Heteroskedastic BART Augmentation

Introduction

We want to extend the BART model so that each observation now can have its own variance (but each observation is still independent). The model becomes:

$$Y = \sum_{t=1}^{m} \mathcal{T}_t\left(X_1, \ldots, X_p\right) + \boldsymbol{\mathcal{E}}, \quad \boldsymbol{\mathcal{E}} \sim \mathcal{N}_n\left(\mathbf{0}, \underbrace{\begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_k^2 \end{bmatrix}}_{\boldsymbol{D}}\right)$$

We take vanilla BART and now condition on the variances and then we augment the last step of the Gibbs sampler to now do n samples — one for σ_1^2, one for σ_2^2, ..., and one for σ_n^2. Remember that E is the n-vector of residuals, Y minus the best guess inside the trees:

$$\begin{array}{rcl}
\mathcal{T}_1 & \mid & R_1, \sigma_1^2, \ldots, \sigma_n^2 \\
M_1 & \mid & \mathcal{T}_1, R_1, \sigma_1^2, \ldots, \sigma_n^2 \\
\vdots & & \\
\mathcal{T}_m & \mid & R_m, \sigma_1^2, \ldots, \sigma_n^2 \\
M_m & \mid & \mathcal{T}_m, R_m, \sigma_1^2, \ldots, \sigma_n^2 \\
\sigma_1^2 & \mid & E \\
\vdots & & \\
\sigma_n^2 & \mid & E
\end{array}$$

Thus, there are many pieces in the BART model that will now have to be altered. Each subsection below is devoted to each piece.

Prior on the variance

We present two possible designs below for the prior on the σ^2's:

- We keep the same prior as before, now it's on each of the observation's variances:

$$\sigma_1^2, \ldots, \sigma_n^2 \overset{iid}{\sim} \text{InvGamma}\left(\frac{\nu}{2}, \frac{\nu\lambda}{2}\right)$$

  We may have to rethink the above, but it seems like a good start. Once again we pick $\nu = 3$ and we pick λ so that 90% of the prior's density is below s^2, our estimate from the data which is either just the sample variance, or the RMSE squared from a linear model.

- We also can have individual priors:

$$\begin{aligned} \sigma_1^2 &\sim \text{InvGamma}\left(\frac{\nu}{2}, \frac{\nu\lambda_1}{2}\right) \\ &\vdots \\ \sigma_n^2 &\sim \text{InvGamma}\left(\frac{\nu}{2}, \frac{\nu\lambda_n}{2}\right) \end{aligned}$$

  where λ_i is picked based on the weighted least squares residual for each observation.

Changes in Gibbs Sampler for M

The other step that has to change is sampling the leaf μ's. Remember that $M_t \,|\, \mathcal{T}_1, R_1, \sigma_1^2, \ldots, \sigma_n^2$ is actually the sampling for all leaves where each leaf is considered independent:

$$\begin{aligned} \mu_{t1} &\;|\; \mathcal{T}_t, R_{t_1}, \sigma_1^2, \ldots, \sigma_n^2 \\ &\vdots \\ \mu_{tb_t} &\;|\; \mathcal{T}_t, R_{t_{b_t}}, \sigma_1^2, \ldots, \sigma_n^2 \end{aligned}$$

The subscripts on the R term indicates we only consider the data that falls into the leaf. So remember that the prior on the leaf value was normal, and the likelihood was assumed normal as well. The only thing that changes is we now have a different variance estimate for each observation.

We derive the correct posterior distribution below. We drop the subscripts just for convenience since it will be the same for each of the above and denote k as the number of data records that fell into this leaf. $\{\sigma_1^2, \ldots, \sigma_k^2\} \subset \{\sigma_1^2, \ldots, \sigma_n^2\}$ for the data in this leaf as well:

$$\mathbb{P}\left(\mu \,\middle|\, R, \sigma_1^2, \ldots, \sigma_k^2\right) \;\propto\; \mathbb{P}\left(R \,\middle|\, \mu, \sigma_1^2, \ldots, \sigma_n^2\right) \mathbb{P}\left(\mu \,\middle|\, \sigma_1^2, \ldots, \sigma_k^2\right)$$

