Comparison of Approaches for Fitting Generalized Additive Models

Harald Binder
Freiburg Center for Data Analysis and Modeling
binderh@fdm.uni-freiburg.de

Generalized Additive Models

- Response y taken to be from an exponential family (e.g. binary or Poisson) with $E(y|x) = \mu = g(\eta)$
- Covariates $x' = (x_1, \ldots, x_p)$
- Model $\eta = f_1(x_1) + \cdots + f_p(x_p)$
- Based on univariate scatterplot smoothers for unknown smooth functions $f_1, \ldots, f_p$

Methods for Fitting

A) Backfitting (`gam:gam`) with stepwise selection of degrees of freedom (df)

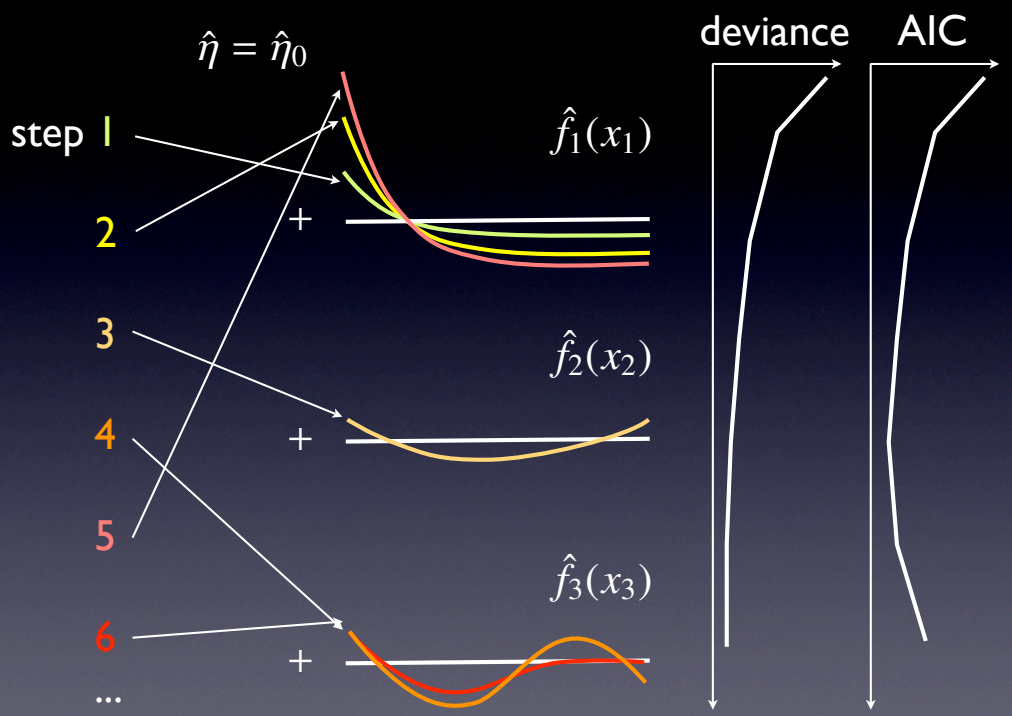
B) Simultaneous estimation with optimization in smoothing parameter space (`mgcv:gam`)

C) Using a mixed-model representation (e.g. `mgcv:gamm`) – does not require selection of df or smoothing parameters

D) Likelihood-based boosting (`GAMBoost`)

GAMBoost

- Stepwise boosting procedure
- Repeated fitting of residuals in the GLM framework: incorporation of previous boosting steps as η-offset
- Implicit variable and smoothness selection by updating only one smooth function in each step (by adding a penalized B-spline fit)
- R package `GAMBoost`
- see Tutz, G. & Binder, H., Biometrics, 2006



Simulation comparison

true structure			(A) backfitting	(B) sm. par. optim.	(C) mixed model	(D) GAMBoost
linear			o	o	-	+
smooth	n small	p small	o	o	-	+
		p large	-	-	-	o
	n large	p small	+	++	+	+
		p large	o	-	-	+

optimal performance: ++ good: + still reasonable: o problematic: -