รายวิชา 09131201 ระเบียบวิธีเชิงตัวเลขทางด้านคอมพิวเตอร์ (Numerical Methods for Computers) บทที่ 4 สมการถดถอยอันดับสองน้อยที่สุด (Least Squares Regression)

ดร.วงศ์วิศรุต เขื่องสตุ่ง

สาขาวิชาคณิตศาสตร์
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

July 28, 2022

Outline

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Outline

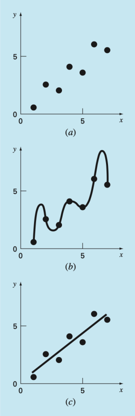
บทนำ

บทนำ

บทนำ

วิธีหาเส้นสมการที่เหมาะสมกับข้อมูล โดยเส้นสมการที่เหมาะสมนั้นไม่จำเป็นต้องผ่านทุกๆ จุดของข้อมูล อาจเป็นการพลอตข้อมูล และเขียนเส้นที่เห็นว่าดีที่สุด แต่วิธีการนี้ไม่ใช่วิธีที่ดี เพราะแต่ละครั้งจะได้เส้นสมการที่ต่างกัน วิธีหนึ่งที่จะได้เส้นสมการที่เหมาะสม คือใช้ความแตกต่างระหว่างจุดของข้อมูลเพื่อหาเส้นสมการที่เหมาะสมที่มีค่าผิดพลาดน้อยที่สุด เรียกวิธีการหาเส้นสมการที่เหมาะสมนี้ว่า สมการถดถอยอันดับสองน้อยที่สุด

บทนำ



รูปที่ 1: (a) Data exhibiting significant error. (b) Polynomial fit oscillating beyond the range of the data. (c) More satisfactory result using the least-squares fit.

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

การหาเส้นสมการถดถอยเชิงเส้นตรง เป็นระเบียบวิธีในการประมาณค่าการหาเส้นสมการที่เหมาะสมด้วยการสร้างเส้นตรงเพื่อประมาณค่าเซตของจุดหรือข้อมูล: $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ ซึ่งรูปแบบทั่วไปของสมการเส้นตรงคือ

$$y = a_0 + a_1 x + e \tag{4.1}$$

เมื่อ a_0, a_1 คือค่าคงที่ และ e คือ ความคลาดเคลื่อน

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

พิจารณาค่าคลาดเคลื่อน หรือค่าผิดพลาด คือ เป็นค่าความแตกต่างระหว่างค่าจริงและค่าประมาณ และในทางสถิติค่าคลาดเคลื่อน เรียกว่า **Residual** จากสมการ (4.2) สามารถเขียนได้ในรูป

$$e = y - a_0 - a_1 x \tag{4.2}$$

การหาเส้นสมการถดถอยเชิงเส้นตรง (Linear Regression)

การหาเส้นสมการที่เหมาะสมที่สุด คือการทำให้ผลรวมของค่าคลาดเคลื่อน (S_r) มีค่าน้อยที่สุด นั่นคือ

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \tag{4.3}$$

เมื่อ n คือจำนวนจุดข้อมูล

- ในที่นี้เรายกกำลังสองของค่าคลาดเคลื่อนเพื่อกำจัดค่าที่อาจมีเครื่องหมายเป็นลบ

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

เพื่อหาค่าผลรวมของค่าคลาดเคลื่อน (S_r) ที่มีค่าน้อยที่สุด โดยใช้หลักการการหาค่าต่ำสุด (Minimization) ดังนั้น การหาค่าต่ำสุดของ (S_r) เปรียบเทียบกับตัวไม่ทราบค่า a_0 และ a_1 จะได้

$$\frac{\partial S_r}{\partial a_0} = 0 \tag{4.4}$$

และ

$$\frac{\partial S_r}{\partial a_1} = 0 \tag{4.5}$$

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

จาก (4.3) จะได้

$$\frac{\partial S_r}{\partial a_0} = -2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i) \tag{4.6}$$

