CURVE FITTING

Least Squares Line, Least Squares Polynomial Fitting

LEAST SQUARE LINE

Consider the class of linear function of the form

$$y = f(x) = Ax + B \qquad (1)$$

In previous chapter, we saw how to construct a polynomial that passes through a set of points. If all numerical values $\{x_k\}, \{y_k\}$ are known to several significant digits of accuracy, then the polynomial interpolation can be used successfully ; other wise it can not.
For values $\{x_k\}, \{y_k\}$, it realize that the true value $f(x_k)$ satisfies

$$f(x_k) = y_k + e_k \qquad (2)$$

where e_k is the measurement error.

How do find the best approximation of the form (1) that goes near (not always through) the points?
To answer this question we need to discuss the errors

$$e_k = f(x_k) - y_k \qquad (3)$$

There are several norms that can be used with the residuals in (3) to measure how far the curve $y = f(x)$ lies from the data

$$MAXIMUM\ \ ERROR : E_\infty(f) = \max_{1\le k\le N}\{|f(x_k) - y_k|\} = \max_{1\le k\le N}\{|e_k|\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

$$AVERAGE\ \ ERROR : E_1(f) = \frac{1}{N}\sum_{K=1}^{N}|f(x_k) - y_k| = \frac{1}{N}\sum_{K=1}^{N}|e_k| \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

$$ROOT\ \ MEAN\ \ SQUARE\ ERROR\ (RMS) : E_2(f) = \left(\frac{1}{N}\sum_{K=1}^{N}|f(x_k) - y_k|^2\right)^{\frac{1}{2}} = \left(\frac{1}{N}\sum_{K=1}^{N}|e_k|^2\right)^{\frac{1}{2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

Example: Compare the maximum error, average error and RMS error for the linear approximation $f(x) = 8.6 - 1.6x$ to the data points (-1,10) , (0,9), (1,7), (2,5), (3,4), (4,3), (5,0) and (6,-1).

x_k	y_k	$f(x_k) = 8.6 - 1.6x_k$	$\|e_k\| = \|f(x_k) - y_k\|$	$\|e_k\|^2$
-1	10	10.2	0.2	0.04
0	9	8.6	0.4	0.16
1	7	7.0	0	0
2	5	5.4	0.4	0.16
3	4	3.8	0.2	0.04
4	3	2.2	0.8	0.64
5	0	0.6	0.6	0.36
6	-1	-1.0	0	0
			+	+
			2.6	1.4

N=8

$$MAXIMUM\ \ ERROR \quad : E_\infty(f) = \max\{0.2, 0.4, 0.04, 0.2, 0.8, 0.6, 0\} = 0.8$$

