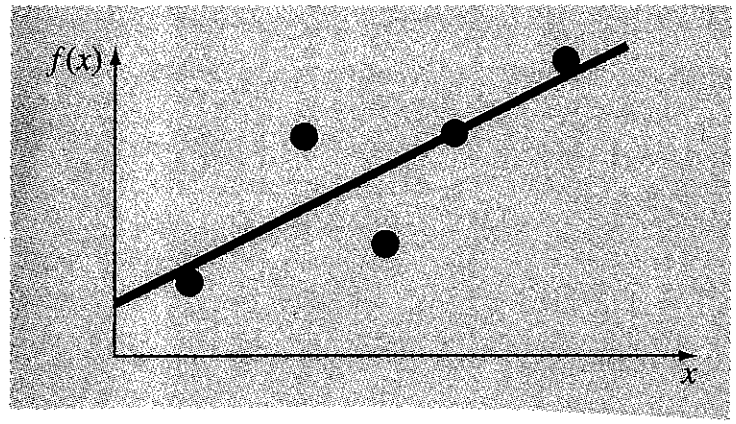
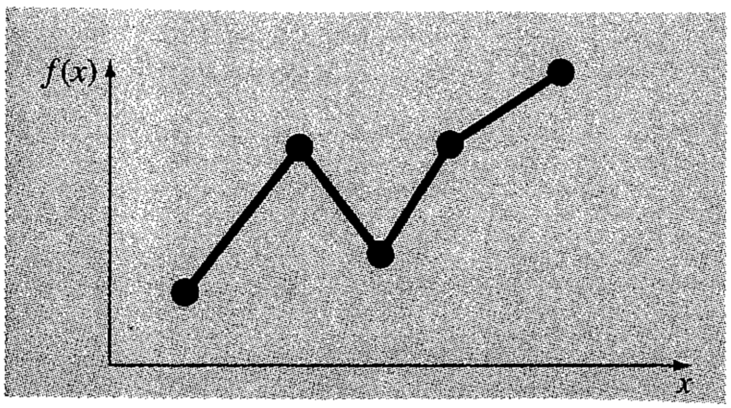
PART 4 CURVE FITTING

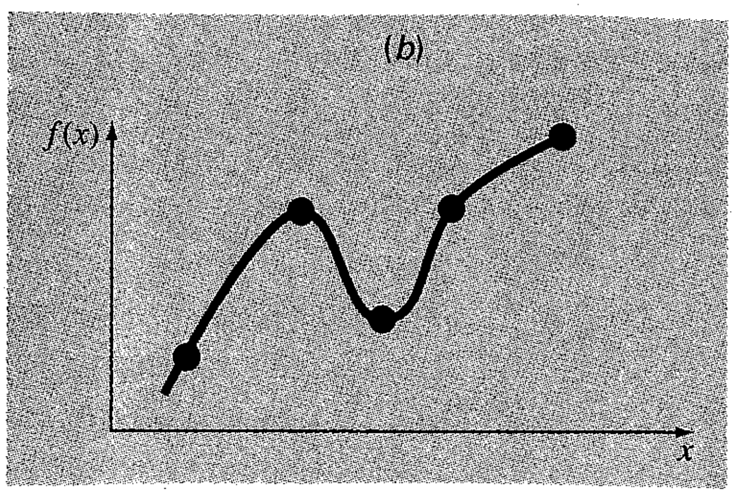
Least square



linear interpolation



Curvilinear Interpolation



Applications

1. Trend analysis
2. Hypothesis testing : compare predicted value and observed value

13. Linear Regression

Motivation:

- overdetermined system
- best straight line

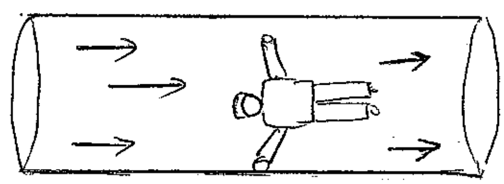
Drag Coefficient of Bungee Jumper

$$F_u = \underset{\sim}{C_d} \cdot v^2$$

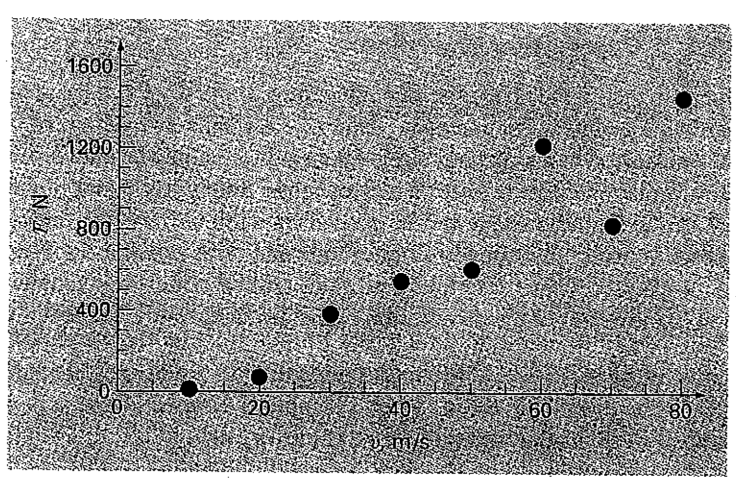
$$\therefore C_d = \frac{F_u}{v^2}$$

Fluid Mechanics.

