Interpolation and approximation with polynomials

The theory (short excerpts from lectures)

Polynomial interpolation

Values of some real function $y(x)$ at a finite set of *distinct* points are prescribed and we want to interpolate them by a polynomial $p(x)$, so that we can estimate intermediate values of the function $y(x)$. Let us denote by $x_i,\ i = 0, 1, 2, \dots n$ the values of independent variable x and by y_i the prescribed values of the function $y(x)$ at x_i and summarize all given values in a table:

x	x_0	x_1	x_2	$\dots$	x_n
y	y_0	y_1	y_2	$\dots$	y_n

Generally, if we have $n+1$ data points $[x_i, y_i]$, there is exactly one polynomial

$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$$

of degree at most n going through all the data points. Its coefficients $a_0, a_1, \dots a_n$ are determined by $n+1$ linear equations $p(x_i) = y_i,\ i = 0, 1, \dots n$:

$$\begin{aligned}
a_0 + a_1 x_0 + a_2 x_0^2 + \dots + a_n x_0^n &= y_0 \\
a_0 + a_1 x_1 + a_2 x_1^2 + \dots + a_n x_1^n &= y_1 \\
&\dots \\
a_0 + a_1 x_n + a_2 x_n^2 + \dots + a_n x_n^n &= y_n
\end{aligned}$$

This can be expressed in matrix form as $\mathbf{Q}\,\mathbf{a} = \mathbf{y}$:

$$\mathbf{Q}\,\mathbf{a} \equiv \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \dots \\ y_n \end{bmatrix} \equiv \mathbf{y} \qquad (1)$$

The matrix is called Vandermonde matrix and for distinct values of x_i it is invertible, so the system (1) has unique solution for unknown vector of coefficients $\mathbf{a}$.

Remark: The system (1) can be ill conditioned and solving it is not an efficient way how to compute the coefficients of the interpolation polynomial. Using *Lagrange interpolation polynomials* can be recommended instead.

Polynomial approximation

As before, we are given $n+1$ discrete values of some real function $y(x)$. Now we want to *approximate* them by a polynomial $p(x)$, so that we can estimate intermediate values of the function. We do not require the polynomial going exactly *through* the data points, we are satisfied if it is just *close* to them. However, we usually require the order m of the polynomial to be rather small, often m is less than 3.

Remark: The requirement of x_i to be distinct values is not needed here.

Method of least squares

In this approximation method, the term *close* to the data points means that we want to minimize the sum of squares of vertical distances $r_i = p(x_i) - y_i$ of the given points from the graph of the polynomial.

Let us start again with the requirement that the polynomial goes through all the given data points: $p(x_i) = y_i,\ i = 0, 1, \dots n$. The result will be a system of equations similar to (1). Now the system matrix

$\mathbf{Q}$ is not square any more, however. Now it has m columns and n rows, m being much smaller than n, so that the system is over-determined and existence of solution generally cannot be expected:

$$\mathbf{Q}\,\mathbf{a} \equiv \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^m \\ 1 & x_1 & x_1^2 & \dots & x_1^m \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \dots \\ \dots \\ \dots \\ y_n \end{bmatrix} \equiv \mathbf{y} \quad . \tag{2}$$

This problem can be fixed by solving a system of *normal equations* $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\,\mathbf{a} = \mathbf{Q}^{\mathrm{T}}\mathbf{y}$ instead:

$$\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\,\mathbf{a} \equiv \begin{bmatrix} n+1 & \sum_{i=0}^n x_i & \dots & \sum_{i=0}^n x_i^m \\ \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 & \dots & \sum_{i=0}^n x_i^{m+1} \\ \sum_{i=0}^n x_i^2 & \sum_{i=0}^n x_i^3 & \dots & \sum_{i=0}^n x_i^{m+2} \\ \dots & \dots & \dots & \dots \\ \sum_{i=0}^n x_i^m & \sum_{i=0}^n x_i^{m+1} & \dots & \sum_{i=0}^n x_i^{2m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n y_i \\ \sum_{i=0}^n x_i y_i \\ \sum_{i=0}^n x_i^2 y_i \\ \dots \\ \sum_{i=0}^n x_i^m y_i \end{bmatrix} \equiv \mathbf{Q}^{\mathrm{T}}\mathbf{y} \tag{3}$$

