

Posterior computation for BART

- ▶ In order to generate the posterior for f , we sample the structure of all the trees \mathcal{T}_h , for $h = 1, \dots, H$; the values of all leaves μ_{hn} for $n \in \mathcal{L}_h$ within tree h ; and, the error variance σ^2
- ▶ Additionally, with the sparsity prior, there are samples of the vector of splitting variable selection probabilities $[s_1, \dots, s_P]$ and, when the sparsity parameter is random, samples of θ
- ▶ The leaf and variance parameters are sampled from the posterior using Gibbs sampling
- ▶ Since the priors on these parameters are conjugate, the Gibbs conditionals are specified analytically
- ▶ The leaves, μ_{hn} , are drawn from a Normal conditional density
- ▶ The error variance, σ^2 , is drawn from a scaled inverse Chi-square conditional (equivalent to inverse Gamma)

Posterior computation for BART

- ▶ Drawing a tree from the posterior requires a Metropolis-within-Gibbs sampling scheme, i.e., a Metropolis-Hastings (MH) step within Gibbs sampling
- ▶ For single-tree models, four different proposals are defined in Chipman George 1998 *JASA*
- ▶ The complementary BIRTH/DEATH proposals are essential (the others are CHANGE and SWAP which are optional)
- ▶ This is only one of several schemes that have been proposed; see Pratola 2016 *Bayesian Analysis*
- ▶ For programming simplicity, the **BART/BART3** package only implements BIRTH and DEATH with equal probability
- ▶ BIRTH selects a leaf and turns it into a branch: a new variable and cut-point with two leaves “born” as its descendants
- ▶ DEATH selects a branch leading to two terminal leaves and “kills” the branch by replacing it with a single leaf

Posterior computation for BART: the BIRTH proposal

- ▶ we present the acceptance probability for a BIRTH proposal
- ▶ a DEATH proposal is the reversible inverse of a BIRTH
- ▶ The algorithm assumes a fixed discrete set of possible split values for each x_j denoted by c_{jk}
- ▶ the leaf values, μ_{hn} , are integrated out so this search in tree space is over a large discrete set of possibilities
- ▶ At the m th MCMC step, let \mathcal{T}_h^m denote the current state of tree h while \mathcal{T}_h^* denotes the tree proposal
- ▶ \mathcal{T}_h^* are identical \mathcal{T}_h^m except that one terminal leaf of \mathcal{T}_h^m is replaced by a branch of \mathcal{T}_h^* with two terminal leaves

Posterior computation for BART: the BIRTH proposal

- The proposed tree is accepted with the following probability:

$$\pi_{\text{BIRTH}} = \min \left(1, \frac{[\mathcal{T}_h^* | y]}{[\mathcal{T}_h^m | y]} \frac{[\mathcal{T}_h^m | \mathcal{T}_h^*]}{[\mathcal{T}_h^* | \mathcal{T}_h^m]} \right)$$

- $[\mathcal{T}_h^* | \mathcal{T}_h^m]$ is the probability of proposing \mathcal{T}_h^* (a BIRTH) given current state \mathcal{T}_h^m
- while $[\mathcal{T}_h^m | \mathcal{T}_h^*]$ is the reverse
the probability of proposing \mathcal{T}_h^m (a DEATH) given \mathcal{T}_h^*
- Now we delve into the calculations for $[\mathcal{T}_h^* | y]$ and $[\mathcal{T}_h^m | y]$

Posterior computation for BART: the BIRTH proposal

- ▶ First, we describe the likelihood contribution to the posterior
- ▶ Let y_n denote the partition of y corresponding to the leaf node n given the tree \mathcal{T}_h
- ▶ Because the leaf values are a priori conditionally independent, we have $[y|\mathcal{T}_h] = \prod_n [y_n|\mathcal{T}_h]$ where $[y_n|\mathcal{T}_h] = \int [\mu_n] [y_n|\mathcal{T}_h, \mu_n] d\mu_n$ (we will return to this later)
- ▶ After cancellation of terms in the ratio $\frac{[\mathcal{T}_h^*|y]}{[\mathcal{T}_h^m|y]}$ we have the likelihood contributions