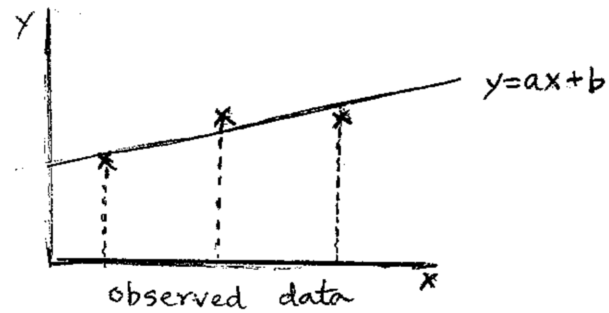
Wind Tunnel Experiment.



v	10	20	30	40	50	60	70	80
F_u	25	70	380	550	610	1220	830	1450



Least-Square Method

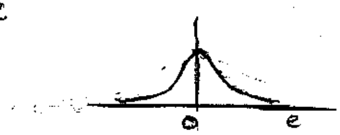


Note overdetermined → least square method.

Goal: minimize the sum of the errors.

$$\text{Min} \sum_{k=1}^{m} |\overbrace{a x_k + b - y_k}^{\text{error}}| \quad : l_1 \text{ approximation} \rightarrow \text{Linear Programming}$$

- $$\phi(ab) = \sum_{k=1}^{m} (\underbrace{a x_k + b - y_k}_{e})^2 \quad : l_2 \text{ approximation (if error follows normal distribution)}$$



Make $\phi(ab)$ minimum $\Rightarrow \frac{\partial \phi}{\partial a} = 0, \frac{\partial \phi}{\partial b} = 0$.

$$\begin{cases} \sum_{k=1}^{m} 2(a x_k + b - y_k) x_k = 0 \\ \sum_{k=1}^{m} 2(a x_k + b - y_k) = 0 \end{cases} \Rightarrow \begin{cases} (\Sigma x_k^2) a + (\Sigma x_k) b = \Sigma y_k x_k & \text{--- ①} \\ (\Sigma x_k) a + m b = \Sigma y_k & \text{--- ②} \end{cases}$$

normal equation

$$\text{①} \times m \qquad (\Sigma x_k^2) a m + (\Sigma x_k) b \cdot m = m \cdot \Sigma y_k x_k$$

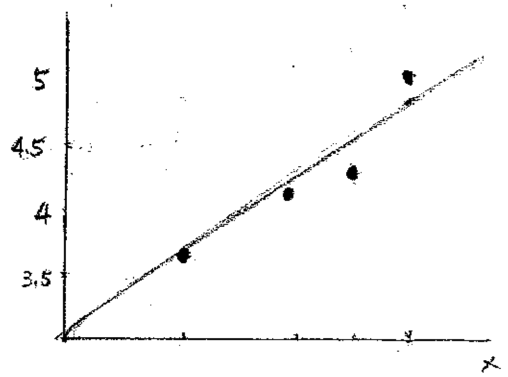
$$\text{②} \times \Sigma x_k \qquad (\Sigma x_k)^2 a + (\Sigma x_k) \cdot m b = \Sigma x_k \cdot \Sigma y_k \quad (-$$

$$\underbrace{[(\Sigma x_k^2) m - (\Sigma x_k)^2]}_{d} a = m \cdot \Sigma x_k y_k - \Sigma x_k \cdot \Sigma y_k$$

$$a = \frac{1}{d}(m \cdot \Sigma x_k y_k - \Sigma x_k \cdot \Sigma y_k)$$

$$b = \frac{1}{d}(\Sigma x_k^2 \cdot \Sigma y_k - \Sigma x_k \cdot \Sigma x_k y_k) \text{ or } \bar{y} - a \cdot \bar{x}$$

Example 1. Polynomial Case: $y = \underline{a}x + \underline{b}$



X	1.0	2.0	2.5	3.0
Y	3.7	4.1	4.3	5.0

$m = 4$

$\Sigma X = 8.5$

$\Sigma Y = 17.1$

$\Sigma XY = 3.7 + 8.2 + 10.75 + 15 = 37.65$

$\Sigma X^2 = 1 + 4 + 6.25 + 9 = 20.25$

$$\therefore \begin{cases} 20.25a + 8.5b = 37.65 \\ 8.5a + 4b = 17.1 \end{cases} \Rightarrow \begin{array}{l} a = 0.6, \\ b = 3.0 \end{array}$$

