Bayesian Adaptive Regression Splines (BARS) †

Soma S Dhavala

† Based on Bayesian Curve-Fitting with free-knot splines, Robert. E. Kass et al Biometrika, 88(4), pp.1055-71. 2001

Introduction

We want to fit a function to data, i.e.,

- $y_i = f(x_i)$
- y_i is the data
- x_i is the covariate (like time)
- f is some function that we want to estimate

In a more general setting (e.g. Generalized linear model)

- $y_i \sim p(\theta_i)$ (p is some pdf)
- $\theta_i = f(x_i)$
- x_i is the covariate (like time)
- f can be viewed as an unknown link function

Introduction

Function can be described in terms of parameters, e.g:

- $f(x_i) = sin(2\pi\theta x_i)$
- $f(x_i) = exp((x_i - \theta)^2)$

But then the choice is *limited* to the type of data under study
We can use non-parametric functions that adapt to the data
A function is represented as a weighed combination of basis functions/ expansion sets. Some popular choices are:

- modulated sinusoids (Fourier basis functions)
- wavelets
- splines

Non-parametric curve/function estimation

$f(x) = \sum_{j=1}^{M} \beta_j b_j(x)$

- $b_j(x)$ is the jth basis function/ expansion set
- β_j is the weight of the $b_j(x)$
- estimate the function $\Longrightarrow$ estimate β_js

Choice of $b_j(x)$

- Gaussian wavelet: $\propto exp[(\frac{x-\theta_1}{\theta_2})^2]$
- Fourier basis: $\propto exp[2\pi\theta x]$
- Cubic B-spline: some form of piecewise polynomials (for details see De Boor, C. (2001), A Practical Guide to Splines (rev. ed.), New York: Springer)

Non-parametric curve/function estimation

Fourier basis functions

- good for stationary signals
- need large number of them if there are any transients/ jumps

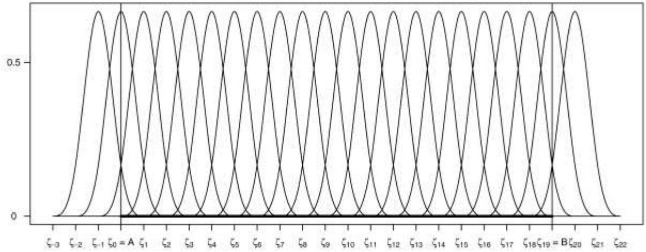
Wavelets

- can adapt to transients/ jumps
- many families to choose from

Splines

- can adapt to transients/ jumps
- in curve-fitting, sometime can result in smaller MSE than wavelets
- many families to choose from
- popular in non-parametric regression

Cubic B splines (order 3)



Curve fitting using Splines

Spline specification requires

- knot locations
- number of knots

If the above are known, standard regression techniques can be used

If not

- one can view the problem as model selection
- each model consisting of knots at different locations and/or with different number of knots

Reversible jump MCMC comes to the rescue

BARS Set-up

Model

- $Y_i = f(x_i) + \epsilon_i (i = 1, 2, ..., n)$
- $\epsilon_i \sim N(0, \sigma^2)$

Assumptions

- f is well-approximated by a cubic spline in the interval $[a, b]$

Cubic spline Specificaiton

- k: number of knots
- ψ_j: jth knot
- $\psi = (\psi_1, \psi_2, ..., \psi_k)$
- $a < x_{(1)} \leq \psi_1 \leq \psi_2 \leq ...\psi_k \leq x_{(n)} < b$

BARS Prior specification

$f(x) = \sum_{j=1}^{k+2} \beta_j b_j(x)$

- $\beta = (\beta_1, \beta_2, ..., \beta_k)$
- $B_{k,\psi}$ be a matrix whose (i,j)th element is $b_j(x_i)$
- then $f(x) = B_{k,\psi}\beta$

Prior on k (number of knots)

- uniform in (k_{min}, k_{max}) or
- Poisson with some λ or

Prior on ψ (knot locations)

- uniform in (a, b) or

Prior on β (spline weights)

- Normal with zero mean

Prior on σ^2 (measurement error variance)

- Jeffrey's improper prior

BARS MCMC

Denision, Mallick and Holmes used "least-squared" plug-in estimates(making it a quasi-Bayesian approach)

With the choice of these priors, β, σ^2 can be analytically integrated out, i.e., $p(y|k,\psi) = \int p(y|\beta,k,\psi)\pi(\beta|k,\psi)\pi(\sigma^2)\, d\beta\, d\sigma^2$

RJMCMC can now be used to switch between different models (that differ in the knots)

