Last time : smoothing splines

$$\Delta(f,\lambda) = \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int_{-\infty}^{+\infty} f''(t)\,dt$$

Multidimensional splines

$$\iint_{\mathbb{R}\,\mathbb{R}} \left[\left(\frac{\partial^2 f(x)}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 f(x)}{\partial x_2^2} \right)^2 \right] dx_1 dx_2$$

The non-parametric techniques to relate y and x suffers from the curse of dimensionality

- good job in 1 dimension;
- fail for $p > 2$.

We know that parametric techniques do not suffer too much because we impose a strong structure

- parametric family
- use all information to estimate only a few parameters
- loose flexibility

Trade-off between flexibility and structure:

Additive models

- only impose additive structure, keeping the flexibility of a non-parametric approach

$$f(x) = f(x_1, \dots, x_p) = \beta_0 + \sum_{j=1}^{p} f_j(x_j)$$

where $f_j(x_j)$ are smooth functions of one variable and β_0 is the intercept.

$$f(x) = f(x_1,\dots,x_p) = \beta_0 + \sum_{j=1}^{p} f_j(x_j)$$

To make the model identifiable, we need to add one constrain, that is the functions $f_j(x_j)$ are centred

$$\sum_{i=1}^{n} f_j(x_{ij}) = 0 \quad \forall j = 1,\dots,p$$

$\Rightarrow \exists$ only one $\hat\beta_0 = \bar y$

E.g. $f(x) = \beta_0 + f_1(x_1)$

$$f(x) = \underbrace{(\beta_0 + 2)}_{\beta_0'} + \underbrace{f_1'(x_1)}_{f_1(x_1) - 2}$$

To fit the model we use the backfitting algorithm

1 - initialize, $\hat\beta_0 = \bar y$, $\hat f_j(x_{ij}) = 0 \quad \forall j$

2 - in turn, $j=1, j=2, \dots, j=p, j=1, j=2, \dots, j=p, \dots, j=1, \dots$

$$\hat f_j \leftarrow S\left(\underbrace{\left\{ y_i - \hat\beta_0 - \sum_{k \neq j} \hat f_k(x_{ik}) \right\}_1^n}\right)$$

$\to$ fix from the previous step

$$\hat f_j \leftarrow \hat f_j - n^{-1} \sum_{i=1}^{n} \hat f_j(x_{ij})$$

until $\hat f_j$ stabilize.

E.g.

	1 step		2 step		3 step		4 step		...
$\hat\beta_0 = \bar y$	$\bar y$		$\bar y$		$\bar y$		$\bar y$		$\bar y$
$\hat f_1 = 0$ update $\to$	$\hat f_1(x_1)$		$\hat f_1(x_1)$	update $\to$	$\hat f_1(x_1)$		$\hat f_1(x_1)$	difference ε ...	$\hat f_1(x_1)$
$\hat f_2 = 0$	0	update $\to$	$\hat f_2(x_2)$		$\hat f_2(x_2)$	update $\to$	$\hat f_2(x_2)$	difference ε ..	$\hat f_2(x_2)$

Generalization 1: add interactions term

$$f(x) = \beta_0 + \sum_{j=1}^{p} f_j(x_j) + \underbrace{\sum_{j=1}^{p} \sum_{k<j} f_{jk}(x_j, x_k)}_{\text{interaction between pairs of variables}} + \underbrace{\sum_{j=1}^{p} \sum_{k<j} \sum_{h<k} f_{jkh}(x_j, x_k, x_h)}_{\text{triplets}}$$

In figure 4.13

left side: $y = f(x_1) + f(x_2)$ splines

right side: $y = f(x_1, x_2)$ non-parametric density

Generalization 2: additive model → generalized additive model

similar to move from lm to glm

$$g\left(E[Y \mid X_1, \ldots, X_p]\right) = \beta_0 + \sum_{j=1}^{p} f_j(x_j)$$

where g is the link function (as in GLM), which is problem-specific. E.g., for logistic regression, logit

$$\text{logit}(\pi) = \beta_0 + \sum_{j=1}^{p} f_j(x_j)$$

To estimate it, we use a combination of backfitting algorithm and weighted least squares

Projection pursuit

We apply an additive model on transformed variables

$$f(x_1, \dots, x_p) = \beta_0 + \sum_{k=1}^{K} f_k(\beta_k^T x)$$

This is called projection pursuit regression model:

- K is the number of projections;
- β_k are the projection vectors (must be estimated)
- $f_k(\cdot)$ are the ridge functions (are constant in all directions but those defined by β_k)
- the derived feature $V_k = \beta_k^T x$ is the projection of x onto the vector β_k, which is chosen in order for the model to fit well (pursuit)
- $K \neq p$

  can be computed by cross-validation

Important: for K large enough and appropriate choice of f_k the model can approximate arbitrarily well any function of the covariates (i.e., any function in $\mathbb{R}^p$)

universal approximator (in contrast to additive model, where this is not true)

Obviously, generality comes at a price: the model is impossible to interpret → only used for prediction