which means that the error $\mathbf{r} = \mathbf{Q}\,\mathbf{a} - \mathbf{y}$ should be orthogonal to column space of the matrix $\mathbf{Q}$. The matrix $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}$ is symmetric positive semidefinite, and if the columns of $\mathbf{Q}$ are linearly independent, then it is positive definite (and consequently invertible).

Note: $n+1$ is the number of the given data points.

Theorem 1: approximation by a polynomial using the least squares method

Suppose data points $x_i, y_i,\ i = 0, \dots n$ are given so that the matrix $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}$ of the system (3) is invertible. Then the solution of (3) represents coefficients of the polynomial $p_m(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m$ which minimizes the norm $\|\mathbf{r}\|_2$ of the vector of residuals $r_i = p_m(x_i) - y_i$ (called *quadratic deviation*) among all polynomials of degree at most m.

For better insight, two different ways of proving this theorem follow.

Proof 1

We want to minimize the quadratic deviation δ, or δ^2:

$$\delta^2 \equiv \|\mathbf{r}\|_2^2 = \sum_{i=0}^n r_i^2 = \sum_{i=0}^n (p_m(x_i) - y_i)^2 \quad \to \quad min\ .$$

The minimum of this sum, with respect to the unknown coefficients $a_0, a_1, \dots a_m$ of the polynomial $p_m(x)$, is found by setting the gradient of the function $\delta^2 \equiv S(a_0, a_1, \dots a_m)$ to zero. Since it has $m+1$ variables, there are $m+1$ gradient equations

$$\begin{aligned} 0 = \frac{\partial S}{\partial a_0} &= \sum_{i=0}^n 2r_i \frac{\partial r_i}{\partial a_0} = \sum_{i=0}^n 2r_i \\ 0 = \frac{\partial S}{\partial a_1} &= \sum_{i=0}^n 2r_i \frac{\partial r_i}{\partial a_1} = \sum_{i=0}^n 2r_i x_i \\ 0 = \frac{\partial S}{\partial a_2} &= \sum_{i=0}^n 2r_i \frac{\partial r_i}{\partial a_2} = \sum_{i=0}^n 2r_i x_i^2 \\ &\dots \\ 0 = \frac{\partial S}{\partial a_m} &= \sum_{i=0}^n 2r_i \frac{\partial r_i}{\partial a_m} = \sum_{i=0}^n 2r_i x_i^m \end{aligned}$$

After substituting of $r_i = p_m(x_i) - y_i = a_0 + a_1 x + \dots + a_m x^m - y_i$ and rearanging, system (3) is obtained.

Proof 2

is based on **Theorem 2:**

Suppose $\mathbf{A}$ is a symmetric and positive definite (*spd*) matrix, $\mathbf{b}$ is a vector and $\mathbf{F}(\mathbf{x})$ is a quadratic functional $\mathbf{F}(\mathbf{x}) = \mathbf{x}^T\,\mathbf{A}\,\mathbf{x} - 2\,\mathbf{x}^T\,\mathbf{b}$. Then $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b} \;\Leftrightarrow\; \mathbf{F}(\bar{\mathbf{x}}) < \mathbf{F}(\mathbf{x}) \quad \forall\, \mathbf{x} \neq \bar{\mathbf{x}}$.

The proof of Theorem 1:

Let $\mathbf{A} = \mathbf{Q}^{\mathrm{T}}\mathbf{Q}$, $\mathbf{b} = \mathbf{Q}^{\mathrm{T}}\mathbf{y}$ and $\mathbf{a}$ is the solution of (3), i.e. $\mathbf{A}\,\mathbf{a} = \mathbf{b}$.