The likelihood ratio (to be used in the M-H step) when moving from k to $k+1$ can then be obtained as

$$\frac{p(y|k^c,\psi^c)}{p(y|k,\psi)} = \frac{1}{\sqrt{n+1}} \left(\frac{y^T\{I_n - \frac{n}{n+1} B_{k,\psi}(B_{k,\psi}^T B_{k,\psi})^{-1} B_{k,\psi}^T\} y}{y^T\{I_n - \frac{n}{n+1} B_{k^c,\psi^c}(B_{k^c,\psi^c}^T B_{k^c,\psi^c})^{-1} B_{k^c,\psi^c}^T\} y} \right)^{\frac{n}{2}}$$

BARS Reversible-Jump MCMC

Let M_k represent the model parametrized by k, ψ. Then possible moves are:

Addition

- jump from M_k to M_{k+1}
- with probability $b_k = c \min(1, p(k+1)/p(k))$
- draw new candidate knot from a proposal density $q(M_{k+1}|M_k) = b_k \frac{1}{k} \sum_i h_B(\psi_{cand}|\psi, \tau_B)$ (mixture distribution)
- $h_B(|\psi)$ some density centered around ψ
- so, new candidate knot $\psi_{cand}$is chosen from any of the k distributions

BARS Reversible-Jump MCMC

Deletion

- jump from M_k to M_{k-1}
- with probability$d_k = c \min(1, p(k-1)/p(k))$
- remove the knot from a proposal density $q(M_{k-1}|M_k) = d_k \frac{1}{k}$ (uniform distribution)

BARS Reversible-Jump MCMC

Relocation (death + birth)

- jump from M_k to M_k
- with probability $\eta_k = 1 - b_k - d_k$
- first choose a knot $\psi*$ from ψ, then generate new candidate knot from a $h_R(|\psi*)$ centered around $\psi*$
- remove $\psi*$ from ψ
- proposal density is then $q(M_k|M_k) = \eta_k \frac{1}{k} h_R(|\psi*)$

BARS Reversible-Jump MCMC

The M-H step acceptance probability combines

- likelihood ratio $LR = \frac{p(y|k^c, \psi^c)}{p(y|k, \psi)}$
- prior ratio $PR = \frac{\pi(k^c, \psi^c)}{\pi(k, \psi)}$ and
- asymmetry correction for the moves $QR = \frac{q(M_k|M_k^c)}{q(M_k^c|M_k)}$

Then

$$\alpha = \min\left(1, LR * PR * QR\right)$$

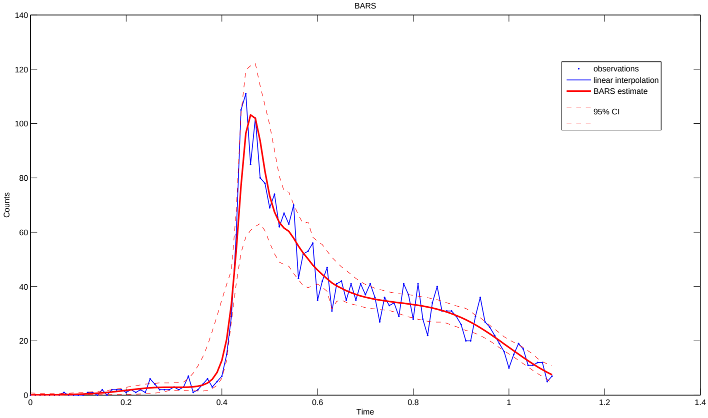
BARS: General Case

$$y_i \sim p(\theta_i)$$
$$\theta_i = f(x_i)$$

- analytical integration may not be possible
- approximation of Likelihood ratios in terms of BIC is sought
- use importance sampling if some feature of the curve (like minimum or maximum) is required

BARS: Neuron firing example

- $y_i \sim Poisson(\lambda_i)$ (y_i Neuron spike counts)
- $\log(\lambda_i) = B(k, \psi)\beta$ (x_i center of bins (of width 10s))



BARS: Discussion

- Fully Bayesian
- Uncertainties in function estimate (like CIs) can be obtained
- Any feature of the fucntion (like minimum, maximum) can be obtained
- Number of knots need not be known
- Adaptive knot selection
- In a general setting, approximations are sought (they work quite well in practice)
- Performance depends on knot location (not a concern when SNR is large)

BARS: Code

Available at http://lib.stat.cmu.edu/ kass/bars/bars.html

In

- R
- MATLAB

References

- DiMatteo, I., Genovese, C.R., and Kass, R.E. (2001) Bayesian curve-fitting with free-knot splines, Biometrika, 88: 1055-1071.
- Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998), AutomaticBayesian Curve Fitting, Journal of the Royal Statistical Society, Ser. B, 60,333350
- De Boor, C. (2001), A Practical Guide to Splines (rev. ed.), New York: Springer