Nonparametric Curve Smoothing

The goal here is to investigate the relationship between a dependent variable Y and a covariate X (or more generally, a set of covariates), without imposing a global structure such as a polynomial relationship. These methods try to approximate the relationship between Y and X locally for subsets of the data, to reveal how the pattern varies across the covariate space. The text describes the model for a bivariate sample $(X_i, Y_i), i = 1, ..., n$ as:

$$Y = \phi(x) + \varepsilon,$$

where $E(\varepsilon) = 0$. These curve estimators can often be considered as weighted averages of the Y values, typically giving greater weight to values with sample X values close to x. The curve estimate can then be written as:

$$\widehat{\phi}(x) = \frac{\sum_{i=1}^{n} Y_i w(\frac{x - X_i}{\Delta})}{\sum_{i=1}^{n} w(\frac{x - X_i}{\Delta})} = \sum_{i=1}^{n} S_{0i} Y_i,$$

where

$$S_{0i} = \frac{w(\frac{x - X_i}{\Delta})}{\sum_{i=1}^{n} w(\frac{x - X_i}{\Delta})}.$$

The last expression for $\widehat{\phi}$ empasizes that it is a weighted average of the Y_i values, and that it involves the choice of a kernel function $w(u)$ and a bandwidth Δ. Hastie and Tibshirani (1990) refer to the S_{0i} weights as the equivalent kernel at x, which they use to compare different smoothing methods.

In SAS, nonparametric curve smoothing is available with lowess smoothing (LOcally WEighted Scatterplot Smoothing), which is implemented in SAS Proc LOESS. Other smoothers are available in SAS Proc INSIGHT.

Reference

Hastie, T.J., and Tibshirani, R.J. 1990. Generalized Additive Models. London: Chapman & Hall.