2021-11-8 | Natural Splines

$$RSS(f, \lambda) = \sum_{i=1}^{N} \left(y_i - f(x_i)\right)^2 + \lambda \int \left[f''(t)\right]^2 dt$$

Natural Cubic Splines με κόμβους στα $x_1, \dots, x_N$

Άσκηση Δείξτε ότι απαιτούνται N συναρτήσεις βάσης

$$N_1(x) = 1, \quad N_2(x) = x, \quad \underbrace{N_3, \dots, N_N}_{d_k(x) - d_{N-1}(x)} \qquad (k=N)$$

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

Training set

$x_1 \quad y_1$

$\vdots \quad\; \vdots$

$x_N \quad y_N$

$\Longrightarrow$ πίνακας σχεδιασμού (όπως X)

$$\underset{N \times N}{N} = \begin{bmatrix} N_1(x_1) & \cdots & N_N(x_1) \\ \vdots & & \\ N_1(x_N) & \cdots & N_N(x_N) \end{bmatrix}$$

$\boxed{N : \text{αντιστρέψιμος}}$

$\hat{f} = N \cdot \hat{\theta}$, $\quad \hat{\theta}$: εκτιμήτρια των θ.

$$RSS(f) \Rightarrow RSS(\theta)$$

όπου $f(x_i) = \sum_{j=1}^{N} \theta_j N_j(x_i) = N^T(x_i) \cdot \theta$

① $$\sum_{i=1}^{N} \left(y_i - f(x_i)\right)^2 = \left(y - N\theta\right)^T \left(y - N\theta\right)$$

② $$\int \left(f''(t)\right)^2 dt$$

$$f(t) = \sum \theta_j N_j(t) = N^T(t) \cdot \theta \Rightarrow f''(t) = \sum_{j=1}^{N} \theta_j N_j''(t) = \left(N''(t)\right)^T \cdot \theta$$

$$\left[N''(t)\right]^T = \left[N_1''(t), \ldots, N_N''(t)\right]$$

$$\left(f''(t)\right)^2 = \theta^T \underbrace{\underbrace{\left(N''(t)\right)}_{N\times 1} \underbrace{N''(t)^T}_{1\times N}}_{N\times N} \cdot \theta$$

$$\left(N''(t)\right)\left(N''(t)\right)^T = \begin{bmatrix} N_1''N_1'' & N_1''N_2'' & \cdots & N_1''N_N'' \\ N_N''N_1'' & \cdots & \cdots & N_N''N_N'' \end{bmatrix}$$

$$\Rightarrow \int \left(f''(t)\right)^2 dt = \theta^T \underline{\Omega_N} \theta$$

$$\left(\Omega_N\right)_{ij} = \int N_i''(t) N_j''(t)\, dt$$

$$RSS(\theta) = (y - N\theta)^T (y - N\theta) + \lambda \theta^T \Omega_N \theta$$

[αντίστοιχο του ridge regression]

$$\min_{b} (y - Xb)^T (y - Xb) + \lambda b^T b$$

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

$$\hat{f} = N \cdot \hat{\theta} = \underbrace{N (N^T N + \lambda \Omega_N)^{-1} N^T}_{S(\lambda)} y$$

$$\hat{f} = S(\lambda) \cdot y$$

[Στο ordinary regression

$$\hat{f} = \hat{y} = H \cdot y$$

hat matrix]

$$H = X(X^T X)^{-1} X^T$$

$H_{N \times N}$ $\quad$ $X_{N \times (p+1)}$

$r(H) = p+1$, $HH = H$ projection matrix

$trace(H) = p+1 = df_{model}$.

$S(\lambda)$

$$r(S(\lambda)) = N$$

$$S(\lambda) \cdot S(\lambda) \preceq S(\lambda)$$

$$\boxed{\begin{array}{l} S(\lambda) = S(\lambda) \cdot S(\lambda) + R \\ R : \text{pos. definite.} \end{array}}$$

effective degrees of freedom

$$\boxed{df(\lambda) = trace(S(\lambda))}$$

$$S(\lambda) = N (N^T N + \lambda \Omega_N)^{-1} N^T$$

$$\exists\, N^{-1}, (N^T)^{-1}$$

$$N^T N + \lambda \Omega_N = N^T \left[I + \lambda (N^T)^{-1} \Omega_N N^{-1} \right] N$$

