Least Square and Curve fitting.

التربيعات الصغرى والمنحنيات الأوفقية

هناك الكثير من البيانات يمكن الحصول عليها مختبريا او ميدانيا في مختلف المجالات العلمية وتكون هناك حاجة التعبير عنها بمعادلة من اجل الدراسة والتحليل . ان طريقة المربعات الصغرى هي إسنيا طريقة لايجاد احسن دالة تتناسب مع البيانات المعطاة .

ولاستخدام طريقة المربعات الصغرى لابد من تحديد درجة الدالة المتعددة الحدود $y=f(x)$ ثم يتم حساب الفرق بين قيم y البيانات والقيم التقريبية y التي تعطيها هذه المعادلة ويسمى الفرق بين هاتين القيمتين بالخطأ .

في مسائل عديدة من الأندراج تكون لدينا مجموعة من قيم مترادفة لمتغيرين . هناك نموذج رياضي يصف بشكل عام العلاقة بين هاتين المتغيرين ويكون على شكل دالة لاحد المتغيرين وبالاعتماد على بعض الثوابت الغير معلومة مثلا قد تكون صيغة الدالة على الشكل التالي !

$$y = f(x; a,b) = ae^{bx}$$

$$y = f(x; a,b,c) = ax^2 + bx + c$$

$$y = f(x; a,b,c,d) = a\sin bx + c\cos dx$$

ماذا كان منا مطلوب تقدير او تخمين قيمة الدالة في نقاط جديدة علينا اولا ايجاد الثوابت في النماذج الرياضية اعلاه لغرض الحصول على علاقة كاملة بين المتغيرين .

المسألة العامة في المنحنيات الأوفقية كالآتي !

لتكن لدينا القيم المترادفة التالية $(x_1,y_1), (x_2,y_2) --- (x_n,y_n)$

والمطلوب ايجاد العلاقة بين المتغيرين x و y من الشكل

$$y = f(x) = f(x, c_1, c_2, ---, c_m)$$

حيث f معرفة بدلالة x والثوابت $c_1, c_2, ---, c_m$ والتحقق ذلك علينا ايجاد قيم هذه الثوابت التي تجعل الدالة f افضل دالة توافق البيانات المعطاة .

$$e_\infty(f) = \max_{1 \le r \le n} \{ |f(x_r) - y_r| \}$$ Max. errors

$$e_2(f) = \left[\frac{1}{n} \sum_{r=1}^{n} |f(x_r) - y_r|^2 \right]^{1/2}$$ Root mean square errors.

$$e_1(f) = \frac{1}{n} \sum_{r=1}^{n} |f(x_r) - y_r|$$ Average errors.

EX: Let $f(x) = ax + b$. From the following data find $e_1(f)$, $e_2(f)$, $e_\infty(f)$

x_i	-1	0	1	2	3	4	5	6
y_i	10	9	7	5	4	3	0	-1

Sol^n:

$f(x) = P(x_i) = 8.6 - 1.6 x_i$

i	x_i	y_i	$f(x_i)$	$\|e_i\|$	$\|e_i\|^2$
1	-1	10	10.2	0.2	0.04
2	0	9	8.6	0.4	0.16
3	1	7	7	0	0
4	2	5	5.4	0.4	0.16
5	3	4	3.8	0.2	0.04
6	4	3	2.2	0.8	0.64
7	5	0	0.6	0.6	0.36
8	6	-1	-1	0	0
Σ:				2.6	1.4

$e_i = |f(x_i) - P(x_i)|$

$e_\infty(f) = \max |f(x_i) - y_i| = 0.8$

$$e_2(f) = \left(\frac{1}{8} \sum_{i=1}^{8} |f(x_i) - y_i|^2 \right)^{1/2}$$

$$= \frac{1}{8}(1.4)^{1/2} = 0.41$$

$$e_1(f) = \frac{1}{n} \sum_{i=1}^{n} e_i = \frac{1}{8}(2.6)$$

$$= 0.325$$

EX: Let us have the following data:

x_i	-2	-1	0	1	2
y_i	-3.1	-0.9	1	3.2	4.8

It is required to know which of the following f^{n}s represents the relationship between x and y better than the others

