

GAM fitting methods

Matteo Fasiolo (University of Bristol, UK)

matteo.fasiolo@bristol.ac.uk

Material available at:

https://github.com/mfasiolo/workshop_WARSAW19

GAM model fitting

Recall the GAM model structure:

$$y|\mathbf{x} \sim \text{Distr}\{y|\mu(\mathbf{x}), \boldsymbol{\theta}\}$$

where $\mu(\mathbf{x}) = \mathbb{E}(y|\mathbf{x}) = g^{-1}\{\sum_{j=1}^m f_j(\mathbf{x})\}$.

The f_j 's can be

- parametric e.g. $f_j(\mathbf{x}) = \beta_1 x_j + \beta_2 x_j^2$
- random effects
- spline-based smooths such as

$$f_j(x_j) = \sum_{i=1}^r \beta_{ji} b_{ji}(x_j)$$

where β_{ji} are coefficients and $b_{ji}(x_j)$ are known spline basis functions.

NB: we call $\sum_{j=1}^m f_j(\mathbf{x})$ **linear predictor** because it is linear in $\boldsymbol{\beta}$.

GAM model fitting

$\hat{\beta}$ is the maximizer of **penalized** log-likelihood

$$\hat{\beta} = \operatorname{argmax}_{\beta} \operatorname{PenLogLik}(\beta|\gamma) = \operatorname{argmax}_{\beta} \left\{ \overbrace{\log p(\mathbf{y}|\beta)}^{\text{goodness of fit}} - \underbrace{\operatorname{Pen}(\beta|\gamma)}_{\text{penalize complexity}} \right\}$$

where:

- $\log p(\mathbf{y}|\beta) = \sum_i \log p(y_i|\beta)$ is log-likelihood (i.i.d. case)
- $\operatorname{Pen}(\beta|\gamma)$ penalizes the complexity of the f_j 's
- $\gamma > 0$ smoothing parameters ($\uparrow \gamma \uparrow$ smoothness)

GAM model fitting

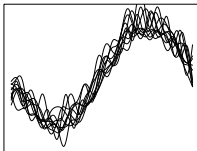
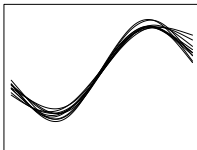
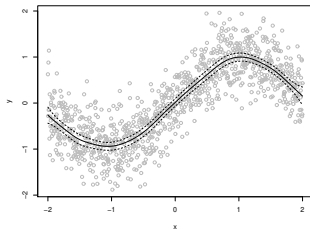
mgcv uses a hierarchical fitting framework:

- 1 Select γ to determine smoothness

$$\hat{\gamma} = \operatorname{argmax}_{\gamma} \text{LAML}(\gamma).$$

- 2 For fixed γ , estimate β to determine actual fit

$$\hat{\beta} = \operatorname{argmax}_{\beta} \text{PenLogLik}(\beta|\gamma).$$



GAM model fitting

Assume smoothing parameter γ are known, so

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \{ \log p(\mathbf{y}|\beta) - \operatorname{Pen}(\beta|\gamma) \}$$

Concrete example:

- $\mathbb{E}(y|x) = f(x)$
- $f(x) = \sum \beta_j b_j(x) = \beta^T \mathbf{b}(x)$
- $f''(x) = \beta^T \mathbf{b}''(x)$
- a cubic spline penalty is

$$\int f''(x)^2 dx = \int \beta^T \mathbf{b}''(x) \underbrace{\beta^T \mathbf{b}''(x)}_{=\mathbf{b}''(x)^T \beta} dx = \beta^T \left[\int \mathbf{b}''(x) \mathbf{b}''(x)^T dx \right] \beta$$

Define $\mathbf{S} = \int \mathbf{b}''(x) \mathbf{b}''(x)^T dx$

GAM model fitting

Hence

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \left\{ \log p(\mathbf{y}|\beta) - \underbrace{\frac{1}{2}\gamma \beta^T \mathbf{S} \beta}_{\operatorname{Pen}(\beta|\gamma)} \right\}$$

In general $\mathbb{E}(y|\mathbf{x}) = \sum_j f_j(\mathbf{x})$ and penalty matrix is $\mathbf{S}_\gamma = \sum_j \gamma_j \mathbf{S}_j$.