$$N^T\left(I+\lambda (N^T)^{-1}\Omega_N N^{-1}\right)N =$$

$$=\left(N^T+\lambda \cancel{N^T (N^T)^{-1}}^{\,I}\,\Omega_N N^{-1}\right)N = \ldots$$

$$S(\lambda) = N(N^TN+\lambda\Omega_N)^{-1}N^T$$

$$=$$

$$\Rightarrow S(\lambda) = N\Big(N^T\big(I+\lambda \underbrace{(N^T)^{-1}\Omega_N N^{-1}}_{K=(N^T)^{-1}\Omega_N N^{-1}}\big)N\Big)^{-1}N^T$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

$$S(\lambda) = N\underline{\left(N^T(I+\lambda K)\cdot N\right)^{-1}}N^T$$

$$= \cancel{N N^{-1}}(I+\lambda K)^{-1}\cancel{(N^T)^{-1} N^T}$$

$$S(\lambda) = (I+\lambda K)^{-1}, \quad K=(N^T)^{-1}\Omega_N N^{-1}$$

Reinsch form of $S(\lambda)$

K δεν εξαρτάται από το λ.

Έστω $K = UDU^T$, $D=\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_N\end{pmatrix}$ ιδιοτιμές του K.

Τότε $S(\lambda) = \boxed{\sum_{j=1}^{N}\frac{1}{1+\lambda d_j}u_j u_j^T} = \sum_{j=1}^{N}\rho_j(\lambda)\,u_j u_j^T$

$$df(\lambda) = \text{trace } S(\lambda) = \sum_{j=1}^{N}\frac{1}{1+\lambda d_j}$$

$\downarrow \lambda$

$$\hat f = S(\lambda)\cdot y = \sum_{j=1}^{N} \frac{1}{1+\lambda d_j}\, u_j \underbrace{u_j^T \cdot y}$$

$$\Rightarrow \hat f = \sum_{j=1}^{N} \frac{u_j^T \cdot y}{1+\lambda d_j}\, u_j$$

$u_j \in \mathbb{R}^N$

j - ιδιοδ. του K.

$$\hat f = \sum_{j=1}^{N} J_j(y)\, \underline{u_j}$$

$(u_1, \dots, u_N)$ βάση του $\mathbb{R}^N$

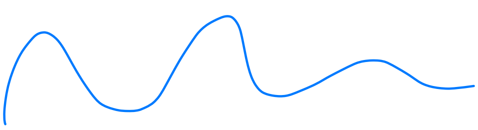
$\boxed{\text{Demmler-Reinsch basis}}$

$$J_j(y) = \frac{u_j^T y}{1+\lambda d_j}$$

$$\left(\text{αν } \lambda = 0 : \quad \hat f = \sum_{j=1}^{N} (u_j^T y)\, u_j \right)$$

<u>Ιδιότητα</u> Τα u_j που αντιστοιχούν σε μικρότερες ιδιοτιμές d_j είναι πιο ομαλά.

Επομένως $\underline{J_j \downarrow_\lambda, \downarrow_{d_j}}$ μικρότερο για τα λιγότερο ομαλά u_j