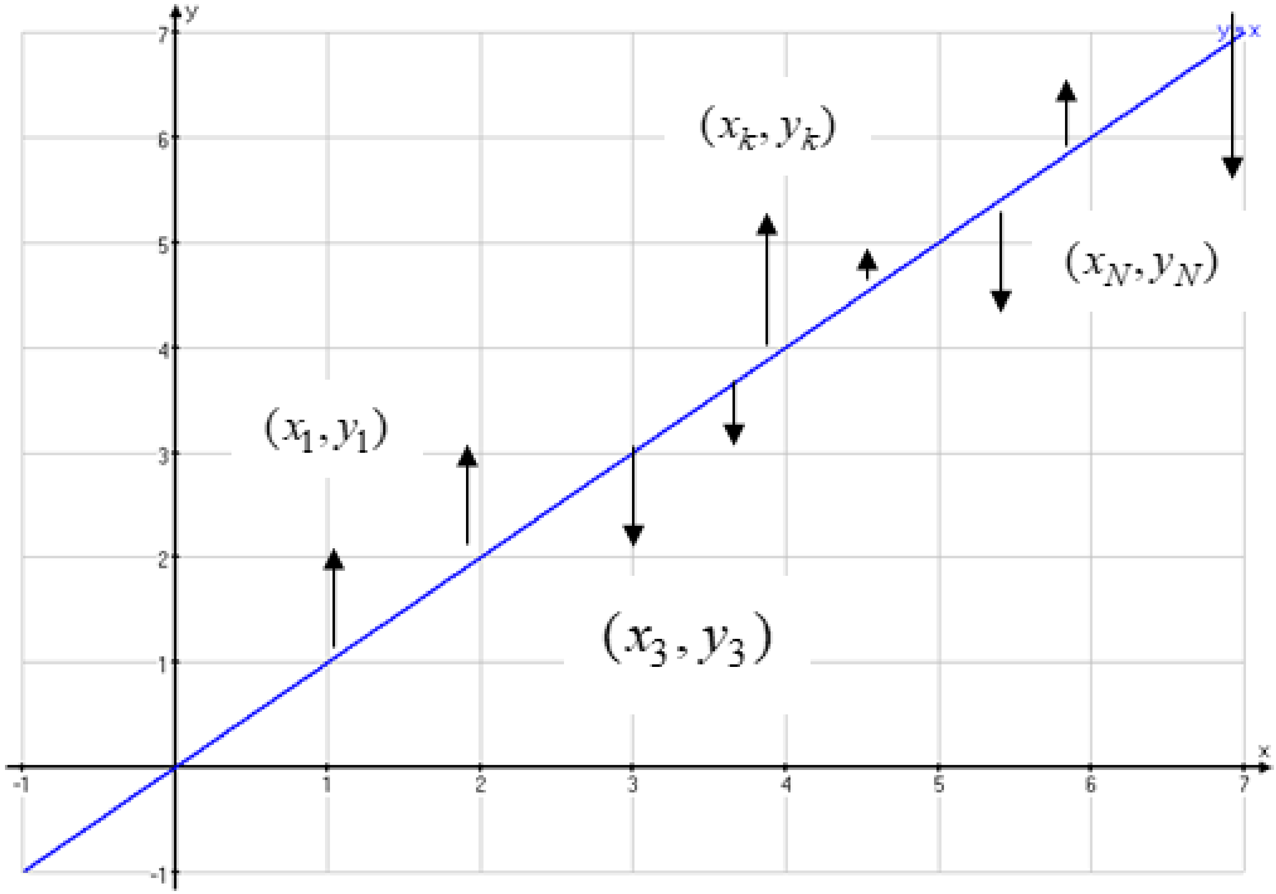
$$AVERAGE\ \ ERROR \quad : E_1(f) = \frac{1}{N}\sum_{K=1}^{N}|e_k| = \frac{1}{8}(2.6) = 0.325$$

$$ROOT\ \ MEAN\ \ SQUARE\ ERROR\ (RMS) : E_2(f) = \left(\frac{1}{N}\sum_{K=1}^{N}|e_k|^2\right)^{\frac{1}{2}} = \left(\frac{1}{8}(1.4)\right)^{\frac{1}{2}} = 0.41833$$

Best fitting line is found by minimizing one of the quantities in equation (4) through (6). Hence there are three best fitting lines that we could find. The third norm $E_2(f)$ is the traditional choice because it is much easier to minimize computationally.

FINDING THE LEAST SQUARES LINES

Let $\{(x_k, y_k)\}_{k=1}^{N}$ be a set of N points (where $\{x_k\}$ are distinct) . The least square line is the line $y = f(x) = Ax + B$ minimize the RMS error $E_2(f)$.



The quantity $E_2(f)$ will be minimum if and only if the quantity

$$N(E_2(f))^2 = \sum_{k=1}^{N} \left[(Ax_k + B) - y_k\right]^2 \text{ is minimum}$$

Theorem

Suppose that $\{(x_k, y_k)\}_{k=1}^{N}$ are N points. The coefficients of the least square line

$$y = Ax + B$$

are the solution of the following linear system

$$\left(\sum_{k=1}^{N} x_k^2\right) A + \left(\sum_{k=1}^{N} x_k\right) B = \sum_{k=1}^{N} x_k\, y_k$$

$$\left(\sum_{k=1}^{N} x_k\right) A + \quad NB \quad = \sum_{k=1}^{N} y_k \qquad (7)$$

Proof:

Where $E(A,B) = \sum_{k=1}^{N}(Ax_k + B - y_k)^2$

The minimum value of $E(A,B)$ is determined by setting the partial derivatives

$$\frac{\partial E(A,B)}{\partial A} \quad and \quad \frac{\partial E(A,B)}{\partial B}$$

equal to zero and solving these equation for A and B.