และ

$$\frac{\partial S_r}{\partial a_1} = -2\sum_{i=1}^{n}([y_i - a_0 - a_1 x_i]x_i) \tag{4.7}$$

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

จาก (4.6) จะได้

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a_0 - \sum_{i=1}^{n} a_1 x_i = 0$$

$$n a_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 = \sum_{i=1}^{n} y_i \tag{4.8}$$

จาก (4.7) จะได้

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} a_0 x_i - \sum_{i=1}^{n} a_1 x_i^2 = 0$$

$$\left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 = \sum_{i=1}^{n} x_i y_i \tag{4.9}$$

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

จาก (4.8) และ (4.9) สามารถเขียนให้อยู่ในรูปเมทริกซ์ ดังนี้

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix} \tag{4.10}$$

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

ใช้กฎของคราเมอร์ จะได้

$$a_0 = \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{4.11}$$

และ

$$a_1 = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{4.12}$$

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

ตัวอย่างที่ 4.1

จงหาสมการเส้นตรงที่เหมาะสมกับจุดข้อมูลที่กำหนดให้

i	1	2	3	4	5
x_i	2.10	6.22	7.17	10.52	13.68
y_i	2.90	3.83	5.98	5.71	7.74

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

ตัวอย่างที่ 4.2

จงหาสมการเส้นตรงที่เหมาะสมกับจุดข้อมูลที่กำหนดให้

i	1	2	3	4	5	6	7
x_i	1	2	3	4	5	6	7
y_i	0.5	2.5	2.0	4.0	3.5	6.0	5.5

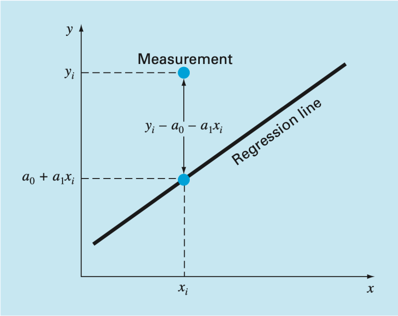
คุณภาพของเส้นสมการที่เหมาะสม (Quantification of Error of Linear Regression)

การพิจารณาคุณภาพของเส้นสมการที่เหมาะสมต้องพิจารณาสิ่งต่อไปนี้

1. ผลรวมกำลังสองของค่าความคลาดเคลื่อนรอบเส้นถดถอย $y = a_0 + a_1 x$ (Sum of Squres of Residuals about Regrssion Line):

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_1)^2 \tag{4.13}$$

คุณภาพของเส้นสมการที่เหมาะสม (Quantification of Error of Linear Regression)



รูปที่ 2: The residual in linear regression represents the vertical distance between a data point and the straight line.

คุณภาพของเส้นสมการที่เหมาะสม (Quantification of Error of Linear Regression)

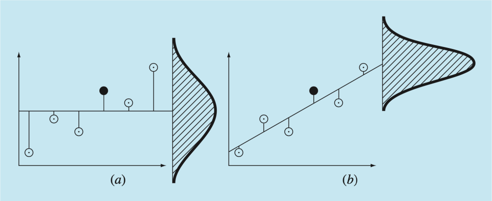
2. ค่าความเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimate):

$$S_{y/x} = \sqrt{\frac{S_r}{n-2}} \tag{4.14}$$

3. ผลรวมกำลังสองของค่าความคลาดเคลื่อนรอบค่าเฉลี่ย $\bar{y}$ (Sum of Squares of Residuals about the Mean $\bar{y}$):

$$S_t = \sum (y_i - \bar{y})^2 \tag{4.15}$$

คุณภาพของเส้นสมการที่เหมาะสม (Quantification of Error of Linear Regression)



รูปที่ 3: Regression data showing (a) the spread of the data around the mean of the dependent variable and (b) the spread of the data around the best-fit line. The reduction in the spread in going from (a) to (b), as indicated by the bell-shaped curves at the right, represents the improvement due to linear regression.

