Appendix A

Least squares polynomials and data fitting

Least square polynomials or polynomial regression is a method of fitting a polynomial to a set of data. Suppose we have a set of n points (x_i,y_i) to which we wish to fit a polynomial of the form

$$y(x) = a_0 + a_1 x + a_2 x^2 + ... a_m x^m. \quad \text{(A.1)}$$

Passing a polynomial through a set of data means selection of the coefficients so as to minimize, in a global sense, the distance between the value of the function $y(x)$ and the values at the points $y(x_i)$. This is done through the least squares method by first writing the "distance" function:

$$S = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2 - .... a_m x_i^m)^2. \quad \text{(A.2)}$$

To minimize this function, we calculate the partial derivatives with respect to each unknown coefficient and set it to zero. For the kth coefficient ($k = 0, 1, 2, \ldots, m$), we write

$$\frac{\partial S}{\partial a^k} = -2\sum_{i=1}^{n} x_i^k (y_i - a_0 - a_1 x_i - a_2 x_i^2 - .... a_m x_i^m) = 0 \quad \text{(A.3)}$$

or

$$\sum_{i=1}^{n} x_i^k (y_i - a_0 - a_1 x_i - a_2 x_i^2 - .... a_m x_i^m) = 0. \quad \text{(A.4)}$$

Repeating this for all m coefficients results in m equations from which the coefficients a_0 through a_m can be evaluated. We show here how to derive the coefficients for a first-order (linear) and second-order (quadratic) polynomial least square fit since these are the most commonly used forms. We assume n data points (x_i, y_i) as above.

A.1 Linear least square data fitting

The polynomial is first order:

$$y(x) = a_0 + a_1 x. \quad \text{(A.5)}$$

The least squares form is

$$S = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2. \quad \text{(A.6)}$$

Taking the partial derivatives with respect to a_0 and a_1 gives

$$\sum_{i=1}^{n} x_i^0 (y_i - a_0 - a_1 x_i) = 0 \quad \text{(A.7)}$$

and

$$\sum_{i=1}^{n} x_i^1 (y_i - a_0 - a_1 x_i) = 0. \quad \text{(A.8)}$$

It is most convenient to expand these and write

$$n a_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad \text{(A.9)}$$

and

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \quad \text{(A.10)}$$

Equations (A.9) and **(A.10)** can be written as a system of equations:

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{Bmatrix} \quad \text{(A.11)}$$

The system may be solved for a_0 and a_1. The result is

$$a_0 = \frac{\left\{\sum_{i=1}^{n} y_i\right\}\left\{\sum_{i=1}^{n} x_i^2\right\} - \left\{\sum_{i=1}^{n} x_i\right\}\left\{\sum_{i=1}^{n} x_i y_i\right\}}{n\sum_{i=1}^{n} x_i^2 - \left\{\sum_{i=1}^{n} x_i\right\}^2}, \; a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left\{\sum_{i=1}^{n} x_i\right\}\left\{\sum_{i=1}^{n} y_i\right\}}{n\sum_{i=1}^{n} x_i^2 - \left\{\sum_{i=1}^{n} x_i\right\}^2}. \quad \text{(A.12)}$$

With these coefficients, **(A.5)** is a linear fit to the data x_i (first-order polynomial fit) and is called a linear best fit or linear least squares fit.

A.2 Parabolic least squares fit

We start with the second-order polynomial

$$y(x) = a_0 + a_1 x + a_2 x^2. \quad \text{(A.13)}$$

The least squares form is

$$S = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2. \quad \text{(A.14)}$$

Taking the partial derivatives with respect to a_0, a_1, and a_2 gives

$$\sum_{i=1}^{n} x_i^0 (y_i - a_0 - a_1 x_i - a_2 x^2) = 0, \quad \text{(A.15)}$$

$$\sum_{i=1}^{n} x_i^1 (y_i - a_0 - a_1 x_i - a_2 x^2) = 0, \quad \text{(A.16)}$$

and

$$\sum_{i=1}^{n} x_i^2 (y_i - a_0 - a_1 x_i - a_2 x^2) = 0. \quad \text{(A.17)}$$

Again, expanding these,

$$n a_0 + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i, \quad \text{(A.18)}$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i, \quad \text{(A.19)}$$

and

$$a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i. \quad \text{(A.20)}$$

Although we could proceed to calculate the coefficients a_0, a_1, and a_2 as for the previous case, the expressions become too complex to handle. It is more practical to write the three equations as a matrix:

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \end{Bmatrix}. \quad \text{(A.21)}$$

To solve for the coefficients, one calculates the various sums first and then proceeds to solve the system of equations. Once the coefficients in **(A.21)** have been obtained, **(A.13)** is the second-order least square fit for the data x_i.

Note also that removing the third row and third column from the matrix in **(A.21)** leads to the

linear best fit in **(A.5)** with the coefficients calculated in **(A.12).**

The extension to higher order polynomials is obvious and simply adds the next terms to the matrix system in **(A.21).** The approximation for order k may be written as:

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \cdots & \sum_{i=1}^{n} x_i^k \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \cdots & \sum_{i=1}^{n} x_i^{k+1} \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 & \cdots & \sum_{i=1}^{n} x_i^{k+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_i^k & \sum_{i=1}^{n} x_i^{k+1} & \sum_{i=1}^{n} x_i^{k+2} & \cdots & \sum_{i=1}^{n} x_i^{2k} \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{k-1} \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \\ \vdots \\ \sum_{i=1}^{n} x_i^k y_i \end{Bmatrix} \quad \text{(A.22)}$$

Finally, note that the coefficients can only be calculated by hand for a limited number of points, n. In most cases, a computation tool such as MATLAB will be found useful.

Notes:

A couple of examples will follow in the near future