$$\frac{\partial E(A,B)}{\partial A} = \sum_{k=1}^{N} 2(Ax_k + B - y_k)x_k = 2\sum_{k=1}^{N}\left(Ax_k^2 + Bx_k - y_k x_k\right) = 0 \quad \ldots\ldots.(8)$$

$$\frac{\partial E(A,B)}{\partial B} = \sum_{k=1}^{N} 2(Ax_k + B - y_k) = 2\sum_{k=1}^{N}(Ax_k + B - y_k) = 0 \qquad \ldots\ldots.(9)$$

from (8) and (9)

$$\left(\sum_{k=1}^{N} x_k^2\right) A + \left(\sum_{k=1}^{N} x_k\right) B = \sum_{k=1}^{N} x_k \, y_k$$

$$\left(\sum_{k=1}^{N} x_k\right) A + \quad NB \quad = \sum_{k=1}^{N} y_k$$

Example:

Find the least squares line $y = Ax + B$ for the following data and evaluate $E_2(f)$.

x_k	-6	-2	0	2	6
y_k	7	5	3	2	0
$f(x_k)$	7	4.6	3.4	2.2	-0.2

According to the above table

x_k	y_k	$f(x_k)$	x_k^2	$x_k y_k$	$e_k = \lvert f(x_k) - y_k \rvert$	$(e_k)^2$
-6	7	7	36	-42	0.0	0
-2	5	4.6	4	-10	0.4	0.16
0	3	3.4	0	0	0.4	0.16
2	2	2.2	4	4	0.2	0.04
6	0	-0.2	36	0	0.2	0.04
+	+		+	+		+
0	17		80	-48		0.4

Where N=5 then

80 A +0B= -48
0A +5B= 17 solve 2x2 system to find A= -0.6 and B= 3.4

$$y = Ax + B \quad \Rightarrow \quad y = -0.6x + 3.4$$

$$and$$

$$E_2(f) = \left(\frac{1}{5} \times 0.4\right)^{\frac{1}{2}} = 0.282$$

THE POWER FIT $y = Ax^m$

Theorem:

Suppose that $\{(x_k, y_k)\}_{k=1}^{N}$ N points, where $\{x_k\}$ are distinct. The coefficients A of the least squares power curve $y = Ax^m$ is given by

$$A = \frac{\left(\sum_{k=1}^{N} x_k^m y_k\right)}{\left(\sum_{k=1}^{N} x_k^{2m}\right)}$$

Proof, where

$$E(A) = \sum_{k=1}^{N} (Ax_k^m - y_k)^2$$

$$\frac{\partial E}{\partial A} = 0 \Rightarrow \sum_{k=1}^{N} 2(Ax_k^m - y_k)\ x_k^m = 0 \Rightarrow A = \frac{\left(\sum_{k=1}^{N} x_k^m y_k\right)}{\left(\sum_{k=1}^{N} x_k^{2m}\right)}$$

Example:

Find the power fits $y = Ax^2$ *and* $y = Ax^3$ for the following data and use $E_2(f)$ to determine which curve is best.

x_k	2	2.3	2.6	2.9	3.2
y_k	5.1	7.5	10.6	14.4	19

$$for \quad y = Ax^2 \quad \Rightarrow A = \frac{\left(\sum_{k=1}^{5} x_k^2 y_k\right)}{\left(\sum_{k=1}^{5} x_k^4\right)} \quad and \quad for \quad y = Ax^3 \quad \Rightarrow A = \frac{\left(\sum_{k=1}^{5} x_k^3 y_k\right)}{\left(\sum_{k=1}^{5} x_k^6\right)}$$

Establish following table

x_k	y_k	x_k^2	x_k^3	x_k^4	x_k^6	$x_k^2 y_k$	$x_k^3 y_k$
2 2.3 2.6 2.9 3.2	5.1 7.5 10.6 14.4 19.0	4 5.29 6.76 8.41 10.24	8 12.167 17.576 24.384 32.768	16 27.984 45.697 70.728 104.74 +	64 148.03 308.915 594.823 1073.74 +	20.4 39.675 71.656 121.104 194.56 +	40.8 91.2525 186.3056 351.2016 622.592 +
				265.266	2189.5	447.395	1292.1516

$$for \quad y = Ax^2 \quad \Rightarrow \quad A = \frac{447.395}{265.266} = 1.687 \quad and \quad for \quad y = Ax^3 \quad \Rightarrow \quad A = \frac{1292.1516}{2189.5} = 0.59$$

