Term selection and shrinkage

Simon N. Wood

University of Bath

Term selection

- Selection between different degrees of complexity for the smooth functions (λ selection) is performed using GCV, AIC etc.
- For consistency it makes sense to select between a smooths inclusion or exclusion from the model using the criterion used for smoothness selection. i.e. GCV, AIC or whatever.
- Less formal guidance on inclusion/exclusion can be based on examination of term-wise confidence bands, approximate p-values for testing a terms equality to zero etc.

Fully automatic selection problems

- Smoothing parameter selection can not usually remove a term from the model altogether.
- This is because wiggliness penalties usually treat some space of functions as *completely smooth* (the *penalty null space*).
- As an example, the penalty $\int f''(x)^2 dx$ is zero for any straight line function, $f(x) = \alpha + \gamma x$.
- Writing $\int f''(x)^2 dx = \boldsymbol{\beta}^{\mathsf{T}} \mathbf{S} \boldsymbol{\beta}$ (where $\boldsymbol{\beta}$ contains only the coefficients for f) this degeneracy corresponds to a rank 2 deficiency in $\mathbf{S}$.

Shrinkage smoothers

- An obvious modification is to remove the penalty matrix rank deficiency from $\mathbf{S}$ so that $\boldsymbol{\beta}^{\mathsf{T}}\mathbf{S}\boldsymbol{\beta} = 0 \Rightarrow \boldsymbol{\beta} = \mathbf{0}$, in which case $\lambda \rightarrow \infty \Rightarrow \hat{f}(x) \rightarrow 0$.
- An obvious way to achieve this is to add a small 'ridge penalty' to $\mathbf{S}$ so that it becomes $\mathbf{S} + \iota\mathbf{I}$ where ι is 'small'.
- Normally the ridge penalty has negligible effect, but when $\lambda \rightarrow \infty$ it will cause the corresponding function to be penalized to zero.
- With such penalties regular GCV/AIC smoothness selection can completely remove model terms.

Shrinkage smooths in `mgv::gam`

- `mgcv::gam` includes `"cs"` and `"ts"` classes of shrinkage smoothers — modifications of the `"cr"` and `"tp"` classes.
- Note that the approach is asking quite alot of the GCV/AIC optimizer, numerically.
- The following slide shows a 6 term additive GAM represented using shrinkage smoothers fitted to data from a 3 term additive truth.

```
gam(y ~ s(x0,bs="ts") + ...
        ... + s(x5,bs="ts"),gamma=1.4)
```

Shrinkage example