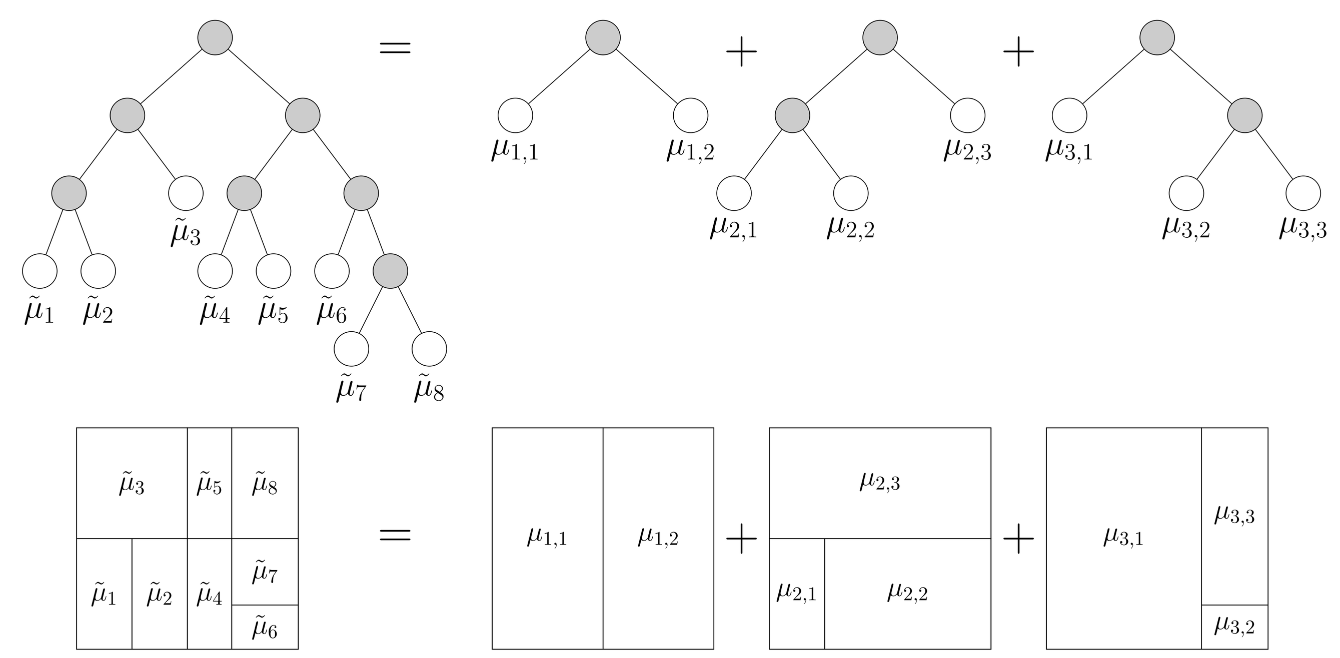
Department of Statistics, University of Wisconsin–Madison