$$f(x) = y = 1.687x^2 \qquad and \qquad g(x) = y = 0.59x^3$$

Find best curve fit using $E_2(f)$

x_k	y_k	$f(x_k)$	$g(x_k)$	$e_{k1} = \lvert f(x_k) - y_k \rvert$	$e_{k2} = \lvert g(x_k) - y_k \rvert$	$(e_{k1})^2$	$(e_{k2})^2$
2 2.3 2.6 2.9 3.2	5.1 7.5 10.6 14.4 19.0	6.748 8.924 11.4 14.19 17.27	4.72 7.18 10.37 14.39 19.33	1.648 1.424 0.8 0.21 1.73	0.38 0.32 0.23 0.01 0.33	2.715 2.027 0.64 0.0441 2.992 +	0.144 0.1024 0.0529 0.0001 0.1089 +
						8.419	0.4083

$$for \;\; y = 1.687x^2 \;\; \Rightarrow \;\; E_2(f) = \left(\frac{1}{5} \times 8.419\right)^{\frac{1}{2}} = 1.297 \;\; and \;\; for \;\; y = 0.59x^3$$

$$\Rightarrow \;\; E_2(f) = \left(\frac{1}{5} \times 0.4083\right)^{\frac{1}{2}} = 0.285 \qquad the\ best\ power\ fit\ is\ \ y = 0.59x^3$$

DATA LINEARIZATION METHOD FOR $y = Ce^{Ax}$

Suppose that we are given points $\{(x_k, y_k)\}_{k=1}^{N}$ and we want to fit an exponential curve

$$y = Ce^{Ax} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

The first step is to take logarithm of both sides

$$\ln y = \ln(Ce^{Ax}) = \ln C + \ln e^{Ax}$$

$$then$$

$$\ln y = Ax + \ln C \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$Then \quad say \quad Y = \ln y \ , \quad X = x \ , \quad and \quad B = \ln C$$

$$from \ (2)$$

$$Y = AX + B$$

Then use

$$\left(\sum_{k=1}^{N} X_k^2\right) A + \left(\sum_{k=1}^{N} X_k\right) B = \sum_{k=1}^{N} X_k Y_k$$

$$\left(\sum_{k=1}^{N} X_k\right) A + \quad NB \quad = \sum_{k=1}^{N} Y_k$$

to find A, B and C

Example:

Use data linearization method and find the exponential fit $y = Ce^{Ax}$ for the following data

x_k	0	1	2	3	4
y_k	1.5	2.5	3.5	5	7.5

$$where, \quad y = Ce^{Ax} \quad then \quad \ln y = Ax + \ln C$$
$$let \quad Y = \ln y \;, \quad x = X \quad and \quad B = \ln C$$

Establish table

$x_k = X_k$	y_k	$Y_k = \ln y_k$	$X_k Y_k$	X_k^2
0	1.5	0.405	0	0
1	2.5	0.916	0.916	1
2	3.5	1.253	2.506	4
3	5	1.609	4.827	9
4	7.5	2.015	8.06	16
+		+	+	+
10		6.196	16.309	30

N=5

30A+10B=16.309

-2x 10A+ 5B=6.196

10A=3.917

A=0.3917

B=0.4558

$$\ln C = B = 0.4558 \quad \Rightarrow \quad C = e^{0.4558} = 1.5774$$

$$y = Ce^{Ax} \quad \Rightarrow \quad y = 1.5774e^{0.3917x}$$

CHANGE OF VARIABLES FOR DATA LINEARIZATION

Function y=f(x)	Linearized form Y=AX+B	Change of variables and constants
$y = A\frac{1}{x} + B$	$y = A\frac{1}{x} + B$	$X = \frac{1}{x}$, $Y = y$
$y = \frac{D}{x + C}$	$\frac{1}{y} = \frac{1}{D}x + \frac{C}{D}$	$Y = \frac{1}{y}$, $A = \frac{1}{D}$, $B = \frac{C}{D}$
$y = \frac{1}{Ax + B}$	$\frac{1}{y} = Ax + B$	$Y = \frac{1}{y}$, $X = x$
$y = \frac{x}{Ax + B}$	$\frac{1}{y} = A + \frac{1}{x}B$	$Y = \frac{1}{y}$, $X = \frac{1}{x}$
$y = A\ln x + B$	$y = A\ln x + B$	$X = \ln x$, $Y = y$
$y = Ce^{Ax}$	$\ln y = Ax + \ln C$	$Y = \ln y$, $X = x$, $B = \ln C$
$y = Cx^{A}$	$\ln y = A\ln x + \ln C$	$Y = \ln y$, $X = \ln x$, $B = \ln C$
$y = (Ax + B)^{-2}$	$y^{\frac{-1}{2}} = Ax + B$	$X = x$, $Y = \frac{1}{\sqrt{y}}$
$y = Cxe^{-Dx}$	$\ln\left(\frac{y}{x}\right) = Ax + B$	$Y = \ln\left(\frac{y}{x}\right)$, $X = x$
$y = \frac{L}{1 + Ce^{Ax}}$	$\ln\left(\frac{L}{y} - 1\right) = Ax + \ln C$	$Y = \ln\left(\frac{L}{y} - 1\right)$, $X = x$, $B = \ln C$

