Mathematical Modeling
Lia Vas

Empirical Models – Interpolation Polynomial Models

Lagrange Polynomial.

Recall that two points (x_1, y_1) and (x_2, y_2) determine a unique line $y = ax + b$ passing them (obtained by solving the system of two equations $ax_1 - y_1 = -b$ and $ax_2 - y_2 = -b$ having unique solution if x-values of the points are different).

Similarly, three points (x_1, y_1), (x_2, y_2) and (x_3, y_3) (let us assume that all the x-values are different again) determine a unique quadratic function $y = ax^2 + bx + c$ passing all three points (write down a linear system of three equations that would produce the coefficients a, b and c). Convince yourself that the desired quadratic is

$$P_2(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)}y_1 + \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)}y_2 + \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}y_3$$

The form of the quadratic above is called the **Lagrangian form of the quadratic polynomial.**

Generalizing this, if $n+1$ points are given, a unique polynomial of degree at most n passing all the points can be obtained by considering the following **Lagrangian polynomial**.

$$P_n(x) = L_1(x)y_1 + \ldots + L_{n+1}y_{n+1} \text{ where}$$

$$L_i(x) = \frac{(x-x_1)\ldots(x-x_{i-1})(x-x_{i+1})\ldots(x-x_{n+1})}{(x_i-x_1)\ldots(x_i-x_{i-1})(x_i-x_{i+1})\ldots(x_i-x_{n+1})}$$

Note that because the Lagrangian polynomial passes all the points *exactly* all the absolute deviations are zero. So, a large **advantage** is that the model is a perfect fit. Also, polynomials can be easily differentiated and integrated. However, there are some **disadvantages** of using the interpolation polynomials.

1. Higher order polynomials have a tendency to oscillate severely near the end points of the interval. For example, consider the set of 17 points such that x is taking integer values between -8 and 8 and y taking value 0 for all 17 points. Graphing the Lagrange polynomial, one can see that the polynomial seems constant and of value 0 on interval [-6, 6]. On intervals [-7, -6] and [6, 7], it is nonzero and concave down with a maximum about 5 and on intervals [-8,-7] and [7,8] it is nonzero, concave up with a minimum of more than 40. This clearly does not seem to follow the data trend although it is a perfect fit for the integer values.

2. The coefficients of the polynomials are very sensitive to the small changes of data. Because we do expect the measurements errors to occur, this should be considered when using the

higher-order polynomials for predictions for values outside of the given interval. For example, let us consider the following data

x_i	0.2	0.3	0.4	0.6	0.9
Case 1 y_i	2.7536	3.2411	3.8016	5.1536	7.8671
Case 2 y_i	2.7536	3.2411	3.8916	5.1536	7.8671
Case 3 y_i	2.754	3.241	3.802	5.154	7.867

Note that the case 2 is the same as case 1 except for the second decimal in y_3 and the case 3 is the same as case 1 except that the values of y_i-values are rounded to three decimals. Consider the values of coefficients of the fourth degree Lagrange polynomial $P_4(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$..

	a_0	a_1	a_2	a_3	a_4
Case 1	2	3	4	-1	1
Case 2	3.458	-13.2	64.75	-91	46
Case 3	2.01238	2.8781	4.4159	-1.5714	1.2698

The significant difference in the coefficients between case 1 and case 2 illustrate well how a small change in data can cause a large change in the coefficients. Note that the graphs of polynomials in cases 1 and 2 still look similar on the interval [0.2, 0.9]. The coefficients of polynomials in cases 1 and 3 are not that different although, if a predictions for large x-values are to be made, these differences should be taken into consideration.

Smoothing. Lower-order polynomials.

One way of correct the disadvantages is to to fit the points choosing a smaller order polynomial (and using the least-squares criterion). This will cause the polynomial not to be a perfect fit but it reduces the tendency of higher order polynomials to oscillate. This technique combines analytic and interpolation methods of model fitting and is called **smoothing**.

When smoothing, two important questions come to mind:

1. Should a polynomial be used?
2. If so, what order of polynomial should be appropriate?

Considering the **differences** can help answer the questions above. This method is based on the fact that n-th derivative of a polynomial degree n will be constant and thus every consecutive derivative will be zero. Computing the differences of $n + 1$ points can be a good indicator if data can fit on a polynomial of smaller degree than n.