$P_1(x) = 1.1 + 1.8x$; $P_2(x) = 1 + 1.99x$, $P_3 = 0.9 + 2x$

Soln:

$$e_K = |y_K(x) - P_K(x)| \quad , \quad K = 1,2,3$$

x_i	y_i	$P_1(x)$	$e_1(x)$	$P_2(x)$	$e_2(x)$	$P_3(x)$	$e_3(x)$
-2	-3.1	-2.5	-0.6	-2.98	-0.12	-3.1	0
-1	-0.9	-0.7	-0.2	-0.99	0.09	-1.1	0.2
0	1	1.1	-0.1	1	0	0.9	0.1
1	3.2	2.9	0.3	2.99	0.21	2.9	0.3
2	4.8	4.7	0.1	4.98	-0.18	4.9	-0.1

$$e_1 = \begin{bmatrix} -0.6 \\ -0.2 \\ -0.1 \\ 0.3 \\ 0.1 \end{bmatrix} , \quad e_2 = \begin{bmatrix} -0.12 \\ 0.09 \\ 0 \\ 0.21 \\ -0.18 \end{bmatrix} , \quad e_3 = \begin{bmatrix} 0 \\ 0.2 \\ 0.1 \\ 0.3 \\ -0.1 \end{bmatrix}$$

ملاحظة: للمقارنة بين هذه المتجهات من حيث المقدار، نستعمل احدى الطرق المعروفة في قياس طول المتجهة وهو ما يعرف باسم المعيار ((Norm)) فاذا استعملنا المعيار الاقليدي نرمز له بالرمز $\|e\|_2$ ويعرف بالشكل التالي:

$$\|e\|_2 = \left[\sum_{i=1}^{m} (y_i - P(x_i))^2 \right]^{1/2}$$

$$\|e_1\|_2 = \left[(-0.6)^2 + (-0.2)^2 + (-0.1)^2 + (0.3)^2 + (0.1)^2 \right]^{1/2} = 0.714$$

$$\|e_2\|_2 = \left[(-0.12)^2 + (0.09)^2 + (0.21)^2 + (-0.18)^2 \right]^{1/2} = 0.315$$

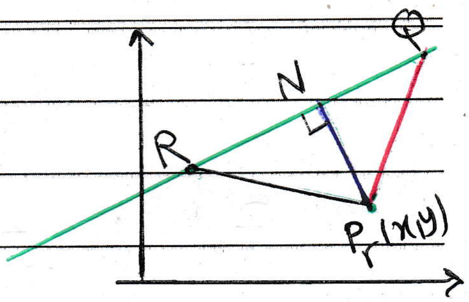
$$\|e_3\|_2 = 0.387$$

since $e_2 < e_1$; $e_2 < e_3 \Rightarrow$ then $P_2(x)$ is the best

$$P_2(x) = 1 + 1.99x$$

افضل مستقيم

Fitting a Straight line



The distance of the point P_r from the given line may be defined in any of the following ways:

(i) P_rN, The perpendicular distance of the point from the line.

(ii) P_rR; The distance parallel to the x-axis.

(iii) P_rQ; The distance " " " y-axis.

To choose the line to satisfy any of the following Conditions:

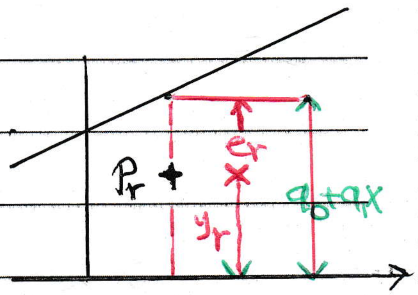
(1) The algebraic sum of the error is to be zero.

i-e $\sum e_r = 0$

(2) The sum of the magnitudes of the errors is to be minimum, i-e $\sum |e_r|$ is minimum.

(3) The sum of the squares of the error is to be a min. i-e $\sum (e_r)^2$ is min.

We will apply the method of least squares, which is that most frequently used.



Suppose $y = a_0 + a_1 x$; a_0, a_1 unknowns.

$e_r = a_0 + a_1 x_r - y_r$

$\sum_r^m (a_0 + a_1 x_r - y_r)^2$ to be min.

$$\frac{\partial}{\partial a_0} \sum (a_0 + a_1 x_r - y_r)^2 = 0$$

$$\sum_{r}^{m} 2(a_0 + a_1 x_r - y_r)(1) = 0$$

$$2a_0 m + 2a_1 \sum_r x_r - 2\sum_r y_r = 0$$

$$a_0 m + a_1 \sum_r x_r = \sum y_r \quad \text{——— (1)}$$

Diff. w.r.to a_1, (a_0 constant)

$$\frac{\partial}{\partial a_1} \sum_r (a_0 + a_1 x_r - y_r)^2 = 0 \Rightarrow \sum_r 2(a_0 + a_1 x_r - y_r)(x_r) = 0$$