Example: Fit a curve of the form $y = \frac{20}{1+Cx^A}$ to the following data

x_k	1	2	3
y_k	5	4	2

and determine $E_2(f)$

$$y = \frac{20}{1+Cx^A} \quad then \quad \frac{20}{y} = 1 + Cx^A \quad \Rightarrow \quad \frac{20}{y} - 1 = Cx^A \quad \Rightarrow \ln\left(\frac{20}{y} - 1\right) = A\ln x + \ln C$$

$$where\ Y = \ln\left(\frac{20}{y} - 1\right)\ ,\ X = \ln x\ ,\ B = \ln C$$

x_k	y_k	$X_k = \ln x_k$	$Y_k = \ln\left(\frac{20}{y_k} - 1\right)$	X_k^2	$X_k Y_k$
1	**5**	**0**	**1.098**	**0**	**0**
2	**4**	**0.693**	**1.386**	**0.48**	**0.961**
3	**2**	**1.098** +	**2.197** +	**1.205** +	**2.412** +
		1.7916	**4.681**	**1.685**	**3.373**

N=3

1.685 A + 1.7916 B = 3.373
1.7916 A + 3 B = 4.681 solve 2x2 system then

-1.8452 A= -1.7325 --------→ A= 0.9389 and B= 0.9996

Where

$$A = 0.9389 \quad and \;\; B = \ln C \quad \Rightarrow \;\; C = e^{B} \quad \Rightarrow \;\; C = e^{0.9996} = 2.7172$$

$$y = f(x) = \frac{20}{1 + 2.7172x^{0.9389}}$$

Now find $E_2(f)$ using f(x) and given data

x_k	y_k	$f(x_k)$	$e_k = \lvert f(x_k) - y_k \rvert$	$(e_k)^2$
1	**5**	**5.38**	**0.38**	**0.1444**
2	**4**	**3.221**	**0.779**	**0.606**
3	**2**	**2.319**	**0.319**	**0.1017**
				+ **0.852**

$$E_2(f) = \left[\frac{1}{3} \times 0.852\right]^{\frac{1}{2}} = 0.532$$

EXERCISE1: Let $y = a_1 x^{\frac{1}{2}} + a_2 x^{-2}$

Find the curve fit for the function y using the following table and determine $E_2(f)$.

x_k	1	1.5	2	2.2	3
y_k	-1	1.1116	2.07084	2.3466	3.1308

EXERCISE2: Find the curve fit $y = f(x) = Cxe^{Ax} + B \quad with \quad f(0) = 1$, for the points (-1,0) , (1, 6) and (2,9) using the least square technique and determine $E_2(f)$.

LEAST SQUARE PARABOLA

Theorem:

Suppose that $\{(x_k, y_k)\}_{k=1}^{N}$ are N points, where the abscissas are distinct. The coefficients of the least squares parabola,

$$y = f(x) = Ax^2 + Bx + C \qquad (1)$$

are the solution values A,B and C of the linear system,

$$\begin{aligned}
\left(\sum_{k=1}^{N} x_k^4\right)A + \left(\sum_{k=1}^{N} x_k^3\right)B + \left(\sum_{k=1}^{N} x_k^2\right)C &= \sum_{k=1}^{N} x_k^2\, y_k \\
\left(\sum_{k=1}^{N} x_k^3\right)A + \left(\sum_{k=1}^{N} x_k^2\right)B + \left(\sum_{k=1}^{N} x_k\right)C &= \sum_{k=1}^{N} x_k\, y_k \qquad (2) \\
\left(\sum_{k=1}^{N} x_k^2\right)A + \left(\sum_{k=1}^{N} x_k\right)B + NC &= \sum_{k=1}^{N} y_k
\end{aligned}$$