$\underline{y = 0.6x + 3}$

Check: $\phi(a, b) = \sum_{k=1}^{4} (\overset{0.6}{a}x_k + \overset{3}{b} - y_k)^2$

$= (0.6 \times 1 + 3 - 3.7)^2 + (0.6 \times 2 + 3 - 4.1)^2 + (0.6 \times 2.5 + 3 - 4.3)^2 + (0.6 \times 3 + 3 - 5.0)^2$

$= 0.01 + 0.01 + 0.04 + 0.04$

$= 0.1$

Overdetermined Linear Systems - Gauss

"Given an $m \times n$ matrix A, where $m \geq n$, and an $m \times 1$ vector b, find a vector x such that Ax is the "best" approximation to b".

Minimize $\|r\|_2 = (r^T r)^{\frac{1}{2}}$, $r = b - Ax$

$$m \overset{n}{\boxed{A}} \overset{x}{[\;]} = \overset{b}{[\;]}$$

Normal Equations

$$\phi(x_1, x_2 \cdots x_n) = \sum_{k=1}^{m} \left(\sum_{j=1}^{n} a_{kj} x_j - b_k \right)^2$$

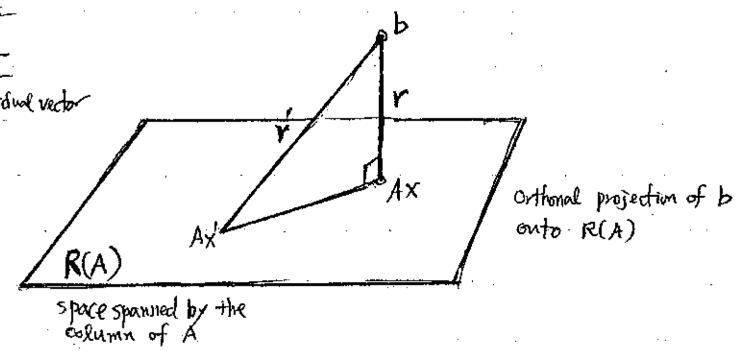
Take partial derivatives with respect to x_i and set them equal to zero,

$$\sum_{j=1}^{n} \left(\sum_{k=1}^{m} a_{ki} a_{kj} \right) x_j = \sum_{k=1}^{m} b_k \cdot a_{ki} \qquad 1 \leq i \leq n$$

In other words, $\boxed{(A^T A)\, x = A^T b}$ <u>normal equations</u> (symmetric $n \times n$ system)

Graph Interpretation

$b = Ax + \underset{\text{residual vector}}{r}$



To minimize $\|r\|_2$, we solve

$$(A^T A)\, x = A^T b$$

$$x = \underbrace{(A^T A)^{-1} A^T}_{A^I} \cdot b$$

: pseudo inverse
or Moore-Penrose generalized inverse.

Ex. 13.2

10	20	30	40	50	60	70	80
25	70	380	550	610	1220	830	1450

i	x_i	y_i	x_i^2	$x_i y_i$
1	10	25	100	250
2	20	70	400	1,400
3	30	380	900	11,400
4	40	550	1,600	22,000
5	50	610	2,500	30,500
6	60	1,220	3,600	73,200
7	70	830	4,900	58,100
8	80	1,450	6,400	116,000
Σ	360	5,135	20,400	312,850

[Method 1]

$$\bar{x} = \frac{360}{8} = 45 \qquad \bar{y} = \frac{5,135}{8} = 641.875$$

$$a = \frac{8(312,850) - 360(5,135)}{8(20,400) - (360)^2} = 19.47024$$

$$b = 641.875 - 19.47024(45) = -234.2857$$

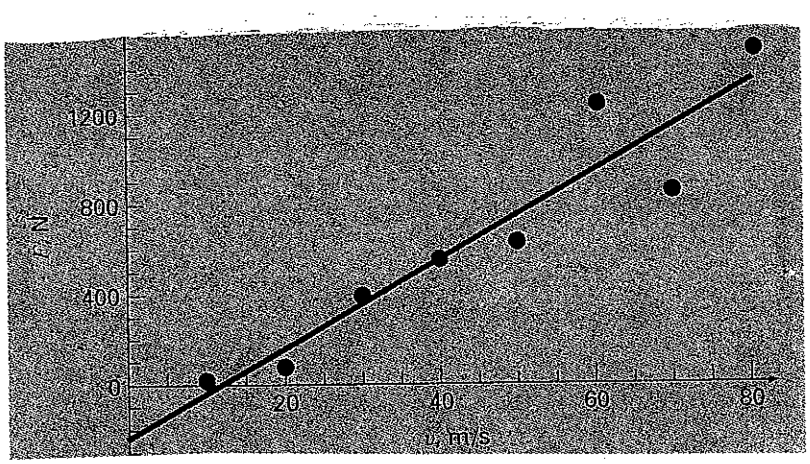
[Method 2] Normal equation

$$\begin{cases} 20400\,a + 360\,b = 312850 & \text{— ①} \\ 360\,a + 8\,b = 5135 & \text{— ②} \end{cases}$$

