Radial Basis Function Networks

Stephen Vardeman
Analytics Iowa LLC
ISU Statistics and IMSE

Smoothing kernels as basis functions

One can use the kernels applied in kernel smoothing as basis functions for prediction. That is, for

$$K_\lambda(\mathbf{x}, \xi) = D\left(\frac{\|\mathbf{x} - \xi\|}{\lambda}\right)$$

one might consider fitting nonlinear predictors of the form

$$f(\mathbf{x}) = \beta_0 + \sum_{j=1}^{M} \beta_j K_{\lambda_j}(\mathbf{x}, \xi_j) \tag{1}$$

where each basis element has prototype parameter ξ and scale parameter λ. A common choice of D for this purpose is the standard normal pdf.

Normalized kernels as basis functions

A version of this with fewer parameters is obtained by restricting to cases where $\lambda_1 = \lambda_2 = \cdots = \lambda_M = \lambda$. This restriction, however, has the potentially unattractive effect of forcing "holes" or regions of $\Re^p$ where (in each) $f(\mathbf{x}) \approx \beta_0$, including all "large" $\mathbf{x}$. A way to replace this behavior with differing values in the former "holes" is to replace the basis functions

$$K_\lambda\left(\mathbf{x}, \xi_j\right) = D\left(\frac{\left\|\mathbf{x} - \xi_j\right\|}{\lambda}\right)$$

with normalized versions

$$h_j^\lambda(\mathbf{x}) = \frac{D\left(\left\|\mathbf{x} - \xi_j\right\| / \lambda\right)}{\sum_{k=1}^{M} D\left(\left\|\mathbf{x} - \xi_k\right\| / \lambda\right)}$$

to produce a form

$$f(\mathbf{x}) = \beta_0 + \sum_{j=1}^{M} \beta_j h_j^\lambda(\mathbf{x}) \qquad (2)$$

Radial basis functions and neural nets

The fitting of form (1) by choice of $\beta_0, \beta_1, \ldots, \beta_M, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_M, \lambda_1, \lambda_2, \ldots, \lambda_M$ or form (2) by choice of $\beta_0, \beta_1, \ldots, \beta_M, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_M, \lambda$ is fraught with all the problems of over-parameterization and lack of identifiability associated with neural networks.

Another way to use radial basis functions to produce flexible functional forms is to replace the forms $\sigma\left(\alpha_{0m} + \boldsymbol{\alpha}_m' \mathbf{x}\right)$ in a neural network with forms $K_{\lambda_m}\left(\boldsymbol{\alpha}_m' \mathbf{x}, \boldsymbol{\xi}_m\right)$ or $h_m^{\lambda}\left(\boldsymbol{\alpha}_m' \mathbf{x}\right)$.