DATA FITTING

1. Data fitting

2. Ordinary Least Square Method (OLS)

Outlines

Data fitting

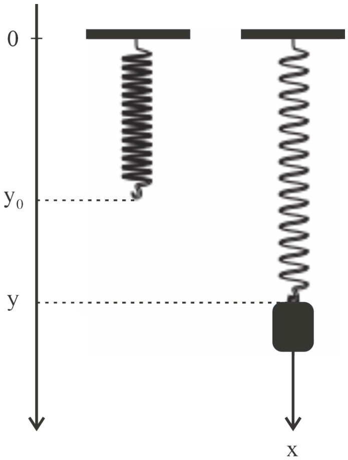
A common problem in experimental work is to obtain a mathematical relationship between a quantity y, called **dependent variable**, and other quantities $x_1, x_2, \cdots, x_n$, called **independent variables**. The relationship will be taken of the form of a function of n-variables, $y = f(x_1, x_2, \cdots, x_n)$.

First, we consider the case $n = 1$ where the relationship between dependent and independent variables is a function of one variable, $y = f(x)$, which belongs to some class of functions.

Example 1. (a) In physics, Hooke's law gives a mathematical relationship between the length y of a uniform spring (the dependent variable) and the force x applied to it (an independent variable). Precisely, the relationship in Hooke's law is of the form

$$y = a + bx, \tag{1}$$

where $y_0 = a$ is the length of the unstretched spring, and b is called the **spring constant**.



(b) The Newton's second law of motion states that a body near the Earth's surface falls vertically downward in accordance with the relationship

$$x = x_0 + v_0 t + \frac{1}{2} g t^2, \tag{2}$$

where x is the height of the body with respect to the Earth surface at time t, for example, x_0 is the height of the body at time $t = 0$ (the initial height), v_0 is the (initial) velocity of the body at time $t = 0$, and g is the acceleration of gravity at the Earth's surface.
In example 1 (a), we consider the class of linear functions in x, and in (b), the class of quadratic functions.
In order to choose a function $y = f(x)$ in the class, i.e. to find the spring constant b in (1), the parameters x_0, v_0, and g in (2), we have to collect data for the values of the independent variable and the corresponding dependent variable. For example, by letting different value for the force x and measuring the corresponding length y of the spring, we may have the following data for (x, y),

x (kg)	1	2	3	4
y (cm)	6.3	6.7	7	7.4

and with example 1 (b), we may have the following data

t (s)	0.1	0.2	0.3	0.4
x (m)	10.1	10.4	10.7	11.2

Outlines

Ordinary Least Square Method (OLS)

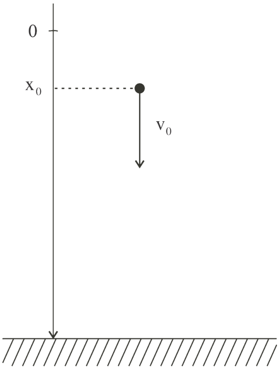
Given data between independent variable x and dependent variable y,

x	x_1	x_2	$\cdots$	x_N
y	y_1	y_2	$\cdots$	y_N

and a function $y = f(x)$, the OLS measures the **fitness** of this function to the observed data by

$$RSS(f) = \sum_{i=1}^{N} (y_i - f(x_i))^2, \tag{3}$$

where $y_i - f(x_i)$, $i = 1, 2, \cdots, N$, are called the **residuals** of the model $y = f(x)$ with respect to the data, and RSS stands for Residuals Sum of Squares.



Given basic functions $f_1(x), f_2(x), \cdots, f_k(x)$, we consider the class of linear combinations of these basic functions,

$$f(x) = b_1 f_1(x) + b_2 f_2(x) + \cdots + b_k f_k(x),$$

with $b_1, b_2, \cdots, b_k \in \mathbb{R}$, i.e. each function in this class is identified by a vector $(b_1, b_2, \cdots, b_k) \in \mathbb{R}^k$.