$$\frac{[y_L, y_R|\mathcal{T}_h^*]}{[y_n|\mathcal{T}_h^m]} = \frac{[y_L|\mathcal{T}_h^*] [y_R|\mathcal{T}_h^*]}{[y_n|\mathcal{T}_h^m]}$$

y_L : partition for the newborn left leaf node

y_R : partition for the newborn right leaf node

Posterior computation for BART

- ▶ Similarly, the terms that the prior contributes to the posterior ratio often cancel since there is only one **area** where the trees differ: either a terminal leaf or a penultimate branch
- ▶ Therefore, the prior contribution to $\frac{[\mathcal{T}_h^*|y]}{[\mathcal{T}_h^m|y]}$ is

$$\frac{P[\psi_n = 1] P[\psi_l = 0] P[\psi_r = 0] s_j}{P[\psi_n = 0]} = \frac{\alpha(d(n) + 1)^{-\beta} [1 - \alpha(d(n) + 2)^{-\beta}]^2 s_j}{1 - \alpha(d(n) + 1)^{-\beta}}$$

- ▶ $P[\psi_n]$ is the branch regularity prior
 s_j is the splitting variable selection probability
 n is the chosen leaf node in tree \mathcal{T}_h^m
 $l = 2n$ is the newborn left leaf node in tree \mathcal{T}_h^*
 $r = 2n + 1$ is the newborn right leaf node in tree \mathcal{T}_h^*

Posterior computation for BART

- Finally, the ratio $\frac{[\mathcal{T}_h^m | \mathcal{T}_h^*]}{[\mathcal{T}_h^* | \mathcal{T}_h^m]}$ is

$$\frac{[\text{DEATH} | \mathcal{T}_h^*] [n | \mathcal{T}_h^*]}{[\text{BIRTH} | \mathcal{T}_h^m] [n | \mathcal{T}_h^m] s_j}$$

- $[n | \mathcal{T}_h]$ is the probability of choosing node n given tree \mathcal{T}_h
- N.B. s_j appears in both the numerator and denominator of the acceptance probability π_{BIRTH} , therefore, canceling which is mathematically convenient

Posterior computation for weighted BART

$$y_i = f(x_i) + \epsilon_i \text{ where } f \stackrel{\text{prior}}{\sim} \text{BART}$$

$$\epsilon_i \sim \text{N}(0, \mathbf{w}_i^2 \sigma^2)$$

BART variant	Weights	Comment
Homoskedastic	$\mathbf{w}_i = 1$	<i>Chipman, George et al. 2010</i>
Weighted		Non-trivial known weights (circa 2015)
Heteroskedastic	$\mathbf{w}_i = s(x_i)$	$s^2 \stackrel{\text{prior}}{\sim} \mathbf{HBART}$ <i>Pratola, Chipman et al. 2020</i>

- ▶ Until now, we have not dwelled on BART variants
- ▶ From here on, we will focus on weighted BART
- ▶ This development extends the homoskedastic variant and naturally incorporates heteroskedastic BART

Posterior computation for weighted BART: BIRTH

- Based on the centered data: $y_i = \tilde{y}_i - \mu$
- Given that the n th leaf only depends on $(y_{i'}, w_{i'})$ where $i' = 1, \dots, m_n$

$$w_n = \sum_{i'} w_{i'} = w_L + w_R$$

$$\xi_n = \sum_{i'} w_{i'} y_{i'} = \xi_L + \xi_R$$

$$\mu_{hl} \stackrel{\text{prior}}{\sim} N(0, \sigma_0^2)$$

$$\sigma_0^2 = \frac{\tau^2}{4H\kappa^2}$$

$$[y_n | \mathcal{T}_h] = \int [\mu_n] [y_n | \mathcal{T}_h, \mu_n] d\mu_n$$

$$\propto (w_n \sigma_0^2 + 1)^{-0.5} \exp \left(\frac{0.5 \xi_n^2 \sigma_0^2}{w_n \sigma_0^2 + 1} \right)$$

$$[y_L | \mathcal{T}_h] \propto (w_L \sigma_0^2 + 1)^{-0.5} \exp \left(\frac{0.5 \xi_L^2 \sigma_0^2}{w_L \sigma_0^2 + 1} \right)$$

$$[y_R | \mathcal{T}_h] \propto (w_R \sigma_0^2 + 1)^{-0.5} \exp \left(\frac{0.5 \xi_R^2 \sigma_0^2}{w_R \sigma_0^2 + 1} \right)$$

