Recap: High-dimensional additive models

$$Y_i = \mu + \sum_{j=1}^{p} f_j(X_i^{(j)}) + \varepsilon_i \ (i = 1, \ldots, n), \quad \sum_{i=1}^{n} f_j(X_i^{(j)}) = 0 \ \forall j$$

parameterization: $$f_j(\cdot) \approx \sum_{k=1}^{K} \beta_{j,k} \underbrace{h_{j,k}(\cdot)}_{\text{basis fct.s}}$$

$$(H_j)_{i,k} = h_{j,k}(X_i^{(j)}),$$

$$\beta_j = (\beta_{j,1}, \ldots, \beta j, K)^T, \ \beta = (\beta_1, \ldots, \beta_p)^T$$

$\rightsquigarrow$ approximation with basis functions at observed data points:

$$\sum_{j=1}^{p} \beta_j \underbrace{H_j}_{n \times K \text{ matrix}}$$

Naive estimation with (prediction) Group Lasso penalty

$$\hat{\beta} = \text{argmin}_{\beta} \|Y - \sum_{j=1}^{p} \beta_j H_j\|_2^2/n + \underbrace{\lambda \sum_{j=1}^{p} \|H_j \beta_j\|_2/\sqrt{n}}_{\text{scaled pred. Group Lasso pen.}}$$

for $f_j = (f_j(X_1^{(j)}), \ldots, f_j(X_n^{(j)}))^T$ and $\|f_j\|_n^2 = \|f_j\|_2^2/n$

$\leadsto$

$$\hat{\beta} = \text{argmin}_{\beta} \|Y - \sum_{j=1}^{p} \beta_j H_j\|_2^2/n + \sum_{j=1}^{p} \|f_j\|_n$$

doesn't take smoothness into account!

Natural cubic splines

the special case with natural cubic splines

(Ch. 5.3.2 in Bühlmann and van de Geer (2011))

consider the estimation problem wit the SSP penalty:

$$\hat{f}_1, \ldots, \hat{f}_p = \text{argmin}_{f_1,\ldots,f_p \in \mathcal{F}} \Big(\| Y - \sum_{j=1}^{p} f_j \|_n^2 + \lambda_1 \| f_j \|_n + \lambda_2 I(f_j) \Big)$$

where $\mathcal{F}$ = Sobolev space of functions on $[a, b]$ that are continuously differentiable with square integrable second derivatives

Proposition 5.1 in Bühlmann and van de Geer (2011)
Let $a, b \in \mathbb{R}$ such that $a < \min_{i,j}(X_i^{(j)})$ and $b > \max_{i,j}(X_i^{(j)})$. Let $\mathcal{F}$ be as above. Then, the $\hat{f}_j$'s are natural cubic splines with knots at $X_i^{(j)}, \ i = 1, \ldots, n$.

implication: the optimization over functions is exactly representable as a parametric problem with dim $\approx 3np$

SSP penalty of group Lasso type

for easier computation: instead of

$$\text{SSP penalty} = \lambda_1 \sum_j \|f_j\|_n + \lambda_2 \sum_j I(fj)$$

one can also use as an alternative:

$$\text{SSP Group Lasso penalty} = \lambda_1 \sum_j \sqrt{\|f_j\|_n^2 + \lambda_2 I^2(f_j)}$$

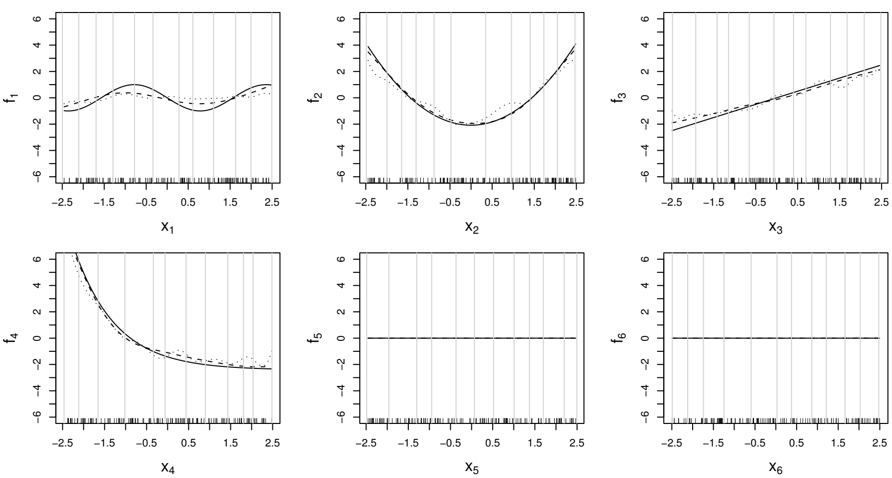
in parameterized form, the latter becomes:

$$\lambda_1 \sum_{j=1}^{p} \sqrt{\|H_j \beta_j\|_2^2 / n + \lambda_2^2 \beta_j^T W_j \beta_j} = \lambda_1 \sum_{j=1}^{p} \sqrt{\beta_j^T (H_j^T H_j / n + \lambda_2^2 W_j) \beta_j}$$