(3) is rewritten as

$$RSS\left(b_1, b_2, \cdots, b_k\right) = \sum_{i=1}^{N} \left[y_i - \left(b_1 f_1\left(x_i\right) + b_2 f_2\left(x_i\right) + \cdots + b_k f_k\left(x_i\right)\right)\right]^2 \tag{4}$$

Example 1 continued. (a) The class of linear functions (1) is generated by two basic functions, $f_1(x) = 1$ and $f_2(x) = x$, i.e.,

$$f(x) = a + bx = af_1(x) + bf_2(x)$$

which is a linear combination of the two basic functions. Under the data, we have

x	1	2	3	4
y	6.3	6.7	7	7.4
$f(x)$	$a+b$	$a+2b$	$a+3b$	$a+4b$

and

$$RSS(f) = RSS(a, b) = (a + b - 6.3)^2 + (a + 2b - 6.7)^2 + (a + 3b - 7)^2$$

(b) The class of quadratic functions (2) is generated by three basic functions, $f_1(x) = 1$, $f_2(x) = x$, and $f_3(x) = x^2$, i.e.,

$$f(x) = a + bx + cx^2 = af_1(x) + bf_2(x) + cf_3(x)$$

and we have

x	0.1	0.2	0.3	0.4
y	10.1	10.4	10.7	11.2
$f(x)$	$a + b \cdot 0.1 + c \cdot 0.1^2$	$a + b \cdot 0.2 + c \cdot 0.2^2$	$a + b \cdot 0.3 + c \cdot 0.3^2$	$a + b \cdot 0.4 + c \cdot 0.4^2$

and

$$RSS(f) = RSS(a, b, c) = \ (a + b \cdot 0.1 + c \cdot 0.1^2 - 10.1)^2 + (a + b \cdot 0$$
$$+ (a + b \cdot 0.3 + c \cdot 0.3^2 - 10.7)^2 + (a + b \cdot 0.4 + c \cdot 0.4^2 - 11.2)^2 .$$

In the ordinary least squares method, the function f in the class is chosen with minimum residual sum of squares,

$$RSS(f) \to \min,$$

i.e., the parameter vector $(b_1, b_2, \cdots, b_k)$, which identifies f, is chosen as the global minimum of the function $RSS(b_1, b_2, \cdots, b_k)$ in (4). By letting

$$\mathbf{X} = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_k(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_k(x_2) \\ \vdots & \vdots & & \vdots \\ f_1(x_N) & f_2(x_N) & \cdots & f_k(x_N) \end{bmatrix} \in \mathbb{R}^{N \times k}, \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^{N \times 1},$$

we get

$$\mathbf{Y}-\mathbf{X}\mathbf{b}=\begin{bmatrix} y_1-\left(b_1 f_1\left(x_1\right)+\cdots+b_k f_k\left(x_1\right)\right) \\ \vdots \\ y_i-\left(b_1 f_1\left(x_i\right)+\cdots+b_k f_k\left(x_i\right)\right) \\ \vdots \\ y_N-\left(b_1 f_1\left(x_N\right)+\cdots+b_k f_k\left(x_N\right)\right) \end{bmatrix}$$

and then (4) becomes

$$RSS\left(\mathbf{b}\right)=\left\|\mathbf{Y}-\mathbf{X}\mathbf{b}\right\|^2. \tag{5}$$

Using the result from convex optimization (Theorem 16), with $k\leq N$ and $rank\left(X\right)=k$, i.e., X is of full rank, the unique solution of (5) is given by

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}=\arg\min_{\mathbf{b}\in\mathbb{R}^k} RSS\left(\mathbf{b}\right). \tag{6}$$

Example 1 continued. (a) Under the data

x	1	2	3	4
y	6.3	6.7	7	7.4

we have

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 6.3 \\ 6.7 \\ 7.0 \\ 7.4 \end{bmatrix},$$

and the best function in the class of linear functions is identified by the parameter vector

$$\mathbf{b} = \begin{bmatrix} a \\ b \end{bmatrix} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}^{-1} \begin{bmatrix} 27.4 \\ 70.3 \end{bmatrix} = \begin{bmatrix} 5.95 \\ 0.36 \end{bmatrix}.$$

Therefore, we get the relation between y and x as

$$y = 5.95 + 0.36x$$

and we conclude that the original length of the spring is 5.95 cm, and the spring constant is 0.36.

(b) With the data

t	0.1	0.2	0.3	0.4
x	10.1	10.4	10.7	11.2

we have

$$\mathbf{X} = \begin{bmatrix} 1 & 0.1 & 0.1^2 \\ 1 & 0.2 & 0.2^2 \\ 1 & 0.3 & 0.3^2 \\ 1 & 0.4 & 0.4^2 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 10.1 \\ 10.4 \\ 10.7 \\ 11.2 \end{bmatrix},$$

and the best function in the class of quadratic functions is characterized by

$$\mathbf{b} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 4 & 1 & 0.3 \\ 1 & 0.3 & 0.1 \\ 0.3 & 0.1 & 0.0354 \end{bmatrix}^{-1} \begin{bmatrix} 42.4 \\ 10.78 \\ 3.272 \end{bmatrix} =$$

i.e., the best relationship between x and t in the class of quadratic functions is

$$x = 9.95 + 1.1t + 5t^2.$$

We conclude that the body's height at $t = 0$ is 9.95 m, the initial velocity is 1.1 m/s, and the acceleration of gravity at the Earth's surface is 10 m/s^2.

