


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 

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

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 constraint, that is the functions $f_j(x_j)$ are centered

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

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

$\Rightarrow \exists$ only one $\hat{\beta}_0 = \bar{y}$

$$f(x) = (\beta_0 + 2) + f_1(x_1)$$

$$\beta_0 \quad f_1(x_1) - 2$$

To fit the model we use the least squares algorithm

1 - initialize, $\hat{\beta}_0 = \bar{y}$, $f_j(x_j) = 0 \quad \forall j$

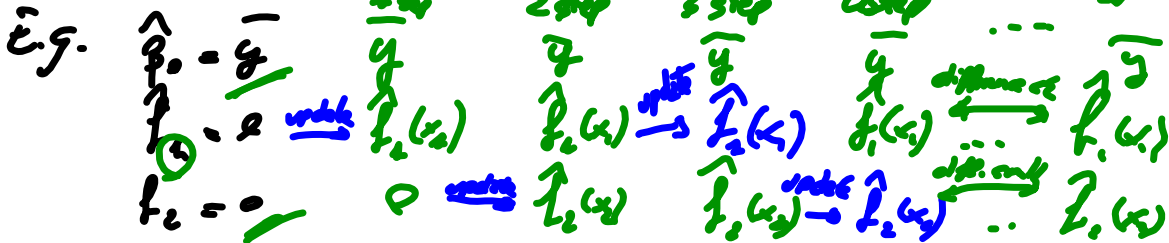
2 - in turn, $j=1, j=2, \dots, j=p, j=1, j=2, \dots, j=p, \dots, j=1, \dots$

$$\hat{\beta}_j \leftarrow \underset{\beta_0}{\text{argmin}} \left\{ \sum_{i=1}^n \left(\hat{y}_i - \beta_0 - \sum_{k \neq j} \hat{f}_k(x_{ik}) \right)^2 \right\}$$

\leftarrow 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.



Generalization 1: add interactions term

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

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 \rightarrow generalized additive model
similar to move from lm to glm

$$g(E[Y | x_1, \dots, x_p]) = \beta_0 + \sum_{j=1}^p f_j(x_j)$$

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)
- R^2/P can be computed by cross-validation

Important: for k large enough and appropriate choice of β_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 predictions