Penalised Splines BART

Adding smoothness to Bayesian Additive Regression Trees

Mateus Maia, Keefe Murphy and Andrew Parnell

Maynooth University, Ireland.

CASI 2023

A joint work

Andrew C. Parnell

Keefe Murphy

Bayesian Additive Regression Trees (BART)

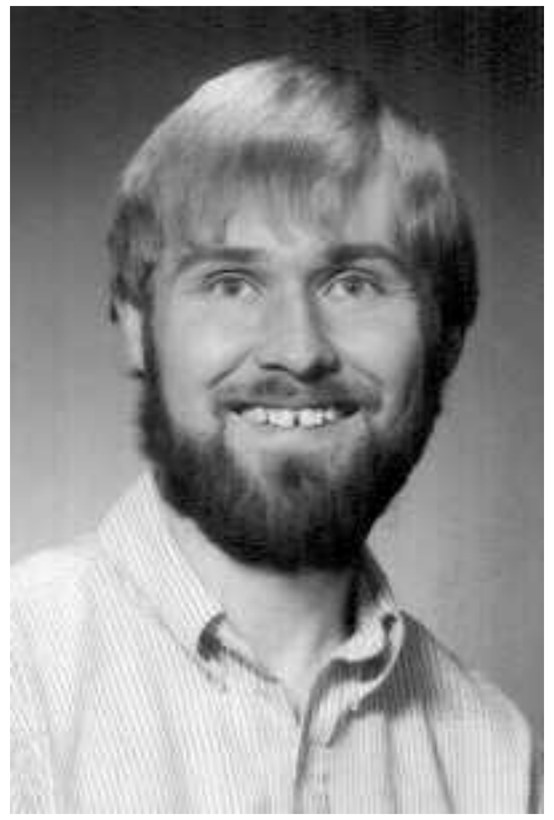
Bayesian Additive Regression Trees (BART)

March 2010

BART: Bayesian additive regression trees

Hugh A. Chipman, Edward I. George, Robert E. McCulloch

Ann. Appl. Stat. 4(1): 266-298 (March 2010). DOI: 10.1214/09-AOAS285



Huge A. Chipman

Edward I. George

Robert E. McCulloch

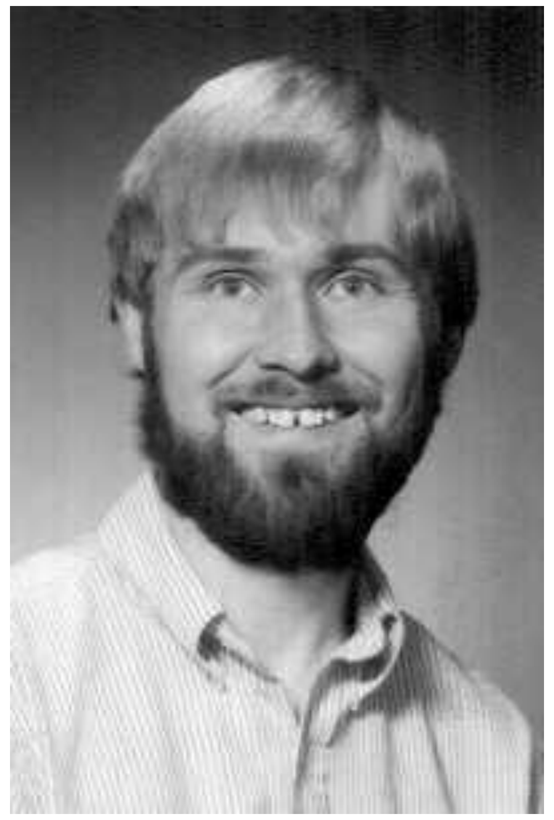
Bayesian Additive Regression Trees (BART)


