Bayesian Additive Regression Trees using Gaussian Processes

Mateus Maia, Keefe Murphy, Andrew Parnell
Maynooth University
mateus.maiamarques.2021@mumail.ie

ArXiv paper:

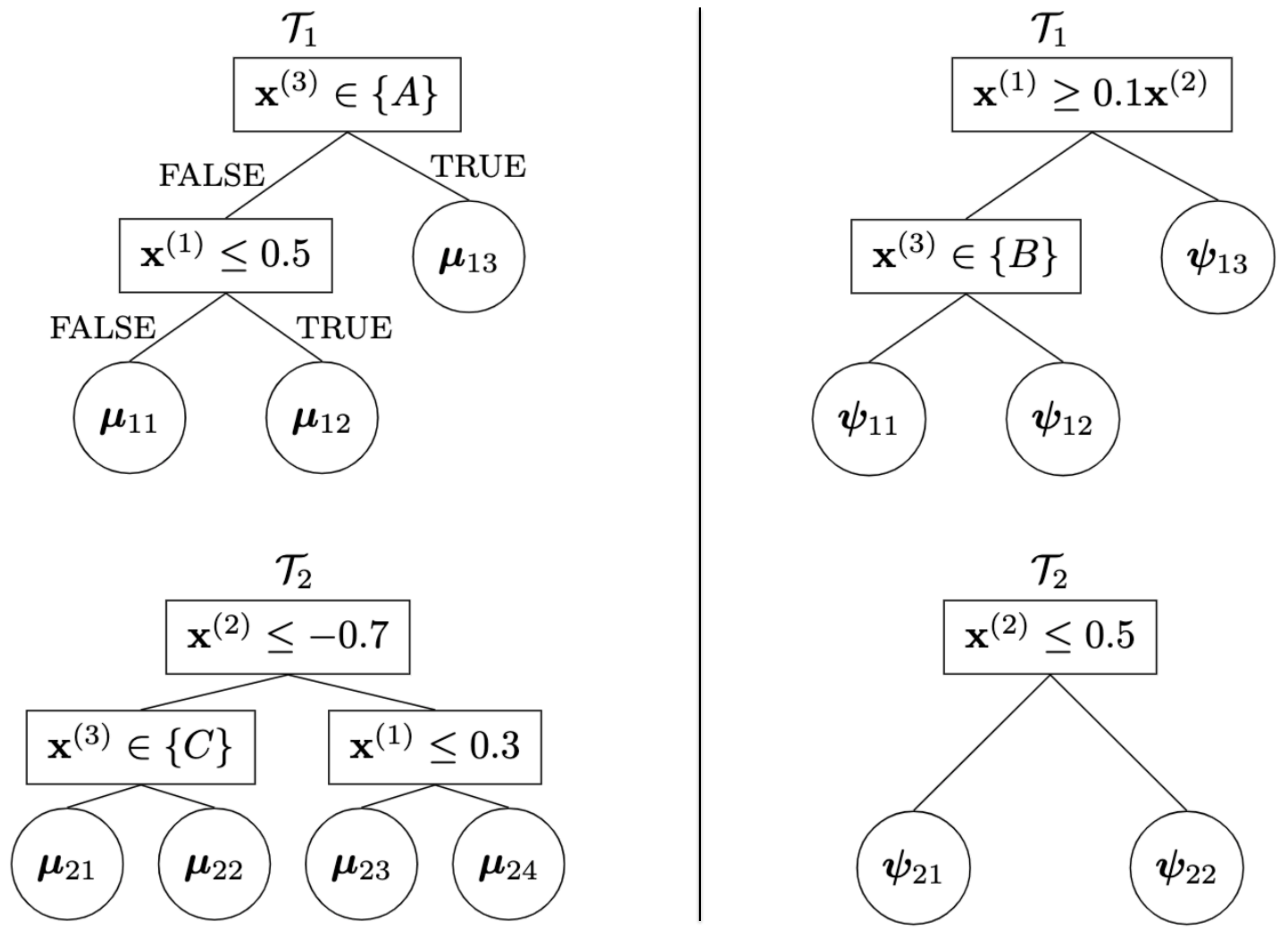
The problem

The Bayesian additive regression trees (BART) approach learns through a **sum of Bayesian trees**, where each terminal node's mean parameter $\mu_{t\ell}$ is constrained by a **regularising prior distribution.** The model is described by:

$$y_i \mid \mathbf{x}_i \sim \mathrm{N}\left(\sum_{t=1}^{T} h\left(\mathbf{x}_i; \mathscr{T}_t, \mathbf{L}_t\right), \tau^{-1}\right)$$

The vector $\mathbf{L}_t = (\mu_{t1}, \ldots, \mu_{tb_t})$ collects all mean parameters $\mu_{t\ell}$ from a tree $\mathscr{T}_t$ and the function h assigns it to an observation $\boldsymbol{x}_i$ that belongs to a terminal node ℓ.