Posterior computation for weighted BART: BIRTH

- After cancellation of terms in the ratio $\frac{[\mathcal{T}_h^*|y]}{[\mathcal{T}_h^m|y]}$ we have the likelihood contributions

$$\begin{aligned}\frac{[y_L, y_R | \mathcal{T}_h^*]}{[y_n | \mathcal{T}_h^m]} &= \frac{[y_L | \mathcal{T}_h^*] [y_R | \mathcal{T}_h^*]}{[y_n | \mathcal{T}_h^m]} \\ &= \frac{(w_L \sigma_0^2 + 1)^{-0.5} (w_R \sigma_0^2 + 1)^{-0.5}}{(w_n \sigma_0^2 + 1)^{-0.5}} \\ &\quad \times \frac{\exp\left(\frac{0.5 \xi_L^2 \sigma_0^2}{w_L \sigma_0^2 + 1} + \frac{0.5 \xi_R^2 \sigma_0^2}{w_R \sigma_0^2 + 1}\right)}{\exp\left(\frac{0.5 \xi_n^2 \sigma_0^2}{w_n \sigma_0^2 + 1}\right)}\end{aligned}$$

Posterior computation for weighted BART: the leaves

- ▶ Based on the centered data: $y_i = \tilde{y}_i - \mu$
- ▶ Given \mathcal{T}_h , suppose that the l th leaf, μ_{hl} , only depends on $(y_{i'}, w_{i'})$ where $i' = 1, \dots, m_{hl}$
- ▶ Define the residuals, $z_{i'}$, due to everything except $(\mathcal{T}_h, \mathcal{M}_h)$

$$z_{i'} = y_{i'} - \sum_{h' \neq h} g(x_{i'}; \mathcal{T}_{h'}, \mathcal{M}_{h'})$$
$$\sim N(\mu_{hl}, w_{i'}^2 \sigma^2)$$

$$\bar{z}_{\cdot} = N(\mu_{hl}, \eta_{hl}^{-1})$$

$$\mu_{hl} \stackrel{\text{prior}}{\sim} N(0, \eta_0^{-1})$$

$$\eta_{hl}^{-1} = \frac{\sigma^2}{m_{hl}^2} \sum_{i'} w_{i'}^2$$

$$\eta_0^{-1} = \frac{\tau^2}{4H\kappa^2}$$

$$\mu_{hl} | (\mathcal{T}_h, y) \sim N\left(\frac{\eta_{hl} m_{hl} \bar{z}_{\cdot}}{\eta_0 + \eta_{hl} m_{hl}}, \frac{1}{\eta_0 + \eta_{hl} m_{hl}}\right)$$

Posterior computation for BART: uncertainty intervals

- ▶ Suppose that we want to estimate f at some value x whether from the training or testing
- ▶ The standard Bayesian estimate is $\hat{f}(x) = M^{-1} \sum_m f_m(x)$
- ▶ Construct a Bayesian $(1 - \alpha) \times 100\%$ credible interval from the $\alpha/2$ and the $1 - \alpha/2$ quantiles of the posterior
- ▶ $(f_{(\alpha/2)}(x), f_{(1-\alpha/2)}(x))$
- ▶ Construct a Bayesian $(1 - \alpha) \times 100\%$ prediction interval
- ▶ Generate \tilde{y} from the predictive distribution:
 $\tilde{y}_m \sim N(f_m(x), \sigma_m^2)$
- ▶ Select the $\alpha/2$ and the $1 - \alpha/2$ quantiles of \tilde{y}
- ▶ $(\tilde{y}_{(\alpha/2)}, \tilde{y}_{(1-\alpha/2)})$