① − 45×②

$$4200\,a = 81775 \Rightarrow a = 19.47024$$

From ②, $b = \frac{1}{8}(5135 - 360(19.47024)) \Rightarrow b = -234.2858$



Linearization of Nonlinear Relations

(1) Exponential $\quad y = \alpha e^{\beta x}$

- population growth

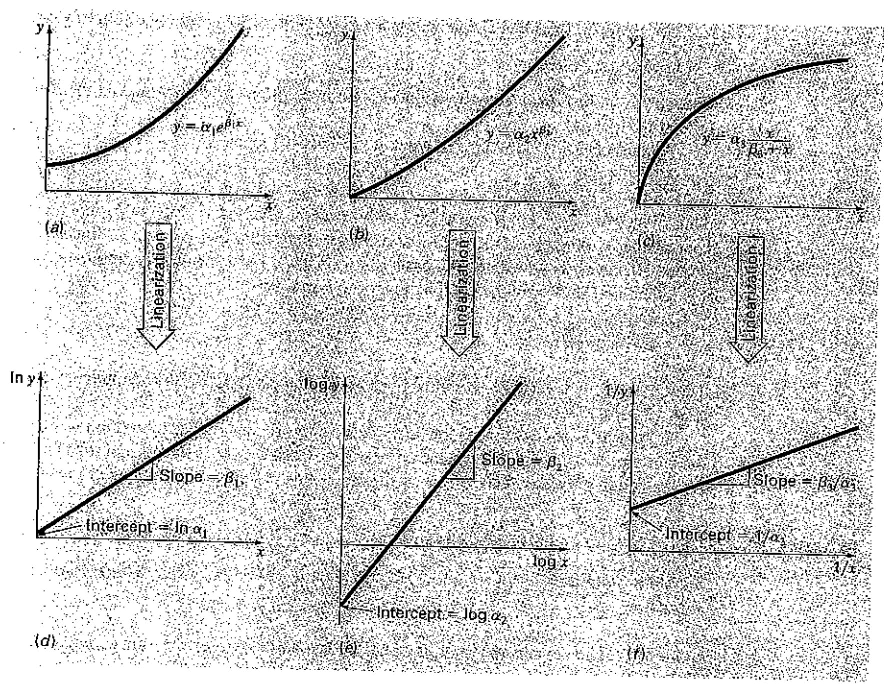
$$\underline{\ln y} = \ln \alpha + \beta \underline{x}$$

(2) Power equation $\quad y = \alpha x^{\beta}$

$$\underline{\log y} = \log \alpha + \beta \cdot \underline{\log x}$$

(3) Saturation-growth-ratio $\quad y = \alpha \dfrac{x}{\beta + x}$

$$\underline{\frac{1}{y}} = \frac{1}{\alpha} + \frac{\beta}{\alpha} \cdot \underline{\frac{1}{x}}$$



Example [8].

$$A = \begin{bmatrix} 2 & 3 \\ 1 & -4 \\ 2 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ -9 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & 2 \\ 3 & -4 & -1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & -4 \\ 2 & -1 \end{bmatrix} = \overset{A^TA}{\begin{bmatrix} 9 & 0 \\ 0 & 26 \end{bmatrix}}, \qquad \begin{bmatrix} 2 & 1 & 2 \\ 3 & -4 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ -9 \\ -1 \end{bmatrix} = \overset{A^Tb}{\begin{bmatrix} -9 \\ 40 \end{bmatrix}}$$

Normal equation:

$$\begin{bmatrix} 9 & 0 \\ 0 & 26 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -9 \\ 40 \end{bmatrix} \Rightarrow x = \begin{bmatrix} -1 \\ \frac{20}{13} \end{bmatrix}$$

Better Algorithm: $Ax = b$

modified Gram-Schmidt method

(1) $A = QR$ (QR-factorization)

Q : orthonormal columns. (ⅰ $Q^TQ = I$, $Q^T = Q^{-1}$)

R : upper-triangular matrix with $r_{ii} > 0$

(2) $y = Q^Tb$

(3) $Rx = y$