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

ผศ.ดร.วงศ์วิศรุต เขื่องสตุ่ง และ ดร.รัฐพรหม พรหมคำ

สาขาวิชาคณิตศาสตร์
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

September 2, 2024

Outline

บทที่ 4 สมการถดถอยอันดับสองน้อยที่สุด (Least Squares Regression)
1.1 บทนำ
1.2 สมการการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)
สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์ (Simple Linear Regression in Matrix Form)
1.3 สมการถดถอยพหุนาม (Polynomial Regression)
สมการถดถอยพหุนามในรูปแบบเมทริกซ์ (Polynomial Regression in Matrix Form)
1.4 สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)
สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์ (Multiple Linear Regression in Matrix Form)

Table of Contents

บทที่ 4 สมการถดถอยอันดับสองน้อยที่สุด (Least Squares Regression)
1.1 บทนำ
1.2 สมการการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)
1.3 สมการถดถอยพหุนาม (Polynomial Regression)
1.4 สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

Outline

บทที่ 4 สมการถดถอยอันดับสองน้อยที่สุด (Least Squares Regression)

1.1 บทนำ

1.2 สมการการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์ (Simple Linear Regression in Matrix Form)

1.3 สมการถดถอยพหุนาม (Polynomial Regression)

สมการถดถอยพหุนามในรูปแบบเมทริกซ์ (Polynomial Regression in Matrix Form)

1.4 สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์ (Multiple Linear Regression in Matrix Form)

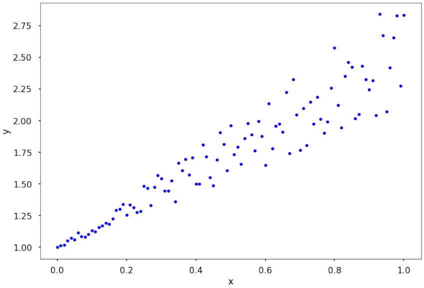
บทนำ

บทนำ

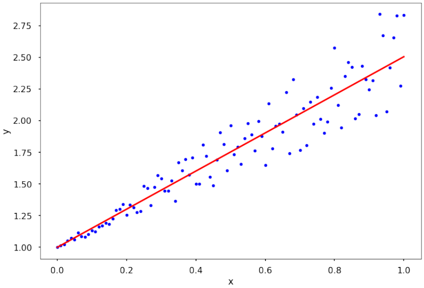
บทนำ

การสร้างเส้นโค้งที่เหมาะสม (Curve fitting) คือกระบวนการทางคณิตศาสตร์ที่ใช้ในการหาเส้นโค้งหรือฟังก์ชันที่เหมาะสมที่สุดในการแสดงผลชุดของจุดข้อมูลที่ได้รับจากการสังเกตหรือการทดลองทางวิทยาศาสตร์ต่างๆ โดยจุดมุ่งหมายคือการหาสมการทางคณิตศาสตร์ที่สอดคล้องกับข้อมูลให้มากที่สุด

บทนำ



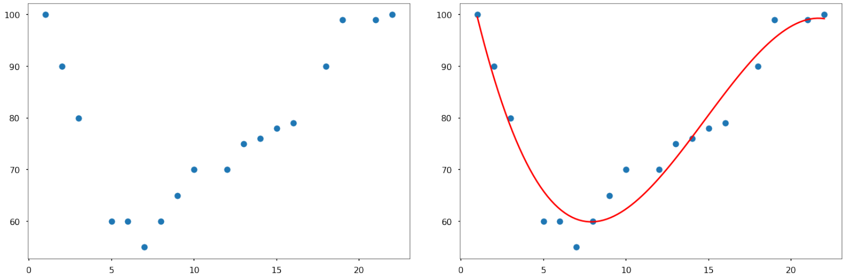
(a)



(b)

รูปที่ 1: แผนภาพการกระจาย (Scatter Plot)

บทนำ



(a) (b)

รูปที่ 2: แผนภาพการกระจาย (Scatter Plot)

บทนำ

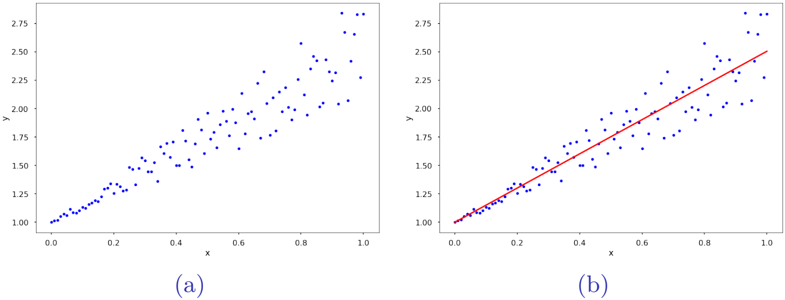
Curve fitting แบ่งได้เป็นสองประเภทหลัก ดังนี้