$$\sum_r 2a_0 x_r + 2\sum a_1 x_r^2 - 2\sum_r y_r x_r = 0$$

$$\Rightarrow \sum_r a_0 x_r + \sum_r a_1 x_r^2 = \sum_r y_r x_r \quad \text{——— (2)}$$

$$a_0 m + a_1 \sum_r x_r = \sum_r y_r \quad \text{——— (1)}$$

$$a_0 \sum_r x_r + \sum_r a_1 x_r^2 = \sum_r x_r y_r \quad \text{——— (2)}$$

$$\begin{bmatrix} m & \sum x_r \\ \sum x_r & \sum x_r^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum y_r \\ \sum x_r y_r \end{bmatrix}$$

$$D = \begin{vmatrix} m & \sum x_r \\ \sum x_r & \sum x_r^2 \end{vmatrix} \neq 0 \Rightarrow m\sum x^2 - (\sum x)^2 \neq 0$$

m عدد النقاط

$$D_{a_0} = \begin{bmatrix} \sum y_r & \sum x_r \\ \sum x_r y_r & \sum x_r^2 \end{bmatrix}, \quad D_{a_1} = \begin{bmatrix} m & \sum y_r \\ \sum x_r & \sum x_r y_r \end{bmatrix}$$

$$a_0 = \frac{D_{a_0}}{D} = \frac{\sum y_r \sum x_r^2 - \sum x \sum xy}{m\sum x^2 - (\sum x)^2}$$

a_0

$$a_1 = \frac{m\sum xy - \sum y \sum x}{m\sum x^2 - (\sum x)^2}$$

a_1

نعوض هذه القيم a_1, a_0 في الخط المستقيم .

EX: To Fit a straight line to the following data

x_r	0.5	1	1.5	2	2.4	2.6	2.8	3
y_r	0.22	.24	.27	.33	0.35	0.35	0.38	0.39

Sol^n:

x	y	x^2	xy
0.5	0.22	0.25	0.110
1	0.24	1	0.240
1.5	0.27	2.25	0.405
2	0.33	4	0.660
2.4	0.35	5.76	0.840
2.6	0.35	6.76	0.910
2.8	0.38	7.84	1.064
3	0.39	9	1.170
Σ: 15.8	2.53	36.86	5.399

$$a_0 = \frac{\sum y \sum x^2 - \sum x \sum xy}{m \sum x^2 - (\sum x)^2}$$

$$= \frac{(2.53)(36.86) - (15.8)(5.399)}{8(36.86) - (15.8)^2}$$

$$= \frac{7.95}{45.24} = 0.176$$

$$a_1 = \frac{m \sum xy - \sum y \sum x}{m \sum x^2 - (\sum x)^2}$$

$$= \frac{8(5.399) - (2.53)(15.8)}{8(36.86) - (15.8)^2}$$

$$= 0.071$$

Fit line → $P_1(x) = a_0 + a_1 x = 0.176 + 0.071x$

: تطابق

x_r	0.5	1	1.5	2	2.4	2.6	2.8	3
y_r	0.22	0.24	0.27	0.33	0.35	0.35	0.38	0.39
$y = P(x)$	0.21	0.25	0.28	0.32	0.35	0.36	0.38	0.39

EX: H.w : Fit a straight line agree with the following data

x_r	-2	-1	0	1	2
y_r	-3.1	-0.9	1	3.2	4.8

(2)

Fitting a Polynomial

Let $y = a_0 + a_1 x + a_2 x^2$.

$e_r = y_r - f(x_r) = a_0 + a_1 x_r + a_2 x_r^2 - f(x_r)$ be min.

$$S = \sum_{r=1}^{4} (e_r)^2 = \sum_{r=1}^{4} \left(y_r - f(x_r)\right)^2 = \sum \left[(a_0 + a_1 x_r + a_2 x_r^2) - f(x_r)\right]^2$$

ولكي يكون المقدار اقل ما يمكن فإن قيمة تكون

$$\frac{\partial S}{\partial a_0} = 0, \quad \frac{\partial S}{\partial a_1} = 0, \text{ and } \frac{\partial S}{\partial a_2} = 0$$

$$\frac{\partial S}{\partial a_0} = 2\sum_{r=1}^{4} \left[(a_0 + a_1 x_r + a_2 x_r^2) - f(x_r)\right] = 0$$

$$\sum_{r=1}^{4} \left[(a_0 + a_1 x_r + a_2 x_r^2)\right] = \sum f(x_r)$$

$$4a_0 + a_1 \sum x_r + a_2 \sum x_r^2 = \sum f(x_r) \quad \text{—— (1)}$$

$$\frac{\partial S}{\partial a_1} = 2\sum_{r=1}^{4} \left[(a_0 + a_1 x_r + a_2 x_r^2) - f(x_r)\right](x_r) = 0$$