March 2010

BART: Bayesian additive regression trees

Hugh A. Chipman, Edward I. George, Robert E. McCulloch

Ann. Appl. Stat. 4(1): 266-298 (March 2010). DOI: 10.1214/09-AOAS285



Huge A. Chipman

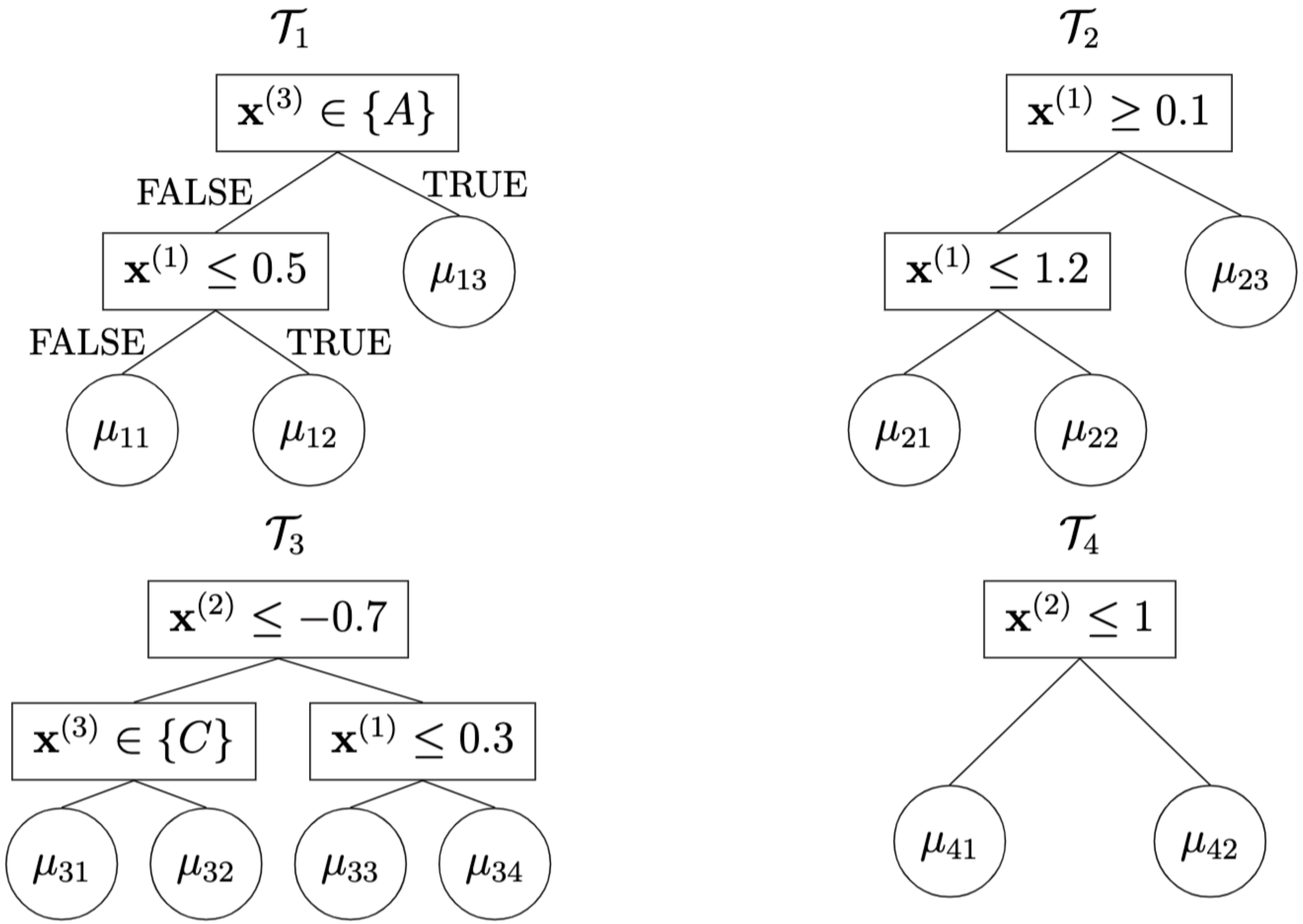
Edward I. George

Robert E. McCulloch

BART and 2016 Causal Inference Data Analysis Challenge (Dorie et. al, 2019)

BART

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



$\mu_{\ell t}$ priors:

$$\mu_{t\ell} \,|\, \mathscr{T}_t, \tau \sim \mathrm{N}(0, \tau_\mu^{-1})$$

$$\tau_\mu = 4k^2T$$

BART

Algorithm 1: BART sampling algorithm