คุณภาพของเส้นสมการที่เหมาะสม (Quantification of Error of Linear Regression)

4. สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination):

$$r^2 = \frac{S_t - S_r}{S_t} \tag{4.16}$$

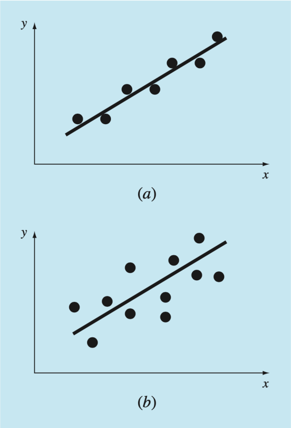
5. สัมประสิทธ์สหสัมพันธ์ (Correlation Coefficient):

$$r = \sqrt{= \frac{S_t - S_r}{S_t}} \tag{4.17}$$

หมายเหตุ

สำหรับเส้นสมการที่เหมาะสมที่สุด จะได้ $S_r = 0, r = r^2 = 1$ แสดงว่าเส้นสมการที่มีค่าต่อเนื่องที่ได้สามารถใช้แทนข้อมูลได้ 100%

คุณภาพของเส้นสมการที่เหมาะสม (Quantification of Error of Linear Regression)



รูปที่ 4: Examples of linear regression with (a) small and (b) large residual errors.

การหาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Linear Regression)

ตัวอย่างที่ 4.3

จากข้อมูลที่กำหนดให้ จงคำนวณหาสัมประสิทธิ์สหสัมพันธ์

i	1	2	3	4	5	6	7
x_i	1	2	3	4	5	6	7
y_i	0.5	2.5	2.0	4.0	3.5	6.0	5.5

Algorithm for linear regression

```
SUB Regress(x, y, n, a1, a0, syx, r2)

  sumx = 0: sumxy = 0: st = 0
  sumy = 0: sumx2 = 0: sr = 0
  DOFOR i = 1, n
    sumx = sumx + x_i
    sumy = sumy + y_i
    sumxy = sumxy + x_i*y_i
    sumx2 = sumx2 + x_i*x_i
  END DO
  xm = sumx/n
  ym = sumy/n
  a1 = (n*sumxy − sumx*sumy)/(n*sumx2 − sumx*sumx)
  a0 = ym − a1*xm
  DOFOR i = 1, n
    st = st + (y_i − ym)^2
    sr = sr + (y_i − a1*x_i − a0)^2
  END DO
  syx = (sr/(n − 2))^0.5
  r2 = (st − sr)/st

END Regress
```

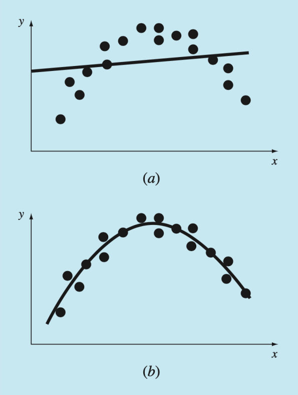
การประยุกต์หาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Application of Linear Regression)

การประยุกต์หาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Application of Linear Regression)

การประยุกต์หาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Application of Linear Regression)

ถ้าหากพิจารณาความสัมพันธ์ ระหว่างข้อมูลให้มีลักษณะไม่เชิงเส้น สามารถทำได้โดยการประยุกต์นำหลักการประมาณค่าการหาเส้นสมการที่เหมาะสมที่เป็นเส้นตรงมาใช้ ซึ่งลักษณะเช่นนี้เรียกว่า Linearization of Nonlinear Relationships แล้วหาความสัมพันธ์ระหว่างตัวแปรไม่เชิงเส้นได้

การประยุกต์หาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Application of Linear Regression)



รูปที่ 6: Data that are ill-suited for linear least-squares regression. (b) Indication that a parabola is preferable.

