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.

For the differentiation above the second derivative of $(x-\xi_l)_+^3$ equals $6(x-\xi_l)_+$. Therefore, for $x \leq \xi_l$ all terms in the second derivative are 0 and for $x \geq \xi_K$ the x's in each of the fractions cancel each other and then both fractions are seen to be equal to 1, thus the difference is 0.

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

$$(\theta_1 \ \theta_2) \begin{pmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = 0,$$

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 result above can also be proved simply by proving directly that $\mathbf{N}$ has full rank N whenever $x_1, \ldots, x_N$ are all different. Then $\mathbf{N}^T \mathbf{N}$ is positive definite. It is actually straight forward to see that it has rank at least $N-1$. The $(N-1) \times (N-1)$ upper left block matrix is lower triangular with non-zero numbers in the diagonal, which implies that the last $N-1$ columns must be linearly independent. However, it is not a priory crystal clear that the first column – the column of ones – is also always linearly independent of the others. Anyway there is a good point in observing that $\mathbf{\Omega}_N$ in itself is only positive semidefinite, and in such a way that the two paremeters corresponding to a linear fit are not penalized.

To understand the question of whether $\mathbf{N}$ has full rank it is useful to take a slightly more abstract point of view. The function space of natural cubic splines with knots $\xi_1 < \ldots < \xi_K$

is a K dimensional vector space. If we take *any* basis $\varphi_1, \ldots, \varphi_K$ of functions we know that the K functions are linearly independent – as functions. A recurring problem is whether the vectors $\varphi_1(x), \ldots, \varphi_K(x)$ where $\varphi_i(x) = (\varphi_i(x_1), \ldots, \varphi_K(x_N))^T$ are also linearly independent as N dimensional vectors if $x = (x_1, \ldots, x_N)^T$ is an N-vector with at least K different coordinates. If we take these points to be precisely the K knots, this is equivalent to asking if the vectors span a K dimensional space, which means that for any $y_1, \ldots, y_K$ there are $\beta_1, \ldots, \beta_K$ such that

$$\sum_{i=1}^{K} \beta_i \varphi_i(\xi_j) = y_j$$

for $j = 1, \ldots, K$. Since $\sum_{i=1}^{K} \beta_i \varphi_i$ is a natural cubic spline and $\varphi_1, \ldots, \varphi_K$ span the space of natural cubic splines with knots $\xi_1 < \ldots \xi_K$ we are actually asking whether there is a natural cubic spline that *interpolates* the points $(\xi_1, y_1), \ldots, (\xi_K, y_K)$. This interpolation property is a well established property of splines (for $K \geq 2$), and we provide a reference below.

Due to the interpolation property of natural cubic splines we conclude that *for any basis* $\varphi_1, \ldots, \varphi_K$ of the space of natural cubic splines with knots $\xi_1 < \ldots < \xi_K$ the vectors $\varphi_1(\xi_1), \ldots, \varphi_K(\xi_K)$ are linearly independent. This holds in particular for the previously considered specific basis, which implies that $\mathbf{N}$ always has full rank N if the x_i's are all different.

A splendid reference for many more details on splines is *Nonparametric Regression and Generalized Linear Models* by Green and Silverman. Here you can also find details on fast, linear algebra algorithms for computing with splines and spline bases. Theorem 2.2 gives the interpolation property of natural cubic splines.

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

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

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

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