Local Regression

Methodology

The central idea of local regression is to take a scatter plot of data and create a trendline based on data within the immediate vicinity of a given point while also minimizing error. The algorithm is iterative and does not generate a function. The algorithm relies on two inputs; a bandwidth (h), which acts as a smoothing coefficient and the degree (p) of polynomial used for the interpolation.

The points we wish to interpolate exist in the following form:

$$A = \begin{bmatrix} x_0 & y_0 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}$$

The algorithm can use any degree polynomial but too large a degree will result in wild variations while lower degrees may not accurately interpolate the fluctuations present in a normal data set.

For our purposes we use a third degree polynomial of the for:.

$$u(x, x_i) = a_0 + a_1(x_i - x) + \frac{1}{2}a_2(x_i - x)^2 + \frac{1}{6}a_3(x_i - x)^3$$

A general polynomial of degree p would look like this:

$$u(x, x_i) = \sum_{k=0}^{p} a_k \frac{(x_k - x)^k}{k!}$$

In the above equation each a_i is an unknown coefficient and needs to be solved for in order to interpolate the local regression value at x. To do this we use the following formula:

$$\begin{bmatrix} a_0 \\ \vdots \\ a_p \end{bmatrix} = (X^T W X)^{-1} X^T W Y$$

Now to explain what this all means. X is a Design Matrix:

$$X = \begin{bmatrix} 1 & \cdots & \frac{(x_0 - x)^n}{n!} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & \frac{(x_n - x)^n}{n!} \end{bmatrix}$$

The design matrix we use when using a third degree polynomial:

$$X = \begin{bmatrix} 1 & x_0 - x & \frac{(x_0 - x)^2}{2} & \frac{(x_0 - x)^3}{6} \\ 1 & x_1 - x & \frac{(x_1 - x)^2}{2} & \frac{(x_1 - x)^3}{6} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n - x & \frac{(x_n - x)^2}{2} & \frac{(x_n - x)^3}{6} \end{bmatrix}$$

W is a diagonal matrix with each entry represented by a separate weighting function w:

$$w(s) = \begin{cases} (1-|s|^3)^3 & -1 \le x \le 1 \\ 0 & -1 > x > 1 \end{cases}$$

$$W = \begin{bmatrix} w\left(\frac{x_0 - x}{h}\right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w\left(\frac{x_n - x}{h}\right) \end{bmatrix}$$

The Y matrix is simply the y values of our data points:

$$Y = \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix}$$

Sample Calculation

Suppose we want to use a third degree polynomial and a bandwidth of 5. Given the matrix below we will walk through the local regression process:

$$A = \begin{bmatrix} 0 & 36 \\ 1 & 29 \\ 2 & 40 \\ 3 & 39 \\ 4 & 42 \end{bmatrix}$$

We start by noting that our process will interpolate for all values x where $0 \le x \le n$. The starting x value does not matter so we'll start at $x = 2$ because it's a little more exciting than $x = 0$, but not too exciting. The matrices needed for the calculation are shown below starting with the W matrix:

$$W = \begin{bmatrix} w\left(\frac{0-2}{5}\right) & 0 & 0 & 0 & 0 \\ 0 & w\left(\frac{1-2}{5}\right) & 0 & 0 & 0 \\ 0 & 0 & w\left(\frac{2-2}{5}\right) & 0 & 0 \\ 0 & 0 & 0 & w\left(\frac{3-2}{5}\right) & 0 \\ 0 & 0 & 0 & 0 & w\left(\frac{4-2}{5}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \left(1-\left|-\frac{2}{5}\right|^3\right)^3 & 0 & 0 & 0 & 0 \\ 0 & \left(1-\left|-\frac{1}{5}\right|^3\right)^3 & 0 & 0 & 0 \\ 0 & 0 & w\left(\frac{2-2}{5}\right) & 0 & 0 \\ 0 & 0 & 0 & \left(1-\left|\frac{1}{5}\right|^3\right)^3 & 0 \\ 0 & 0 & 0 & 0 & \left(1-\left|\frac{2}{5}\right|^3\right)^3 \end{bmatrix} = \begin{bmatrix} .82 & 0 & 0 & 0 & 0 \\ 0 & .976 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & .976 & 0 \\ 0 & 0 & 0 & 0 & .82 \end{bmatrix}$$