- **การถดถอยอันดับสองน้อยที่สุด (Least Squares Regression)**: เป็นการสร้างเส้นโค้งที่เหมาะสมที่ไม่จำเป็นต้องผ่านทุกจุดของข้อมูล โดยวิธีนี้จะพยายามหาฟังก์ชันที่ลดค่าความแตกต่างระหว่างจุดข้อมูลจริงกับค่าที่ฟังก์ชันคาดการณ์ให้มีค่าน้อยที่สุดในเชิงกำลังสอง (เช่น เส้นตรงหรือเส้นโค้งที่แสดงถึงแนวโน้มของข้อมูล)
- **การประมาณค่าในช่วง (Interpolation)**: เป็นการสร้างเส้นโค้งที่เหมาะสมซึ่งต้องผ่านทุกจุดของข้อมูล วิธีนี้มักใช้ในกรณีที่ข้อมูลมีความถูกต้องสูงและต้องการฟังก์ชันที่สามารถประมาณค่าได้อย่างแม่นยำระหว่างจุดข้อมูลต่าง ๆ

สมการการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)

สมการการถดถอยเชิงเส้นอย่างง่าย

การหาเส้นสมการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) เป็นระเบียบวิธีในการประมาณค่าการหาเส้นสมการที่เหมาะสมด้วยการสร้างเส้นตรงเพื่อประมาณค่าเซตของจุดหรือข้อมูล



รูปที่ 3: แผนภาพการกระจาย (Scatter Plot)

สมการการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)

สำหรับ $i = 1, 2, 3, ..., n$

- กำหนดให้ (x_i, y_i) เป็นเซตของจุดข้อมูล
- กำหนดให้ $\hat{y}_i = f(x)$ เป็นเส้นโค้งที่เหมาะสมกับเซตของจุดข้อมูล
- ณ ตำแหน่ง $x = x_i$ พิกัดที่กำหนดคือ y_i นั่นคือคู่อันดับ (x_i, y_i) และค่าของฟังก์ชันที่สอดคล้องกันบนเส้นโค้งที่เหมาะสมคือ $f(x_i)$
- ถ้า e_i คือค่าคลาดเคลื่อนของการประมาณค่า ณ ตำแหน่ง $x = x_i$ แล้ว

$$e_i = y_i - f(x_i) \tag{1.1}$$

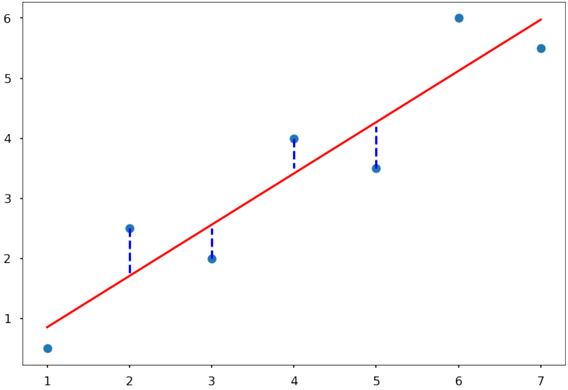
สมการการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)

ถ้าเรากำหนดให้

$$
\begin{aligned}
S_r &= [y_1 - f(x_1)]^2 + [y_2 - f(x_2)]^2 + \cdots + [y_n - f(x_n)]^2 \\
&= e_1^2 + e_2^2 + \cdots + e_n^2 \\
&= \sum_{i=1}^{n} e_i^2 \qquad (1.2)
\end{aligned}
$$

แล้ว **วิธีกำลังสองน้อยที่สุด (Least Squares Method)** เป็นการทำให้ผลรวมกำลังสองของค่าคลาดเคลื่อน (S_r) มีค่าน้อยที่สุด

สมการการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)



รูปที่ 4

สมการการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)

จากรูปที่ 4 กำหนดให้

$$\hat{y} = a_0 + a_1 x \tag{1.3}$$

เป็น**สมการการถดถอยเชิงเส้นอย่างง่าย**ที่เหมาะสมที่สุดกับเซตของจุดข้อมูล (x_i, y_i) สำหรับ $i = 1, 2, 3, ..., n$

พิจารณา ค่าคลาดเคลื่อนของการประมาณค่า ณ ตำแหน่ง x_i นั่นคือ

$$e_i = y_i - (a_0 + a_1 x_i)$$

จาก (1.4) จะได้ว่า

$$\begin{aligned} S_r &= [y_1 - (a_0 + a_1 x_1)]^2 + [y_2 - (a_0 + a_1 x_2)]^2 + \cdots + [y_n - (a_0 + a_1 x_n)]^2 \\ &= e_1^2 + e_2^2 + \cdots + e_n^2 \\ &= \sum_{i=1}^{n} e_i^2 \end{aligned} \tag{1.4}$$