Bayesian view: consider *smoothing prior* $\beta \sim N(\mathbf{0}, \mathbf{S}_\gamma^{-1})$ call it $p(\beta)$.

By Bayes theorem

$$p(\beta|\mathbf{y}) = \frac{p(\mathbf{y}|\beta)p(\beta)}{p(\mathbf{y})}$$

or

$$\log p(\beta|\mathbf{y}) = \log p(\mathbf{y}|\beta) + \log p(\beta) - \log p(\mathbf{y}).$$

But $\log p(\beta) = -\frac{1}{2}\beta^T \mathbf{S}_\gamma \beta + \text{const}$, so

$$\log p(\beta|\mathbf{y}) = \log p(\mathbf{y}|\beta) - \frac{1}{2}\beta^T \mathbf{S}_\gamma \beta + \text{const}.$$

GAM model fitting

So $\text{PenLogLik}(\beta|\gamma) \propto \log p(\beta|\mathbf{y})$: we are doing **Maximum a Posteriori** (MAP) estimation!

How to select smoothing parameters γ ?

Recall

$$p(\beta|\mathbf{y}) = \frac{p(\mathbf{y}|\beta)p(\beta)}{p(\mathbf{y})}$$

where

$$p(\mathbf{y}) = \int p(\mathbf{y}|\beta) \underbrace{p(\beta)}_{p(\beta|\gamma)} d\beta = p(\mathbf{y}|\gamma).$$

We want to maximize $p(\mathbf{y}|\gamma)$ wrt γ .

Integral is intractable \rightarrow use Laplace Approximate Marginal Likelihood

$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmax}} \text{LAML}(\gamma).$$

GAM model fitting

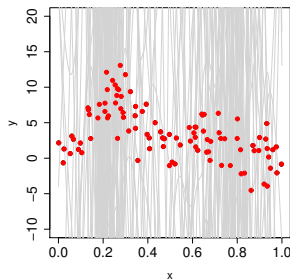
Why do we want to maximize

$$\text{LAML}(\gamma) \approx p(\mathbf{y}|\gamma) = \int p(\mathbf{y}|\beta)p(\beta|\gamma)d\beta$$

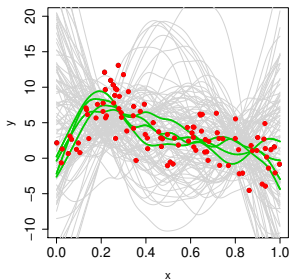
wrt γ ?

Let $\lambda = \gamma$

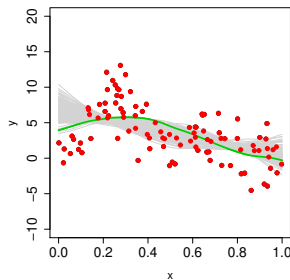
λ too low, prior variance too high



λ and prior variance about right



λ too high, prior variance too low



GAM model fitting

Alternatives LAML for γ selection:

- Generalized Cross-Validation (GCV)
- Akaike Information Criterion (AIC)

but LAML is most widely applicable in `mgcv`.

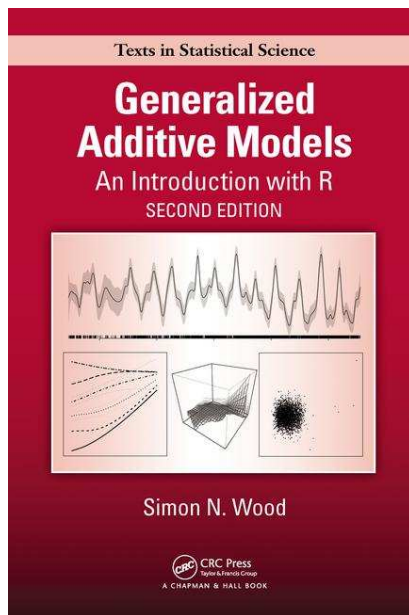
Variance parameters of random effects can be included in γ and estimated by LAML.

To choose γ estimation method in `mgcv`

```
fit <- gam(y ~ ..., method = "REML")
```

see `?gam`.

Further reading



References I

- Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models*, Volume 43. CRC Press.
- Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Number 12. Cambridge university press.
- Wood, S. (2017). *Generalized additive models: an introduction with R*. CRC press.