Input: $\mathbf{X}$, $\mathbf{y}$, T, N_{MCMC}, and all hyperparameters of the priors.
Initialise: T tree stumps, $\tau = 1$.
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, change, or prune move;
    Accept and update $\mathcal{T}_t = \mathcal{T}_t^\star$ with probability

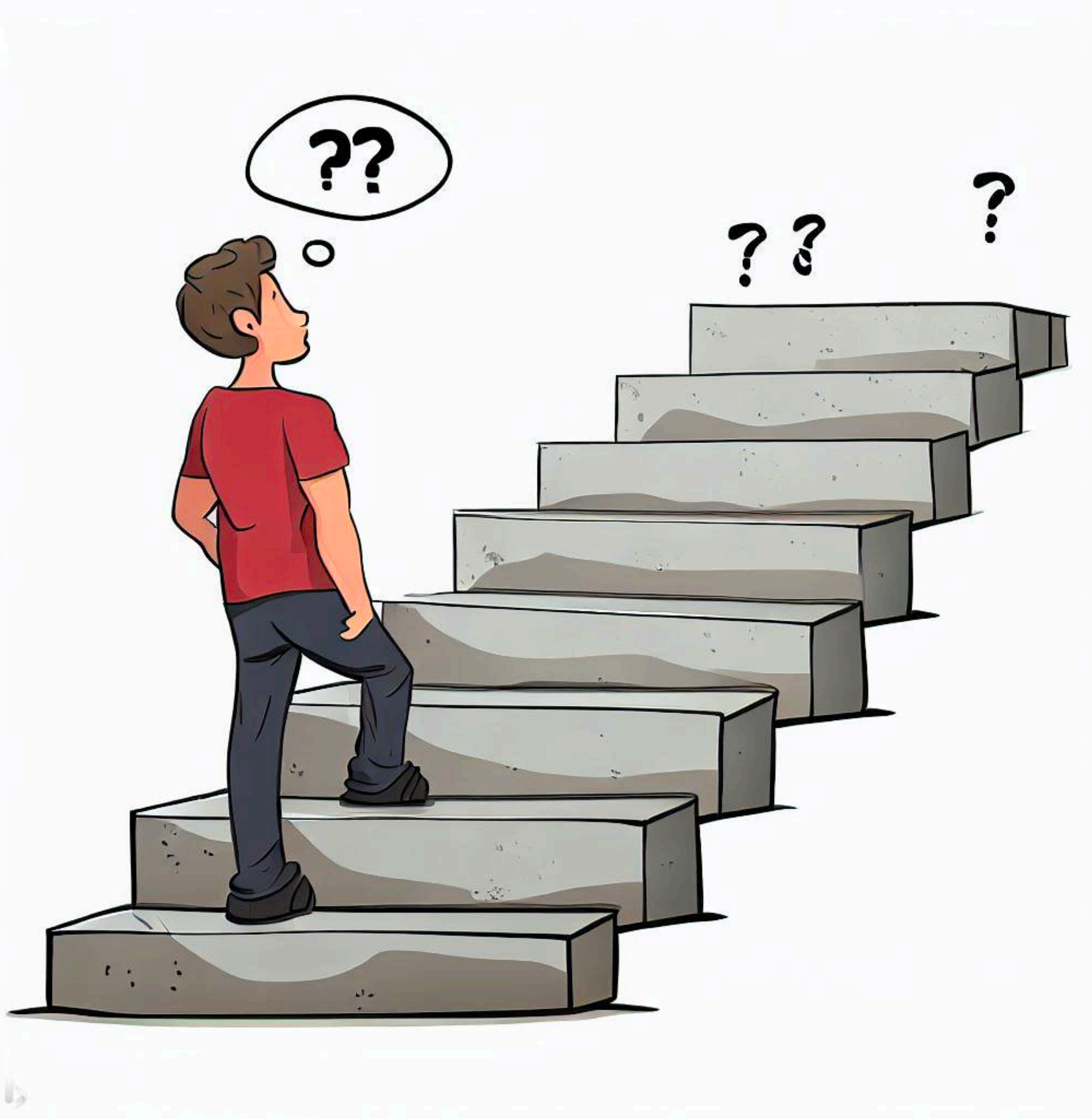
$$\gamma^\star\left(\mathcal{T}_t, \mathcal{T}_t^\star\right) = \min\left\{1, \frac{\pi\left(\mathbf{R}_t \mid \mathcal{T}_t^\star, \tau_\mu, \tau\right)\pi\left(\mathcal{T}_t^\star\right)}{\pi\left(\mathbf{R}_t \mid \mathcal{T}_t, \tau_\mu, \tau\right)\pi\left(\mathcal{T}_t\right)}\right\}$$