$$
\begin{aligned}
&= \mathcal{N}_k\left(\mu\mathbf{1},\, \boldsymbol{D}\right)\mathcal{N}\left(0,\, \sigma_\mu^2\right)\\
&\vdots\\
&= \mathcal{N}\left(\frac{\frac{k^2\bar{R}}{\sum_{i=1}^k \sigma_i^2}}{\frac{1}{\sigma_\mu^2}+\frac{k^2}{\sum_{i=1}^k \sigma_i^2}},\, \frac{1}{\frac{1}{\sigma_\mu^2}+\frac{k^2}{\sum_{i=1}^k \sigma_i^2}}\right)
\end{aligned}
$$

Changes in Gibbs Sampler for σ_i^2

In vanilla BART, the posterior of σ^2 was:

$$
\sigma^2 \mid e_1, \ldots, e_n \sim \text{InvGamma}\left(\frac{\nu+n}{2},\, \frac{\lambda\nu+\sum_{i=1}^n e_i^2}{2}\right)
$$

Now, we have to sample each $\sigma_1^2, \ldots, \sigma_n^2$ from their individual posteriors. The analagous expression is that the SSE is based on only one residual:

$$
\sigma_i^2 \mid e_i \sim \text{InvGamma}\left(\frac{\nu+1}{2},\, \frac{\lambda\nu+e_i^2}{2}\right)
$$

Changes in Likelihood for R

In vanilla BART, we had:

$$
\begin{aligned}
&\frac{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T^*,\sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T,\sigma^2\right)}\\
=\;& \frac{\int_{\mathbb{R}}\mathbb{P}\left(R_{\ell_{L,1}},\ldots,R_{\ell_{L,n_{\ell,L}}}\,\middle|\,\mu,\sigma^2\right)\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu \quad \int_{\mathbb{R}}\mathbb{P}\left(R_{\ell_{R,1}},\ldots,R_{R,\ell_{n_{\ell,R}}}\,\middle|\,\mu,\sigma^2\right)\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu}{\int_{\mathbb{R}}\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,\middle|\,\mu,\sigma^2\right)\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu}\\
\vdots\;&\\
=\;& \sqrt{\frac{\sigma^2\left(\sigma^2+n_\ell\sigma_\mu^2\right)}{\left(\sigma^2+n_{\ell_L}\sigma_\mu^2\right)\left(\sigma^2+n_{\ell_R}\sigma_\mu^2\right)}}\,\exp\left(\frac{\sigma_\mu^2}{2\sigma^2}\left(\frac{\left(\sum_{i=1}^{n_{\ell_L}} R_{\ell_L,i}\right)^2}{\sigma^2+n_{\ell_L}\sigma_\mu^2}+\frac{\left(\sum_{i=1}^{n_{\ell_R}} R_{\ell_R,i}\right)^2}{\sigma^2+n_{\ell_R}\sigma_\mu^2}-\frac{\left(\sum_{i=1}^{n_\ell} R_{\ell,i}\right)^2}{\sigma^2+n_\ell\sigma_\mu^2}\right)\right)
\end{aligned}
$$

We cannot use this expression as before since σ^2 is now non-constant. Instead we have to use:

$$
\begin{aligned}
&\frac{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T^*,\sigma_1^2,\ldots,\sigma_n^2\right)}{\mathbb{P}\left(\boldsymbol{R}\,\middle|\,T,\sigma_1^2,\ldots,\sigma_n^2\right)}\\
=\;& \frac{\int_{\mathbb{R}}\mathbb{P}\left(R_{\ell_{L,1}},\ldots,R_{\ell_{L,n_{\ell,L}}}\,\middle|\,\mu,\sigma_1^2,\ldots,\sigma_n^2\right)\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu \quad \int_{\mathbb{R}}\mathbb{P}\left(R_{\ell_{R,1}},\ldots,R_{R,\ell_{n_{\ell,R}}}\,\middle|\,\mu,\sigma_1^2,\ldots,\sigma_n^2\right)\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu}{\int_{\mathbb{R}}\mathbb{P}\left(R_{\ell_1},\ldots,R_{\ell_{n_\ell}}\,\middle|\,\mu,\sigma_1^2,\ldots,\sigma_n^2\right)\mathbb{P}\left(\mu\right)\,\mathrm{d}\mu}
\end{aligned}
$$