สมการการถดถอยเชิงเส้นอย่างง่าย

นั่นคือ

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{1.5}$$

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

$$\frac{\partial S_r}{\partial a_0} = 0 \tag{1.6}$$

และ

$$\frac{\partial S_r}{\partial a_1} = 0 \tag{1.7}$$

สมการการถดถอยเชิงเส้นอย่างง่าย

จาก (1.5) จะได้

$$\frac{\partial S_r}{\partial a_0} = -2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i) \tag{1.8}$$

และ

$$\frac{\partial S_r}{\partial a_1} = -2\sum_{i=1}^{n}([y_i - a_0 - a_1 x_i]x_i) \tag{1.9}$$

สมการการถดถอยเชิงเส้นอย่างง่าย

จาก (1.8) และ $\sum_{i=1}^{n} a_0 = na_0$ จะได้

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

$$na_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 = \sum_{i=1}^{n} y_i \qquad (1.10)$$

จาก (1.9) จะได้

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

$$\left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 = \sum_{i=1}^{n} x_i y_i \qquad (1.11)$$

สมการการถดถอยเชิงเส้นอย่างง่าย

ดังนั้น

$$\begin{cases} na_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 & = \sum_{i=1}^{n} y_i \\ \left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 & = \sum_{i=1}^{n} x_i y_i \end{cases} \tag{1.12}$$

เราจะเรียกสมการ (1.12) ว่า **สมการปกติ (normal equations)**

สมการการถดถอยเชิงเส้นอย่างง่าย

โดยการแก้ระบบสมการ (1.12) เพื่อหาค่า a_0 และ a_1 จะได้

$$a_0 = \bar{y} - a_1\bar{x} \tag{1.13}$$

และ

$$a_1 = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{1.14}$$

เมื่อ $\bar{y}$ และ $\bar{x}$ คือ ค่าเฉลี่ยของ y และ x ตามลำดับ

สมการการถดถอยเชิงเส้นอย่างง่าย

ตัวอย่างที่ 1.1

จงหาสมการการถดถอยเชิงเส้นอย่างง่ายที่ใช้ประมาณค่าของข้อมูลต่อไปนี้

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

สมการการถดถอยเชิงเส้นอย่างง่าย

x_i	y_i	x_i^2	$x_i y_i$
1	0.5	1	
2	2.5	4	
3	2.0	9	
4	4.0	16	
5	3.5	25	
6	6.0	36	
7	5.5	49	

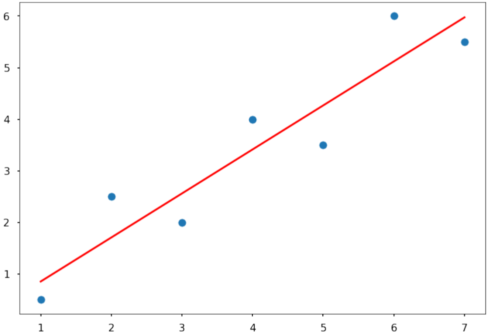
ตาราง 1: แสดงค่าจากการคำนวณของตัวอย่างที่ 1.2

สมการการถดถอยเชิงเส้นอย่างง่าย

วิธีทำ จากโจทย์จะได้

- $n = 7$
- $\sum_{i=1}^{7} x_i y_i = 119.5$
- $\sum_{i=1}^{7} x_i^2 = 140$
- $\sum_{i=1}^{7} x_i = 28$ ดังนั้น $\bar{x} = \dfrac{28}{7} = 4$
- $\sum_{i=1}^{7} y_i = 24$ ดังนั้น $\bar{y} = \dfrac{24}{7} = 3.428571$

สมการการถดถอยเชิงเส้นอย่างง่าย



รูปที่ 5

สมการการถดถอยเชิงเส้นอย่างง่าย

ตัวอย่างที่ 1.2

จงหาสมการการถดถอยเชิงเส้นอย่างง่ายที่ใช้ประมาณค่าของข้อมูลต่อไปนี้

$$(1, 0.6), (2, 2.4), (3, 3.5), (4, 4.8), (5, 5.7)$$

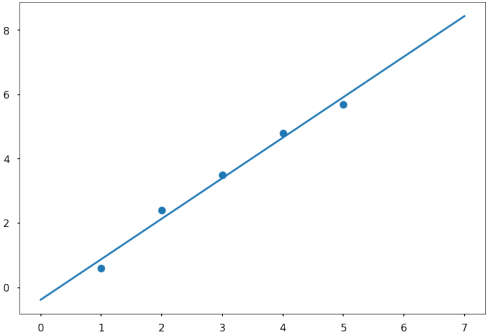
สมการการถดถอยเชิงเส้นอย่างง่าย

x_i	y_i	x_i^2	$x_i y_i$
1	0.6	1	0.6
2	2.4	4	4.8
3	3.5	9	10.5
4	4.8	16	19.2
5	5.7	25	28.5
15	17.0	55	63.6