This holds if and only if $\mathbf{a}$ minimizes $\mathbf{F}(\mathbf{a}) = \mathbf{a}^{\mathrm{T}}\,\mathbf{A}\,\mathbf{a} - 2\,\mathbf{a}^T\,\mathbf{b}$ (using Theorem 2).

And this is equivalent to minimization of the norm

$$\begin{aligned}\|\mathbf{r}\|_2^2 = \|\mathbf{Q}\,\mathbf{a}-\mathbf{y}\|_2^2 &= (\mathbf{Q}\,\mathbf{a}-\mathbf{y})^{\mathrm{T}}(\mathbf{Q}\,\mathbf{a}-\mathbf{y}) = \mathbf{a}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\mathbf{a}-\mathbf{y}^{\mathrm{T}}\mathbf{Q}\mathbf{a}-\mathbf{a}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{y}+\mathbf{y}^{\mathrm{T}}\mathbf{y} = \mathbf{a}^{\mathrm{T}}\mathbf{A}\,\mathbf{a}-2\mathbf{a}^{\mathrm{T}}\mathbf{b}+\|\mathbf{y}\|_2^2\\ &= \mathbf{F}(\mathbf{a}) + \|\mathbf{y}\|_2^2\end{aligned}$$

Illustration of Theorem 2 in R^2, recapitulation from Calculus:

$$\mathbf{F}(\mathbf{x}) = \mathbf{x}^T\,\mathbf{A}\,\mathbf{x} - 2\,\mathbf{x}^T\,\mathbf{b} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} a & c \\ c & d \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} =$$

$$= a\,x_1^2 + 2c\,x_1x_2 + d\,x_2^2 - 2x_1b_1 - 2x_2b_2$$

$$\operatorname{grad}\,\mathbf{F}(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial\mathbf{F}}{\partial x_1} \\[2ex] \dfrac{\partial\mathbf{F}}{\partial x_2} \end{bmatrix} = 2\begin{bmatrix} a\,x_1 + c\,x_2 - b_1 \\[2ex] c\,x_1 + d\,x_2 - b_2 \end{bmatrix} = 2\,(\mathbf{A}\mathbf{x} - \mathbf{b})$$

it follows that $\mathbf{F}$ can have one critical point only − the solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ (the solution is unique, because the matrix $\mathbf{A}$ is spd and so it is invertible)

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2\mathbf{F}}{\partial x_1^2} & \dfrac{\partial^2\mathbf{F}}{\partial x_1 x_2} \\[2ex] \dfrac{\partial^2\mathbf{F}}{\partial x_2 x_1} & \dfrac{\partial^2\mathbf{F}}{\partial x_2^2} \end{bmatrix} = \mathbf{A}$$

A is spd, so the critical point represents minimum and the graph of $\mathbf{F}$ is an elliptic paraboloid oriented upwards.

Example 1

Consider the following table of data points

x	-1	1	2
y	8	4	5

Find the interpolation polynomial and use it for estimating the value at $x = 0.5$.

The solution

There are 3 data points, it follows that we are seeking the polynomial of the second order $p_2(x) = a_0 + a_1 x + a_2 x^2$ which has three unknown coefficients. Let us put together the equations (1):

$$\begin{aligned} a_0 + a_1 \cdot (-1) + a_2 \cdot (-1)^2 &= 8 \\ a_0 + a_1 \cdot 1 + a_2 \cdot 1^2 &= 4 \\ a_0 + a_1 \cdot 2 + a_2 \cdot 2^2 &= 5 \end{aligned}$$

or, in the matrix form

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_n \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \\ 5 \end{bmatrix} \quad .$$

The solution is $a_0 = 5$, $a_1 = -2$ and $a_2 = 1$, so the interpolating polynomial is $p_2(x) = 5 - 2x + x^2$.

Value at 0.5 can be estimated as $p_2(0.5) = 5 - 2 \cdot 0.5 + 0.5^2 = 4.25$

Example 2

Consider the following table of seven data points

x	-1	0	0	1	1	2	4
y	5	6	5	7	6	8	11

Find the approximation polynomials of the first and the second order and the standard deviations.