$\leadsto$ for every λ_2: a generalized Group Lasso penalty

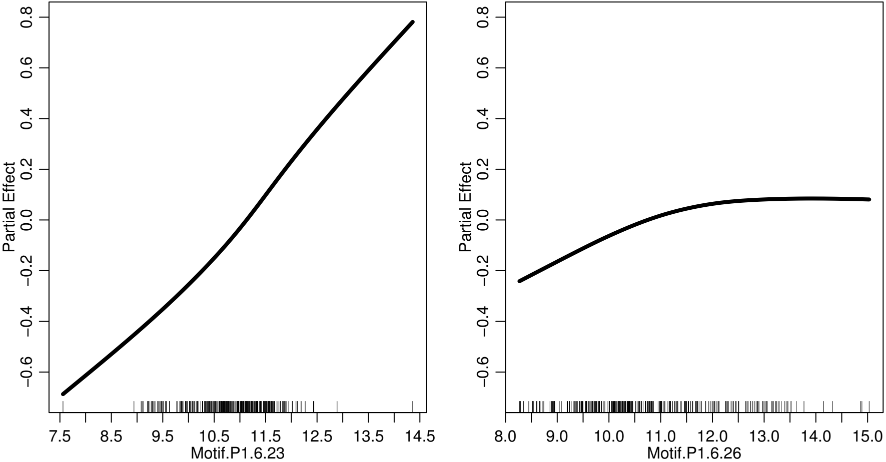
`R`-package `hgam`

simulated example: $n = 150, p = 200$ and 4 active variables



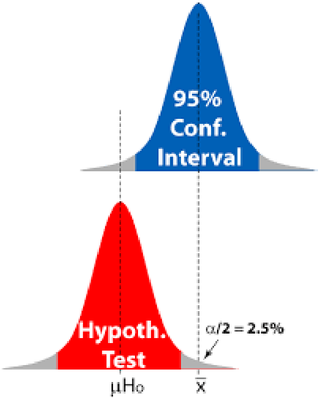
dotted line: $\lambda_2 = 0$

$\rightsquigarrow$ λ_2 seems not so important: just consider a few candidates motif regression: $n = 287,\ p = 195$



$\rightsquigarrow$ a linear model would be "fine as well"

Uncertainty quantification:
p-values and confidence intervals (slides, denoted as Ch. 10)



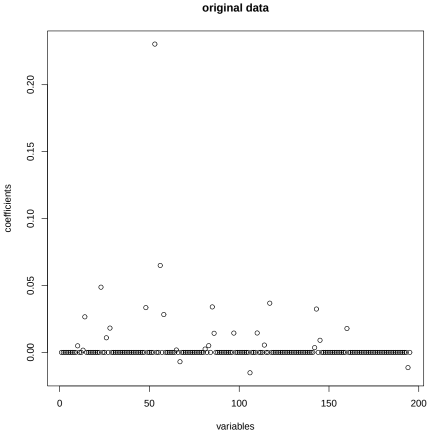
frequentist
uncertainty quantification

(in contrast to Bayesian inference)

classical concepts but in very high-dimensional settings

Toy example: Motif regression ($p = 195, n = 143$)

Lasso estimated coefficients $\hat{\beta}(\hat{\lambda}_{\mathrm{CV}})$



p-values/quantifying uncertainty would be very useful!

$$Y = X\beta^0 + \varepsilon \quad (p \gg n)$$

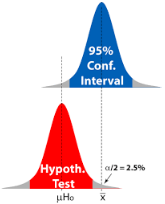
classical goal: statistical hypothesis testing

$$H_{0,j} : \beta_j^0 = 0 \text{ versus } H_{A,j} : \beta_j^0 \neq 0$$

or $$H_{0,G} : \beta_j^0 = 0 \ \forall\, j \in \underbrace{G}_{\subseteq\{1,\ldots,p\}} \text{ versus } H_{A,G} : \exists j \in G \text{ with } \beta_j^0 \neq 0$$

background: if we could handle the asymptotic distribution of the Lasso $\hat{\beta}(\lambda)$ under the null-hypothesis

$\leadsto$ could construct p-values



this is very difficult!
asymptotic distribution of $\hat{\beta}$ has some point mass at zero,...
Knight and Fu (2000) for $p < \infty$ and $n \to \infty$

because of “non-regularity” of sparse estimators
“point mass at zero” phenomenon $\rightsquigarrow$ “super-efficiency”


(Hodges, 1951)

$\rightsquigarrow$ standard bootstrapping and subsampling should not be used

$\rightsquigarrow$ de-sparsify/de-bias the Lasso instead