Generalization to the case $n > 1$. Each basic function is a function of n-variables, $f_i\left(x_1, x_2, \cdots, x_n\right)$, $i = 1, 2, \cdots, k$, and we consider the class of linear combination of these basic functions,

$$f\left(\mathbf{x}\right) = b_1 f_1\left(\mathbf{x}\right) + b_2 f_2\left(\mathbf{x}\right) + \cdots + b_k f_k\left(\mathbf{x}\right), \mathbf{x} = \left(x_1, x_2, \cdots, x_n\right).$$

The observed data have to be of the form

$\mathbf{x} = (x_1, \ldots, x_n)$	$\mathbf{x}_1 = (x_{1,1}, x_{2,1}, \ldots, x_{n,1})$	$\mathbf{x}_2 = (x_{1,2}, x_{2,2}, \ldots, x_{n,2})$	$\cdots$	$\mathbf{x}_N = (x_{1,N}, x_{2,N}, \ldots$
y	y_1	y_2	$\cdots$	y_N

In the OLS algorithm, we try to find the global minimum of the residual sum of squares,

$$RSS(\mathbf{b}) = RSS(b_1, b_2, \cdots, b_k)$$
$$= \sum_{i=1}^{N} [y_i - (b_1 f_1(\mathbf{x}_i) + b_2 f_2(\mathbf{x}_i) + \cdots + b_k f_k(\mathbf{x}_i))]^2$$

As in the case $n = 1$, by letting

$$\mathbf{X} = \begin{bmatrix} f_1(\mathbf{x}_1) & f_2(\mathbf{x}_1) & \cdots & f_k(\mathbf{x}_1) \\ f_1(\mathbf{x}_2) & f_2(\mathbf{x}_2) & \cdots & f_k(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ f_1(\mathbf{x}_N) & f_2(\mathbf{x}_N) & \cdots & f_k(\mathbf{x}_N) \end{bmatrix} \in \mathbb{R}^{N \times k}$$

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^{N\times 1}, \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} \in \mathbb{R}^{k\times 1}$$

we get

$$RSS\left(\mathbf{b}\right) = \left\|\mathbf{Y} - \mathbf{X}\mathbf{b}\right\|^2,$$

and the unique solution of the corresponding optimization problem, when $k \leq N$ and $rank\left(X\right) = k$, is also given by

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} = \underset{\mathbf{b}\in\mathbb{R}^k}{\arg\min} RSS\left(\mathbf{b}\right). \tag{7}$$

Example 2. Suppose that we try to find the relationship between the selling price of a house, y, with the area x_1 and the number of bedrooms x_2 of this house from the following observations

x_1 (m^2)	x_2	y (billion VND)
50	1	1.5
60	2	3.2
70	3	4.5
80	3	5.8
100	3	6.5

Model 1. Consider the class of linear functions $f(x_1, x_2) = a + bx_1 + cx_2$, which are linear combinations of three basic functions, $f_1(x_1, x_2) = 1$, $f_2(x_1, x_2) = x_1$, and $f_3(x_1, x_2) = x_2$. Let

$$\mathbf{X} = \begin{bmatrix} f_1(50,1) & f_2(50,1) & f_3(50,1) \\ f_1(60,2) & f_2(60,2) & f_3(60,2) \\ f_1(70,3) & f_2(70,3) & f_3(70,3) \\ f_1(80,3) & f_2(80,3) & f_3(80,3) \\ f_1(100,3) & f_2(100,3) & f_3(100,3) \end{bmatrix} = \begin{bmatrix} 1 & 50 & 1 \\ 1 & 60 & 2 \\ 1 & 70 & 3 \\ 1 & 80 & 3 \\ 1 & 100 & 3 \end{bmatrix}.$$

$$\mathbf{Y} = \begin{bmatrix} 1.5 \\ 3.2 \\ 4.5 \\ 5.8 \\ 6.5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

The best function in this class is given by

$$\mathbf{b} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 5 & 360 & 12 \\ 360 & 27400 & 920 \\ 12 & 920 & 32 \end{bmatrix}^{-1} \begin{bmatrix} 21.5 \\ 1696 \\ 58.3 \end{bmatrix}$$

$$= \begin{bmatrix} -2.57 \\ 0.0615 \\ 1.0175 \end{bmatrix}$$

, i.e., $f\left(x_1, x_2\right) = -2.57 + 0.0615x_1 + 1.0175x_2$.