ตาราง 2: แสดงค่าจากการคำนวณของตัวอย่างที่ 1.2

สมการการถดถอยเชิงเส้นอย่างง่าย

สมการการถดถอยเชิงเส้นอย่างง่าย



รูปที่ 6: กราฟของสมการถดถอยเชิงเส้นอย่างง่ายของตัวอย่างที่ 1.2

คุณภาพของเส้นสมการที่เหมาะสม

เนื่องจากเส้นโค้งของสมการถดถอยเชิงเส้นไม่ผ่านทุกจุดของข้อมูล จึงจำเป็นต้องมีการประเมินสมการถดถอยเชิงเส้น ดังนั้น **การพิจารณาคุณภาพของสมการถดถอยเชิงเส้น (Quantification of Error of Linear Regression)** ต้องพิจารณาสิ่งต่อไปนี้

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

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \tag{1.15}$$

คุณภาพของเส้นสมการที่เหมาะสม

2. ค่าความเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimate):

$$S_{y/x} = \sqrt{\frac{S_r}{n-2}} \tag{1.16}$$

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

$$S_t = \sum (y_i - \bar{y})^2 \tag{1.17}$$

คุณภาพของเส้นสมการที่เหมาะสม

4. สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination):

$$r^2 = \frac{S_t - S_r}{S_t} \tag{1.18}$$

5. สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient):

$$r = \sqrt{\frac{S_t - S_r}{S_t}} \tag{1.19}$$

หมายเหตุ

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

สมการการถดถอยเชิงเส้นอย่างง่าย

ตัวอย่างที่ 1.3

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

สมการการถดถอยเชิงเส้นอย่างง่าย

วิธีทำ จากตัวอย่างที่ 1.3 จะได้ สมการถดถอยเชิงเส้นอย่างง่ายที่ใช้ประมาณค่า คือ $\hat{y} = 0.07142857 + 0.8392857x$ ซึ่งสามารถนำมาสร้างตารางได้ดังตารางที่ 3

x_i	y_i	$(y_i - \bar{y})$	$(y_i - a_0 - a_1 x_i)^2$
1	0.5	8.5765	0.1687
2	2.5	0.8622	0.5625
3	2.0	2.0408	0.3473
4	4.0	0.3265	0.3265
5	3.5	0.0051	0.5896
6	6.0	6.6122	0.7972
7	5.5	4.2908	0.1993
$\sum$	24.0	22.7143	2.9911

ตาราง 3: แสดงค่าจากการคำนวณของตัวอย่างที่ 1.3

สมการการถดถอยเชิงเส้นอย่างง่าย

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์ (Simple Linear Regression in Matrix Form)

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

ในหัวข้อนี้จะกล่าวถึงการหา**สมการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression in Matrix Form)** โดยอาศัยหลักของเมทริกซ์ เมื่อข้อมูลขนาดใหญ่ การนำเมทริกซ์มาประยุกต์ในการหาสมการถดถอยจะช่วยให้การคำนวณง่ายขึ้นเมื่อนำไปเขียนโปรแกรมทางคอมพิวเตอร์

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

จากความสัมพันธ์ระหว่างสมการการถดถอยเชิงเส้นอย่างง่าย และค่าคลาดเคลื่อน นั่นคือ

$$e_i = y_i - (a_0 + a_1 x_i) \tag{1.20}$$

จะได้ว่า

$$y_i = a_0 + a_1 x_i + e_i \tag{1.21}$$

สำหรับ $i \in \{1, 2, ..., n\}$ จากสมการ (1.21) จะได้

$$\begin{aligned} y_1 &= a_0 + a_1 x_1 + e_1 \\ y_2 &= a_0 + a_1 x_2 + e_2 \\ &\vdots \\ y_n &= a_0 + a_1 x_n + e_n \end{aligned} \tag{1.22}$$

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

จากสมการ (1.22) เขียนให้อยู่ในรูปแบบเมทริกซ์ จะได้

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_0 + a_1 x_1 \\ a_0 + a_1 x_2 \\ \vdots \\ a_0 + a_1 x_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{1.23}$$