$$\sum_{r=1}^{4} a_0 x_r + a_1 \sum_{r=1} x_r^2 + a_2 \sum_{r=1} x_r^3 = \sum f(x_r) x_r \quad \text{—— (2)}$$

$$\frac{\partial S}{\partial a_2} = 2\sum \left[(a_0 + a_1 x_r + a_2 x_r^2) - f(x_r)\right](x_r^2) = 0$$

$$\sum \left[a_0 + a_1 x_r + a_2 x_r^2\right] = \sum f(x_r)(x_r^2)$$

$$a_0 \sum x_r^2 + a_1 \sum x_r^3 + a_2 \sum x_r^4 = \sum f(x_r) x_r^2 \quad \text{—— (3)}$$

نحصل على نظام (3x3) معادلات من المجاهيل a_0، a_1، a_2.

$$\begin{bmatrix} m & \sum x_r & \sum x_r^2 \\ \sum x_r & \sum x_r^2 & \sum x_r^3 \\ \sum x_r^2 & \sum x_r^3 & \sum x_r^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_r \\ \sum x_r y_r \\ \sum x_r^2 y_r \end{bmatrix}$$

in general! Let the equation of the required poly. be $y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$
and the given points (x_r, y_r), $r = 1, 2, \cdots, m$, $n < m$
where again the x values are assumed to be exact and only the y values in error. Then the error at the r-th points is

$e_r = a_0 + a_1 x_r + a_2 x_r^2 + \cdots + a_n x_r^n - y_r$, we want to find the values of $a_0, a_1, a_2, \cdots, a_n$ which are min. value of $\sum e_r^2$.

Using partial diff. for each of the variables $a_0, a_1, a_2, \cdots, a_n$, we get

$$a_0 m + a_1 \sum_{r=1}^{m} x_r + a_2 \sum_{r=1}^{m} x_r^2 + \cdots + a_n \sum_{r=1}^{m} x_r^n = \sum_{r=1}^{m} y_r$$

$$a_0 \sum x_r + a_1 \sum x_r^2 + a_2 \sum x_r^3 + \cdots + a_n \sum x_r^{n+1} = \sum_{r=1} x_r y_r$$

$$a_0 \sum x_r^2 + a_1 \sum x_r^3 + a_2 \sum x_r^4 + \cdots + a_n \sum x_r^{n+2} = \sum_{r=1}^{m} x_r^2 y_r$$

$$\vdots$$

$$a_0 \sum_{r=1}^{m} x_r^n + a_1 \sum_{r=1}^{m} x_r^{n+1} + a_2 \sum_{r=1}^{m} x_r^{n+2} + \cdots + a_n \sum_{r=1}^{m} x_r^{2n} = \sum_{r=1}^{m} x_r^n y_r$$

$$\begin{bmatrix} m & \sum x_r & \sum x_r^2 & \cdots & \sum x_r^n \\ \sum_{r=1}^{m} x_r & \sum x_r^2 & \sum x_r^3 & \cdots & \sum x_r^{n+1} \\ \vdots & & & & \\ \sum_{r=1}^{m} x_r^n & \sum x_r^{n+1} & \sum x_r^{n+2} & \cdots & \sum_{r=1}^{m} x_r^{2n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum_{r=1}^{m} y_r \\ \sum_{r=1}^{m} x_r y_r \\ \vdots \\ \sum_{r=1}^{m} x_r^n y_r \end{bmatrix}$$

Note: Change of origin.

1) In some cases it may be possible to simplify the working by changing the origin of either or both of the variables. For example, if a curve of the form $y = f(x)$ is required to fit the following data:

x	14.2	14.4	14.6	14.8	15
y	53.1	57.6	61.3	65.9	69.4

the working is easier if new variables

$$X = x - 14 \; ; \text{ and, } Y = y - 50$$

are used. i.e Find a curve of the form $Y = F(X)$ to fit the data.

X	0.2	0.4	0.6	0.8	1
Y	3.1	7.6	11.3	15.9	19.4

$X = 14.2 - 14 = 0.2$
$X = 14.4 - 14 = 0.4$
$\vdots$
$X = 15 - 14 = 1$

$Y = y - 50 = 53.1 - 50 = 3.1$, $Y = 57.6 - 50 = 7.6$, ...

