

# BART with an Unobserved Predictor

Sameer K. Deshpande

May 6, 2019

## 1 Setup

Suppose that we observe  $n$  independent pairs of data  $(\mathbf{x}_i, y_i)$  and that the usual non-parametric regression model  $y_i = f(\mathbf{x}_i) + \sigma\varepsilon_i, \varepsilon_i \sim N(0, 1)$  is mis-specified. In particular, suppose that there are  $p_u$  missing predictors  $u_i$  and that in actuality  $y_i = f(\mathbf{x}_i, u_i) + \sigma\varepsilon$ . We use BART (?) to make inference about  $f$ .

If we had in fact observed each  $u_i$ , we may use the Gibbs sampler introduced in (?) to simulate approximate posterior draws of  $f$ . Since we have no observed  $u_i$ , we instead place a standard normal prior  $u_i \sim N(0, I_{p_u})$  on the latent predictors and augment the Gibbs sampler to sample from the posterior distribution of  $u_i$  before updating the remaining parameters. To this end, first note that the conditional posterior distribution of  $\mathbf{u} = \{u_1, \dots, u_n\}$  can be written as

$$\pi(\mathbf{u} \mid \sigma^2, \mathbf{f}, \mathbf{y}) \propto \prod_{i=1}^n \exp \left\{ -\frac{1}{2\sigma^2} (y_i - \mathbf{f}(\mathbf{x}_i, u_i))^2 - \frac{1}{2} u_i^\top u_i \right\}$$

This means that in our Gibbs sampler, we may update each  $u_i$  independently. To do so, we use a Metropolis step in which we accept a randomly drawn proposal  $u_i^* \sim Q(u_i)$  with probability

$$\alpha(u_i^*, u_i) = \min \left\{ \frac{\exp \left\{ -\frac{(y_i - \mathbf{f}(\mathbf{x}_i, u_i^*))^2}{2\sigma^2} - \frac{u_i^{*\top} u_i^*}{2} \right\}}{\exp \left\{ -\frac{(y_i - \mathbf{f}(\mathbf{x}_i, u_i))^2}{2\sigma^2} - \frac{u_i^\top u_i}{2} \right\}}, 1 \right\}$$

Letting  $\Sigma_i$  denote the current estimate of  $\text{Cov}(u_i)$ , we follow ? and use the following jump distribution

$$Q(u_i) = (1 - \beta)N(u, (2.38)^2 \Sigma_i) + \beta N(u, (0.1)^2 I_{p_u})$$

where  $\beta = 0.05$ .

In particular, suppose that there is a missing prediction  $u_i$  and that in actuality  $y_i = f(\mathbf{x}_i, u_i) + \sigma\varepsilon_i$ . In other words the conditional mean function  $\mathbb{E}[Y \mid X, U]$  is some non-linear

function that may involve interactions between  $X$  and  $U$ . When we fully observe each covariate, we can use the usual Gibbs sampler for BART. In this setting, where we do not observe the missing predictor, we augment this Gibbs sampler by simulating the values of  $\mathbf{u} = (u_1, \dots, u_n)$ . This involves placing a prior on  $\mathbf{u}$  and then updating its values according to the appropriate posterior.

## 2 Updating $\mathbf{u}$

One useful property of BART is that the fits are invariant to monotone transformations of each predictor. This is because the partition induced by each decision tree is invariant to monotone transformation of the predictors. We start by placing a standard normal prior  $u_i \sim N(0, 1)$  on the unknown predictors.

Now observe

$$\pi(\mathbf{u}, | \sigma^2, \mathbf{f}, y) \propto \prod_{i=1}^n \left[ \exp \left\{ -\frac{(y_i - f(\mathbf{x}_i, u_i))^2}{2\sigma^2} \right\} \times \exp \left\{ -\frac{u_i^2}{2} \right\} \right]$$

Since the  $y_i$  does not depend on  $u_{i'}$  for  $i \neq i'$  and since we are modeling the  $u_i$ 's as *a priori* independent, the expression above tells us that we may update the  $u_i$ 's independently in our Gibbs sampler. We specifically do this with a Metropolis step: given the current value  $u_i$ , we draw  $u_i^* \sim Q(u_i)$  and accept this new value with probability

$$\alpha(u_i) = \min \left\{ 1, \frac{\exp \left\{ -\frac{(y_i - f(\mathbf{x}_i, u_i^*))^2}{2\sigma^2} - \frac{u_i^{*2}}{2} \right\}}{\exp \left\{ -\frac{(y_i - f(\mathbf{x}_i, u_i))^2}{2\sigma^2} - \frac{u_i^2}{2} \right\}} \right\}$$

We use the proposal distribution recommended by ?

$$Q(u) = (1 - \beta)N(u, (2.38)\hat{\sigma}_u^2) + \beta N(u, 0.01)$$

where  $\hat{\sigma}_u$  is the current estimate of the standard deviation of  $u_i$ .