จะได้

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{1.24}$$

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

ดังนั้น **สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์ (Simple Linear Regression in Matrix Form)** คือ

$$\mathbf{Y} = \mathbf{X}a + e \tag{1.25}$$

เมื่อ

- $\mathbf{X}$ ถูกเรียกว่า the design matrix.
- a คือ the vector of parameters
- e คือ the error vector
- $\mathbf{Y}$ คือ the response vector

โดยที่

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X}_{n\times 2} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, a_{2\times 1} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}, e_{n\times 1} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{1.26}$$

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

กำหนดให้ $\mathbf{X}^T$ เป็นเมทริกซ์สลับเปลี่ยน (transpose of a matrix) จากสมการ (1.26) จะได้

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots 1 \\ x_1 & x_2 & \cdots x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$(\mathbf{X}^T\mathbf{X})[a] = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

และ

$$\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots 1 \\ x_1 & x_2 & \cdots x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

จะสังเกตเห็นว่า $(\mathbf{X}^T\mathbf{X})a = \mathbf{X}^T\mathbf{Y}$ สอดคลองกับสมการปกติ นั่นคือ

$$
\begin{aligned}
na_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 &= \sum_{i=1}^{n} y_i \\
\left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 &= \sum_{i=1}^{n} x_i y_i
\end{aligned}
$$

ดังนั้น จะพิจารณา $(\mathbf{X}^T\mathbf{X})a = \mathbf{X}^T\mathbf{Y}$ โดยนำอินเวอร์สของ $(\mathbf{X}^T\mathbf{X})$ นั่นคือ $(\mathbf{X}^T\mathbf{X})^{-1}$ คูณตลอดสมการ จะได้

$$(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})a = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y})$$

เนื่องจาก $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X} = I$ ดังนั้น

$$a = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{1.27}$$

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

จากสมการ (1.26) และสมการ (1.27) จะได้

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots 1 \\ x_1 & x_2 & \cdots x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix}$$

และ

$$\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots 1 \\ x_1 & x_2 & \cdots x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

ต่อไป จะแสดงว่า $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = a$ จากสมการ(1.27) จะได้

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix} \tag{1.28}$$

จาก (1.28) (1.13) และ (1.14) จะได้

$$\begin{aligned}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} &= \begin{bmatrix} \bar{y} - a_1\bar{x} \\ \dfrac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \end{bmatrix} \\ &= \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \\ &= a\end{aligned}$$

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

ดังนั้น สมการปกติของสมการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์ คือ

$$a = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{1.29}$$

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

ตัวอย่างที่ 1.4

จงหาสมการการถดถอยเชิงเส้นอย่างง่ายที่ใช้ประมาณค่าของข้อมูลต่อไปนี้

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

สมการการถดถอยเชิงเส้นอย่างง่ายในรูปแบบเมทริกซ์

แบบฝึกหัด

1. จากข้อมูลต่อไปนี้ จงหาสมการถดถอยเชิงเส้นอย่างง่าย

x	1	2	3	4	5	6
y	2.4	3.1	3.5	4.2	5.0	6.0

2. จากข้อมูลต่อไปนี้ จงหาสมการถดถอยเชิงเส้นอย่างง่าย

x	0	1	2	3	4
y	1.0	2.9	4.8	6.7	8.6

3. จากแบบฝึกหัดข้อที่ 1 จงหาค่าสัมประสิทธ์สหสัมพันธ์
4. จากแบบฝึกหัดข้อที่ 2 จงหาค่าสัมประสิทธ์สหสัมพันธ์
5. จากแบบฝึกหัดข้อที่ 1 จงเขียนภาษาโปรแกรมไพธอนเพื่อหาสมการถดถอยเชิงเส้นอย่างง่าย และค่าสัมประสิทธ์สหสัมพันธ์
6. จากแบบฝึกหัดข้อที่ 2 จงเขียนภาษาโปรแกรมไพธอนเพื่อหาสมการถดถอยเชิงเส้นอย่างง่าย และค่าสัมประสิทธ์สหสัมพันธ์

เฉลยแบบฝึกหัด

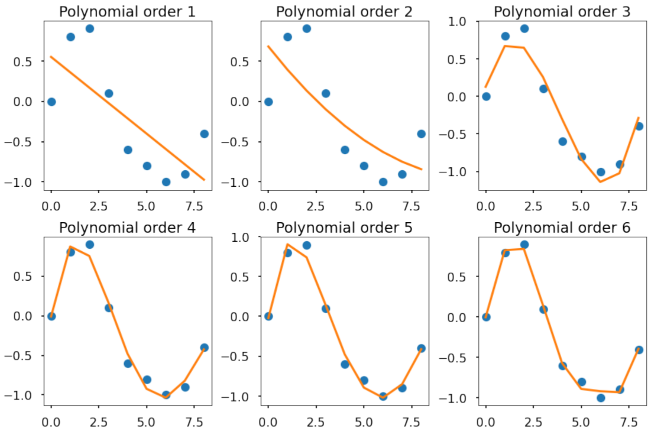
1. $\hat{y} = 1.59333333 + 0.69714286x$
2. $\hat{y} = 1 + 1.9x$
3. $-5.70434783x_2$
4. $r = 0.99140039$
5. $r = 1$

