

- finish characterising splines
- smoothing splines
- splines in higher dimensions
- Additive Models (GAM)
- (maybe introduce statistical trees)

$$(x - \xi)_+$$

In general, we splines of order M , with K knots $\xi_1, \dots, \xi_K$

$$f(x) = \sum_{m=1}^{M+K} \beta_m h_m(x)$$

$$h_m(x) = x^{m-1} \quad m=1, \dots, M \quad \rightarrow \text{for cubic splines, } M=4$$

$$h_{M+l}(x) = (x - \xi_l)_+^{M-1} \quad l=1, \dots, K$$

Normally M is fixed in advance, let us say $M=4$ (cubic splines). We need to decide:

- how many knots (K)
- where to place the knots

How many knots is related to the complexity: larger K , more complex model.

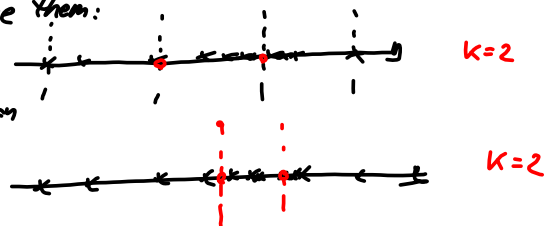
The idea is to find K that assures the right complexity, i.e. find the K that assures the best balance in terms of bias-variance trade-off

→ in practice: use cross-validation

Once we decide the value of K , we can place them:

→ uniformly in the x range

→ using the quantiles of the empirical distribution of x



Smoothing splines

→ we do not want to decide how many knots and where to place them. We place the knots on the unique x and penalise the "bumpiness"

Consider the penalised least squares criterion

$$RSS(\beta, \lambda) = \underbrace{\sum_{i=1}^n (y_i - f(x_i))^2}_{\text{how close we are to the data}} + \underbrace{\lambda \int_{-\infty}^{+\infty} (f''(t))^2 dt}_{\text{penalty for the function curvature}}$$

λ is the tuning parameter (smoothing parameter). Choosing λ decides how much we allow the function $f(x)$ to vary ("be bumpy")

→ $\lambda \rightarrow 0$, more curvature is allowed for the function $f(x)$, until $\lambda = 0$ when $\lambda = 0$ the function $f(x)$ interpolates the points

→ $\lambda \rightarrow \infty$, the curvature of $f(x)$ is penalised more and more, until the point (λ is sufficiently large) that any move from the straight line is not convenient

The key result (Green & Silverman, 1994) is that the minimizer ($\hat{f}(x)$) of $RSS(\beta, \lambda)$ is a NATURAL CUBIC SPLINES,

$$\hat{f}(x) = \sum_{j=1}^{n_0} \theta_j N_j(x)$$

where

n_0 = number of unique x

N_j = basis-functions of the natural ^{cubic} splines

$$N_1(x) = 1, \quad N_2(x) = x, \quad N_{e+2} = d_e(x) - d_{e-1}(x)$$

$$\text{where } d = \frac{(x - \bar{x}_e)_+^3 - (x - \bar{x}_k)_+^3}{\bar{x}_k - \bar{x}_e}$$



The $RSS(\beta, \lambda)$ can be rewritten in the form

$$RSS(\beta, \lambda) = (y - N\theta)^T (y - N\theta) + \lambda \theta^T \Omega \theta$$

where N is a matrix with elements $N_{[i,j]} = N_j(x_i)$ and Ω is a matrix with elements $\Omega_{jk} = \int N_j'(t) N_k'(t) dt$

This reminds the ridge regression, and indeed

$$\hat{\theta} = (N^T N + \lambda \Omega)^{-1} N^T y$$

Obviously the solution depends on λ , we can choose by cross-validation or by fixing the degrees of freedom and λ accordingly

Effective degrees of freedom

- linear method

$$\hat{y} = X (X^T X)^{-1} X^T y$$

↳ projection matrix
hat matrix

independent on y

→ we apply it to y to get $\hat{y}$

Smoothing approaches like smoothing splines and ridge regression are also linear methods

$$\hat{y}_{\text{ridge}} = X (X^T X + \lambda I_p)^{-1} X^T y$$

p is the number of columns of X

measure
of complexity

↳ smoother matrix H_λ

effective degrees of freedom : trace of H_λ , $\text{tr}(H_\lambda)$

$$\text{tr}(H_\lambda) = \sum_{j=1}^p H_{\lambda, [j,j]}$$

$$\begin{bmatrix} a_{11} & & \\ & a_{22} & \\ & & \ddots \\ & & & a_{pp} \end{bmatrix}$$

for ridge regression $\text{tr}(H_\lambda) = \sum_{j=1}^p \frac{d_j^2}{d_j^2 + \lambda}$, where d_j are the eigenvalues of the singular value decomposition of X

Note that:
larger λ , smaller $\frac{d_j^2}{d_j^2 + \lambda}$, less complexity

smaller λ , larger $\frac{d_j^2}{d_j^2 + \lambda}$, higher complexity

$$\lambda=0 \quad \text{tr}(H_\lambda) = \sum_{j=1}^p \frac{d_j^2}{d_j^2} = p$$

degrees of freedom of
the linear model ($\lambda=0$)

- for smoothing splines, $\hat{f}_\lambda(x) = N (N^T N + \lambda \Omega_n)^{-1} N^T y$
 S_λ smoother matrix

effective degrees of freedom : $\text{tr}(S_\lambda)$

Since there is a monotone relationship between $\text{tr}(S_\lambda)$ and λ ($\lambda \uparrow$, $\text{tr}(S_\lambda) \downarrow$), instead of choosing λ , we can fix the effective degrees of freedom and find the corresponding λ

Multi dimensional splines

It is not straightforward to extend splines in $p > 1$ dimensions.

It is almost never used $p > 2$
 Splines with $p=2$ are called "thin plates"

The penalty term to penalise the curvature for $p=2$

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

The solution has form

$$f(x) = \beta_0 + \beta^T x + \sum_{j=1}^n \alpha_j h_j(x)$$

where $h_j(x) = \|x - x_j\|^2 \log \|x - x_j\|$