    for *terminal nodes* ℓ *from* 1 *to* b_t **do**
      Update $\mu_{t\ell}$
    end
  end
  Update τ
end
Output: Approx. samples from $\pi\left((\mathcal{T}_1, \mathbf{G}_1), \ldots, (\mathcal{T}_T, \mathbf{G}_T), \tau \mid \mathbf{y}\right)$.

BART

$$y = sin(x) + N(0, \sigma^2)$$

How to overcome the inherent step-wise behaviour from BART?



BART extensions

- **Piecewise-constant nature from the model ($\mu_{t\ell}$)**
  - **Soft-BART** - Linero, A. R., & Yang, Y. (2018). Bayesian regression tree ensembles that adapt to smoothness and sparsity Series B Statistical methodology.
  - **MOTR-BART** - Prado, E. B., Moral, R. A., & Parnell, A. C. (2021). Bayesian additive regression trees with model trees. Statistics and Computing, 31(3), 1-13.
  - **GP-BART** - Maia, M., Murphy, K., & Parnell, A. C. (2022). GP-BART: a novel Bayesian additive regression trees approach using Gaussian processes. arXiv preprint arXiv:2204.02112.

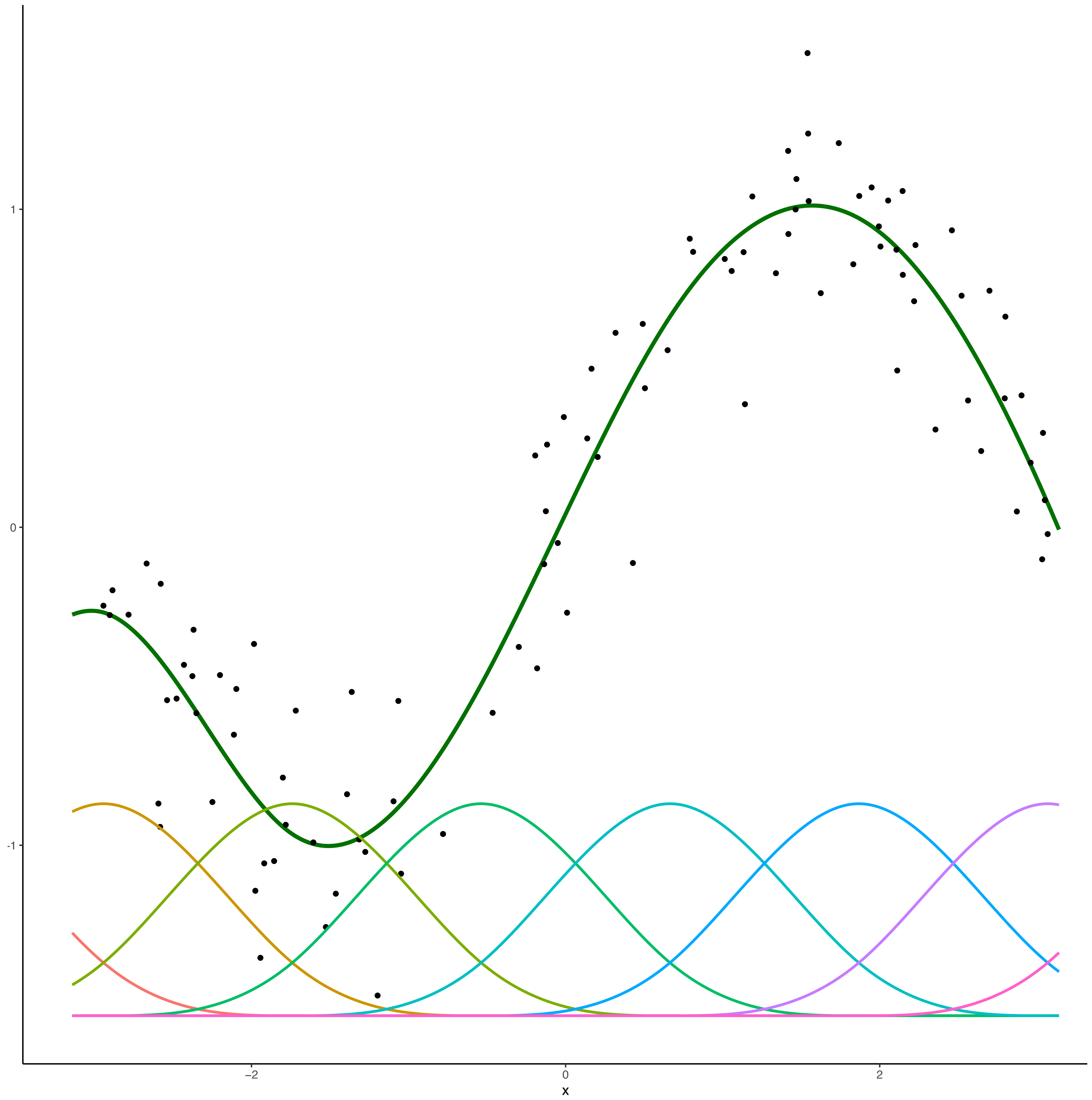
BART extensions

- **Piecewise-constant nature from the model ($\mu_{t\ell}$)**
    - **Soft-BART** - Linero, A. R., & Yang, Y. (2018). Bayesian regression tree ensembles that adapt to smoothness and sparsity Series B Statistical methodology.
    - **MOTR-BART** - Prado, E. B., Moral, R. A., & Parnell, A. C. (2021). Bayesian additive regression trees with model trees. Statistics and Computing, 31(3), 1-13.
    - **GP-BART** - Maia, M., Murphy, K., & Parnell, A. C. (2022). GP-BART: a novel Bayesian additive regression trees approach using Gaussian processes. arXiv preprint arXiv:2204.02112.

