4.1.4.4. LOESS (aka LOWESS)

Useful When $f(\vec{x};\vec{\beta})$

LOESS is one of many "modern" modeling methods that build on "classical" methods, such as linear and nonlinear least squares regression. Modern regression methods are designed to address situations in which the classical procedures do not perform well or cannot be effectively applied without undue labor. LOESS combines much of the simplicity of linear least squares regression with the flexibility of nonlinear regression. It does this by fitting simple models to localized subsets of the data to build up a function that describes the deterministic part of the variation in the data, point by point. In fact, one of the chief attractions of this method is that the data analyst is not required to specify a global function of any form to fit a model to the data, only to fit segments of the data.

The trade-off for these features is increased computation. Because it is so computationally intensive, LOESS would have been practically impossible to use in the era when least squares regression was being developed. Most other modern methods for process modeling are similar to LOESS in this respect. These methods have been consciously designed to use our current computational ability to the fullest possible advantage to achieve goals not easily achieved by traditional approaches.

Definition of a LOESS Model

LOESS, originally proposed by Cleveland (1979) and further developed by Cleveland and Devlin (1988), specifically denotes a method that is (somewhat) more descriptively known as locally weighted polynomial regression. At each point in the data set a low-degree polynomial is fit to a subset of the data, with explanatory variable values near the point whose response is being estimated. The polynomial is fit using weighted least squares, giving more weight to points near the point whose response is being estimated and less weight to points further away. The value of the regression function for the point is then obtained by evaluating the local polynomial using the explanatory variable values for that data point. The LOESS fit is complete after regression function values have been computed for each of the n data points. Many of the details of this method, such as the degree of the polynomial model and the weights, are flexible. The range of choices for each part of the method and typical defaults are briefly discussed next.

Localized Subsets of Data

The subsets of data used for each weighted least squares fit in LOESS are determined by a nearest neighbors algorithm. A user-specified input to the procedure called the "bandwidth" or "smoothing parameter" determines how much of the data is used to fit each local polynomial. The smoothing parameter, q, is a number between $(d+1)/n$ and 1, with d denoting the degree of the local polynomial. The value of q is the proportion of data used in each fit. The subset of data used in each weighted least squares fit is comprised of the nq (rounded to the next largest integer) points whose explanatory variables values are closest to the point at which the response is being estimated.

q is called the smoothing parameter because it controls the flexibility of the LOESS regression function. Large values of q produce the smoothest functions that wiggle the least in response to fluctuations in the data. The smaller q is, the closer the regression function will conform to the data. Using too small a value of the smoothing parameter is not desirable, however, since the regression function will eventually start to capture the random error in the data. Useful values of the smoothing parameter typically lie in the range 0.25 to 0.5 for most LOESS applications.

Degree of Local Polynomials

The local polynomials fit to each subset of the data are almost always of first or second degree; that is, either locally linear (in the straight line sense) or locally quadratic. Using a zero degree polynomial turns LOESS into a weighted moving average. Such a simple local model might work well for some situations, but may not always approximate the underlying function well enough. Higher-degree polynomials would work in theory, but yield models that are not really in the spirit of LOESS. LOESS is based on the ideas that any function can be well approximated in a small neighborhood by a low-order polynomial and that simple models can be fit to data easily. High-degree polynomials would tend to overfit the data in each subset and are numerically unstable, making accurate computations difficult.

Weight Function

As mentioned above, the weight function gives the most weight to the data points nearest the point of estimation and the least weight to the data points that are furthest away. The use of the weights is based on the idea that points near each other in the explanatory variable space are more likely to be related to each other in a simple way than points that are further apart. Following this logic, points that are likely to follow the local model best influence the local model parameter estimates the most. Points that are less likely to actually conform to the local model have less influence on the local model parameter estimates.

The traditional weight function used for LOESS is the tri-cube weight function,

$$w(x) = \begin{cases} (1 - |x|^3)^3 & \text{for } |x| < 1 \\ 0 & \text{for } |x| \geq 1 \end{cases}$$

However, any other weight function that satisfies the properties listed in Cleveland (1979) could also be used. The weight for a specific point in any localized subset of data is obtained by evaluating the weight function at the distance between that point and the point of estimation, after scaling the distance so that the maximum absolute distance over all of the points in the subset of data is exactly one.

Examples

A simple computational example is given here to further illustrate exactly how LOESS works. A more realistic example, showing a LOESS model used for thermocouple calibration, can be found in Section 4.1.3.2

Advantages of LOESS

As discussed above, the biggest advantage LOESS has over many other methods is the fact that it does not require the specification of a function to fit a model to all of the data in the sample. Instead the analyst only has to provide a smoothing parameter value and the degree of the local polynomial. In addition, LOESS is very flexible, making it ideal for modeling complex processes for which no theoretical models exist. These two advantages, combined with the simplicity of the method, make LOESS one of the most attractive of the modern regression methods for applications that fit the general framework of least squares regression but which have a complex deterministic structure.

Although it is less obvious than for some of the other methods related to linear least squares regression, LOESS also accrues most of the benefits typically shared by those procedures. The most important of those is the theory for computing uncertainties for prediction and calibration. Many other tests and procedures used for validation of least squares models can also be extended to LOESS models.

Disadvantages of LOESS

Although LOESS does share many of the best features of other least squares methods, efficient use of data is one advantage that LOESS doesn't share. LOESS requires fairly large, densely sampled data sets in order to produce good models. This is not really surprising, however, since LOESS needs good empirical information on the local structure of the process in order perform the local fitting. In fact, given the results it provides, LOESS could arguably be more efficient overall than other methods like nonlinear least squares. It may simply frontload the costs of an experiment in data collection but then reduce analysis costs.

Another disadvantage of LOESS is the fact that it does not produce a regression function that is easily represented by a mathematical formula. This can make it difficult to transfer the results of an analysis to other people. In order to transfer the regression function to another person, they would need the data set and software for LOESS calculations. In nonlinear regression, on the other hand, it is only necessary to write down a functional form in order to provide estimates of the unknown parameters and the estimated uncertainty. Depending on the application, this could be either a major or a minor drawback to using LOESS.

Finally, as discussed above, LOESS is a computational intensive method. This is not usually a problem in our current computing environment, however, unless the data sets being used are very large. LOESS is also prone to the effects of outliers in the data set, like other least squares methods. There is an iterative, robust version of LOESS [Cleveland (1979)] that can be used to reduce LOESS' sensitivity to outliers, but extreme outliers can still overcome even the robust method.

NIST SEMATECH HOME TOOLS & AIDS SEARCH BACK NEXT