สมการถดถอยพหุนาม (Polynomial Regression)

สมการถดถอยพหุนาม (Polynomial Regression)

ในหัวข้อก่อนหน้านี้ ได้กล่าวถึงวิธีการสร้างสมการเชิงเส้นอย่างง่ายโดยใช้สมการถดถอย เมื่อกราฟของข้อมูลมีแนวโน้มเป็นเส้นตรง แต่ในบางครั้งการใช้สมการเชิงเส้นกับข้อมูลในทางวิศวกรรมหรือวิทยาศาสตร์อาจให้ผลลัพธ์ที่ไม่ดีมากนัก ดังรูปที่ 7 ดังนั้น ในหัวข้อนี้จะกล่าวถึงสมการถดถอยพหุนามอันดับที่ m (m^{th} Polynomial Regression)

สมการถดถอยพหุนาม (Polynomial Regression)



รูปที่ 7: สมการถดถอยพหุนาม

สมการถดถอยพหุนาม (Polynomial Regression)

ในหัวข้อนี้จะกล่าวถึงสมการถดถอยพหุนามอันดับที่ m (m^{th} Polynomial Regression) **รูปทั่วไปของสมการถดถอยพหุนามอันดับที่** m คือ

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

ซึ่งเป็นสมการที่เหมาะสมที่สุดกับเซตของจุดข้อมูล (x_i, y_i) สำหรับ $i = 1, 2, 3, ..., n$
โดยที่ ค่าคลาดเคลื่อนของการประมาณค่า ณ ตำแหน่ง x_i นั่นคือ

$$e_i = y_i - (a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m)$$

ซึ่งสามารถเขียน ผลรวมกำลังสองของค่าคลาดเคลื่อนได้ ดังต่อไปนี้

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

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

$$a_0 n + a_1 \sum x_i + a_2 \sum x_i^2 + \cdots + a_m \sum x_i^m = \sum y_i \tag{1.30}$$

$$a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 + \cdots + a_m \sum x_i^{m+1} = \sum x_i y_i \tag{1.31}$$

$$a_0 \sum x_i^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4 + \cdots + a_m \sum x_i^{m+2} = \sum x_i^2 y_i \tag{1.32}$$

$$\vdots \tag{1.33}$$

$$a_0 \sum x_i^m + a_1 \sum x_i^{m+1} + a_2 \sum x_i^{m+2} + \cdots + a_m \sum x_i^{2m} = \sum x_i^m y_i \tag{1.34}$$

$$\tag{1.35}$$

สมการถดถอยพหุนาม (Polynomial Regression)

ข้อสังเกต 1.1

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 1.5

จากข้อมูลที่กำหนดให้ จงหาสมการถดถอยพหุนามอันดับสองที่เหมาะสมกับข้อมูลที่กำหนดให้

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)

จากโจทย์ จะได้

$$m = 2 \quad \sum x_i = 15 \quad \sum x_i^4 = 979$$
$$n = 6 \quad \sum y_i = 152.6 \quad \sum x_i y_i = 585.6$$
$$\bar{x} = 2.5 \quad \sum x_i^2 = 55 \quad \sum x_i^2 y_i = 2488.8$$
$$\bar{y} = 25.433 \quad \sum x_i^3 = 225$$

จากสมการ (1.30) จะได้

$$\begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 152.6 \\ 585.6 \\ 2488.8 \end{bmatrix}$$

สมการถดถอยพหุนาม (Polynomial Regression)

โดยใช้ระเบียบวิธีการกำจัดแบบเกาส์ จะได้
$a_0 = 2.47857, a_1 = 2.35929, a_2 = 1.86071$
ดังนั้น สมการถดถอยพหุนามอันดับสอง คือ
$\hat{y} = 2.47857 + 2.35929x + 1.86071x^2$

x_i	y_i	$(y_i - \bar{y})$	$(y_i - a_0 - a_1 x_i - a_2 x_i^2)^2$
0	2.1	544.44	0.14332
1	7.7	314.47	1.00286
2	13.6	140.03	1.08158
3	27.6	3.12	0.80491
4	40.9	239.22	0.61951
5	61.1	1272.11	0.09439
$\sum$	152.6	22.7143	3.74657