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

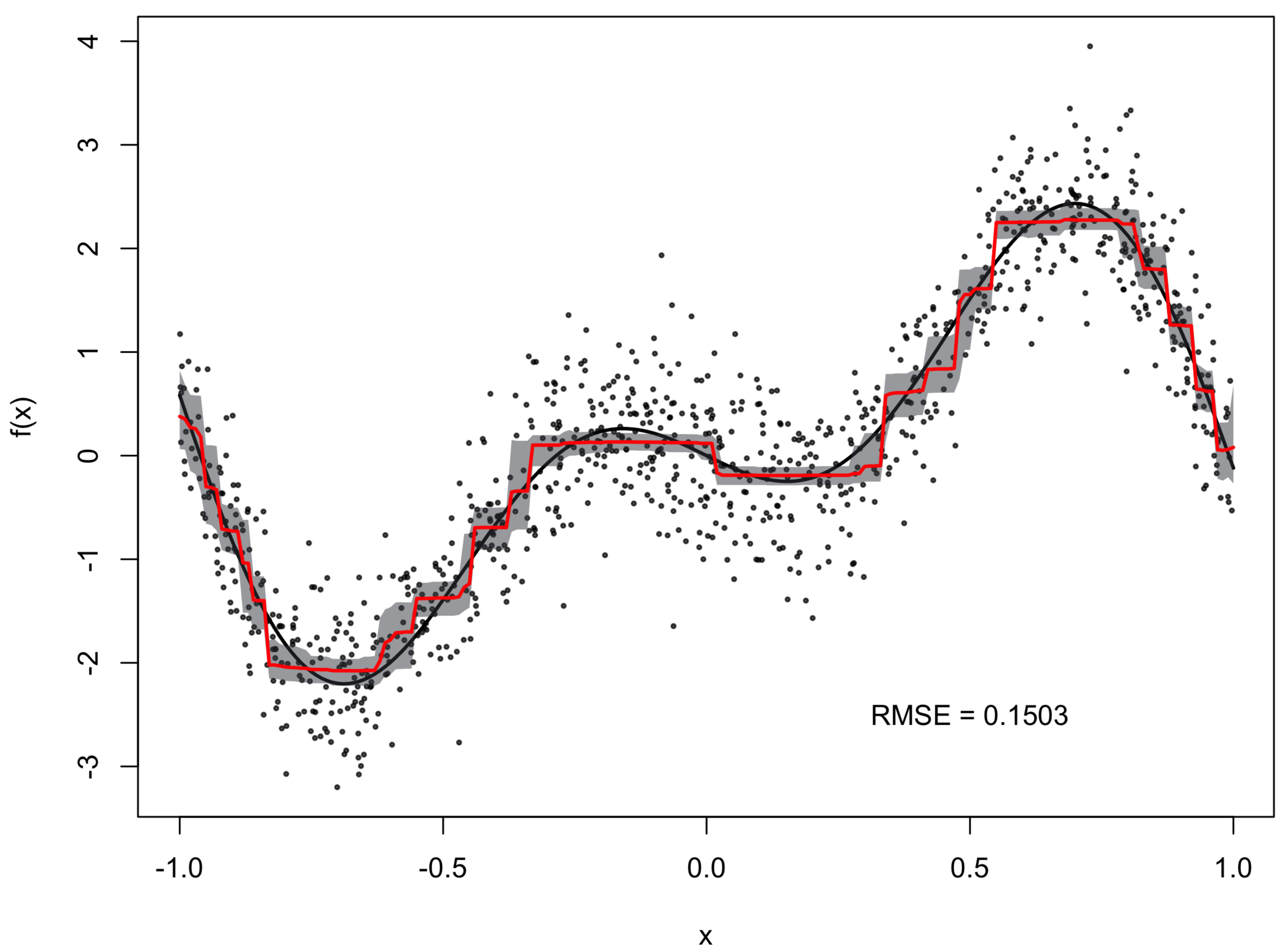
Introduction to BART

- **Problem:** non-parametric regression: $y_n \sim \mathcal{N}(f(\mathbf{x}_n), \sigma)$
- **Main Idea:** approximate $f(\mathbf{x})$ with step-function (i.e., tree)

- **Insight:** express complicated tree as sum of simpler trees



- **Result:** function estimation with uncertainty quantification



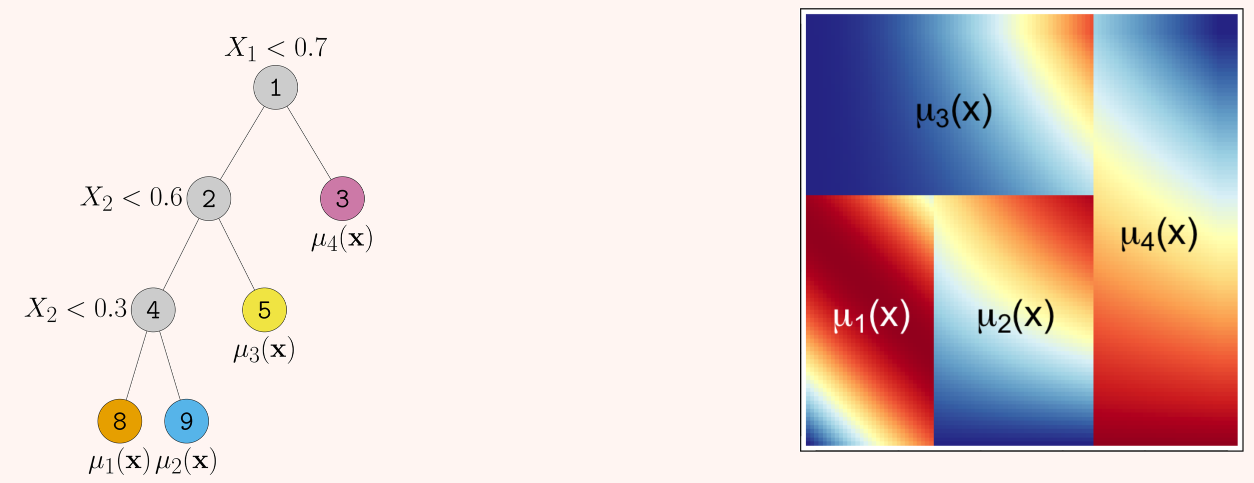
- **Issue:** BART inherently produces non-smooth outputs

