RKHSs and Smoothing Splines

Stephen Vardeman

Analytics Iowa LLC

ISU Statistics and IMSE

RKHSs and penalization in prediction

Reproducing kernel Hilbert spaces (RKSHs) and penalized fitting provide a unifying theory for many important prediction methods. There is a very nice 2012 *Statistics Surveys* paper by Nancy Heckman "The Theory and Application of Penalized Least Squares Methods or Reproducing Kernel Hilbert Spaces Made Easy," that is an excellent exposition of the connection of this material to splines. Parts of what follows are borrowed shamelessly from her paper. There is also some very helpful stuff in CFZ Section 3.5 and scattered through Izenman about RKHSs. See also "Penalized Splines and Reproducing Kernels" by Pearce and Wand in *The American Statistician* (2006). "Kernel Methods in Machine Learning" by Hofmann, Scholköpf, and Smola in *The Annals of Statistics* (2008) addresses other methods besides splines.

The *p=1* smoothing spline case

To provide motivation for a somewhat more general discussion, consider again the smoothing spline problem. We consider here the function space

$$\mathcal{A} = \left\{ h : [0,1] \rightarrow \Re \,\middle|\, \begin{array}{l} h \text{ and } h' \text{ are absolutely continuous} \\ \text{and } \int_0^1 \left(h''(x)\right)^2 dx < \infty \end{array} \right\}$$

as a Hilbert space (a linear space with inner product where Cauchy sequences have limits) with inner product

$$\langle f, g \rangle_{\mathcal{A}} \equiv f(0)\, g(0) + f'(0)\, g'(0) + \int_0^1 f''(x)\, g''(x)\, dx$$

(and corresponding norm $\|h\|_{\mathcal{A}} = \langle h, h \rangle_{\mathcal{A}}^{1/2}$).

The function evaluation functional

With this definition of inner product, for $x \in [0,1]$ the (linear) functional (a mapping $\mathcal{A} \to \Re$)

$$F_x[f] \equiv f(x)$$

is continuous. Thus the so-called "Riesz representation theorem" says that there is an $R_x \in \mathcal{A}$ such that

$$F_x[f] = \langle R_x, f \rangle_{\mathcal{A}} = f(x) \quad \forall f \in \mathcal{A}$$

(R_x is called the **representer of evaluation** at x.) In fact, for $z \in [0,1]$ and

$$R_{1x}(z) \equiv xz \min(x,z) - \frac{x+z}{2}(\min(x,z))^2 + \frac{1}{3}(\min(x,z))^3$$

the representer function is

$$R_x(z) = 1 + xz + R_{1x}(z)$$

The reproducing kernel and differential operator

The function of two variables defined by

$$R(x, z) \equiv R_x(z)$$

is called the **reproducing kernel** for $\mathcal{A}$, and $\mathcal{A}$ is **reproducing kernel Hilbert space** (RKHS) (of functions).

One (linear) differential operator on elements of $\mathcal{A}$ (a map from $\mathcal{A}$ to some appropriate function space) is

$$L[f](x) = f''(x)$$

Then the optimization problem solved by smoothing splines is

minimization (over choices of $h \in \mathcal{A}$) of

$$\sum_{i=1}^{N} (y_i - F_{x_i}[h])^2 + \lambda \int_0^1 (L[h](x))^2 \, dx \tag{1}$$

Reduction to a matrix calculus problem

It is possible to show that the minimizer of the quantity (1) is necessarily of the form

$$h(x) = \alpha_0 + \alpha_1 x + \sum_{i=1}^{N} \beta_i R_{1x_i}(x)$$

and that for such h, the criterion (1) is

$$(\mathbf{Y} - \mathbf{T}\boldsymbol{\alpha} - \mathbf{K}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{T}\boldsymbol{\alpha} - \mathbf{K}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}' \mathbf{K} \boldsymbol{\beta}$$

for

$$\underset{N \times 2}{\mathbf{T}} = (\mathbf{1}, \mathbf{X}), \boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}, \text{ and } \mathbf{K} = (R_{1x_i}(x_j))$$

So this has finally produced a matrix calculus problem.