The solution

The **first-order** polynomial is a line $p_1(x) = a_0 + a_1 x$, coefficients of which are determined by the linear system (3) with $m = 1$, $n = 6$:

$$\mathbf{A} = \begin{bmatrix} n+1 & \sum_{i=0}^n x_i \\ \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \sum_{i=0}^n y_i \\ \sum_{i=0}^n x_i y_i \end{bmatrix}$$

Let us extend the given table of data in order to compute the elements of the matrix and the right hand side

								sum
x	-1	0	0	1	1	2	4	7
y	5	6	5	7	6	8	11	48
xy	-5	0	0	7	6	16	44	68
$\mathbf{x^2}$	1	0	0	1	1	4	16	23

which gives two linear equations for unknowns a_0 and a_1:

$$\begin{bmatrix} 7 & 7 \\ 7 & 23 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 48 \\ 68 \end{bmatrix}.$$

The solution is $a_0 = 5.6071$, $a_1 = 1.25$ and the linear approximation of the given data is
$p_1(x) = 5.6071 + 1.25\,x$.

$\delta^2 = \sum_{i=0}^{7} (p_1(x_i) - y_i)^2 =$
$= (4.3571-5)^2+(5.6071-6)^2+(5.6071-5)^2+(6.8571-7)^2+(6.8571-6)^2+(8.1071-8)^2+(10.6071-11)^2 =$
$= (-0.6429)^2 + (-0.3929)^2 + 0.6071^2 + (-0.1429)^2 + 0.8571^2 + 0.1071^2 + (-0.3929)^2 =$
$= 0.4133 + 0.1543 + 0.3685 + 0.0204 + 0.7346 + 0.0115 + 0.1544 = 1.8571$

$\delta = \sqrt{\delta^2} = 1.3628$

The **second-order** polynomial is a parabola $p_2(x) = a_0 + a_1 x + a_2 x^2$, coefficients of which are determined by the linear system (3) with $m = 2$, $n = 6$:

$$\mathbf{A} = \begin{bmatrix} n+1 & \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 \\ \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} x_i y_i \\ \sum_{i=0}^{n} x_i^2 y_i \end{bmatrix}$$

Let us extend the table by three more rows:

								sum
x	-1	0	0	1	1	2	4	7
y	5	6	5	7	6	8	11	48
xy	-5	0	0	7	6	16	44	68
$\mathbf{x^2}$	1	0	0	1	1	4	16	23
$\mathbf{x^2y}$	5	0	0	7	6	32	176	226
$\mathbf{x^3}$	-1	0	0	1	1	8	64	73
$\mathbf{x^4}$	1	0	0	1	1	16	256	275

which gives three linear equations for unknowns a_0, a_1 and a_2:

$$\begin{bmatrix} 7 & 7 & 23 \\ 7 & 23 & 73 \\ 23 & 73 & 275 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 48 \\ 68 \\ 226 \end{bmatrix}.$$

The solution is $a_0 = 5.5856$, $a_1 = 0.8313$, $a_2 = 0.1340$ and the quadratic approximation of the given data is $p_1(x) = 5.5856 + 0.8313\,x + 0.1340\,x^2$.

$\delta^2 = (4.8883-5)^2 + (5.5856-6)^2 + (5.5856-5)^2 + (6.5509-7)^2 + (6.5509-6)^2 + (7.7842-8)^2 +$
$+(11.0548-11)^2 = 1.0819$

$\delta = \sqrt{\delta^2} = 1.0401$

Another way, how to obtain the linear system above:

$$\mathbf{Q} = \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ \dots & \dots & \dots \\ 1 & x_n & x_n^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{bmatrix}, \ \mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \begin{bmatrix} 7 & 7 & 23 \\ 7 & 23 & 73 \\ 23 & 73 & 275 \end{bmatrix}, \ \mathbf{Q}^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} 48 \\ 68 \\ 226 \end{bmatrix}.$$