Linear Regression with fixed points

Stancho Pavlov (stancho_pavlov@yahoo.com) Adnan Sharaf (engadnansharaf@yahoo.com)

May 24, 2014

Lets we have two sets of coordinates- S_n and S'_n:
$S_n = \{(x_i, y_i)/i = 1 \cdots n \;\; x_i \neq x_j \text{ if } i \neq j\}$ $\quad S'_n = \{(x'_i, y'_i)/i = 1 \cdots m\}$
For the first one we demand $x_i \neq x_j$ if $i \neq j$
The main aim of this article is to find a polynomial with a minimal power $f(x)$ for which $f(x_i) = y_i$ for the points of the first set and $f(x)$ best fits the points of the another.
Let denote by $W(x)$ a function which satisfy the property $W(x_i) = 0$ for the points of the first set.
Here we will choose $W_n(x) = (x - x_1)(x - x_2) \cdots (x - x_n)$.
An another possible choice[1] is $W(x) = (\varphi_1(x) - \varphi_1(x_1))(\varphi_2(x) - \varphi_2(x_2)) \cdots (\varphi_n(x) - \varphi_n(x_n))$
Using the notation $L_n(x)$ for the Lagrang's interpolation polynomial for the first set of points, we will search a polynomial $f(x)$ of the form $f(x) = L_n(x) + P(x)W(x)$.
Here P(x) is a polynomial of degree d, whose coefficients we have to determine.

$$P(x) = a_0 x^d + a_1 x^{d-1} + \cdots + a_n$$

If x_i is the is the first coordinate of an arbitrary point from the first set then
$f(x_i) = L_n(x_i) + P(x_i)W_n(x_i) = L_n(x_i) = y_i$.
The searched coefficients of $P(x)$ we will find by well known least squares method.
The sum of squared residuals

$$\sum_{i=1}^{m} \left(L_n(x'_i) + P(x'_i)W_n(x'_i) - y'_i\right)^2$$

must be minimized.
Determining partial derivatives and zeroing them we get:

$$\sum_{i=1}^{m} \left(L_n(x'_i) + P(x'_i)W_n(x'_i) - y'_i\right) x'^k_i W_n(x'_i) \quad k = 0 \cdots d$$

or the equivalent system:

$$\left| \sum_{i=1}^{m} P(x'_i) x'^k_i W_n^2(x'_i) = \sum_{i=1}^{m} \left(y'_i - L_n(x'_i)\right) x'^k_i W_n(x'_i) \quad k = 0 \cdots d \right.$$

If d=1 we get the system

$$\left| \begin{array}{l} \left[\sum\limits_{i=1}^{m} x'_i W_n^2(x'_i)\right] a_0 + \left[\sum\limits_{i=1}^{m} W_n^2(x'_i)\right] a_1 = \sum\limits_{i=1}^{m} \left(y'_i - L_n(x'_i)\right) W_n(x'_i) \\ \\ \left[\sum\limits_{i=1}^{m} x'^2_i W_n^2(x'_i)\right] a_0 + \left[\sum\limits_{i=1}^{m} x'_i W_n^2(x'_i)\right] a_1 = \sum\limits_{i=1}^{m} \left(y'_i - L_n(x'_i)\right) x'_i W_n(x'_i) \end{array} \right.$$

Let us consider an example with d=2.
Initial data is:

x_i	0	1
y_i	0	1

for the set S_2 and

x'_i	2	3	4	5	6
y'_i	2.5	2.8	3.0	4.7	6.2

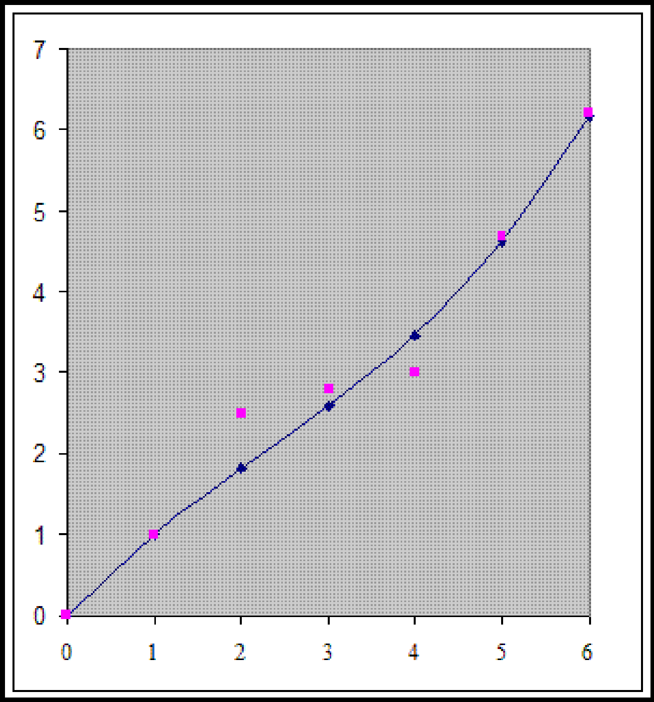
for another set S'_5. $Then \mathrm{L}_2(x) = x$ and $W_2(x) = x(x-1)$.
The calculations where maid in "EXCEL" and are shown on tha table bellow:

x_1	0.00	0.00							
x_2	1.00	1.00		W(x')	W(x')^2	x'W(x')^2	x'^2W(x')^2	(y'-L(x'))W(x')	(y'-L(x'))x'W(x')
x'_1	2.00	2.50	y'_1	2.00	4.00	8.00	16.00	1	2
x'_2	3.00	2.80	y'_2	6.00	36.00	108.00	324.00	-1.2	-3.6
x'_3	4.00	3.00	y'_3	12.00	144.00	576.00	2304.00	-12	-48
x'_4	5.00	4.70	y'_4	20.00	400.00	2000.00	10000.00	-6	-30
x'_5	6.00	6.20	y'_5	30.00	900.00	5400.00	32400.00	6	36
			SUM	70.00	1484.00	8092.00	45044.00	-12.20	-43.60

The system to solve is:

$$8092.00a_0 + 1484.00a_1 = -12.20$$
$$45044.00a_0 + 8092.00a_1 = -43.60$$

from where we conclude $(a_0, a_1) = (0.024926, -0.144139)$.
The results in graphical form



References

[1] Stancho Pavlov, Adnan Sharaf Generalized Lagrange's Interpolation Polynomial , 2014.
http://ek.roncho.net/Science/GenInterpolation/GenInterpolation.pdf