Now onto the X Matrix:

$$X=\begin{bmatrix}1 & x_0-x & \frac{(x_0-x)^2}{2} & \frac{(x_0-x)^3}{6}\\ 1 & x_1-x & \frac{(x_1-x)^2}{2} & \frac{(x_1-x)^3}{6}\\ 1 & x_2-x & \frac{(x_2-x)^2}{2} & \frac{(x_2-x)^3}{6}\\ 1 & x_3-x & \frac{(x_3-x)^2}{2} & \frac{(x_3-x)^3}{6}\\ 1 & x_4-x & \frac{(x_4-x)^2}{2} & \frac{(x_4-x)^3}{6}\end{bmatrix}=\begin{bmatrix}1 & 0-2 & \frac{(0-2)^2}{2} & \frac{(0-2)^3}{6}\\ 1 & 1-2 & \frac{(1-2)^2}{2} & \frac{(1-2)^3}{6}\\ 1 & 2-2 & \frac{(2-2)^2}{2} & \frac{(2-2)^3}{6}\\ 1 & 3-2 & \frac{(3-2)^2}{2} & \frac{(3-2)^3}{6}\\ 1 & 4-2 & \frac{(4-2)^2}{2} & \frac{(4-2)^3}{6}\end{bmatrix}=\begin{bmatrix}1 & -2 & 2 & 1.33\\ 1 & 1 & .5 & .167\\ 1 & 0 & 0 & 0\\ 1 & -1 & .5 & -.167\\ 1 & 2 & 2 & -1.33\end{bmatrix}$$

And the Y matrix, the easiest of them all:

$$Y=\begin{bmatrix}36\\ 29\\ 40\\ 39\\ 42\end{bmatrix}$$

At this point we can do the calculation, but first take a second to notice the symmetry of the X and W matrices. The calculation is straightforward and yields the a_i coefficients. Don't' forget to transpose the X matrix.

$$(X^TWX)^{-1}=\left(\begin{bmatrix}1 & -2 & 2 & 2.667\\ 1 & -1 & .5 & .333\\ 1 & 0 & 0 & 0\\ 1 & 1 & .5 & .333\\ 1 & 2 & 2 & 2.667\end{bmatrix}^T\begin{bmatrix}.82 & 0 & 0 & 0 & 0\\ 0 & .976 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & .976 & 0\\ 0 & 0 & 0 & 0 & .82\end{bmatrix}\begin{bmatrix}1 & -2 & 2 & 2.667\\ 1 & -1 & .5 & .333\\ 1 & 0 & 0 & 0\\ 1 & 1 & .5 & .333\\ 1 & 2 & 2 & 2.667\end{bmatrix}\right)^{-1}$$

$$X^TWY=\begin{bmatrix}1 & -2 & 2 & 2.667\\ 1 & -1 & .5 & .333\\ 1 & 0 & 0 & 0\\ 1 & 1 & .5 & .333\\ 1 & 2 & 2 & 2.667\end{bmatrix}^T\begin{bmatrix}.82 & 0 & 0 & 0 & 0\\ 0 & .976 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & .976 & 0\\ 0 & 0 & 0 & 0 & .82\end{bmatrix}\begin{bmatrix}36\\ 29\\ 40\\ 39\\ 42\end{bmatrix}$$

$$(X^TWX)^{-1}X^TWY=\begin{bmatrix}36.125\\ 6.167\\ 1.044\\ 7\end{bmatrix}=\begin{bmatrix}a_0\\ a_1\\ a_2\\ a_3\end{bmatrix}$$

At this point we have our coefficients and can obtain our regression value at $x=2$. Recall:

$$u(x,x_i)=a_0+a_1(x_i-x)+\frac{1}{2}a_2(x_i-x)^2+\frac{1}{6}a_3(x_i-x)^3$$

Plugging in our obtained values yields:

$$u(2,2)=36.125+6.167(2-2)+\frac{1}{2}1.044(2-2)^2+\frac{1}{6}(-3.5)(2-2)^3$$
$$u(2,2)=36.125$$