ตาราง 4: แสดงค่าจากการคำนวณของตัวอย่างที่ 1.5

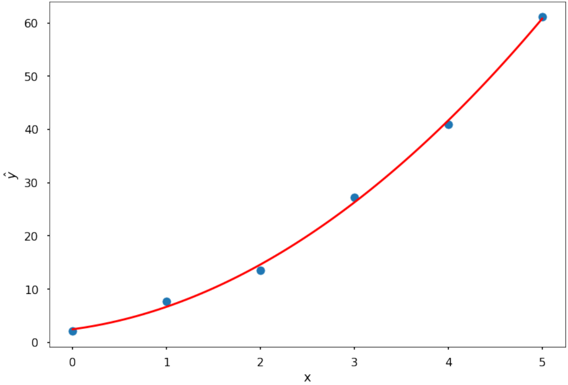
สมการถดถอยพหุนาม (Polynomial Regression)

จากตารางที่ 4 จะได้ สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination) คือ

$$r^2 = \frac{2513.39 - 3.74657}{2513.39} = 0.99851$$

หรือสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) คือ $r = 0.99925$

สมการถดถอยพหุนาม (Polynomial Regression)



รูปที่ 8: กราฟของสมการถดถอยพหุนามอันดับที่ 2 ของตัวอย่างที่ 1.5

สมการถดถอยพหุนามในรูปแบบเมทริกซ์ (Polynomial Regression in Matrix Form)

สมการถดถอยพหุนามในรูปแบบเมทริกซ์

จากความสัมพันธ์ระหว่างสมการถดถอยพหุนาม และค่าคลาดเคลื่อน นั่นคือ

$$e_i = y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_m x_i^m) \tag{1.36}$$

จะได้ว่า

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_m x_i^m + e_i \tag{1.37}$$

สำหรับ $i \in \{1, 2, ..., n\}$ จากสมการ (1.37) จะได้

$$\begin{aligned} y_1 &= a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_m x_1^m + e_1 \\ y_2 &= a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_m x_2^m + e_2 \\ &\vdots \\ y_n &= a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_m x_n^m + e_n \end{aligned} \tag{1.38}$$

สมการถดถอยพหุนามในรูปแบบเมทริกซ์

จากสมการ (1.39) เขียนให้อยู่ในรูปแบบเมทริกซ์ จะได้

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_m x_1^m \\ a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_m x_2^m \\ \vdots \\ a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_m x_n^m \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{1.39}$$

จะได้

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{1.40}$$

สมการถดถอยพหุนามในรูปแบบเมทริกซ์

ดังนั้น สมการถดถอยพหุนามในรูปแบบเมทริกซ์ (Polynomial Regression in Matrix Form) คือ

$$\mathbf{Y} = \mathbf{X}a + e \tag{1.41}$$

โดยที่

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X}_{n\times(m+1)} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}, \tag{1.42}$$

$$a_{(m+1)\times 1} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}, e_{n\times 1} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{1.43}$$

สมการถดถอยพหุนามในรูปแบบเมทริกซ์

โดยใช้กระบวนการเดียวกันกับหัวข้อ 1 จะได้ **สมการปกติของสมการถดถอยพหุนามในรูปแบบเมทริกซ์ (Polynomial Regression in Matrix Form)** คือ

$$a = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

เมื่อ $\mathbf{X}^T$ คือ transpose of the matrix

สมการถดถอยพหุนาม (Polynomial Regression)

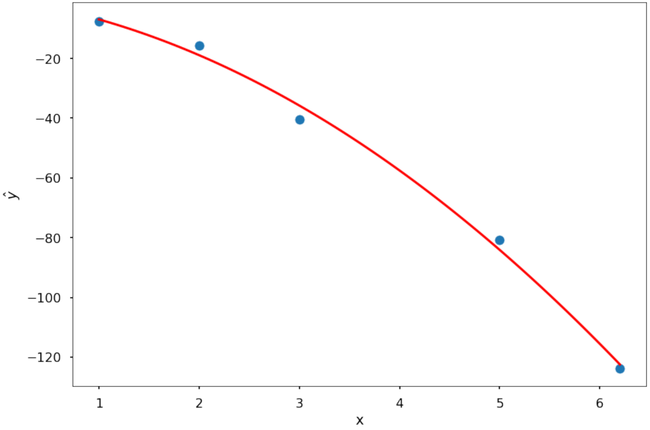
ตัวอย่างที่ 1.6

จากข้อมูลที่กำหนดให้ จงหาสมการถดถอยพหุนามอันดับสองที่เหมาะสมกับข้อมูลที่กำหนดให้

x_i	1	2	3.5	5	6.2
y_i	-7.500	-15.600	-40.500	-80.700	-123.876

สมการถดถอยพหุนาม (Polynomial Regression)

สมการถดถอยพหุนาม (Polynomial Regression)