Model 2 : Consider the class of polynomial functions $f(x_1, x_2) = a + bx_1 + cx_1^2 + dx_2$ with four basic functions, $f_1(x_1, x_2) = 1,\ f_2(x_1, x_2) = x_1,\ f_3(x_1, x_2) = x_1^2$, and $f_4(x_1, x_2) = x_2$. Let

$$\mathbf{X} = \begin{bmatrix} f_1(50,1) & f_2(50,1) & f_3(50,1) & f_4(50,1) \\ f_1(60,2) & f_2(60,2) & f_3(60,2) & f_4(60,2) \\ f_1(70,3) & f_2(70,3) & f_3(70,3) & f_4(70,3) \\ f_1(80,3) & f_2(80,3) & f_3(80,3) & f_4(80,3) \\ f_1(100,3) & f_2(100,3) & f_3(100,3) & f_4(100,3) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 50 & 50^2 & 1 \\ 1 & 60 & 60^2 & 2 \\ 1 & 70 & 70^2 & 3 \\ 1 & 80 & 80^2 & 3 \\ 1 & 100 & 100^2 & 3 \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} 1.5 \\ 3.2 \\ 4.5 \\ 5.8 \\ 6.5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}.$$

The best function in this model is given by

$$\mathbf{b} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}$$

$$= \begin{bmatrix} 5 & 360 & 27400 & 12 \\ 360 & 27400 & 2196000 & 920 \\ 27400 & 2196000 & 184180000 & 73600 \\ 12 & 920 & 73600 & 32 \end{bmatrix}^{-1} \begin{bmatrix} 21.5 \\ 1696 \\ 139440 \\ 58.3 \end{bmatrix}$$

$$= \begin{bmatrix} -19.25 \\ 0.5775 \\ -0.003 \\ -0.6625 \end{bmatrix},$$

i.e., $f(x_1, x_2) = -19.25 + 0.5775x_1 - 0.003x_1^2 - 0.6625x_2$.

Model 3 : linear functions with "interaction term"
$f(x_1, x_2) = a + bx_1 + cx_2 + dx_1x_2$ with four basic functions,
$f_1(x_1, x_2) = 1$, $f_2(x_1, x_2) = x_1$, $f_3(x_1, x_2) = x_2$, and $f_4(x_1, x_2) = x_1x_2$.
Let

$$\mathbf{X} = \begin{bmatrix} f_1(50,1) & f_2(50,1) & f_3(50,1) & f_4(50,1) \\ f_1(60,2) & f_2(60,2) & f_3(60,2) & f_4(60,2) \\ f_1(70,3) & f_2(70,3) & f_3(70,3) & f_4(70,3) \\ f_1(80,3) & f_2(80,3) & f_3(80,3) & f_4(80,3) \\ f_1(100,3) & f_2(100,3) & f_3(100,3) & f_4(100,3) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 50 & 1 & 1\cdot 50 \\ 1 & 60 & 2 & 2\cdot 60 \\ 1 & 70 & 3 & 3\cdot 70 \\ 1 & 80 & 3 & 3\cdot 80 \\ 1 & 100 & 3 & 3\cdot 100 \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} 1.5 \\ 3.2 \\ 4.5 \\ 5.8 \\ 6.5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}.$$

The best function in this model is given by

$$\mathbf{b} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}$$

$$= \begin{bmatrix} 5 & 360 & 12 & 920 \\ 360 & 27400 & 920 & 73600 \\ 12 & 920 & 32 & 2540 \\ 920 & 73600 & 2540 & 208600 \end{bmatrix}^{-1} \begin{bmatrix} 21.5 \\ 1696 \\ 58.3 \\ 4746 \end{bmatrix} = \begin{bmatrix} -3.58571 \\ 0.08143 \\ 1.33571 \\ -0.00643 \end{bmatrix},$$

i.e., $f\left(x_1, x_2\right) = -3.58571 + 0.08143x_1 + 1.33571x_2 - 0.00643x_1x_2$.