Bayesian Projection Pursuit Regression

Gavin Collins☆ ◊, Devin Francom☆, Kellin Rumsey☆
☆Los Alamos National Laboratory, ◊Sandia National Laboratories
DATAWorks 2024
1. Regression

2. Bayesian Nonlinear Regression

3. Bayesian Projection Pursuit Regression
Regression

Data
• \((x_1, y_1), \ldots, (x_n, y_n)\)
• input \(x_i \in \mathbb{R}^p \rightarrow response \ y_i \in \mathbb{R}\)

Regression Model
• \(y_i \mid x_i \sim N(f(x_i), \sigma^2), i = 1, \ldots, n\) (ind)
• \(f: \mathbb{R}^p \rightarrow \mathbb{R}\) (mean function)
• \(\sigma^2 \geq 0\) (residual variance)

Hurricane Example:
• \(p = 6\)
• \(f =\) Nature/Physics
• \(n = 4,000\)

Input \(x\)
• \(x_1 =\) Initial Sea Level
• \(x_2 =\) Hurricane Heading
• \(x_3 =\) Velocity of the Eye
• \(x_4 =\) Max Wind Speed
• \(x_5 =\) Min Pressure
• \(x_6 =\) Landfall Location

Response \(y\)
\(y =\) Maximum Water Level at a Particular Location
Regression

Data
- \((x_1, y_1), ..., (x_n, y_n)\)
- input \(x_i \in \mathbb{R}^p \rightarrow\) response \(y_i \in \mathbb{R}\)

Regression Model
- \(y_i | x_i \sim N(f(x_i), \sigma^2), i = 1, ..., n\) (ind)
- \(f: \mathbb{R}^p \rightarrow \mathbb{R}\) (mean function)
- \(\sigma^2 \geq 0\) (residual variance)

“Lim” Example:
- \(p = 2\)
- \(f = \text{“The Lim Function” (Lim et al., 2002)}\)
- \(n = 350\)
- \(x_1, ..., x_{350} \sim \text{Unif}(0,1)^2\) (iid)
- \(\sigma^2 = 1\)
Bayesian Nonlinear Regression

\[ f(x) = \sum_{j=1}^{m} g(x; \theta_j) \]

Additive Structure
- Basis function \( g \) chosen ahead of time (I’ll show examples)
- Number of basis functions \( m \) either user-chosen or learned from data
- Parameters \( \theta_1, \ldots, \theta_m \) learned from data

Bayesian Model-Fitting
- Sample from the posterior distribution of \( \theta_1, \ldots, \theta_m \)
- Provides uncertainty intervals for \( f(x) \)
Bayesian Additive Regression Trees (BART)

\[ f(x) = \sum_{j}^{m} g(x; T_j) \]

**Additive Structure**
- Parameter is a decision tree \( T_j \)
- Number of trees \( m \) typically user-chosen (e.g., \( m = 200 \))
- My dissertation develops a method to learn \( m \) from data

**Bayesian Model-Fitting**
- Estimate the posterior distribution of \( T_1, \ldots, T_m \) via MCMC

Hugh A. Chipman, Edward I. George, Robert E. McCulloch "BART: Bayesian additive regression trees," The Annals of Applied Statistics, Ann. Appl. Stat. 4(1), 266-298, (March 2010)
Bayesian Additive Regression Trees (BART)

\[ g(x; T) \]

\[ g((x_1, x_2); T) \]

\[ \sum_{j=1}^{m} g((x_1, x_2); T_j) \]
Bayesian Additive Regression Trees (BART)

$R^2: 0.972$
Coverage of 95% CI: 0.987
Bayesian Adaptive Spline Surfaces (BASS)

\[ f(x) = \sum_{j}^{m} g(x; S_j) \]

Additive Structure
- Parameter is a set \( S_j \) of 1D tensors
- Basis function \( g \) is a tensor product of the tensors in \( S_j \)
- Number of basis functions \( m \) learned from data

Bayesian Model-Fitting
- Estimate the posterior distribution of \( S_1, ..., S_m \) via MCMC

DENISON, D.G.T., MALLICK, B.K. & SMITH, A.F.M. Bayesian MARS. Statistics and Computing 8, 337–346 (1998).

