$$Y_i = \mu + \sum_{j=1}^{p} f_j^0 (X_i^{(j)}) + \varepsilon_i$$

$f_j^0 : \mathbb{R} \to \mathbb{R}$ smooth

$$f_j(\cdot) = \sum_{k=1}^{K} \beta_{j,k} \underbrace{h_{j,k}(\cdot)}_{\text{specified basis functions e.g.:}} \qquad j = 1, \to p$$

- splines
- orthogonal polynomials
- cosine / sine functions (Fourier)
- wavelets

Denote by $\underset{n \times K}{(H_j)}_{i,k} = h_{j,k}(X_i^{(j)})$

$\underset{K \times 1}{\beta_j} = (\beta_{j,1}, \dots, \beta_{j,K})^T$, $\underset{pK \times 1}{\beta} = (\beta_1, \dots, \beta_p)^T$

$\rightsquigarrow \left\{ \sum_{j=1}^{p} f_j(X_i^{(j)}) \right\}_{i=1}^{n} = \sum_{j=1}^{p} H_j \beta_j$

<u>Motivation</u>: want that $X^{(j)}$ is selected or not

$\Longleftrightarrow$ $\hat{\beta}_j \equiv 0$ or $\hat{\beta}_j \not\equiv 0$
and even: $\hat{\beta}_{j,k} \neq 0 \ \forall k$

naive proposal: (SpAM: Ravikumar, Lafferty, Liu & Wasserman (2009))

$$\rightsquigarrow \hat{\beta} = \arg\min_{\beta} \Big(\| Y - \underbrace{\sum_{j=1}^{p} H_j \beta_j}_{H\beta} \|_2^2 / n + \mathrm{pen}(\beta) \Big)$$

$$H\beta, \quad \underset{n \times pK}{H} = \begin{bmatrix} H_1 & H_2 & \cdots & H_p \end{bmatrix}$$

$$\mathrm{pen}(\beta) = \lambda \sum_{j=1}^{p} \| H_j \beta_j \|_2 / \sqrt{n}$$

$$=: \lambda \sum_{j=1}^{p} \| f_j \|_n$$

$$f_j = (f_j(X_1^{(j)}), \ldots, f(X_n^{(j)}))^T = H_j \beta_j$$

$\|f_j\|_n^2 := f_j^T f_j / n$

this is really Group Lasso with groupwise prediction penalty

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left(\|Y - H\beta\|_2^2 / n + \lambda \sum_{j=1}^{p} \|H_j \beta_j\|_2 / \sqrt{n} \right)$$

$$\beta = (\beta_1 \ldots \beta_p)^T, \quad H = [H_1 \; H_2 \ldots H_p]$$

<u>but</u>: does not take the smoothness of $f_j(\cdot)$ explicitly into account $\rightsquigarrow$ could be done choosing different K for different $f_j(\cdot)$'s $\rightsquigarrow$ very cumbersome ...!

taking smoothness into account by looking at smoothing splines

V.2. Natural cubic splines and Sobolev spaces

(Ch. 5.3.2 in PB & vdG (2011))

instead of "explicite" basis expansion (a-priori):

$$\hat{f}_1, \hat{f}_2, \ldots \hat{f}_p = \underset{\substack{f_1, f_2, \ldots f_p \\ \in \mathcal{F}}}{\operatorname{argmin}} \left(\| Y - \sum_{j=1}^{p} f_j (X_i^{(j)}) \|_2^2 / n \right.$$

$$\left. + \lambda_1 \underbrace{\sum_{j=1}^{p} \| f_j \|_n}_{\text{sparsity penalty}} + \lambda_2 \underbrace{\sum_{j=1}^{p} I(f_j)}_{\text{smoothness penalty}} \right)$$

$$\| f_j \|_n = \| (f_j(X_1^{(j)}), \ldots f_j(X_n^{(j)}))^T \|$$

$$\mathcal{F} = \Big\{ f : [a,b] \to \mathbb{R},\ f \text{ cont. twice differentiable and } \underbrace{\int_a^b (f''(x))^2\,dx}_{=: I^2(f)} < \infty \Big\}$$