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

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

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

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

where:

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

GAM model fitting

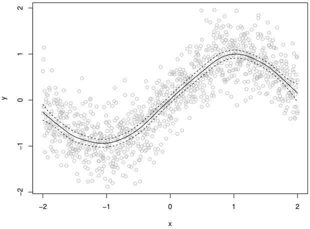
`mgcv` uses a hierarchical fitting framework:

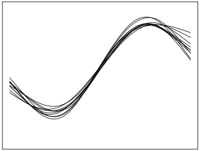
1. Select γ to determine smoothness

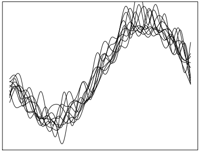
$$\hat{\gamma} = \underset{\gamma}{\text{argmax}} \; \text{LAML}(\gamma).$$

2. For fixed γ, estimate $\boldsymbol{\beta}$ to determine actual fit

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmax}} \; \text{PenLogLik}(\boldsymbol{\beta}|\gamma).$$







GAM model fitting

Assume smoothing parameter γ are known, so

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmax}} \left\{ \log p(\mathbf{y}|\boldsymbol{\beta}) - \text{Pen}(\boldsymbol{\beta}|\gamma) \right\}$$

Concrete example:

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

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

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

GAM model fitting

Hence

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

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* $\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{S}_\gamma^-)$ call it $p(\boldsymbol{\beta})$.

By Bayes theorem

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

or

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

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

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

GAM model fitting

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

How to select smoothing parameters $\boldsymbol{\gamma}$?

Recall

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

where

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

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

Integral is intractable $\rightarrow$ use Laplace Approximate Marginal Likelihood

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

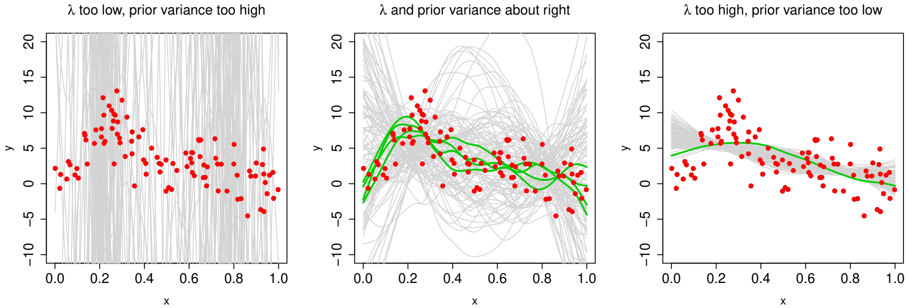
GAM model fitting

Why do we want to maximize

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

wrt $\boldsymbol{\gamma}$?

Let $\boldsymbol{\lambda} = 1/\boldsymbol{\gamma}$



GAM model fitting

Alternatives LAML for γ selection:

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

but LAML is most widely applicable in `mgcv`.

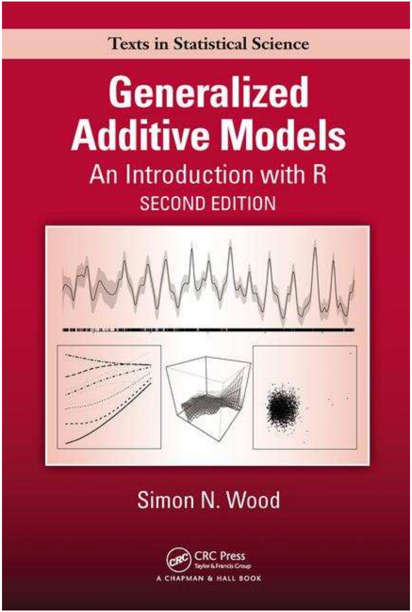
Variance parameters of random effects can be included in γ and estimated by LAML.

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.