This assignment of a single predicted value for all observations sharing a terminal node introduces a **stepwise-continuous** nature, such that the estimated functions from BART are generally **non-smooth**.



On the left, graphical representation of two example trees from a BART model. On the right, the graphical representation for GP-BART trees, with a GP prior $\boldsymbol{\psi}_{t\ell}$ at the terminal node level.

The solution

The main contribution of this work is to define a **Gaussian process (GP) prior** in the terminal nodes.

$$\boldsymbol{\psi}_{t\ell} \mid \mathscr{T}_t, \mu_{t\ell}, \boldsymbol{\phi}_t, \nu \sim \mathrm{GP}\left(\boldsymbol{\mu}_{t\ell} = \mu_{t\ell}\mathbf{1}_{n_{t\ell}}, \nu^{-1}\exp\left\{-\frac{1}{2}\sum_{j=1}^{p}\frac{\left(x_i^{(j)} - x_k^{(j)}\right)^2}{\phi_{tj}^2}\right\}\right)$$

- To prevent any single tree from dominating the ensemble, the ν parameter fixes the GP's precision at $\nu = 8\kappa^2 T$, which regularises all trees.
- A gamma prior over the length parameters ϕ_{tj} allows Automatic Relevance Determination (ARD) and variable selection within the GP.

The algorithm

Algorithm 1: GP-BART sampling algorithm

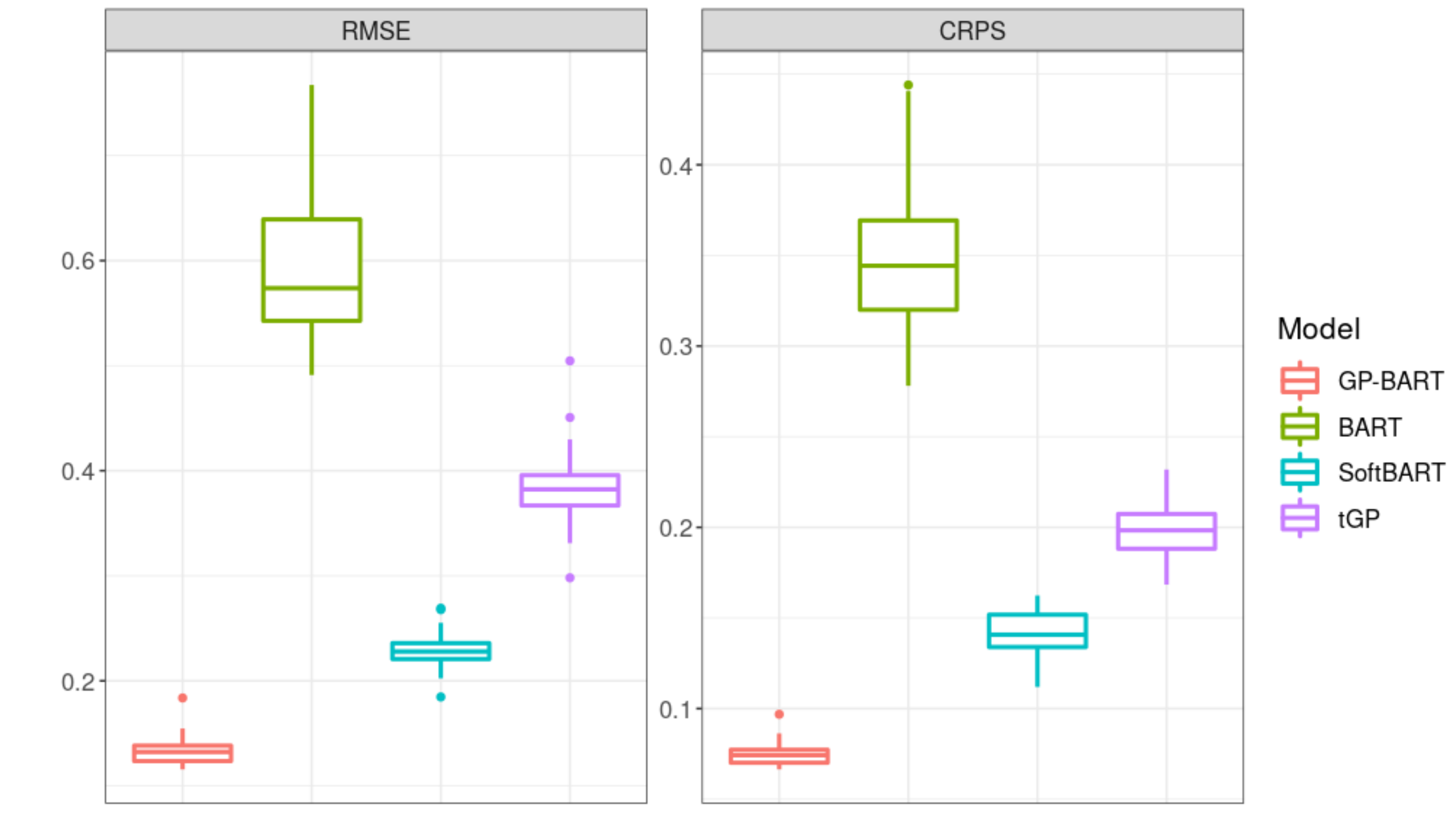
Input: $\mathbf{X}$, $\mathbf{y}$, T, N_{MCMC}, and all hyperparameters of the priors.
Initialise: T tree stumps, $\tau = 1$, and $\phi_{tj} = 1\,\forall\,(t, j)$.
for *iterations m from 1 to N_{MCMC}* **do**
  for *trees t from 1 to T* **do**
    Calculate the partial residuals $\mathbf{R}_t$;
    Propose a new tree $\mathcal{T}_t^\star$ by a grow, grow-project, change, change-project, or prune move;
    Accept and update $\mathcal{T}_t = \mathcal{T}_t^\star$ with probability

