Multi-Dimensional Smoothing Splines

Stephen Vardeman
Analytics Iowa LLC
ISU Statistics and IMSE

p=2 function optimization problem

If $p = 2$ (with $\mathbf{x} = (x_1, x_2)$), one might propose to seek

$$\hat{f}_\lambda = \underset{\text{functions } h \text{ with 2 derivatives}}{\arg\min} \left(\sum_{i=1}^{N} (y_i - h(\mathbf{x}_i))^2 + \lambda J[h] \right)$$

for

$$J[h] \equiv \iint_{\Re^2} \left(\frac{\partial^2 h}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 h}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 h}{\partial x_2^2} \right)^2 dx_1 dx_2$$

An optimizing $\hat{f}_\lambda : \Re^2 \longrightarrow \Re$ can be identified and is called a "thin plate spline." As $\lambda \longrightarrow 0$, $\hat{f}_\lambda$ becomes an interpolator, as $\lambda \longrightarrow \infty$ it defines the OLS plane through the data in 3-space. In general, it can be shown to be of the form

$$f_\lambda(\mathbf{x}) = \beta_{0\lambda} + \boldsymbol{\beta}'_\lambda \mathbf{x} + \sum_{i=1}^{N} \alpha_{i\lambda} g_i(\mathbf{x}) \qquad (1)$$

where $g_i(\mathbf{x}) = \eta(\|\mathbf{x} - \mathbf{x}_i\|)$ for $\eta(z) = z^2 \ln z^2$.

Radial basis functions and large N

The $g_i(\mathbf{x})$ are "radial basis functions" (radially symmetric basis functions) and fitting is accomplished much as for the $p = 1$ case. The form (1) is plugged into the optimization criterion and a discrete penalized least squares problem emerges (after taking account of some linear constraints that are required to keep $J[f_\lambda] < \infty$). HTF seem to indicate that in order to keep computations from exploding with N, it usually suffices to replace the N functions $g_i(\mathbf{x})$ in (1) with $K \ll N$ functions $g_i^*(\mathbf{x}) = \eta(\|\mathbf{x} - \mathbf{x}_i^*\|)$ for K potential input vectors $\mathbf{x}_i^*$ placed on a rectangular grid covering the convex hull of the N training data input vectors $\mathbf{x}_i$.

Large p smoothing spline strategies

For large p, one might simply declare that attention is going to be limited to predictors of some restricted form, and for h in that restricted class, seek to optimize

$$\sum_{i=1}^{N} (y_i - h(\mathbf{x}_i))^2 + \lambda J[h]$$

for $J[h]$ some appropriate penalty on h intended to regularize/restrict its wiggling. For example, one might assume that a form

$$g(\mathbf{x}) = \sum_{j=1}^{p} g_j(x_j)$$

will be used and set

$$J[g] = \sum_{j=1}^{p} \int \left(g_j''(x)\right)^2 dx$$

and be led to additive splines.

Large p smoothing spline strategies cont.

Or, one might assume that

$$g(\mathbf{x}) = \sum_{j=1}^{p} g_j(x_j) + \sum_{j,k} g_{jk}(x_j, x_k) \quad (2)$$

and invent an appropriate penalty function. It seems like a sum of 1-D smoothing spline penalties on the g_j and 2-D thin plate spline penalties on the g_{jk} is the most obvious starting point. Details of fitting are a bit murky (though I am sure that they can be found in book on generalized additive models). Presumably one cycles through the summands in (2) iteratively fitting functions to sets of residuals defined by the original y_i minus the sums of all other current versions of the components until some convergence criterion is satisfied. (2) is a kind of "main effects plus 2-factor interactions" decomposition, but it is (at least in theory) possible to also consider higher order terms in this kind of expansion.