u_j  $\Rightarrow J_j \downarrow$

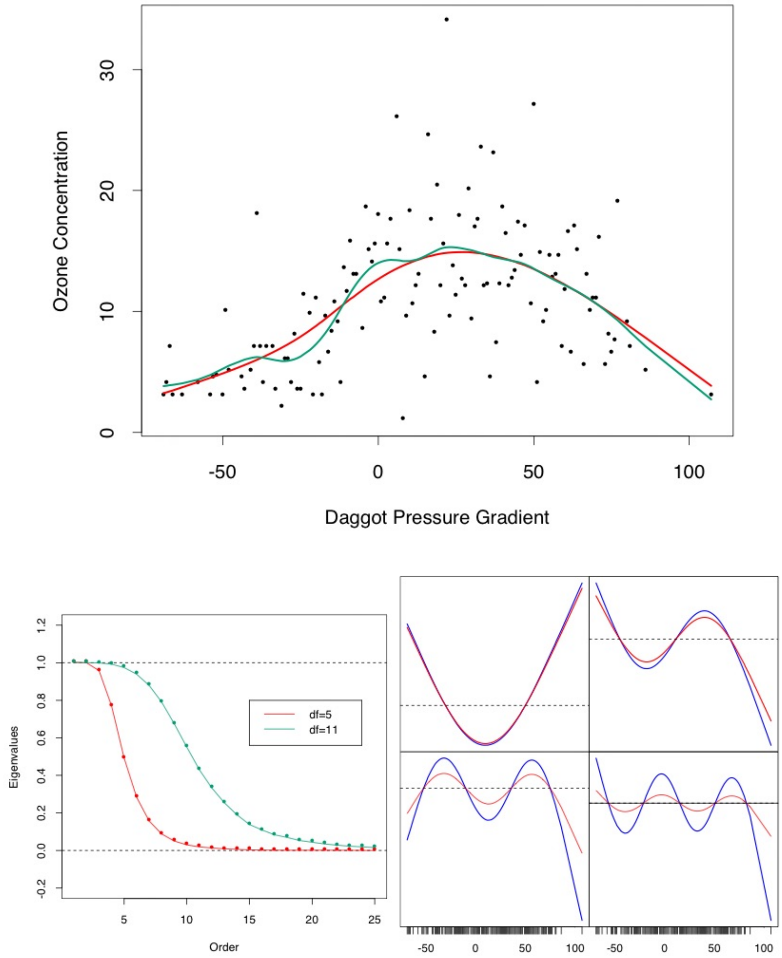


FIGURE 5.7. *(Top:) Smoothing spline fit of ozone concentration versus Daggot pressure gradient. The*

$df(\lambda) = \mathrm{trace}(S(\lambda))$

training set $\begin{pmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_N & y_N \end{pmatrix}$

$\begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \Rightarrow N = (N_j(x_i))_{i,j}$

$\Rightarrow \Omega_N$

$\Rightarrow K = (N^{-1})^{T} \Omega_N N^{T}$

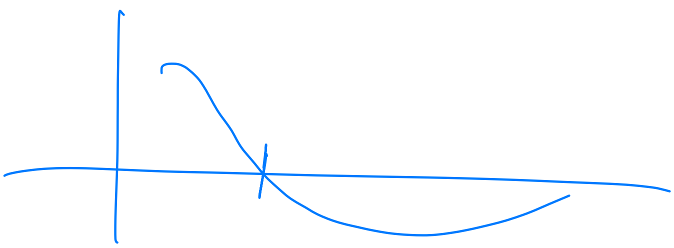
$\Rightarrow K = UDU^T \Rightarrow d_1, \dots, d_N$, $u_1, \dots, u_N$

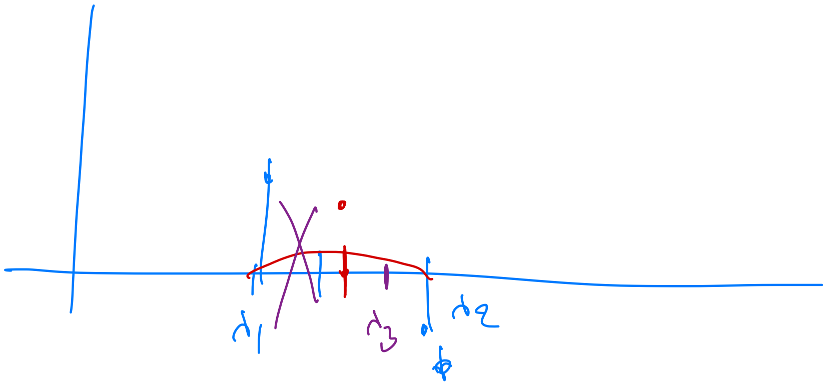
$S(\lambda) = (I + \lambda K)^{-1}$ $\boxed{df(\lambda) = \mathrm{trace}(S(\lambda))}$!!

Αν ζητήσουμε $df(\lambda) = m \Rightarrow$ λύνουμε ως προς λ

$\mathrm{trace}(S(\lambda)) = \sum_{j=1}^{N} \frac{1}{1+\lambda d_j}$

d_j ιδιοτιμές του K





$\partial f(\alpha) = (2 = 0$

$g(\alpha) = 0$