Penalised Splines BART

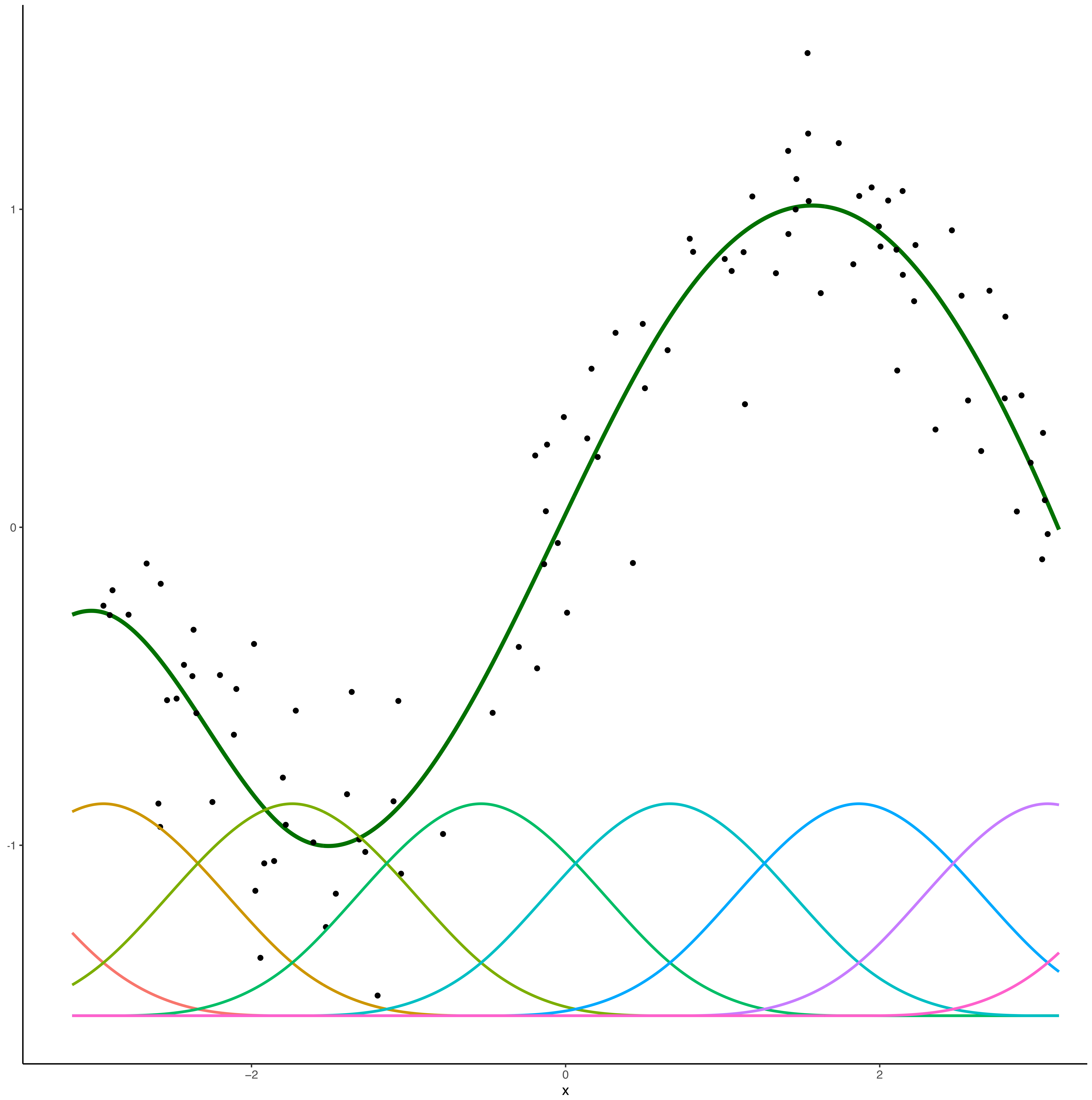
Splines

- "In words, a $\mathbf{k}$**-th** order spline is a piecewise polynomial function of **degree** $k-1$, that is continuous and has **continuous derivatives** of orders $1, \ldots, k$, at its **knot points**".



Splines

- "In words, a $\mathbf{k}$**-th** order spline is a piecewise polynomial function of **degree** $k-1$, that is continuous and has **continuous derivatives** of orders $1,\ldots,k$, at its **knot points**".



P-Splines

- Eilers, P. H., & Marx, B. D. (1996).
- "…choosing the optimal number and positions of knots is a **complex task.** "

Splines BART (sBART)

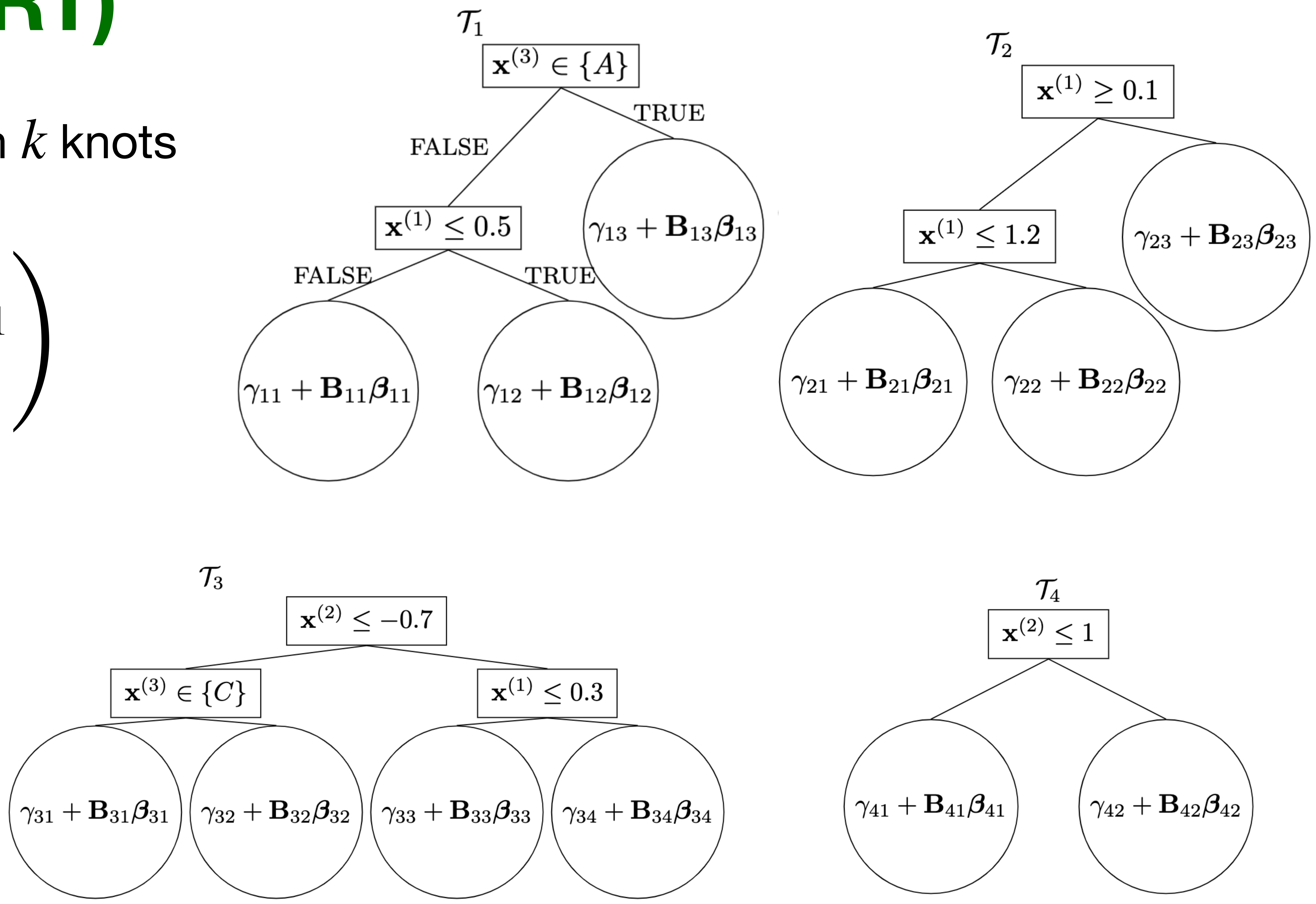
Build a natural B-spline $\mathbf{B}$ matrix with k knots