1. สมการเอกซ์โปเนนเชียล (Exponential Equation)

รูปทั่วไปของสมการเอกซ์โปเนนเชียล คือ $y = ae^{bx}$
เมื่อ a และ b เป็นค่าคงที่
ทำการ Take Logarithm ฐาน e หรือ Natural Logarithm ทั้งสองด้าน จะได้

$$\begin{aligned} \ln y &= \ln a + \ln(e^{bx}) \\ \ln y &= \ln a + bx \\ z &= c + bx \end{aligned}$$

เมื่อ $z = \ln y$ และ $c = \ln a$ จะเห็นว่า z และ x มีความสัมพันธ์เชิงเส้น สามารถใช้เส้นสมการถดถอยอันดับหนึ่งกับเซ็ตของจุด (x_i, z_i) หรือ $(x_i, \ln_{y_i})$

2. สมการกำลัง (Power Equation)

รูปทั่วไปของสมการเอกซ์โปแนนเชียล คือ $y = ax^b$
เมื่อ a และ b เป็นค่าคงที่
ทำการ Take Logarithm ฐาน 10 ทั้งสองด้าน จะได้

$$\begin{aligned} \log y &= \log a + \log(x^b) \\ \log y &= \log a + b \log x \\ z &= c + bw \end{aligned}$$

เมื่อ $z = \log y$, $c = \log a$ และ $w = \log x$ จะเห็นว่า z และ w มีความสัมพันธ์เชิงเส้น สามารถใช้เส้นสมการถดถอยอันดับหนึ่งกับเซ็ตของจุด (w_i, z_i) หรือ $(\log x_i, \log y_i)$

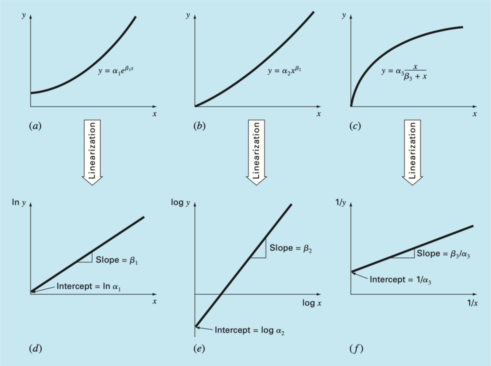
3. สมการแสดงอัตราการเจริญแบบอิ่มตัว (Saturation Growth Rate Equation)

รูปทั่วไปของสมการเอกซ์โปแนนเชียล คือ $y = a\dfrac{x}{b+x}$
เมื่อ a และ b เป็นค่าคงที่
สามารถแปลงให้อยู่ในรูปเชิงเส้นได้โดยการกลับเศษเป็นส่วน จะได้

$$\begin{aligned} \frac{1}{y} &= \frac{b+x}{ax} \\ &= \frac{b}{a}\cdot\frac{1}{x} + \frac{1}{a} \\ z &= c + dw \end{aligned}$$

เมื่อ $z = \dfrac{1}{y}$, $c = \dfrac{1}{a}$, $d = \dfrac{b}{a}$ และ $w = \dfrac{1}{x}$ จะเห็นว่า z และ w มีความสัมพันธ์เชิงเส้น สามารถใช้เส้นสมการถดถอยอันดับหนึ่งกับเซ็ตของจุด (w_i, z_i) หรือ $(\dfrac{1}{x_i}, \dfrac{1}{y_i})$

การประยุกต์หาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Application of Linear Regression)



รูปที่ 7: (a) The exponential equation, (b) the power equation, and (c) the saturation-growth-rate equation. Parts (d), (e), and (f) are linearized versions of these equations that result from simple transformations

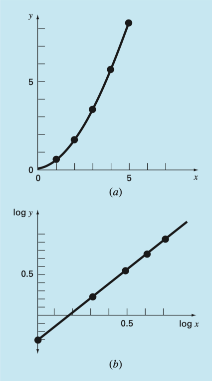
การประยุกต์หาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Application of Linear Regression)

ตัวอย่างที่ 4.4

จากข้อมูลที่กำหนดให้ จงหาสมการที่เหมาะสมโดยใช้ฟังก์ชันกำลัง

i	1	2	3	4	5
x_i	1	2	3	4	5
y_i	0.5	1.7	3.4	5.7	8.4

การประยุกต์หาเส้นสมการถดถอยอันดับหนึ่ง หรือเชิงเส้นตรง (Application of Linear Regression)



รูปที่ 8: (a) Plot of untransformed data with the power equation that fits these data. (b) Plot of transformed data used to determine the coefficients of the power equation.