Proof:

The coefficients A,B and C will minimize the quantity,

$$E(A,B,C) = \sum_{k=1}^{N} \left(Ax_k^2 + Bx_k + C - y_k\right)^2$$

The partial derivatives

$$0 = \frac{\partial E}{\partial A} = 2\sum_{k=1}^{N} \left(Ax_k^2 + Bx_k + C - y_k\right)\left(x_k^2\right) \quad , \quad \left(\sum_{k=1}^{N} x_k^4\right)A + \left(\sum_{k=1}^{N} x_k^3\right)B + \left(\sum_{k=1}^{N} x_k^2\right)C = \sum_{k=1}^{N} x_k^2\, y_k$$

$$0 = \frac{\partial E}{\partial B} = 2\sum_{k=1}^{N} \left(Ax_k^2 + Bx_k + C - y_k\right)\left(x_k\right) \quad , \quad \left(\sum_{k=1}^{N} x_k^3\right)A + \left(\sum_{k=1}^{N} x_k^2\right)B + \left(\sum_{k=1}^{N} x_k\right)C = \sum_{k=1}^{N} x_k\, y_k$$

$$0 = \frac{\partial E}{\partial C} = 2\sum_{k=1}^{N} \left(Ax_k^2 + Bx_k + C - y_k\right) \quad , \quad \left(\sum_{k=1}^{N} x_k^2\right)A + \left(\sum_{k=1}^{N} x_k\right)B + NC = \sum_{k=1}^{N} y_k$$

Example: Find the least squares parabola for the four points $(-3,3), (0,1), (2,1)$ *and* $(4,3)$

	xk	yk	xk^2	xk^3	xk^4	(xk)(yk)	(xk^2)(yk)
	-3	3	9	-27	81	-9	27
	0	1	0	0	0	0	0
	2	1	4	8	16	2	4
	4	3	16	64	256	12	48
$\sum$	3	8	29	45	353	5	79

The linear system for finding A,B and C becomes

$$\left(\sum_{k=1}^{N} x_k^4\right)A + \left(\sum_{k=1}^{N} x_k^3\right)B + \left(\sum_{k=1}^{N} x_k^2\right)C = \sum_{k=1}^{N} x_k^2\, y_k$$

$$\left(\sum_{k=1}^{N} x_k^3\right)A + \left(\sum_{k=1}^{N} x_k^2\right)B + \left(\sum_{k=1}^{N} x_k\right)C = \sum_{k=1}^{N} x_k\, y_k$$

$$\left(\sum_{k=1}^{N} x_k^2\right)A + \left(\sum_{k=1}^{N} x_k\right)B + NC = \sum_{k=1}^{N} y_k$$

$$353A + 45B + 29C = 79$$
$$45A + 29B + 3C = 5$$
$$29A + 3B + 4C = 8$$

Now we should solve 3x3 system. Using Cramer's Rule,

$$353A + 45B + 29C = 79$$
$$45A + 29B + 3C = 5$$
$$29A + 3B + 4C = 8$$

$$\varDelta = \begin{vmatrix} 353 & 45 & 29 \\ 45 & 29 & 3 \\ 29 & 3 & 4 \end{vmatrix} = 13112$$

$$\varDelta_1 = \begin{vmatrix} 79 & 45 & 29 \\ 5 & 29 & 3 \\ 8 & 3 & 4 \end{vmatrix} = 2340$$

$$A = \frac{\varDelta_1}{\varDelta} = \frac{2340}{13112} = 0.178462$$

$$\varDelta_2 = \begin{vmatrix} 353 & 79 & 29 \\ 45 & 5 & 3 \\ 29 & 8 & 4 \end{vmatrix} = -2524$$

$$B = \frac{\varDelta_2}{\varDelta} = -\frac{2524}{13112} = -0.192495$$

$$\varDelta_3 = \begin{vmatrix} 353 & 45 & 79 \\ 45 & 29 & 5 \\ 29 & 3 & 8 \end{vmatrix} = 11152$$

$$C = \frac{\varDelta_3}{\varDelta} = \frac{11152}{13112} = 0.850519$$

Thus, the least-square parabola is $y = 0.178462x^2 - 0.192495x + 0.850519$