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

$\boldsymbol{\beta}_{\ell t}$ and $\gamma_{t\ell}$ priors:

$$\boldsymbol{\beta}_{t\ell} \sim \mathrm{N}\left(0, \tau_\beta^{-1} \mathbf{I}_p\right)$$

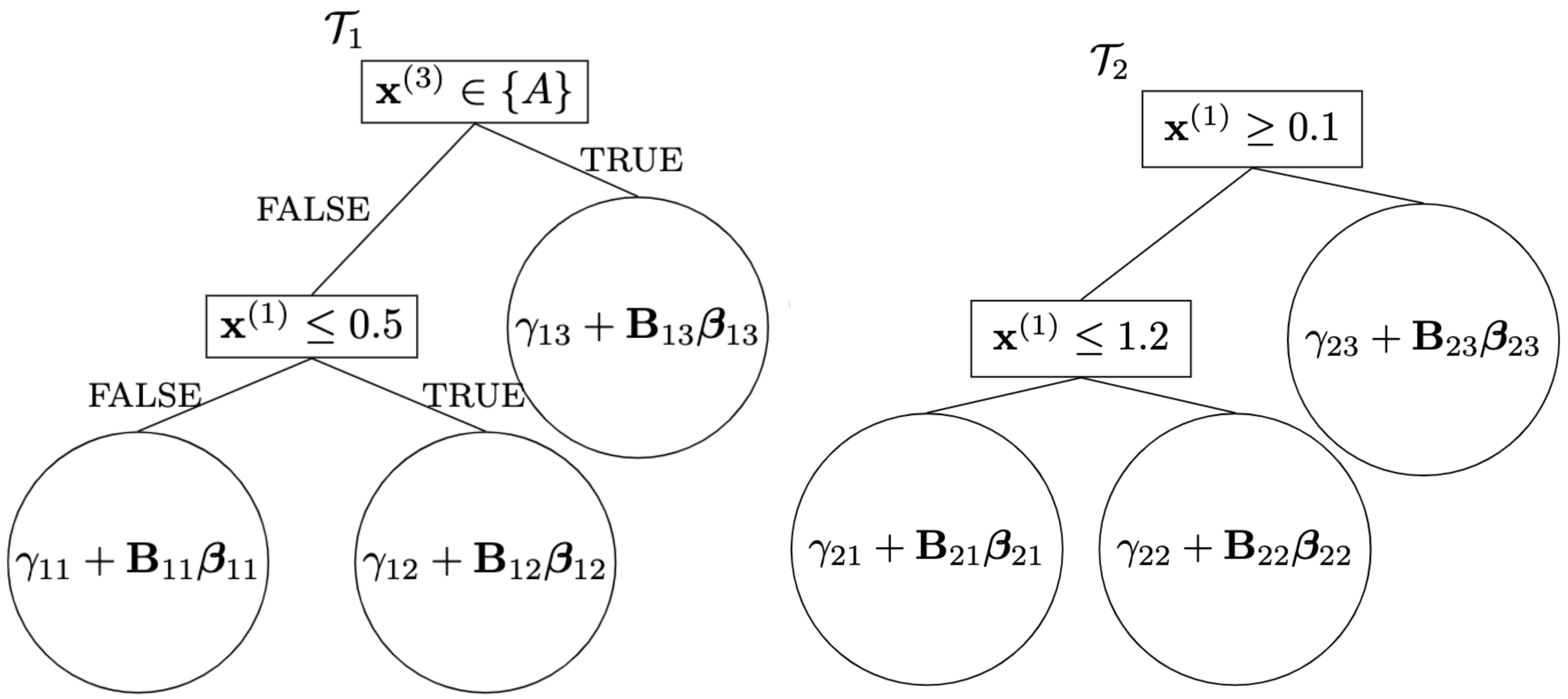
$$\gamma_{t\ell} \sim \mathrm{N}\left(0, \tau_\gamma^{-1}\right)$$



Penalised Splines BART (sBART)

Build a natural B-spline $\mathbf{B}$ matrix with k knots, penalising the $\boldsymbol{\beta}_{t\ell}$ differences

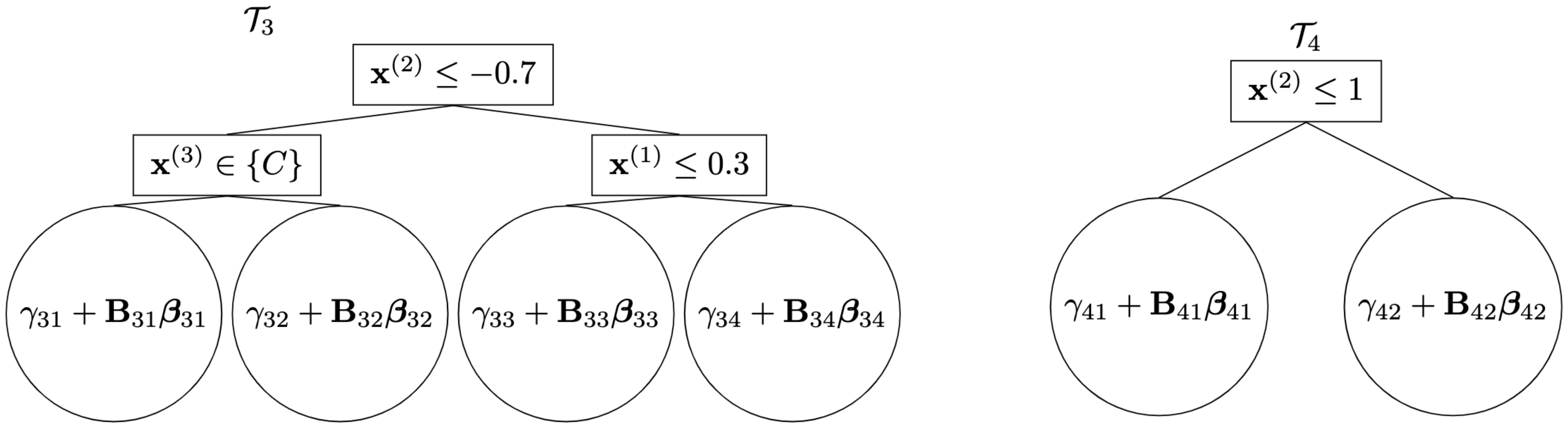
$$y_i \,|\, \mathbf{x}_i \sim \mathrm{N}\left(\sum_{t=1}^{T} b\left(\mathbf{x}_i; \mathcal{T}_t, \mathbf{L}_t\right), \tau^{-1} \right)$$