สมการถดถอยพหุนาม (Polynomial Regression)

สมการถดถอยพหุนาม (Polynomial Regression)

สมการถดถอยพหุนาม (Polynomial Regression)

ในหัวข้อนี้จะกล่าวถึงสมการถดถอยพหุนามอันดับที่ m (m^{th} Polynomial Regression) รูปทั่วไปของสมการพหุนามอันดับที่ m คือ

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

ซึ่งผลรวมกำลังสองของค่าคลาดเคลื่อน คือ

$$S_r = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots - a_m x_i^m \right)^2$$

สมการถดถอยพหุนาม (Polynomial Regression)

เพื่อหาค่าต่ำสุด ดังนั้น

$$\frac{\partial S_r}{\partial a_0} = -2\sum\left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots - a_m x_i^m\right) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_i\left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots - a_m x_i^m\right) = 0$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_i^2\left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots - a_m x_i^m\right) = 0$$

$$\vdots$$

$$\frac{\partial S_r}{\partial a_m} = -2\sum x_i^m\left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots - a_m x_i^m\right) = 0$$

สมการถดถอยพหุนาม (Polynomial Regression)

ดังนั้นสมการปกติ คือ

$$
\begin{aligned}
a_0 n + a_1 \sum x_i + a_2 \sum x_i^2 + \cdots + a_m \sum x_i^m &= \sum y_i \\
a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 + \cdots + a_m \sum x_i^{m+1} &= \sum x_i y_i \\
a_0 \sum x_i^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4 + \cdots + a_m \sum x_i^{m+2} &= \sum x_i^2 y_i \\
&\vdots \\
a_0 \sum x_i^m + a_1 \sum x_i^{m+1} + a_2 \sum x_i^{m+2} + \cdots + a_m \sum x_i^{2m} &= \sum x_i^m y_i
\end{aligned}
$$

สมการถดถอยพหุนาม (Polynomial Regression)

ข้อสังเกต 4.1

ค่าคลาดเคลื่อนมาตรฐานของค่าประมาณของสมการถดถอยพหุนาม คือ

$$S_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}; \quad r^2 = \frac{s_t - S_r}{S_t}$$

เมื่อ m คือ อันดับของสมการพหุนาม

สมการถดถอยพหุนาม (Polynomial Regression)

ตัวอย่างที่ 4.5

จากข้อมูลที่กำหนดให้ จงหาสมการพหุนามอันดับสองที่เหมาะสมกับข้อมูลที่กำหนดให้

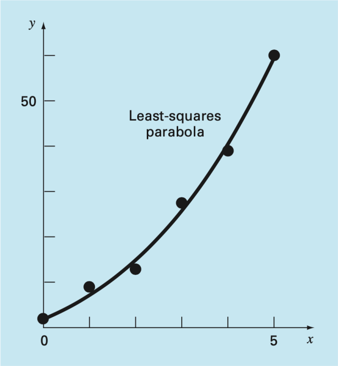
x_i	0	1	2	3	4	5
y_i	2.1	7.7	13.6	27.2	40.9	61.1

สมการถดถอยพหุนาม (Polynomial Regression)

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1x_i - a_2x_i^2)^2$
0	2.1	544.44	0.14332
1	7.7	314.47	1.00286
2	13.6	140.03	1.08158
3	27.2	3.12	0.80491
4	40.9	239.22	0.61951
5	61.1	1272.11	0.09439
Σ	152.6	2513.39	3.74657

รูปที่ 9: Computations for an error analysis of the quadratic least-squares fit.

สมการถดถอยพหุนาม (Polynomial Regression)



รูปที่ 10: Fit of a second-order polynomial.

