Classical least squares

Consider a discrete points t_i and a corresponding observations b_i:

$$D = \{(t_1, b_1), (t_2, b_2), \dots, (t_m, b_m)\}$$

Based on the observations we need to estimate or predicate at $\hat{t}$, that are beyond the observation t_i.

A standard approach is to find $y = f(t)$ that closely fits the points D so that we can estimate any non-observation $\hat{t}$ with the value $\hat{y} = f(\hat{t})$.

We will start with a stright line, i.e $y = \alpha + \beta t$ and we need to determine the coffi‍cients α & β, in the sense that the sum of the squares of the vertical errors $\varepsilon_1, \dots, \varepsilon_m$, is minimal. That is

at $(t_i, b_i) \Rightarrow f(t_i) = \alpha + \beta t_i$

$$\Rightarrow \varepsilon_i = \left| f(t_i) - b_i \right| = \left| \alpha + \beta t_i - b_i \right|$$

So our objective is to find α and β s.th.

$$\sum_{i=1}^{m} \varepsilon_i^2 = \sum_{i=1}^{m} (\alpha + \beta t_i - b_i)^2 \quad \text{is minimal}$$

From Calculus α & β can be found by:

$$0 = \frac{\partial \left[\sum_{i=1}^{m} (\alpha + \beta t_i - b_i)^2 \right]}{\partial \alpha} = 2 \sum_{i=1}^{m} (\alpha + \beta t_i - b_i)$$

$$0 = \frac{\partial \left[\sum_{i=1}^{m} (\alpha + \beta t_i - b_i)^2 \right]}{\partial \beta} = 2 \sum_{i=1}^{m} (\alpha + \beta t_i - b_i) t_i$$

$$\Rightarrow \left(\sum_{i=1} 1\right)\alpha + \left(\sum_{i=1} t_i\right)\beta = \sum_{i=1} b_i$$

$$\text{\&} \quad \left(\sum_{i=1}^{m} t_i\right)\alpha + \left(\sum_{i=1}^{m} t_i^2\right)\beta = \sum t_i b_i$$

let

$$A = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \quad \text{and} \quad x = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

so we have $A^T A x = A^T b \quad (Ax = b)$.

If t_i are distinct then $\text{Rank}(A) = 2$ and insures that the normal equations have a unique solution given by:

$$x = (A^T A)^{-1} A^T b$$

and the total sum of squares of the error

$$\sum_{i=1}^{m} \varepsilon_i^2 = (Ax - b)^T (Ax - b)$$

<u>Example:</u> A small company has been in business for four years and has recorded annual sales (in tens of thousands of dollars) as follows:

Year	1	2	3	4
Sales	23	27	30	34

Predict the sales for any future year if this trend continues.

if
$$A=\begin{pmatrix} 1 & 1 & -10 \\ 1 & 1 & -5 \\ 1 & 1 & 0 \\ 1 & 2 & -10 \\ 1 & 2 & -5 \\ 1 & 2 & 0 \\ 1 & 3 & -10 \\ 1 & 3 & -5 \\ 1 & 3 & 0 \end{pmatrix}, \quad X=\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}, \quad b=\begin{pmatrix} 0.15 \\ 0.18 \\ 0.20 \\ 0.17 \\ 0.19 \\ 0.22 \\ 0.20 \\ 0.23 \\ 0.25 \end{pmatrix}$$

$\Rightarrow Ax=b$

The Gauss–Markov thm states under certain reasonable assumptions concerning ε, the best estimates for α_i's are obtained by the least square estimate.

We can verify that $b \notin R(A) \Rightarrow Ax=b$ inconsistent $\Rightarrow$ we can't determine exact values for α_0, α_1 & α_2. the best we can do is determine least square estimates for the α_i's by solving $A^TAx=A^Tb$,

$$\Rightarrow \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} 0.174 \\ 0.025 \\ 0.005 \end{pmatrix} \Rightarrow y = .174 + .025t_1 + .005t_2$$

For example the mean weight loss of a pint of ice-cream that stored for nine weeks at $-35°F$ is estimated

$$\hat{y} = .174 + 0.025(9) + 0.005(-35) = 0.224 \text{ grams.}$$

Least Squares Curve Fitting

Find a polynomial

$$p(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_n t^{n-1}$$

that pass through $D = \{(t_1, b_1), \ldots, (t_m, b_m)\}$

where t_i's are distinct numbers and $n \leq m$.