$\boldsymbol{\beta}_{\ell t}$ and $\gamma_{t\ell}$ priors:

$$\boldsymbol{\beta}_{t\ell} \sim \mathrm{N}\left(0, \tau_\beta^{-1}\mathbf{P}^{-1} + \tau_0^{-1}\mathbf{P}_0^{-1}\right)$$

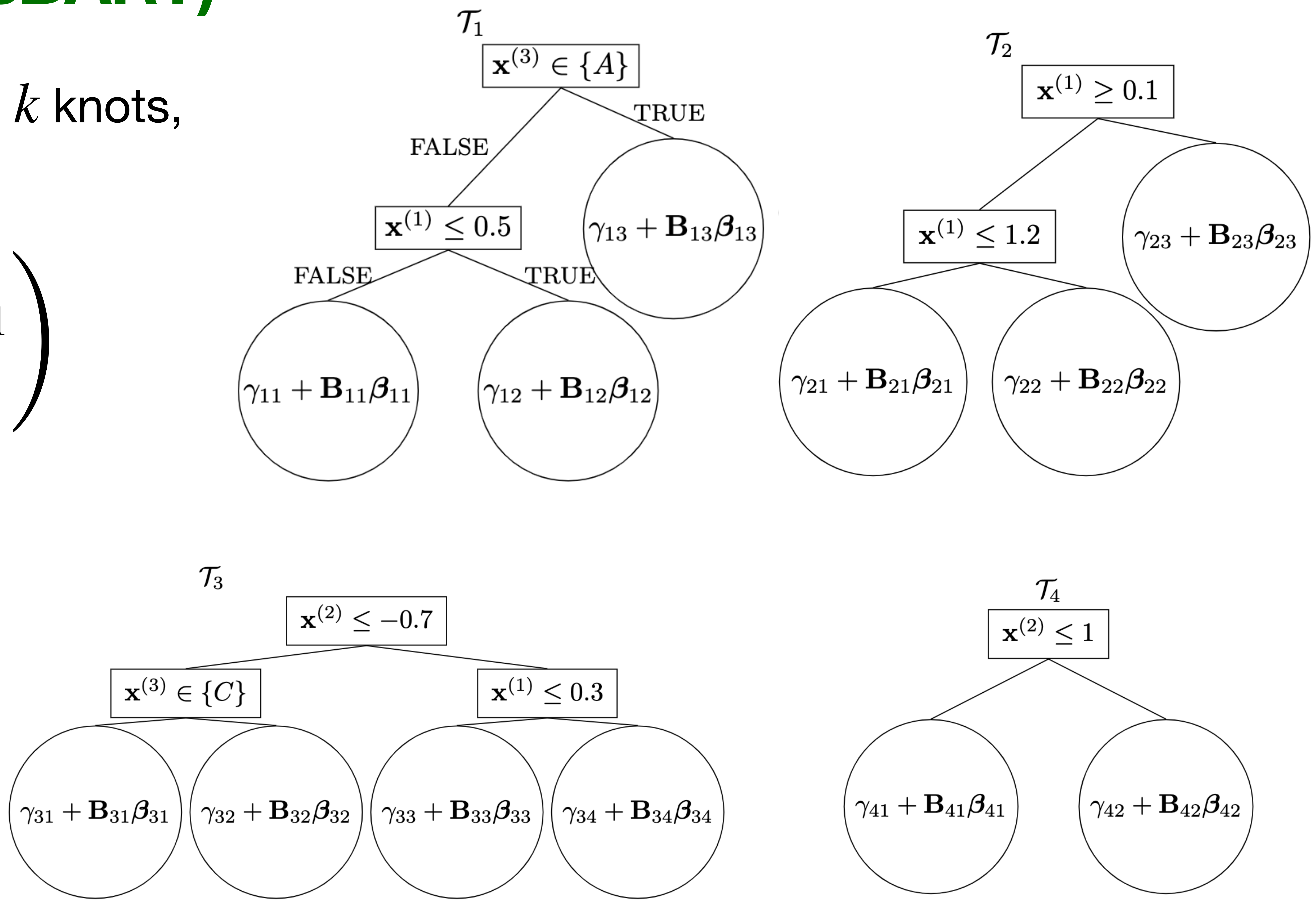
$$\gamma_{t\ell} \sim \mathrm{N}\left(0, \tau_\gamma^{-1}\right)$$



