BART update for tree changes

Andrew C. Parnell

February 22, 2018

Suppose there are n observations in a terminal node and suppose that the (partial) residuals in this terminal node are denoted $R_1, \ldots, R_n$. The prior distribution for these residuals is:

$$R_1, \ldots, R_n | \mu, \tau \sim N(\mu, \tau^{-1})$$

Furthermore the prior on μ is:

$$\mu \sim N(0, \tau_\mu^{-1})$$

Using π to denote a probability distribution, we want to find:

$$\begin{aligned}
\pi(R_1, \ldots, R_n | \sigma) &= \int \pi(R_1, \ldots, R_n | \mu, \tau) \pi(\tau) \partial \mu \\
&\propto \prod_{i=1}^{n} \tau^{1/2} e^{-\frac{\tau}{2}(R_i - \mu)^2} \tau_\mu^{1/2} e^{-\frac{\tau_\mu}{2}\mu^2} \partial \mu \\
&= \int \tau^{n/2} e^{-\frac{\tau}{2}\sum(R_i - \mu)^2} \tau_\mu^{1/2} e^{-\frac{\tau_\mu}{2}\mu^2} \partial \mu \\
&= \int \tau^{n/2} \tau_\mu^{1/2} e^{-\frac{1}{2}\left[\tau\left\{\sum R_i^2 + n\mu^2 - 2\mu n \bar{R}\right\} + \tau_\mu \mu^2\right]} \partial \mu \\
&= \tau^{n/2} \tau_\mu^{1/2} e^{-\frac{1}{2}\left[\tau \sum R_i^2\right]} \int e^{-\frac{1}{2}Q} \partial \mu
\end{aligned}$$

where

$$
\begin{aligned}
Q =& \tau n\mu^2 - 2\tau n\mu\bar{R} + \tau_\mu\mu^2 \\
=& (\tau_\mu + n\tau)\mu^2 - 2\tau n\mu\bar{R} \\
=& (\tau_\mu + n\tau)\left[\mu^2 - \frac{2\tau n\mu\bar{R}}{\tau_\mu + n\tau}\right] \\
=& (\tau_\mu + n\tau)\left[\left(\mu - \frac{2\tau n\bar{R}}{\tau_\mu + n\tau}\right)^2 - \left(\frac{\tau n\bar{R}}{\tau_\mu + n\tau}\right)^2\right] \\
=& (\tau_\mu + n\tau)\left(\mu - \frac{2\tau n\bar{R}}{\tau_\mu + n\tau}\right)^2 - \frac{(n\tau\bar{R})^2}{\tau_\mu + n\tau}
\end{aligned}
$$

so therefore:

$$
\begin{aligned}
\int e^{-\frac{1}{2}Q}\partial\mu =& \int \exp\left[-\frac{\tau_\mu + n\tau}{2}\left(\mu - \frac{2\tau n\bar{R}}{\tau_\mu + n\tau}\right)^2 + \frac{(n\tau\bar{R})^2}{2(\tau_\mu + n\tau)}\right]\partial\mu \\
\propto& \exp\left[\frac{1}{2}\frac{(\tau n\bar{R})^2}{\tau_\mu + n\tau}\right](\tau_\mu + n\tau)^{-1/2}
\end{aligned}
$$

And finally:

$$
\begin{aligned}
\pi(R_1, \ldots, R_n|\tau) \propto& (\tau_\mu + n\tau)^{-1/2}\tau^{n/2}\tau_\mu^{1/2}\exp\left[\frac{1}{2}\frac{(\tau n\bar{R})^2}{\tau_\mu + n\tau}\right]\exp\left[-\frac{\tau}{2}\sum R_i^2\right] \\
=& \tau^{n/2}\left(\frac{\tau_\mu}{\tau_\mu + n\tau}\right)^{1/2}\exp\left[-\frac{\tau}{2}\left\{\sum R_i^2 - \frac{\tau(n\bar{R})^2}{\tau_\mu + n\tau}\right\}\right]
\end{aligned}
$$

Including multiple terminal nodes

When we put back in terminal nodes we write R_{ji} where j is the terminal node and i is still the observation, so in terminal node j we have partial residuals $R_{j1}, \ldots, R_{jn_j}$. When we have $j = 1, \ldots, b$ terminal nodes the full conditional distribution is then:

$$
\prod_{j=1}^{b}\pi(R_{j1}, \ldots, R_{jn_j}|\tau) \propto \prod_{j=1}^{b}\left\{\tau^{n_j/2}\left(\frac{\tau_\mu}{\tau_\mu + n_j\tau}\right)^{1/2}\exp\left[-\frac{\tau}{2}\left\{\sum_{i=1}^{n_j}R_{ji}^2 - \frac{\tau(n_j\bar{R}_j)^2}{\tau_\mu + n_j\tau}\right\}\right]\right\}
$$

which on the log scale gives:

$$\sum_{j=1}^{b}\left\{\frac{n_j}{2}\log(\tau)+\frac{1}{2}\log\left(\frac{\tau_\mu}{\tau_\mu+n_j\tau}\right)-\frac{\tau}{2}\left[\sum_{i=1}^{n_j}R_{ji}^2-\frac{\tau(n_j\bar{R}_j)^2}{\tau_\mu+n_j\tau}\right]\right\}$$