We now have the result of our local regression around the point x_2 for $x=2$. If we ran it again about $x=2.1$ we would get a different result, but it would closely fit our value for $x=2$. Also notice that when $x_i=x$, $u(x,x_i)$ minimizes to compute a_0

Variance

The variance for each point x, can be found using the following formula:

$$V(x) = u(x) \pm c\sigma \|l(x)\|$$

The value c, represents a confidence interval corresponding to the standard distribution; for a 95% confidence interval $c = 1.96$, which is also what we use in our model. The standard deviation σ is then multiplied by the length of the l matrix. The l matrix is defined as:

$$l(x) = \begin{bmatrix} l_0 \\ \vdots \\ l_p \end{bmatrix} = e_1^T (X^T W X)^{-1} X^T W$$

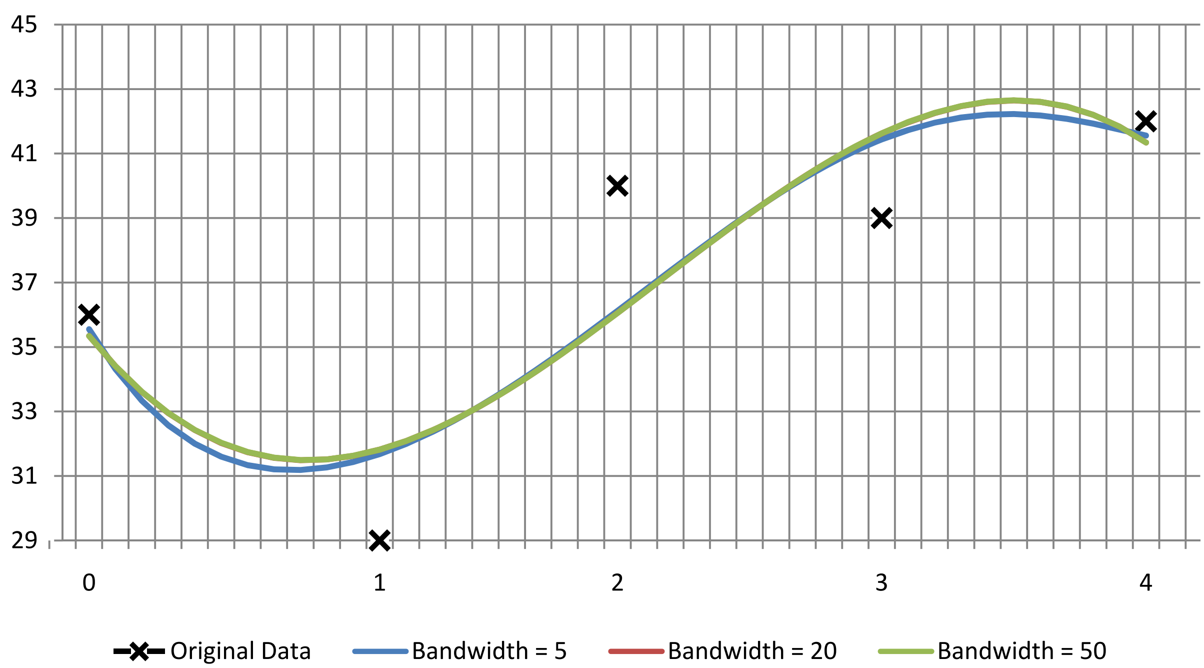
Where e_1^T is equal to the unit vector $[1 \quad 0 \quad \cdots \quad 0]$ with p columns.

The length of $l(x)$ is found by dotting $l(x)$ with $l(x)$ and taking the square root of the result:

$$\|l(x)\| = \sqrt{l(x) \cdot l(x)}$$

Bandwidth Selection

There are numerous methods of bandwidth selection; the most commonly used method is Mean Integrated Squared Error. The bandwidth choice affects the extent to which neighboring points are weighted. The graph below shows the points used in our example calculation. Notice that bandwidth values of 20 and 50 are nearly identical while the choice of 5 causes the trend to respond more strongly to neighboring points.



References

Loader, C. (1999). *Local Regression and Likelihood.* New York, New York: Springer.

Loader, C. (2004). Smoothing: Local Regression Techniques. *Handbook of Computational Statistics* .