Penalised Splines BART (sBART)

Build a natural B-spline $\mathbf{B}$ matrix with k knots, penalising the $\boldsymbol{\beta}_{t\ell}$ differences

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



$\boldsymbol{\beta}_{\ell t}$ and $\gamma_{t\ell}$ priors:

$$\boldsymbol{\beta}_{t\ell} \sim \mathrm{N}\left(0, \tau_\beta^{-1}\mathbf{P}^{-1} + \tau_0^{-1}\mathbf{P}_0^{-1}\right)$$

$$\gamma_{t\ell} \sim \mathrm{N}\left(0, \tau_\gamma^{-1}\right)$$

$$\tau_\gamma = 4k^2T$$

Penalised Splines BART

(Priors)

Penalised Splines BART

(Priors)

Penalised Splines BART

(Priors)

$\boldsymbol{\beta}_{t\ell}$ priors

$$\boldsymbol{\beta}_{t\ell} \sim \mathrm{N}\left(0, \tau_\beta^{-1}\mathbf{P}^{-1} + \tau_0^{-1}\mathbf{P}_0^{-1}\right)$$

$$\tau_\beta \,|\, \delta \sim \mathrm{Gamma}\left(\frac{\nu}{2}, \frac{\delta\nu}{2}\right)$$

$$\delta \sim \mathrm{Gamma}\left(a_\delta, d_\delta\right)$$

$$\tau_0 \sim \mathrm{Gamma}\left(a_\eta, d_\eta\right)$$

$\gamma_{t\ell}$ priors

$$\gamma_{t\ell} \sim \mathrm{N}\left(0, \tau_\gamma^{-1}\right)$$

$$\tau_\gamma = 4k^2T$$

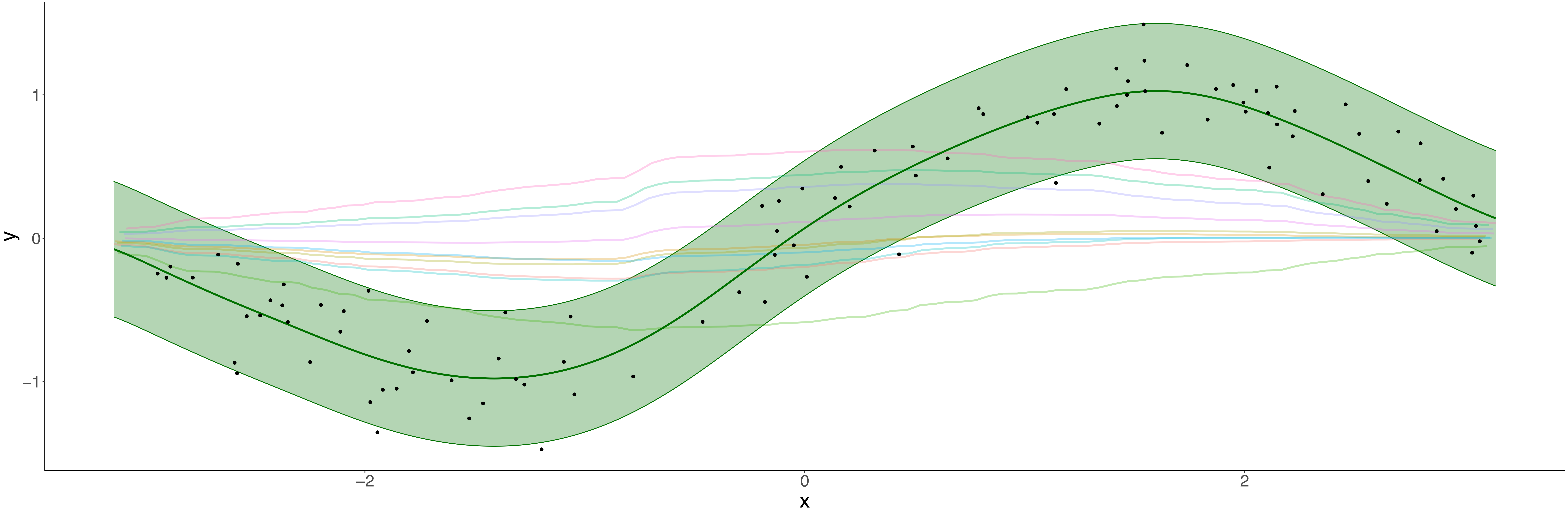
Illustrating psBART in action

BARE

$$y = sin(x) + N(0,\sigma^2)$$

psBART

psBART



A spline model would be enough?

psBART

Final Remarks

Remarks

- The extension of BART using splines seems to be a good choice to tackle the lack of smoothness from the standard formulation.

- Despite satisfactory results on first trials over the real dataset benchmarking, a extensive and exhaustive comparison using a broader range of datasets is required.

- The choice of priors hyper parameters and the effect over the interactions in the model is still a ongoing challenge to be investigated.

Thank you for the attention!
Questions?

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.

Fondúireacht Eolaíochta Éireann
Dá bhfuil romhainn
Science Foundation Ireland
For what's next

References

- Dorie, Vincent, et al. **"Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition."** (2019): 43-68.

- Eilers, Paul HC, and Brian D. Marx. **"Flexible smoothing with B-splines and penalties."** Statistical science 11.2 (1996): 89-121.

- Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. "**BART: Bayesian additive regression trees.**" (2010): 266-298.