V. Additive models and many smooth univariate functions

(Ch. 5 in book)

V. 1. Model and estimation (Ch. 5.2)

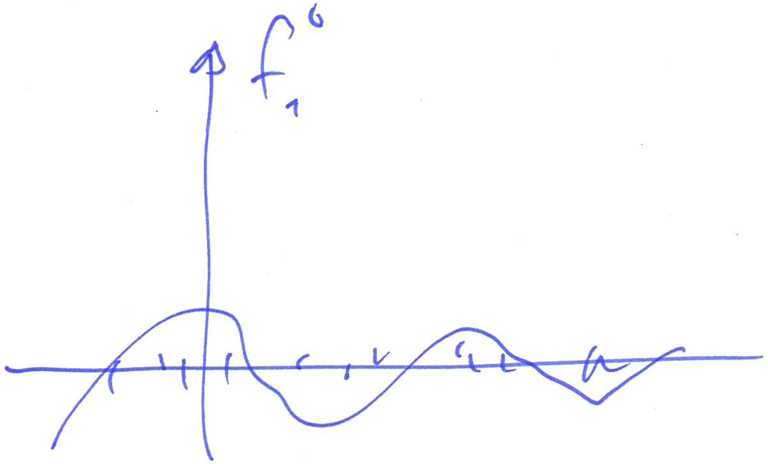
additive model:

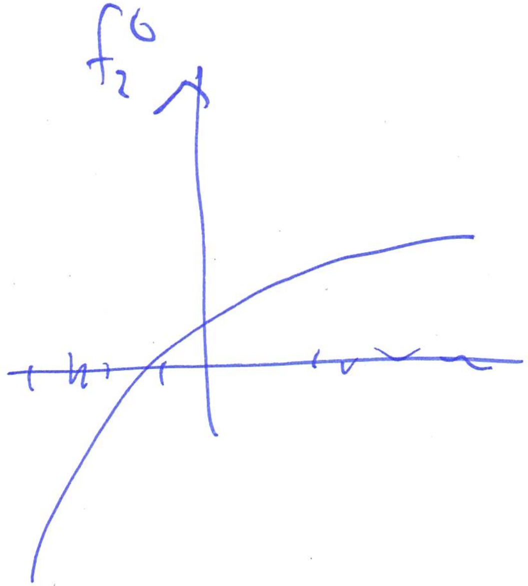
$$Y_i = \mu + \sum_{j=1}^{p} f_j^0(X_i^{(j)}) + \varepsilon_i$$

smooth functions

$E[\varepsilon_i] = 0$, ε_i indep. of X_i if X_i is random

identification: $\sum_{i=1}^{n} f_j^0(X_i^{(j)}) \equiv 0 \quad \forall j$





for $p < n$: classical additive model (e.g. Hastie & Tibshirani, 1986)

escapes curse of dimensionality:

$$\cancel{Y_i = \underbrace{f^0}_{\text{smooth}}(X_i^{(1)}, \ldots, X_i^{(p)}) + \varepsilon_i}$$

because of curse of dim.

$p \gg n$: need to add sparsity among

$$\{ f_j^0(\cdot) \, ; \, j = 1, \to p \}$$

it will "work" and

because it's additive: avoids curse of dimensionality

approach: use basis expansion

$$f_j(\cdot) = \sum_{k=1}^{K} \underbrace{\beta_{j,k}}_{\text{unknown parameters}} \underbrace{h_{j,k}(\cdot)}_{\text{specified basis functions}} \approx f_j^0(\cdot)$$

if $K \to \infty$: "="

specified basis functions

- splines
- Fourier
- wavelets

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

$$\beta_j = (\beta_{j,2}, \dots, \beta_{j,K})^T, \quad \underset{pK \times 1}{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$$

$$\rightsquigarrow \sum_{j=1}^{p} \underbrace{f_j(X_i^{(j)})}_{\approx f_j^0(X_i^{(j)})} = \sum_{j=1}^{p} H_j \beta_j \quad \Big(= \underset{[H_1, H_2 \dots H_p]}{H} \beta \Big)$$

estimation: want that $X^{(j)}$ is selected or not

$$\Longleftrightarrow \hat{\beta}_j \equiv 0 \quad \text{or} \quad \hat{\beta}_j \not\equiv 0 \text{ with } \hat{\beta}_{j,k} \neq 0 \ \forall k$$

~~naive~~ simple proposal (SpAM: sparse additive model)

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

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

right scaling

groupwise prediction penalty with groups of cardinality K

K should be less than n

($K \asymp n^{2/5}$ is often good choice)

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

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

$$f_j = \left(f_j(X_1^{(j)}), f_j(X_2^{(j)}), \ldots, f_j(X_n^{(j)}) \right)^T; \quad \|f_j\|_n^2 = f_j^T f_j / n$$

$$= \|f_j\|_2^2 / n$$

this simple approach does not take smoothness of different $f_j^0(\cdot)$ into account (could in principle be addressed by different K for different f_j) $\rightsquigarrow$ too hard!

take smoothness into account: convenient with splines

V.2. Natural cubic splines and Sobolev spaces

(Ch. 5.3.2 in book)

instead of explicit basis expansion:

$$(\hat{f}_1, \hat{f}_2, \dots, \hat{f}_p) = \underset{f_1, f_2, \dots, f_p \in \mathcal{F}}{\operatorname{argmin}} \Big\{ \|Y - \sum_{j=1}^{p} f_j\|_2^2 / n + \underbrace{\lambda_1 \sum_{j=1}^{p} \|f_j\|_n}_{\text{sparsity penalty}} + \underbrace{\lambda_2 \sum_{j=1}^{p} I(f_j)}_{\text{smoothness penalty}} \Big\}$$

$\rightarrow$ choose λ_1, λ_2 (instead of $\cancel{K_1, K_2, \dots, K_p}$)

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