$$\gamma^\star(\mathcal{T}_t, \mathcal{T}_t^\star) = \min\left\{1, \frac{\pi(\mathbf{R}_t \mid \mathcal{T}_t^\star, \boldsymbol{\phi}_t, \nu, \tau_\mu, \tau)\pi(\mathcal{T}_t^\star)}{\pi(\mathbf{R}_t \mid \mathcal{T}_t, \boldsymbol{\phi}_t, \nu, \tau_\mu, \tau)\pi(\mathcal{T}_t)}\right\}.$$

    for *terminal nodes ℓ from 1 to b_t* **do**
      Update $\boldsymbol{\psi}_{t\ell}$ using Gibbs.
    end
    Update $\boldsymbol{\phi}_t$ using MH.
  end
  Update τ using Gibbs.
end
Output: Posterior samples from $\pi\left((\mathcal{T}_1, \mathbf{G}_1), \ldots, (\mathcal{T}_T, \mathbf{G}_T), \tau \mid \mathbf{y}\right)$.

Results

- GP-BART was compared against other tree-based methods (BART, SoftBART, tGP).
- Simulation scenarios employed in the study used the Friedman data [1] with 5 additional noise variables.



Comparison of RMSE and CRPS over the 25 test folds from 5 repetitions of 5-fold cross-validation for the Friedman data set with $n = 500$ and $p = 10$.

- For a comparison using real data, four **spatial** datasets were analysed and GP-BART **outperformed** the competitors by presenting the lowest values of RMSE and CRPS in three of the four applications.
- Overall, GP-BART presents a promising extension of BART by imposing covariance structure to achieve **smoothness**.
- An R implementation of GP-BART is available at github.com/MateusMaiaDS/gpbart

References

[1] Friedman, Jerome H. "Multivariate adaptive regression splines." The Annals of Statistics 19.1 (1991): 1-67.

Acknowledgements

Mateus Maia's work was supported by Science Foundation Ireland Career Development Award grant number 17/CDA/4695 and SFI research centre award 12/RC/2289P2.