Thus, we have to figure out the following below. I'm going to drop the subscripts for convenience:

$$
\begin{aligned}
&\int_{\mathbb{R}} \mathbb{P}\left(\boldsymbol{R} \,\middle|\, \mu, \sigma_1^2, \ldots, \sigma_n^2\right) \mathbb{P}(\mu) \, \mathrm{d}\mu \\
=\ & \int_{\mathbb{R}} \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{D}|^{-\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{R}-\mu\mathbf{1})^\top \boldsymbol{D}^{-1}(\boldsymbol{R}-\mu\mathbf{1})\right) \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right) \mathrm{d}\mu \\
=\ & \int_{\mathbb{R}} \left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(R_i-\mu)^2\right)\right) \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{1}{2\sigma_\mu^2}\mu^2\right) \mathrm{d}\mu \\
=\ & \frac{1}{(2\pi)^{\frac{n}{2}+1}\sqrt{\sigma_\mu^2 \prod_{i=1}^n \sigma_i^2}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\mu^2 + \sum_{i=1}^{n}(R_i-\mu)^2\right)\right) \mathrm{d}\mu \\
=\ & \frac{1}{(2\pi)^{\frac{n}{2}+1}\sqrt{\sigma_\mu^2 \prod_{i=1}^n \sigma_i^2}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\frac{\mu^2}{\sigma_\mu^2} + \sum_{i=1}^{n}\frac{R_i^2}{\sigma_i^2} - \frac{2\mu R_i}{\sigma_i^2} + \frac{\mu^2}{\sigma_i^2}\right)\right) \mathrm{d}\mu \\
=\ & \frac{1}{(2\pi)^{\frac{n}{2}+1}\sqrt{\sigma_\mu^2 \prod_{i=1}^n \sigma_i^2}} \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\frac{\mu^2}{\sigma_\mu^2} + \sum_{i=1}^{n}\frac{R_i^2}{\sigma_i^2} - 2\mu\sum_{i=1}^{n}\frac{R_i}{\sigma_i^2} + \mu^2\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)\right) \mathrm{d}\mu \\
=\ & \frac{1}{(2\pi)^{\frac{n}{2}+1}\sqrt{\sigma_\mu^2 \prod_{i=1}^n \sigma_i^2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{R_i^2}{\sigma_i^2}\right) \underbrace{\int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\frac{\mu^2}{\sigma_\mu^2} - 2\mu\sum_{i=1}^{n}\frac{R_i}{\sigma_i^2} + \mu^2\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)\right) \mathrm{d}\mu}
\end{aligned}
$$

Now we can solve the underbraced integral above using Mathematica:

```
In[6]:= ∫_{-∞}^{∞} e^{-1/2 * (x^2 / a - 2 * x * b + c * x^2)} dx
```

```
Out[6]= {ConditionalExpression[
```

$$
\frac{e^{\frac{a\,b^2}{2+2\,a\,c}}\sqrt{2\,\pi}}{\sqrt{\frac{1}{a}+c}}, \ \mathrm{Re}\left[\frac{1}{a}+c\right] > 0\,]\}
$$

Plugging this result into the equation where the integral once was, we obtain:

$$= \frac{1}{(2\pi)^{\frac{n}{2}+1}\sqrt{\sigma_\mu^2 \prod_{i=1}^n \sigma_i^2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n \frac{R_i^2}{\sigma_i^2}\right) \sqrt{\frac{2\pi}{\frac{1}{\sigma_\mu^2} + \sum_{i=1}^n \frac{1}{\sigma_i^2}}} \exp\left(\frac{\sigma_\mu^2 \left(\sum_{i=1}^n \frac{R_i}{\sigma_i^2}\right)^2}{2 + 2\sigma_\mu^2 \sum_{i=1}^n \frac{1}{\sigma_i^2}}\right)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\left(\prod_{i=1}^n \sigma_i^2\right)\left(1 + \sigma_\mu^2 \sum_{i=1}^n \frac{1}{\sigma_i^2}\right)}} \exp\left(-\frac{1}{2}\sum_{i=1}^n \frac{R_i^2}{\sigma_i^2}\right) \exp\left(\frac{\sigma_\mu^2}{2}\left(\frac{\left(\sum_{i=1}^n \frac{R_i}{\sigma_i^2}\right)^2}{1 + \sigma_\mu^2 \sum_{i=1}^n \frac{1}{\sigma_i^2}}\right)\right)$$

