More on Splines

Recall the basis

$$N_1(x) = 1, \quad N_2(x) = x$$

and

$$N_{2+l}(x) = \frac{(x-\xi_l)_+^3 - (x-\xi_K)_+^3}{\xi_K - \xi_l} - \frac{(x-\xi_{K-1})_+^3 - (x-\xi_K)_+^3}{\xi_K - \xi_{K-1}}$$

for $l = 1, \ldots, K-2$ for natural cubic splines. Observe that $N_1''(x) = N_2''(x) = 0$ and

$$N_{2+l}''(x) = \begin{cases} 6\frac{x-\xi_l}{\xi_K-\xi_l} & x \in (\xi_l, \xi_{K-1}] \\ 6\frac{(\xi_{K-1}-\xi_l)(\xi_K-x)}{(\xi_K-\xi_l)(\xi_K-\xi_{K-1})} & x \in (\xi_{K-1}, \xi_K) \\ 0 & x \leq \xi_l \text{ and } x \geq \xi_K \end{cases}$$

Assuming that $\xi_1 < \ldots < \xi_K$ the functions $N_3'', \ldots, N_K''$ are linearly independent.

Regularity of the Spline Smoother

If $x_1, \ldots, x_N$ are all different, $N_1, \ldots, N_N$ is the basis for the n.c.s. with knots $x_1, \ldots, x_N$ and $f = \sum_{i=1}^N \theta_i N_i$ we have

$$\theta^T \mathbf{\Omega}_N \theta = \int_a^b f''(x)^2 \mathrm{d}x = 0$$

if and only if $f''(x) = 0$ for all $x \in [a, b]$. Hence

$$\theta_{2+l} = \ldots = \theta_K = 0.$$

If also $\theta^T \mathbf{N}^T \mathbf{N} \theta = 0$ then

$$(\theta_1 \ \theta_2) \left(\begin{array}{cc} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{array} \right) \left(\begin{array}{c} \theta_1 \\ \theta_2 \end{array} \right) = 0,$$

which implies that $\theta_1 = \theta_2 = 0$ if $N \geq 2$. The in general positive semidefinite matrix

$$\mathbf{N}^T \mathbf{N} + \lambda \mathbf{\Omega}_N$$

is thus positive definite for $\lambda > 0$.

The Reinsch Form

Let

$$\mathbf{S}_\lambda = \mathbf{N}(\mathbf{N}^T\mathbf{N} + \lambda\mathbf{\Omega}_N)^{-1}\mathbf{N}^T$$

be the spline smoother and $\mathbf{N} = UDV^T$ the singular value decomposition of $\mathbf{N}$. Since $\mathbf{N}$ is square $N \times N$, U is orthogonal hence invertible with $U^{-1} = U^T$, and D is invertible if $\mathbf{N}$ has full rank N. Then

$$\begin{aligned}
\mathbf{S}_\lambda &= UDV^T(VD^2V^T + \lambda\mathbf{\Omega}_N)^{-1}VDU^T \\
&= U(D^{-1}V^TVD^2V^TVD^{-1} + \lambda D^{-1}V^T\mathbf{\Omega}_N VD^{-1})^{-1}U^T \\
&= U(I + \lambda D^{-1}V^T\mathbf{\Omega}_N VD^{-1})^{-1}U^T \\
&= (U^TU + \lambda U^TD^{-1}V^T\mathbf{\Omega}_N VD^{-1}U)^{-1} \\
&= (I + \lambda \underbrace{U^TD^{-1}V^T\mathbf{\Omega}_N VD^{-1}U}_{\mathbf{K}})^{-1} \\
&= (I + \lambda\mathbf{K})^{-1}
\end{aligned}$$

The Demmler-Reinsch Basis

The matrix $\mathbf{K}$ is positive semidefinite and we write

$$\mathbf{K} = \bar{U} D \bar{U}^T$$

where $D = \mathrm{diag}(d_1, \ldots, d_N)$ with $0 = d_1 = d_2 < d_3 \leq \ldots \leq d_N$ and $\bar{U}$ is orthogonal.

The columns in $\bar{U}$, denoted $\bar{u}_1, \ldots, \bar{u}_N$, are known as the Demmler-Reinsch basis.

The Demmler-Reinsch basis is a (the) orthonormal basis of $\mathbb{R}^N$ with the property that the smoother $\mathbf{S}_\lambda$ is diagonal in this basis:

$$\mathbf{S}_\lambda = \bar{U}(I + \lambda D)^{-1} \bar{U}^T$$

The eigenvalues are in decreasing order

$$\rho_k(\lambda) = \frac{1}{1 + \lambda d_k}$$

for $k = 1, \ldots, N$ – and $\rho_1(\lambda) = \rho_2(\lambda) = 1$.

The Demmler-Reinsch Basis

We may also observe that

$$\mathbf{S}_\lambda \bar{u}_k = \rho_k(\lambda)\bar{u}_k.$$

We think of and visualize $\bar{u}_k$ as a function evaluated in the points $x_1, \ldots, x_N$.

One important consequence of these derivations is that the Demmler-Reinsch basis does not depend upon λ and we can clearly see the effect of λ through the eigenvalues $\rho_k(\lambda)$ that work as shrinkage coefficients multiplied on the basis vectors.

Also

$$\mathrm{trace}(\mathbf{S}_\lambda) = \sum_{k=1}^{N} \frac{1}{1 + \lambda d_k}.$$

Nonparametric Logistic Regression

With

$$\text{logit}\,\Pr(Y = 1 \mid X = x) = f(x)$$

and likelihood loss + a penalty term of the form

$$\lambda \int_a^b f''(x)^2 \mathrm{d}x$$

the minimizer of the penalized minus-log-likelihood is still a spline.

The iterative optimization algorithm operates by the update scheme

$$\mathbf{f}_{i+1} = \mathbf{S}_{\lambda,i} z_i$$

with

$$\mathbf{S}_{\lambda,i} = \mathbf{N}(\mathbf{N}^T \mathbf{W}_i \mathbf{N} + \lambda \mathbf{\Omega}_N)^{-1} \mathbf{N}^T \mathbf{W}_i$$

and

$$z_i = \mathbf{f}_i + \mathbf{W}_i^{-1}(\mathbf{y} - \mathbf{p}_i).$$