2) It can be shown that the straight line obtained by the method of least squares passes through the point $(\bar{x}, \bar{y})$, where $\bar{x}$ and $\bar{y}$ are arithmetic means of the x and y valuse respectively. The origin is changed to $(\bar{x}, \bar{y})$ by the transformation $X = x - \bar{x}$, $Y = y - \bar{y}$, then the eqn. of the required straight line reduces to the form

$$X = x - \bar{x} \text{ and } Y = y - \bar{y}.$$

Ex: Fitting a second degree parabola to the following data by the method of least squares.

x	0.05	0.11	0.15	0.31	0.46	0.52	0.7	0.74	0.82	0.98	1.17
y	0.956	0.890	0.832	0.717	0.571	0.539	0.378	0.370	0.306	0.242	0.104

Solⁿ:

$$m a_0 + a_1 \Sigma x + a_2 \Sigma x^2 = \Sigma y$$

$$a_0 \Sigma x + a_1 \Sigma x^2 + a_2 \Sigma x^3 = \Sigma xy$$

$$a_0 \Sigma x^2 + a_1 \Sigma x^3 + a_2 \Sigma x^4 = \Sigma x^2 y$$

x	y	x^2	x^3	x^4	xy	x^2y
0.05	0.956	0.0002	0.000	0	0.0478	0.0001
0.11	0.890	0.0121	0.0001	0	0.0979	0.0100
0.15	0.832	0.0225	0.0033	0.0004	0.1242	0.0187
0.31	0.717	0.0961	0.297	0.0009	0.2232	0.0689
0.46	0.571	0.2116	0.0973	0.0447	0.2626	0.1228
0.52	0.539	0.2704	0.1426	0.0734	0.2802	0.1457
0.70	0.378	0.4900	0.3430	0.2402	0.2646	0.1852
0.74	0.370	0.5476	0.4052	0.2891	0.2738	0.2036
0.82	0.306	0.6724	0.5511	0.4521	0.2509	0.2050
0.98	0.242	0.9604	0.9411	0.9223	0.2371	0.2324
1.17	0.104	1.3689	1.6016	1.8730	0.1216	0.1433
Σ 6.01	5.905	4.6545	4.1150	3.9161	2.1839	1.3357

$$11 a_0 + 6.01 a_1 + 4.6545 a_2 = 5.905$$

$$6.01 a_0 + 4.6545 a_1 + 4.1150 a_2 = 2.1839$$

$$4.6545 a_0 + 4.1150 a_1 + 3.9161 a_2 = 1.3357$$

Solving the system, we get $a_0 = 0.998$, $a_1 = -1.018$, $a_2 = 0.225$

$$P_2(x) = 0.998 - 1.018x + 0.225x^2$$

Ex:

اذا كانت لدينا البيانات التالية والتي تمثل عدد سكان منطقة في بلد ما مقدرا بالملايين.

t: Year	1986	1987	1988
Y: people	2.1	2.15	2.23

المطلوب ايجاد عدد السكان في سنة 1989.

Soln:

$X_r = t - \bar{t}$, $r = 1, 2, \ldots$ $\bar{t} = \frac{\sum t_i}{n} = 1987$

$x_1 = 1986 - 1987 = -1$

$x_2 = 1987 - 1987 = 0$, $x_3 = 1988 - 1987 = 1$

$\sum x_i = 0$ شرط

$x_1 + x_2 + x_3 = 1 + 0 - 1 = 0 \Rightarrow \sum x_i = 0$

	X	y	x^2	xy
	-1	2.1	1	-2.1
	0	2.15	0	0
	1	2.23	1	2.23
$\sum$	0	6.48	2	0.13

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 6.48 \\ 0.13 \end{bmatrix}$$

$3a_0 = 6.48 \Rightarrow a_0 = 2.16$, $a_1 = 0.065$

$Y = 2.16 + 0.065X$ $t = 1989$

$X = 1989 - 1987 = 2$

$P(2) = 2.16 + 0.065(2)$

$= 2.29$

1- Fit a straight line to the following data

x	0.5	1	1.5	2	2.4	2.6	2.8	3
y	0.22	0.24	0.27	0.33	0.35	0.35	0.38	0.39

2- Fit a curve of the form $y = a_0 + a_1 x + a_2 x^2$ to the following data:

x	1	2	3	4	5	6	7	8	9
y	2.1	2.3	3.9	4.4	4.6	4.8	4.6	4.2	3.4

3- Fit a quadratic to the following data, obtaining the coefficients correct to 3 decimal places.

x	0	2	4	6	8	10	12	14	16
y	101	104	107	107	108	110	109	110	108

\# \# \#