Bayesian Adaptive Regression Splines Using Reversible Jump MCMC

Andy Shen, Devin Francom

Statistical Sciences Group (CCS-6): Los Alamos National Laboratory
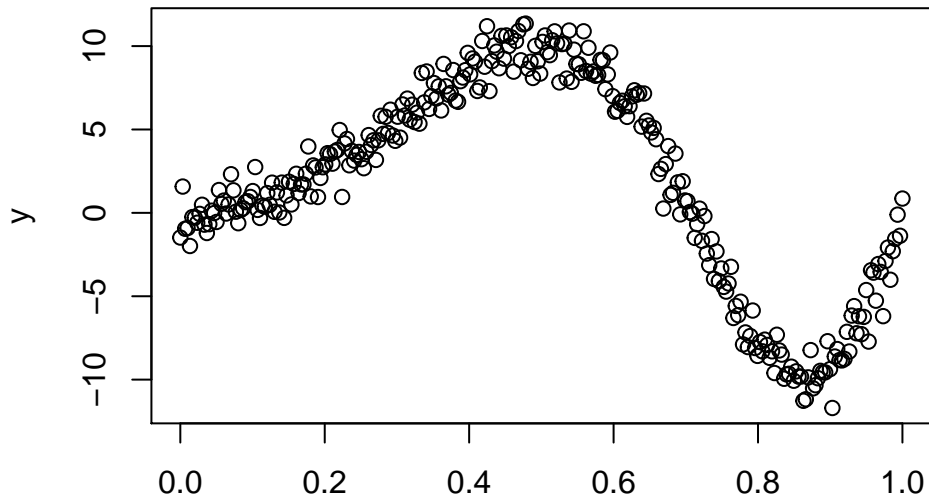
UCLA

Basis Functions and Splines

Let's say you want to fit a model using some wiggly data, such as

One way to fit a model to data like this is to come up with a linear basis and fit a linear model using the basis as the X matrix (which we will call B). Splines are often used as a basis. The simplest set of spline basis functions would be to make the ith basis function (i.e., the ith column of B) look like
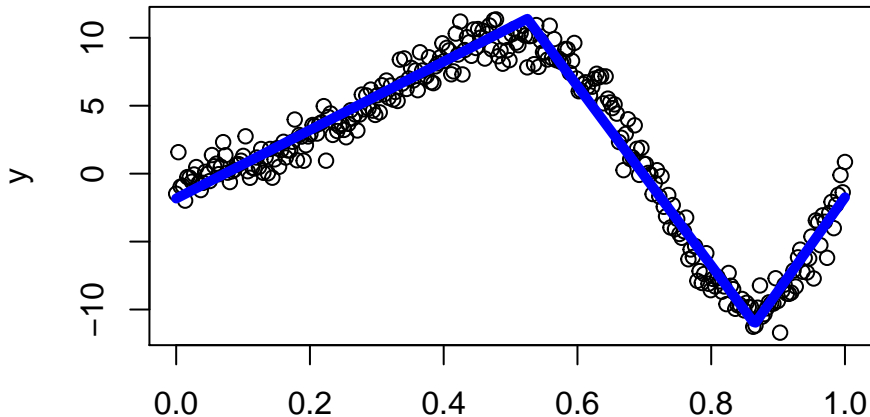
$$B_{ij} = [s_i(x_j - t_i)]_+$$

where $s \in \{-1, 1\}$, is called the sign, t_i (in the domain of x) is called the knot, and $[a]_+ = max(0, a)$.

Spline Function

From inspecting our data, we can manually choose the number of knots and their locations and produce a spline function:

```
t <- c(0, 0.525, 0.865); generate_spline(t)
```

Spline Function

- We may not always be able to visualize our data and manually choose the number of knots and their locations

- We want to use Bayesian techniques (BARS) to generate our spline function

Bayesian Adaptive Regression Splines (BARS)

BARS

- Stands for **B**ayesian **A**daptive **R**egression **S**plines
- We want to use BARS to generate values of the regression coefficients, variance, number of knots, and locations of knots
- **General idea**: find full conditional distributions for your parameters and then use MCMC to sample values from these full conditionals
  - Metropolis Hastings, Gibbs Sampling
- BARS can also be generalized to a multivariate setting (BMARS)

We assume each "piece" of our data follows the traditional linear model likelihood:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \;\; \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

The Bayesian Linear Model simply requires priors for the unknowns, $\boldsymbol{\beta}$ and σ^2, to condition on them

Presentation template courtesy Miles Chen. Used with permission.

9

Prior Distributions

We assume that our regression coefficients $\boldsymbol{\beta}$ follows a Gaussian distribution with mean 0 and covariance matrix $\tau^2 \boldsymbol{I}$ for some τ^2.

We then have that

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{0}, \tau^2 \boldsymbol{I})$$

We assume that σ^2 follows an Inverse Gamma distribution with parameters a and b:

$$\sigma^2 \sim IG(a, b)$$

Prior Distributions

For $\boldsymbol{\beta}$:

$$p(\boldsymbol{\beta}) = (2\pi\tau^2)^{-\frac{p}{2}} exp\left[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}\right]$$

$$\propto (\tau^2)^{-\frac{p}{2}} exp\left[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta}\right]$$

For σ^2:

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-\alpha-1}exp\left[\frac{\beta}{\sigma^2}\right]$$

$$\propto (\sigma^2)^{-\alpha-1}exp\left[\frac{\beta}{\sigma^2}\right]$$

$$L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}) = \prod_{i=1}^{N} p(y_i | \boldsymbol{\beta}, \sigma^2)$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \mu_i)^2 \right]$$

$$\propto (\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \mu_i)^2 \right]$$

$$\propto (\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right]$$

Joint Posterior

$$p(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}) \propto p(\sigma^2)\, p(\boldsymbol{\beta})\, L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y})$$

$$\propto (\sigma^2)^{-\alpha-1} exp\bigg[-\frac{b}{\sigma^2} \bigg](\tau^2)^{-\frac{p}{2}} exp\bigg[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta} \bigg](\sigma^2)^{-\frac{n}{2}} exp\bigg[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \bigg]$$

$$= (\sigma^2)^{-(\frac{n}{2}+\alpha+1)}(\tau^2)^{-\frac{p}{2}} exp\bigg[-\frac{b}{\sigma^2} \bigg] exp\bigg[-\frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta} \bigg] exp\bigg[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \bigg]$$

The joint posterior distribution is effectively an inverse gamma distribution multiplied by two multivariate normal distributions. The full conditionals can then be derived from the joint posterior.