สมการถดถอยพหุนาม (Polynomial Regression)

```
DOFOR i = 1, order + 1
  DOFOR j = 1, i
    k = i + j − 2
    sum = 0
    DOFOR ℓ = 1, n
      sum = sum + x_ℓ^k
    END DO
    a_{i,j} = sum
    a_{j,i} = sum
  END DO
  sum = 0
  DOFOR ℓ = 1, n
    sum = sum + y_ℓ · x_ℓ^(i−1)
  END DO
  a_{i,order+2} = sum
END DO
```

รูปที่ 11: Pseudocode to assemble the elements of the normal equations for polynomial regression.

สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

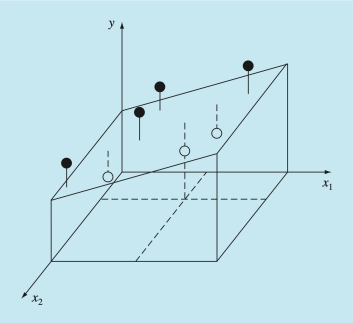
สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

พิจารณากรณีที่ y เป็นฟังก์ชันเชิงเส้นที่ขึ้นอยู่กับ 2 ตัวแปร ตัวอย่างเช่น y ขึ้นอยู่กับตัวแปร x_1 และ x_2 รูปของสมการเชิงเส้น 2 ตัวแปร เมื่อกำหนดให้ตัวแปรอิสระคือx_1 และ x_2 จะได้

$$y = a_0 + a_1x_1 + a_2x_2 + e$$

เมื่อนำมาพลอตกราฟ หรือวาดรูปจะเห็นว่าเป็นรูปใน 3 มิติที่มีลักษณะเป็นระนาบดังรูป

สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)



รูปที่ 12: Graphical depiction of multiple linear regression where y is a linear function of x_1 and x_2.

สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

ซึ่งผลรวมกำลังสองของค่าคลาดเคลื่อน คือ

$$S_r = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} - e\right)^2$$

สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

เพื่อหาค่าต่ำสุด ดังนั้น

$$\frac{\partial S_r}{\partial a_0} = -2\sum (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i}) = 0$$
$$\frac{\partial S_r}{\partial a_1} = -2\sum x_{1i}(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i}) = 0$$
$$\frac{\partial S_r}{\partial a_2} = -2\sum x_{2i}(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i}) = 0$$

สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

ดังนั้นสมการปกติ คือ

$$\begin{aligned} a_0 n + a_1 \sum x_{1i} + a_2 \sum x_{2i} &= \sum y_i \\ a_0 \sum x_{1i} + a_1 \sum x_{1i}^2 + a_2 \sum x_{1i}x_{2i} &= \sum x_{1i}y_i \\ a_0 \sum x_{2i} + a_1 \sum x_{1i}x_{2i} + a_2 \sum x_{2i}^2 &= \sum x_{2i}y_i \end{aligned}$$

สามารถเขียนในรูปเมทริกซ์ จะได้

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{2i} & \sum x_{1i}x_{2i} & \sum x_{2i}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i}y_i \\ \sum x_{2i}y_i \end{bmatrix} \tag{4.18}$$

สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

ตัวอย่างที่ 4.6

จากข้อมูลที่กำหนดให้ จงหาสมการที่เหมาะสม โดยใช้ สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

x_1	0	2	2.5	1	4	7
x_2	0	1	2	3	6	2
y	5	10	9	0	3	27

สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

	y	x_1	x_2	x_1^2	x_2^2	x_1x_2	x_1y	x_2y
	5	0	0	0	0	0	0	0
	10	2	1	4	1	2	20	10
	9	2.5	2	6.25	4	5	22.5	18
	0	1	3	1	9	3	0	0
	3	4	6	16	36	24	12	18
	27	7	2	49	4	14	189	54
Σ	54	16.5	14	76.25	54	48	243.5	100

รูปที่ 13

Table of Contents

Table of Contents

Table of Contents