Recap

High-dimensional additive models

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

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

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

$\rightsquigarrow$ Group Lasso type problem

but: sparsity and smoothness penalty (SpS) is better than Group-sparsity (with respect to basis expansion) alone

Sparsity Smoothness (SpS) penalty with smoothing spline

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

$$\| f_j \|_n^2 = n^{-1} \sum_{i=1}^{n} |f_j(X_i^{(j)})|^2, \; I(f_j)^2 = \int f_j''(x)^2 dx$$

where $\mathcal{F}$ = Sobolev space of functions that 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

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

$\rightsquigarrow$

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