4452 Mathematical Modeling Lecture 19: Modeling of Data: LOWESS

Introduction

We have seen how to fit functions to data using linear regression. This is a powerful and useful technique if you know something about how the data is supposed to behave.

Linear regression works well to fit polynomials to data. Taking logarithms can transform higher order power fits into logarithmic fits, and is a common technique.

But polynomial fits become difficult for more complicated data sets (see Fig. 1).

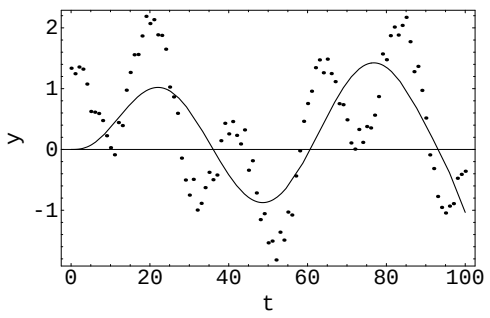

Figure 1: Data set which is not modeled well by a polynomial; linear regression model function is a polynomial of order 8. It would take a very high degree polynomial to fit the data.

A very popular technique for curve fitting complicated data sets is called *lowess*[1, 2] (locally weighted smoothing scatter plots, sometimes called *loess*). In lowess, the data is modeled locally by a polynomial weighted least squares regression, the weights giving more importance to the local data points. This method of approximating data sets is called *locally weighted polynomial regression.*

The power is lowess is that you do not require a fit function to fit the data (a smoothing parameter and degree of the local parameter (usually 1 or 2) is supplied instead).

The disadvantage in using lowess is that you do not end up with an analytic fit function (yes, this was an advantage as well). Also, lowess works best on large, densely sampled data sets.

The Lowess Method

The lowess method requires the use of a weighted least squares linear regression fit. For this weighted fit, the data points each have associated with them a weight, so you are given a set of points and weights (t_i, y_i, w_i), $0 \leq w_i \leq 1$. The difference between weighted least squares and regular least squares is that the function which is minimized is

$$F(\alpha_1, \ldots, \alpha_N) = \sum_{i=1}^{N} w_i \left(\frac{y_i - f(t_i; \alpha_1, \ldots, \alpha_M)}{\sigma_i} \right)^2 \tag{1}$$

The lowess fit is calculated at each data point in the data set. At each point, a local polynomial is fit to a local region of the data using linear least squares regression. The method has two inputs: the smoothing parameter (usually between 0 and 1), and the degree of the local polynomial (usually 1 or 2).

Let's work out lowess fits to N data points, (t_i, y_i), $i = 1, 2, 3, \ldots, N$. We want the fit at the point t_k.

First, we need the weights w_i for the neighbouring data points. Here is how these weights can be calculated.

The user supplies two parameters, α, a smoothing parameter, and the degree of the local polynomial $f(t_i; \alpha_1, \ldots, \alpha_M)$ that is to be fitted to the data. Note that this polynomial is different for each data point (t_k, y_k)!

Then, the following distances are calculated:

$$d_i = |t_k - t_i|, \quad i = 1, 2, \ldots, N,$$

which are then sorted into ascending order.

The quantity q is calculated,

$$q = \max(\text{ Truncate}(\alpha N), 1).$$

This is used to calculate the distance scale

$$D = \begin{cases} d_q & \alpha \leq 1 \\ \alpha d_N & \alpha > 1 \end{cases}$$

The weight function for the data points is defined as

$$T(u) = \begin{cases} (1 - |u|^3)^3 & |u| \leq 1 \\ 0 & |u| \geq 1 \end{cases}$$

The weights for the data points are then given by

$$w_i = T\left(\frac{t_i}{D}\right).$$

Once the weights have been found, the weighted polynomial fit on the points (t_i, y_i) with weights w_i is performed. This fit function is then used to determine the lowess fit at t_k.

This entire procedure is repeated for each data point (t_k, y_k), $k = 1, 2, \ldots, N$.

The effect of the weight function is to make lowess be a local polynomial fit, taking into account the neighbouring points of the point (t_k, y_k). Typically, the smoothing makes only a few neighbouring points contribute; the weights for points far away from (t_k, y_k) are going to be zero.

Since lowess is an intensive computational procedure, and the data sets for which it works well on are dense data sets, the local character of the lowess polynomial is usually taken to be linear or quadratic, although in principle higher order polynomials could be chosen.

The spirit of lowess however, is that simple local functions are used to approximate globally complicated data sets. To use cubic polynomials or other more complex functions for the local approximation, although allowed in the theory, would go against the "simple local function" idea underlying lowess. Figure 2 shows some lowess fits to data sets.

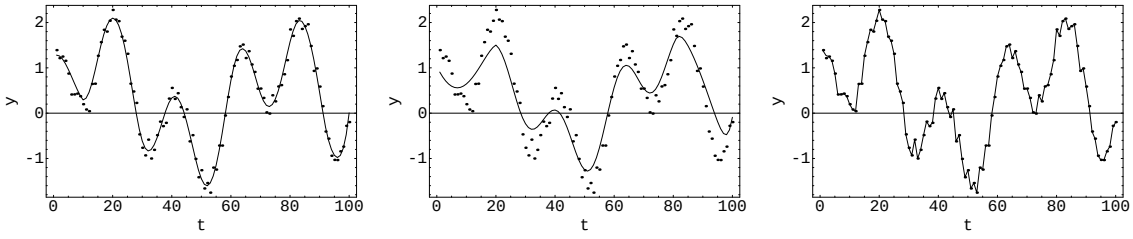

Figure 2: Locally quadratic Lowess fits to data that is not modeled will by a polynomial. The graph on the left is a seemingly good fit ($\alpha = 0.2$); the graph in the middle has been over smoothed ($\alpha = 0.4$); the graph on the right is under smoothed ($\alpha = 0.05$).

Choosing a good smoothing parameter

Choosing the degree of the local polynomial approximation is easy–most applications choose it to be 2. If you have an extremely dense data set, choosing the local polynomial to be linear may also be appropriate. Actually choosing either 1 or 2 to begin is a good idea.

The more difficult task is choosing the smoothing parameter α. We can see that over or under-smoothing the data can make your lowess fit not as good as you may like. Oversmoothing reveals general trends, but obscures the local variations. Under smoothing results in a "choppy" fit, for which there is too much local variation. Neither of these situations is desirable.

So the question becomes, how can we choose the best value for α? Since there is interplay between the local polynomial that is chosen and the smoothing parameter, the first thing we should say is that typically the local polynomial is kept as simple as possible, and the smoothing parameter is then varied. So begin your analysis with a linear local polynomial, and then vary the smoothing parameter until your curve approximates the data well. Typically, smoothing parameters in the range 0.2–0.5 will work well.

We can't measure the distance of the fit to the data points as a measure of how good our fit is, since that would always select a "choppy" fit as the best. We know there are random fluctuations in the data, but quantifying the degree of these fluctuations can be difficult. One way is to define a function to minimize which incorporates, to some degree, the closeness of the fit to the data points and a penalty function which increases for a smoother fit function.

References

[1] Cleveland, W.S. (1979) *Robust Locally Weighted Regression and Smoothing Scatterplots*, Journal of the American Statistical Association, Vol. 74, pp. 829-836.

[2] http://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm

[3] http://www.verbeia.com/mathematica/mathecon/othercode.html