Now we can solve for the likelihood ratio:

$$\frac{\mathbb{P}\left(\boldsymbol{R} \,\middle|\, T^*, \sigma_1^2, \ldots, \sigma_n^2\right)}{\mathbb{P}\left(\boldsymbol{R} \,\middle|\, T, \sigma_1^2, \ldots, \sigma_n^2\right)} = \frac{\mathbb{P}\left(\boldsymbol{R}_L \,\middle|\, \sigma_1^2, \ldots, \sigma_n^2\right) \mathbb{P}\left(\boldsymbol{R}_R \,\middle|\, \sigma^2\right)}{\mathbb{P}\left(\boldsymbol{R} \,\middle|\, \sigma_1^2, \ldots, \sigma_n^2\right)}$$

Note that there's many things that can cancel above. The sum of the R_i^2/σ_i^2's would cancel numerator-denominator since we're adding the same set[1], the product of the σ_i^2's would cancel since we're multiplying over the same set, the 2π's will cancel as well leaving us with:

$$= \sqrt{\frac{\left(1 + \sigma_\mu^2 \sum_{i_L=1}^{n_L} \frac{1}{\sigma_i^2}\right)\left(1 + \sigma_\mu^2 \sum_{i_R=1}^{n_R} \frac{1}{\sigma_i^2}\right)}{1 + \sigma_\mu^2 \sum_{i=1}^n \frac{1}{\sigma_i^2}}} \exp\left(\frac{\sigma_\mu^2}{2}\left(\frac{\left(\sum_{i_L=1}^{n_L} \frac{R_i}{\sigma_i^2}\right)^2}{1 + \sigma_\mu^2 \sum_{i_L=1}^{n_L} \frac{1}{\sigma_i^2}} + \frac{\left(\sum_{i_R=1}^{n_R} \frac{R_i}{\sigma_i^2}\right)^2}{1 + \sigma_\mu^2 \sum_{i_R=1}^{n_R} \frac{1}{\sigma_i^2}} - \frac{\left(\sum_{i=1}^{n} \frac{R_i}{\sigma_i^2}\right)^2}{1 + \sigma_\mu^2 \sum_{i=1}^{n} \frac{1}{\sigma_i^2}}\right)\right)$$

In log form this becomes:

$$\frac{1}{2}\left(\ln\left(1 + \sigma_\mu^2 \sum_{i_L=1}^{n_L} \frac{1}{\sigma_i^2}\right) + \ln\left(1 + \sigma_\mu^2 \sum_{i_R=1}^{n_R} \frac{1}{\sigma_i^2}\right) - \ln\left(1 + \sigma_\mu^2 \sum_{i=1}^{n} + \frac{1}{\sigma_i^2}\right)\right) +$$
$$\frac{\sigma_\mu^2}{2}\left(\frac{\left(\sum_{i_L=1}^{n_L} \frac{R_i}{\sigma_i^2}\right)^2}{1 + \sigma_\mu^2 \sum_{i_L=1}^{n_L} \frac{1}{\sigma_i^2}} + \frac{\left(\sum_{i_R=1}^{n_R} \frac{R_i}{\sigma_i^2}\right)^2}{1 + \sigma_\mu^2 \sum_{i_R=1}^{n_R} \frac{1}{\sigma_i^2}} - \frac{\left(\sum_{i=1}^{n} \frac{R_i}{\sigma_i^2}\right)^2}{1 + \sigma_\mu^2 \sum_{i=1}^{n} \frac{1}{\sigma_i^2}}\right)$$

[1]The set of left leaf values $\cup$ the set of right leaf values has to equal the set of the node's parent's values