Full Conditional Distributions

Full Conditional Distributions

For the full conditional of $\boldsymbol{\beta}$, we have that

$$p(\boldsymbol{\beta}|\cdot) \propto exp\bigg[- \frac{1}{2\tau^2}\boldsymbol{\beta}'\boldsymbol{\beta} - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\bigg]$$

$$\propto exp\bigg[- \frac{1}{2\tau^2\sigma^2}\big[\sigma^2\boldsymbol{\beta}'\boldsymbol{\beta} + \tau^2(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\big]\bigg]$$

$$\propto exp\bigg[- \frac{1}{2\tau^2\sigma^2}\big[\sigma^2\boldsymbol{\beta}'\boldsymbol{\beta} + \tau^2(\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta})\big]\bigg]$$

$$\propto exp\bigg[- \frac{1}{2\tau^2\sigma^2}\big[\boldsymbol{\beta}'(\sigma^2\boldsymbol{I} + \mathbf{X}'\mathbf{X})\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y}\big]\bigg]$$

We set this proportional to

$$\propto exp\bigg[-\frac{1}{2\sigma_\beta^2}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)' \, \boldsymbol{\Sigma}_\beta^{-1} \, (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)\bigg]$$

$$\propto exp\bigg[-\frac{1}{2\sigma_\beta^2}\big[\boldsymbol{\beta}'\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta\big]\bigg]$$

We then get that

$$\boldsymbol{\beta} \mid \cdot \sim \mathcal{N}\bigg\{ \bigg(\boldsymbol{X}'\boldsymbol{X} + \frac{\sigma^2}{\tau^2}\boldsymbol{I} \, \bigg)^{-1}\boldsymbol{X}'\boldsymbol{y}, \bigg(\frac{1}{\sigma^2}(\boldsymbol{X}'\boldsymbol{X}) + \frac{1}{\tau^2}\boldsymbol{I} \, \bigg)^{-1}\bigg\}$$

In R

```r
p_beta <- function(sig_sq, tau_sq = 1000, X) {
  p = ncol(X) - 1
  sig <- solve( (1/sig_sq) * (t(X) %*% X) + (1/tau_sq) * diag(p+1) )
  mu <- (1/sig_sq) * sig %*% t(X) %*% y
  rmvnorm(1, mean = mu, sigma = sig)
} # full conditional for the regression coefficients
```

For σ^2, multiplying the priors by the likelihood function, or looking at the joint posterior, we get that:

$$p(\sigma^2 \mid \cdot) \propto (\sigma^2)^{-\alpha-1} exp\left[-\frac{b}{\sigma^2} \right] (\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right]$$

$$= (\sigma^2)^{-\alpha-1-\frac{n}{2}} exp\left[-\frac{b}{\sigma^2} - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right]$$

$$= (\sigma^2)^{-\alpha-1-\frac{n}{2}} exp\left[-\frac{1}{\sigma^2}\left(\frac{2b + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2} \right) \right]$$

Therefore, we have that

$$\sigma^2 \mid \cdot \sim IG\Big(\alpha + \frac{n}{2}, \ \frac{2b + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2}\Big)$$

In R

```r
p_sig <- function(a = 1, b = 1, X, beta) {
  n <- nrow(X)
  a_term <- a + (n/2)
  b_term <- 0.5 * (2*b + (t(y - (X %*% beta)) %*% (y - (X %*% beta))))
  1 / rgamma(1, shape = a_term, rate = b_term)
} # full conditional for sigma^2 in the regression equation
```

To find the full conditional distribution for t_i, we take an approach similar to the above.

t_i is used to form the basis functions B_{ij}, which functions as our X matrix.

We assume $t_i \sim U(0,1)$ since our domain of x is from 0 to 1.

Our full conditional distribution for t_i is therefore just the likelihood function

$$L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}) \propto (\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right]$$

```r
p_t <- function(sig_sq, betahat, X) {
    (-1/(2*sig_sq)) * t(y - (X %*% betahat)) %*% (y - (X %*% betahat))
}
```

Presentation template courtesy Miles Chen. Used with permission.

22

Reversible Jump Monte Carlo (RJ-MCMC)

The final parameter we must condition on is the number of knots.

There are no prior or posterior distributions for this number.

We must therefore use Reversible Jump Monte Carlo to sample for the number of knots

Metropolis-Hastings Algorithm

- **Goal**: sample from any probability distribution given a function $\pi(x)$ proportional to its PDF

- Generate a candidate value x' from a known proposal distribution $q(x)$ (usually Uniform or Gaussian)

- Given the current state value x, calculate acceptance probability $p = min\{1, r\}$, where

$$r = \frac{\pi(x')}{\pi(x)} \frac{q(x \mid x')}{q(x' \mid x)}$$

- Accept or reject x' given a random number `runif(1)`

RJ-MCMC

RJ-MCMC is a generalized form of the Metropolis-Hastings Algorithm.