In least square sense our objective fn. is to minimize

$$\sum_{i=1}^{m} \varepsilon_i^2 = \sum_{i=1}^{m} (p(t_i) - b_i)^2 = (Ax - b)^T (Ax - b),$$

where
$$A = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ \vdots & & & & \\ 1 & t_m & t_m^2 & \cdots & t_m^{n-1} \end{pmatrix}, \quad x = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{n-1} \end{pmatrix} \; \& \; b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

$$\Rightarrow \quad Ax = b$$

this least squares polynomial is unique because $A_{m\times n}$ is the Vandermonde matrix with $n \leq m$, so $\text{rank}(A) = n$ and $Ax = b$ has a unique least square sol. $x = (A^T A)^{-1} A^T b$.

<u>Example:</u> A missile is fired from enemy territory and its position in flight is observed by radar tracking devices at the following positions:

Position down range (miles)	0	250	500	750	1000
Height (miles)	0	8	15	19	20

Suppose that the missiles are programed to follow a parabolic flight path. Predict how far down range the missile will land?

Solution: $f(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$,

$$\Rightarrow A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & .25 & .0625 \\ 1 & .5 & .25 \\ 1 & .75 & .5625 \\ 1 & 1 & 1 \end{pmatrix}, \quad x = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} \ \& \ b = \begin{pmatrix} 0 \\ .008 \\ 0.015 \\ 0.019 \\ 0.02 \end{pmatrix}$$

$$\Rightarrow \quad x = \begin{pmatrix} -2.286 \times 10^{-4} \\ 3.983 \times 10^{-2} \\ -1.943 \times 10^{-2} \end{pmatrix}$$

$\Rightarrow$ least squares parabola fit is

$$y = -0.000286 + 0.3983t - 0.1943t^2.$$

the missile will land at $t = 0.005755$ and $t = 2.044$

$\Rightarrow$ the missile will land 2044 miles down range.

the sum of squares of the errors $\sum_{i=1}^{5} \varepsilon_i^2 = 4.571 \times 10^{-7}$.

Let $A = [a_{ij}(t)]$, the derivative of A w.r.t. t is defined as:

$$\frac{dA}{dt} = \left[\frac{da_{ij}}{dt}\right]$$

then $$\frac{d(AB)}{dt} = \frac{dA}{dt}B + A\frac{dB}{dt}$$

<u>Proof:</u> $$[AB]_{ij} = \frac{d\left(\sum_k (a_{ik} b_{kj})\right)}{dt}$$

$$= \sum_k \left(\frac{da_{ik}}{dt} b_{kj} + a_{ik} \frac{db_{kj}}{dt} \right) = \sum_k \frac{da_{ik}}{dt} b_{kj} + \sum a_{ik} \frac{db_{kj}}{dt}$$

$$= \left[\frac{dA}{dt}B\right]_{ij} + \left[A\frac{dB}{dt}\right]_{ij} = \left[\frac{dA}{dt}B + A\frac{dB}{dt}\right]_{ij}$$

Linear Independence and matrices

Let $A_{n\times n}$ matrix. Then the following is equivalent to say that the columns of A form a linearly independent set

(i) $N(A) = \{0\}$

(ii) $\text{rank}(A) = n$

<u>Proof</u> The columns of A are linearly indep if $\exists \alpha$'s

$$0 = \alpha_1 A_{*1} + \alpha_2 A_{*2} + \cdots + \alpha_n A_{*n} = (A_1 \cdots A_n)\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_n\end{pmatrix}$$

$\Rightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$ which is equivalent to say that $N(A) = 0$ is equivalent to $\text{rank}(A) = n$.

Vandermonde matrix

the Vandermonde matrix has the form:

$$V_{m\times n} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & & & \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{pmatrix}$$

in which $x_i \neq x_j$ for all $i \neq j$. The columns in V constitute a linearly independent set whenever $n \leq m$.

Proof: V form linearly independent set iff $N(A) = \{0\}$.

If

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & & & \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{n-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \longrightarrow (1)$$

$\Rightarrow \alpha_0 + x_i \alpha_1 + \cdots + x_i^{n-1}\alpha_{n-1} = 0$ for each $i = 1, \ldots, m$

$\Rightarrow P(x) = \alpha_0 + \alpha_1 x + \cdots + x^{n-1}\alpha_{n-1}$

has m distinct roots namely x_i's. However $\deg(P(x)) \leq n-1$

$\Rightarrow$ if $P(x)$ is not the zero polynomial, then $P(x)$ can have at most $n-1$ distinct roots. Therefore (1) holds iff $\alpha_i = 0 \ \forall i$ and thus insures that the columns of V form a linearly independent set.