Our Contribution

We introduce BART with functional outputs of the form:

$$\mu(x) = \sum_{d=1}^{D} \beta_d h(\omega_d^\top x + b_d)$$

where h is a fixed activation function, $\Theta = (\{\omega_d, b_d\})$ collects all *input* weights, and $\boldsymbol{\beta} = (\{\beta_d\})$ is a vector of *output* weights.



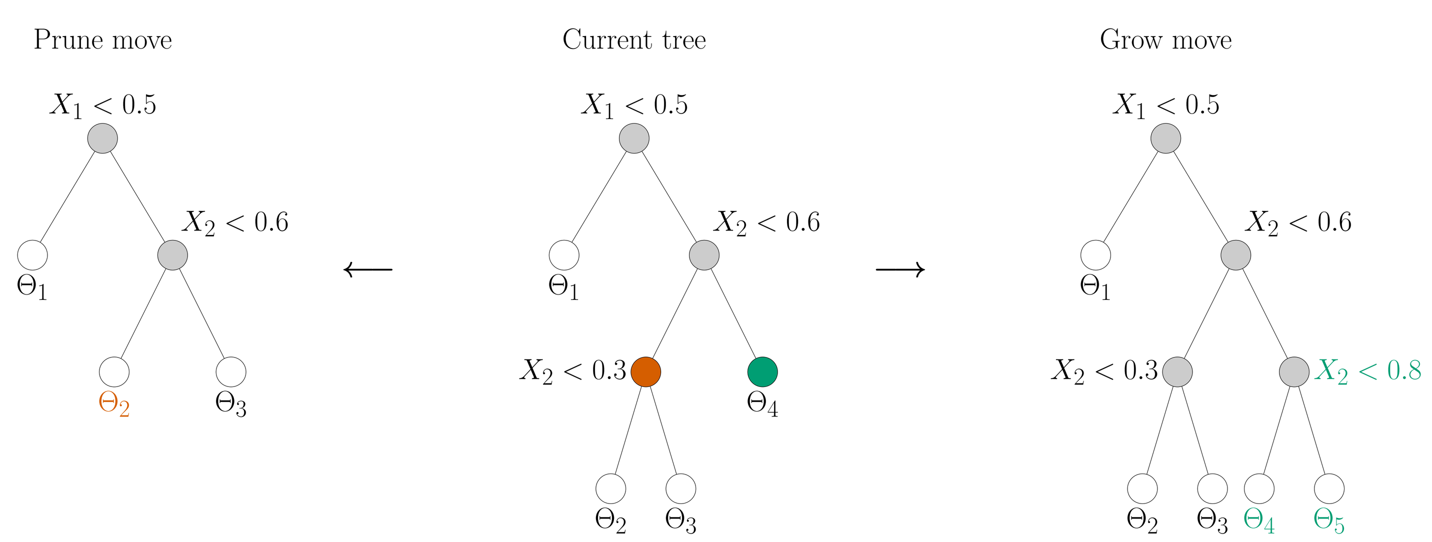
Note: setting $h(x) = \sqrt{2}\cos(x)$ leads to *random Fourier feature-inspired* approximation of treed Gaussian process ensembles.

Sampling

BART utilizes a Metropolis-within-Gibbs sampling scheme to sequentially update trees in the ensemble [1].

Step 1: Update $(\mathcal{T}, \Theta) \to (\mathcal{T}^*, \Theta^*)$ with Metropolis-Hastings.