The first difference Δ is defined to be the difference between consecutive y-values. The second difference Δ^2 is the difference between consecutive values of Δ. The third difference Δ^3 is the difference between consecutive values of Δ^2 and so on. If all the values of Δ^{k+1} is zero (or reasonably close to zero) a polynomial of degree k is a good fit for the data. For example consider the following data.

x	1	2	3	4	5	6
y	1	8	27	64	125	216

Let us compute the differences.

y	1	8	27	64	125	216
Δ	7	19	37	61	91	
Δ^2	12	18	24	30		
Δ^3	6	6	6			
Δ^4	0	0				

As the third order difference is constant (and the fourth zero), we can use the third degree polynomial instead of a fifth degree (as we would based just on the fact that there is 6 points total).

Using equivalent reasoning, we would fit the 17 points with y-values all zero from one of the previous examples with the zero polynomial and thus avoid the oscillations at the end of the intervals.

To avoid possible measuring errors, **divided differences** should be considered instead of differences, in other words the first divided difference is defined to be $\Delta^1 = \frac{\Delta y}{\Delta x}$. The second one to be $\Delta^2 = \frac{\Delta(\Delta^1)}{\Delta x}$. In general, the next divided difference is the quotient of the previous divided difference and the length of the interval over which the change takes place. When using the divided differences, a few issues should be considered:

1. If x-values are close together, dividing by a small number can cause problems.

2. Measurement errors can propagate themselves and cause increase in values of higher order divided differences.

Cubic Spline Models.

Cubic spline interpolation is a very popular method of smoothing polynomials. The idea is to use successive pairs of points and to fit each data with a different cubic polynomial. This reduces the oscillations and the sensitivity to changes of data while keeping the validity.

Before considering the cubic spline, let us look at the linear splines. The idea here is to connect each consecutive pairs of points with a line. For example, to fit the data

x	1	2	3
y	5	8	25

We will find a line connecting (1,5) and (2, 8) and a line connecting (2,8) and (3, 25). These lines turn out to be $y = 3x + 2$ and $y = 17x - 26$. Then to predict a y-value of an x-value between 1 and 2, we would use the first and to predict a y-value of an x-value between 2 and 3, we would use the second line. The resulting piecewise defined function is called a liner spline.

The problem with linear splines is that it is not smooth (i.e. there is a corner at a point of transition from one line to another) and so the derivative does not exist. Considering a polynomial of order higher than linear (but not too high to oscillate significantly) can correct this.

For cubic splines, each consecutive pair of points is considered and a cubic curve is found for each pair on such a way that it the two consecutive cubic curves smoothly fits together at

the endpoints. Note that a linear function, not cubic, can fit two points, the two extra degrees of freedom will allow us to paste the consecutive cubic splines together smoothly so that a derivative will exist.

To explain the method, consider the three points (x_1, y_1), (x_2, y_2) and (x_3, y_3). Let $S_1(x) = a_1+b_1x+c_1x^2+d_1x^3$ be the cubic spline of the first two points and $S_2(x) = a_2+b_2x+c_2x^2+d_2x^3$ be the spline on the second. We will find the 8 unknown coefficients so that the following 6 conditions are satisfied

$$S_1(x_1) = y_1, \;\; S_1(x_2) = y_2, \;\; S_2(x_2) = y_2, \;\; S_2(x_3) = y_3, \;\; S_1'(x_2) = S_2'(x_2), \;\; S_1''(x_2) = S_2''(x_2)$$

To be able to determine the constants uniquely, we require the additional two conditions $S_1''(x_1) = 0$ and $S_2''(x_3) = 0$. The splines with these two conditions are called the **natural splines**.

Alternatively, if the first derivative f' of the endpoints x_1 and x_3 is known, the two additional conditions can be that $S_1'(x_1) = f'(x_1)$ and $S_1'(x_3) = f'(x_3)$. The splines with these two conditions are called the **clamped splines**.

The construction of cubic splines for more data points is done in the same manner. It should be noted that the procedure we explained here is computationally not very efficient. There are more efficient algorithms written for finding the cubic splines.

