Recap

High-dimensional additive models

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

$$f_j : \mathbb{R} \to \mathbb{R} \text{ smooth}, \ \mathbb{E}[f_j(X_i^{(j)})] \equiv 0 \ \forall \ j$$

aim: estimator such that either $\hat{f}_j(.) \equiv 0$ or $\hat{f}_j(.)$ is not the zero function

$\leadsto$ sparsity smoothness (SPS) penalty

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

where $\mathcal{F}$ = Sobolev space of functions ohat are continuously differentiable with square integrable second derivatives

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

is a parametric problem of dimension $d \approx 3pn$, parametrized by natural cubic splines with basis functions encoded in a matrix

$$H_{n \times d} = (H_1, \ldots, H_p)^T$$

and integrated squared second derivatives encoded in a matrix

$$(W_j)_{k,\ell} = \int h''_{j,k}(x) h''_{j,\ell}(x) dx$$

$\rightsquigarrow$

$$\hat{\beta} = \text{argmin}_\beta \left(\| Y - H\beta \|_2^2 / n + \lambda_1 \sum_{j=1}^{p} \sqrt{\beta_j^T H_j^T H_j \beta_j / n} + \lambda_2 \sum_{j=1}^{p} \sqrt{\beta_j^T W_j \beta_j} \right)$$

SPS penalty of group Lasso type

instead of

$$\mathrm{pen}(\beta) = \lambda_1 \sum_{j=1}^{p} \sqrt{\beta_j^T H_j^T H_j \beta_j / n} + \lambda_2 \sum_{j=1}^{p} \sqrt{\beta_j^T W_j \beta_j}$$

use

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

$\leadsto$ for every λ_2: a generalized Group Lasso penalty
can simply use standard Group Lasso software!

high-dimensional additive modeling “works” because with e.g. smoothing splines, the dimension is

$$d \approx 3pn, \quad \log(d)/n \asymp \log(p)/n \; (p \gg n)$$

and assuming spasrity and smoothness

can extend to interaction modeling of first order with functions

$$\sum_{j=1}^{p} f_j(x_j) + \sum_{j \neq r = 1}^{p} f_{j.r}(x_j, x_r)$$

$\leadsto$ dimension

$$d = O(p^2 n^2), \quad \log(d)/n \asymp \log(p)/n \; (p \gg n)$$

but the computation becomes cumbersome!

Inference in high-dimensional linear models: p-values and confidence intervals

one cannot use the bootstrap or subsampling for approximating the distribution of the Lasso
it is inconsistent due to non-Gaussian limiting distribution of the Lasso

$$\text{if} \quad a_n(\hat{\beta}_{\mathrm{Lasso};j} - \beta_j^0) \Longrightarrow Z, Z \sim \text{non-Gaussian distribution } F$$
$$\text{then} \quad a_n(\hat{\beta}^*_{\mathrm{Lasso};j} - \hat{\beta}_{\mathrm{Lasso},j}) \not\Longrightarrow Z \text{ in probability}$$