A new tree is proposed that differs from the current tree in exactly one node by considering either a *GROW* or *PRUNE* move:



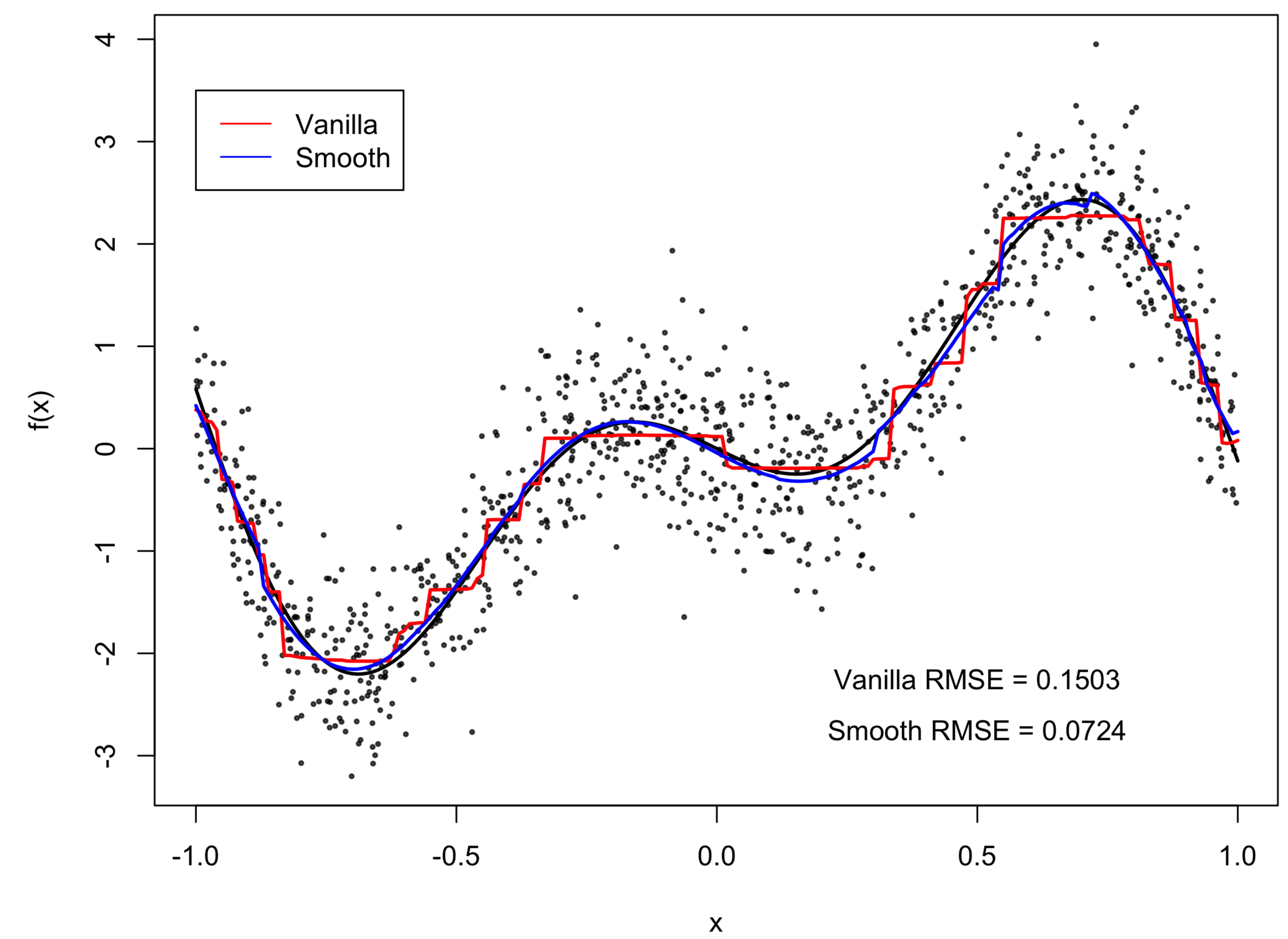
Move to the proposed tree based on a Metropolis-Hastings acceptance probability:

$$\alpha(\mathcal{T}, \Theta \to \mathcal{T}^*, \Theta^*) = \underbrace{\frac{q(\mathcal{T}, \Theta; \mathcal{T}^*, \Theta^*)}{q(\mathcal{T}^*, \Theta^*; \mathcal{T}, \Theta)}}_{\text{transition ratio}} \times \overbrace{\frac{p(\mathbf{R} \mid \mathcal{T}^*, \Theta^*, \sigma^2)}{p(\mathbf{R} \mid \mathcal{T}, \Theta, \sigma^2)}}^{\text{likelihood ratio}} \times \underbrace{\frac{p(\mathcal{T}^*, \Theta^*)}{p(\mathcal{T}, \Theta)}}_{\text{prior ratio}}$$