Francom, D., & Sansó, B. (2020). BASS: An R Package for Fitting and Performing Sensitivity Analysis of Bayesian Adaptive Spline Surfaces. Journal of Statistical Software, 94(8), 1–36.
Bayesian Adaptive Spline Surfaces (BASS)

\[ g(x; S) \]

\[ g((x_1, x_2); S) \]

\[ \sum_{j=1}^{m} g((x_1, x_2); S_j) \]
Bayesian Adaptive Spline Surfaces (BASS)

$R^2$: 0.988
Coverage of 95% CI: 0.867

Observed Response $y$

---

BASS Prediction

BASS Lower Bound

BASS Upper Bound
Bayesian Projection Pursuit Regression (BayesPPR)

\[ f(x) = \sum_{j}^{m} g(x; b_j, \phi_j) \]

Additive Structure
- Parameters are projections \( \phi_j \) and transformations \( b_j \)
- \( g \) is a transformation of a projection \( g(x; b_j, \phi_j) = b_j(x' \phi_j) \)
- Number of basis functions \( m \) learned from data

Bayesian Model-Fitting
- Estimate the posterior distribution of \((b_1, \phi_1), \ldots, (b_m, \phi_m)\) via MCMC

Friedman, Jerome H., and Werner Stuetzle. “Projection Pursuit Regression.” *Journal of the American Statistical Association*, vol. 76, no. 376, 1981, pp. 817–23.

Collins, G., Francom, D. & Rumsey, K. Bayesian projection pursuit regression. *Stat Comput* **34**, 29 (2024).
Bayesian Projection Pursuit Regression (BayesPPR)

Form of Ridge Functions

\[ g_m(x' \theta_m) = \beta'_m b_m(x' \theta_m | t_{m0}) \]

- Coefficient Vector \( \beta'_m \in \mathbb{R}^K \)
- Basis expansion \( b_m : \mathbb{R} \to \mathbb{R}^K \)
- Knot point \( t_{m0} \)
- \( b_m(x' \theta_m | t_{m0}) = n_s_K((x' \theta_m - t_{m0})_+) \)
- \( (s)_+ = s1(s > 0) \) (ReLU)
Bayesian Projection Pursuit Regression (BayesPPR)

Sparse Projection Directions $\theta_m$

- $a_m \sim \text{Unif}\{1, \ldots, A\}$
  - $A \leq p$ user-chosen
  - Default: $A = 3$
- $\theta_m | a_m \sim \text{Unif}\{\theta_m \in S^p : \sum_{j=1}^{p} 1(\theta_{mj} \neq 0) = a_m\}$

Sparsity Example: $a_m = 3$

$$x'\theta_m = x' \begin{pmatrix} 0 \\ -0.47 \\ 0.77 \\ 0 \\ -0.43 \end{pmatrix} = -0.47x_2 + 0.77x_3 - 0.43x_5$$
Bayesian Projection Pursuit Regression (BayesPPR)

\[ g(x; b, \theta) \]

\[ g((x_1, x_2); b, \theta) \]

\[ \sum_{j=1}^{m} g((x_1, x_2); b_j, \theta_j) \]

\[ x_2 \]

\[ x_1 \]

\[ t_0 = 4, 6, 8, 10 \]

\[ x \theta_2 = 0.949x_1 + 0.316x_2 \]
Bayesian Projection Pursuit Regression (BayesPPR)

$R^2$: 0.993
Coverage of 95% CI: 0.974
Bayesian Projection Pursuit Regression
(BayesPPR)

Friedman Function: $f(x) = 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + 0x_6$

Interpretability

Ridge Functions from a single MCMC Iteration:
Comparison of Accuracy

BayesPPR

BASS
Comparison of Uncertainty Quantification

Real Data

Simulated Data

Average Coverage

LAGP
BPPR
BMARS
BART

Average Coverage

SIM
LAGP
BPPR
BMARS
BART
Summary

Bayesian nonlinear regression
• BART, BASS, BayesPPR
• Advantages:
  • Fast (for Bayesian models)
  • Scalable to moderately large n and p
  • Typically work well out of the box
  • Accurate prediction
  • Full UQ

BayesPPR
• Bayesian version of Friedman’s Projection Pursuit Regression
• New form for the ridge functions
• Automatic Variable Selection
• Somewhat interpretable
• Accurate prediction
• Reliable UQ
Thank you!

Distribution and Contact Info

• Collins, G., Francom, D. & Rumsey, K. Bayesian projection pursuit regression. *Stat Comput* **34**, 29 (2024).

• github.com/gqcollins/bayesppr

• gqcolli@sandia.gov