Example Find the natural cubic spline using the data from the previous example. Write down eight linear equations determining two cubic splines for the three data points. Solving the equations yields the cubic splines $S_1(x) = 2 + 10x - 10.5x^2 + 3.5x^3$ and $S_2(x) = 58 - 74x + 31.5x^2 - 3.5x^3$.

Practice Problems.

1. The data below relate the counter on a particular tape recorder and the elapse playing time.

counter reading c	100	200	300	400	500	600	700	800
elapsed playing time (in sec.) t	205	430	677	945	1233	1542	1872	2224

   a) Write down a linear system of 8 equations with 8 unknowns that you would use to find the polynomial of 7th degree that passes all eight points.

   b) Solve the linear system using your calculator or Matlab.

   c) Write down the Lagrangian form of the polynomial.

   d) Find the second, third and fourth order polynomial models using the least-squares fit. Which of the four models (2nd and 7th degree polynomial) do you think is the best to use? Give arguments explaining your choice.

   e) Consider the divided differences of 2nd and 3rd order. Explain if you would change your choice of the polynomial in part d).

2. Consider the data given below.

x	0	1	2	3	4	5	6	7
y	7	15	33	61	99	147	205	273

a) Find the polynomial of 7th degree that first the data using Matlab or calculator.

b) Considering the second, third and fourth order divided differences, determine if a polynomial of smaller degree can be used to fit the data.

3. For the following data sets, write a system of equations that determines the coefficients of the natural cubic splines passing through the given points. Solve the system using calculator or Matlab and graph the splines.

a)

x	2	4	7
y	2	8	12

b)

x	3	4	6
y	10	15	35

Solutions.

1. To lessen the numerical difficulties, you can divide each c-value by 100 and consider the c-values $1, 2, \ldots, 8$. Let $P_7(x)$ denotes the polynomial $a_7x^7 + a_6x^6 + \ldots + a_1x + a_0$. The system of 8 linear equations that computes the eight coefficients $a_0, a_1, \ldots, a_7$ is

$$a_7 1^7 + a_6 1^6 + \ldots a_0 = 205 \quad a_7 2^7 + a_6 2^6 + \ldots a_0 = 430 \quad \ldots \quad a_7 8^7 + a_6 8^6 + \ldots a_0 = 2224$$

Using the calculator or Matlab, you obtain the solutions $a_0 = -14, a_1 = 232.91, \ldots, a_7 = 0.00198$. For part c), use the formula for the Lagrange polynomial. d) All of 2nd, 3rd and 4th degree polynomials have R^2 very close to 1. Considering the efficiency, the quadratic model is a good fit. e) The second order differences are almost constant (just one value is slightly different than other five). This supports the choice of 2nd degree polynomial (the data is essentially quadratic). The third order differences are zero to fourth decimal.

2. a) The system of linear equations that calculates the coefficients of $P_7(x)$ is

$$a_0 = 7 \quad a_7 1^7 + a_6 1^6 + \ldots a_0 = 15 \quad a_7 2^7 + a_6 2^6 + \ldots a_0 = 33 \quad \ldots \quad a_7 7^7 + a_6 7^6 + \ldots a_0 = 273$$

b) The second divided difference is already constant so a quadratic equation is a perfect fit.

3. a) Let $S_1(x) = a_1 + b_1x + c_1x^2 + d_1x^3$ be the cubic spline of the first two points and $S_2(x) = a_2 + b_2x + c_2x^2 + d_2x^3$ be the spline of the second. We will find the 8 unknown coefficients so that the following 8 conditions are satisfied

$$S_1(2) = 2, \;\; S_1(4) = 8, \;\; S_2(4) = 8, \;\; S_2(7) = 12, \;\; S_1'(4) = S_2'(4) = 0, \;\; S_1''(2) = S_2''(7) = 0$$

Solving the 8 linear equations for unknown coefficients, we get the two cubic splines $y = -.0833x^3 + 0.5x^2 + 2.333x - 4$ and $y = .0555x^3 - 1.1677x^2 + 9x - 12.888$.

b) The cubic splines are $y = .833x^3 - 7.5x^2 + 26.667x - 25$ and $y = -.41667x^3 + 7.5x^2 - 33.333x + 55$.