Step 2: Update $\boldsymbol{\beta}^* | \mathbf{R}, (\mathcal{T}^*, \Theta^*) \sim \mathcal{N}_D$

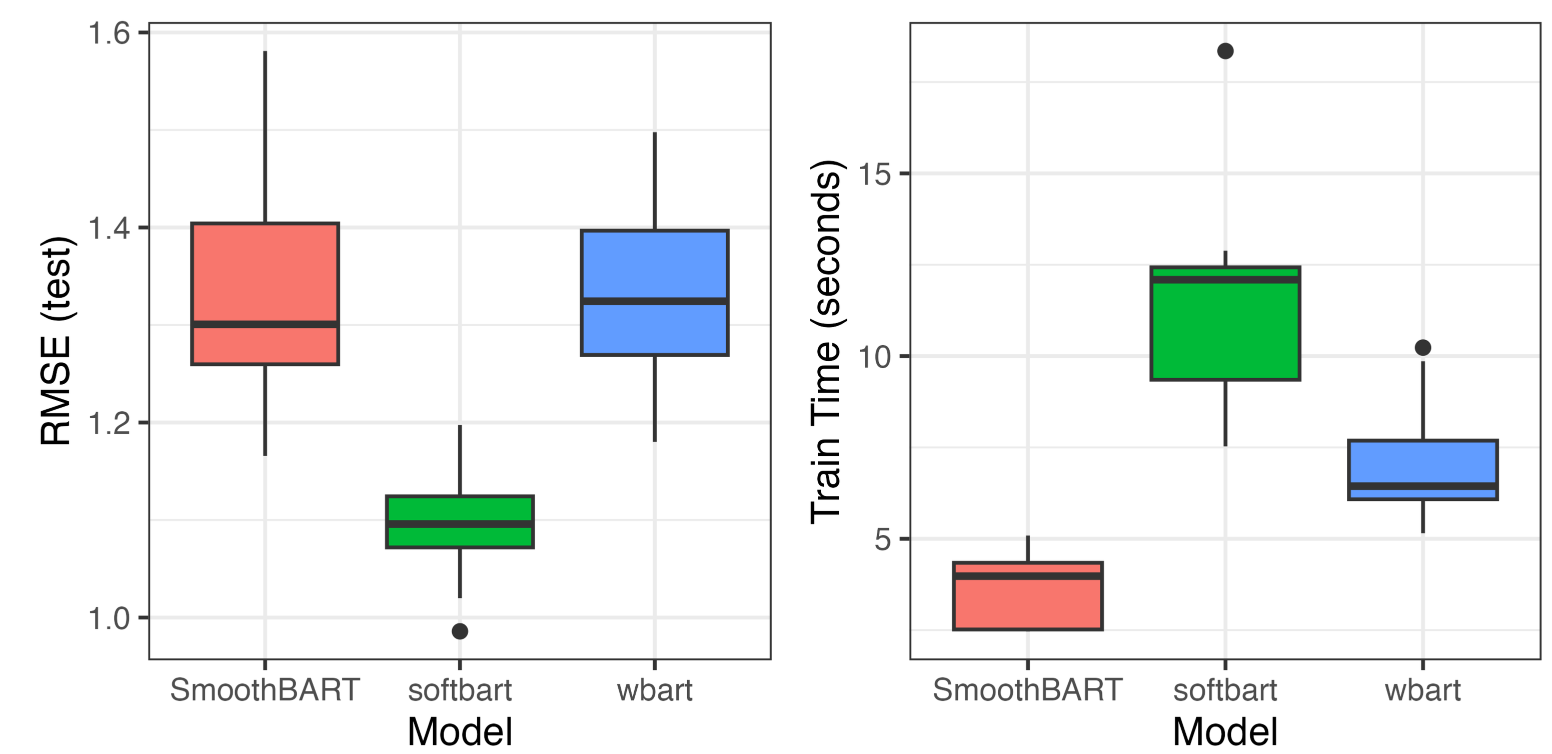
Note: updates only require a *linear* scan of the data

Vanilla vs. Smooth BART



Preliminary Empirical Results

Friedman Function: $f(x) = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5$



References

[1] Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. BART: Bayesian Additive Regression Trees. In *The Annals of Applied Statistics*. 2010.

[2] Ali Rahimi and Benjamin Recht. Random Features for Large-Scale Kernel Machines. In *Advances in Neural Information Processing Systems*. 2007.