รูปที่ 9

สมการถดถอยเชิงเส้นหลายตัวแปร (Multiple Linear Regression)

ในหัวข้อนี้จะกล่าวถึงสมการถดถอยพหุนามอันดับที่ m (m^{th} Polynomial Regression) **รูปทั่วไปของสมการถดถอยพหุนามอันดับที่** m คือ

$$\hat{y} = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$$

ซึ่งเป็นสมการที่เหมาะสมที่สุดกับเซตของจุดข้อมูล (x_i, y_i) สำหรับ $i = 1, 2, 3, ..., n$
โดยที่ ค่าคลาดเคลื่อนของการประมาณค่า ณ ตำแหน่ง x_i คือ

$$e_i = y_i - (a_0 + a_1x + a_2x^2 + \cdots + a_mx^m)$$

ซึ่งสามารถเขียน ผลรวมกำลังสองของค่าคลาดเคลื่อนได้ ดังต่อไปนี้

$$S_r = \sum_{i=1}^{n} \left(y_i - (a_0 + a_1x + a_2x^2 + \cdots + a_mx^m)\right)^2$$

สมการถดถอยเชิงเส้นหลายตัวแปร

สำหรับสมการทั่วไปของสมการถดถอยเชิงเส้นหลายตัวแปรที่มีตัวแปรอิสระ m ตัว ซึ่งอยู่ในรูปแบบ

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m \tag{1.44}$$

เมื่อ $a_0, a_1, ..., a_m$ คือ สัมประสิทธิ์การถดถอยบางส่วน (partial – regression coefficient)
และ ค่าคลาดเคลื่อนของการประมาณค่า คือ

$$e_i = y_i - (a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m)$$

สมการถดถอยเชิงเส้นหลายตัวแปร

พิจารณากรณีที่ y เป็นฟังก์ชันเชิงเส้นที่ขึ้นอยู่กับ 2 ตัวแปร ตัวอย่างเช่น y ขึ้นอยู่กับตัวแปร x_1 และ x_2 รูปของสมการเชิงเส้น 2 ตัวแปร เมื่อกำหนดให้ตัวแปรอิสระ คือ x_1 และ x_2 นั่นคือ จะได้

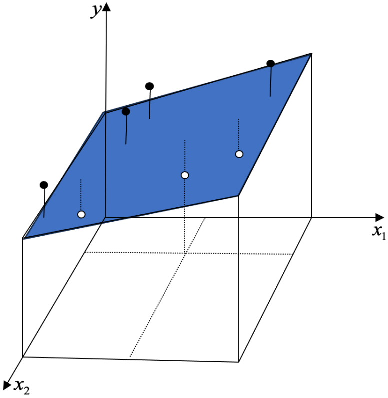
$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 \tag{1.45}$$

จะได้ว่า ค่าคลาดเคลื่อนของการประมาณค่า คือ

$$e_i = y_i - (a_0 + a_1 x_1 + a_2 x_2)$$

เมื่อนำมาพลอตกราฟ หรือวาดรูปจะเห็นว่าเป็นรูปใน 3 มิติที่มีลักษณะเป็นระนาบ ดังรูป

สมการถดถอยเชิงเส้นหลายตัวแปร



รูปที่ 10: กราฟแสดงลักษณะเป็นระนาบของสมการ $y = a_0 + a_1x_1 + a_2x_2 + e$

สมการถดถอยเชิงเส้นหลายตัวแปร

พิจารณา (x_{1i}, x_{2i}, y_i) สำหรับ $i = 1, 2, ..., n$ จะเห็นได้ว่า

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + e_i$$

ซึ่งผลรวมกำลังสองของค่าคลาดเคลื่อน คือ

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

สมการถดถอยเชิงเส้นหลายตัวแปร

เพื่อหาค่าต่ำสุด ดังนั้น

$$\frac{\partial S_r}{\partial a_0} = -2\sum (y_i - a_0 - a_1x_{1i} - a_2x_{2i}) = 0$$
$$\frac{\partial S_r}{\partial a_1} = -2\sum x_{1i}(y_i - a_0 - a_1x_{1i} - a_2x_{2i}) = 0$$
$$\frac{\partial S_r}{\partial a_2} = -2\sum x_{2i}(y_i - a_0 - a_1x_{1i} - a_2x_{2i}) = 0$$

ดังนั้นสมการปกติ คือ

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

สมการถดถอยเชิงเส้นหลายตัวแปร

สามารถเขียนในรูปเมทริกซ์ จะได้

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{2i} & \sum x_{1i}x_{2i} & \sum x_{2i}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i}y_i \\ \sum x_{2i}y_i \end{bmatrix} \tag{1.46}$$

ข้อสังเกต 1.2

ค่าคลาดเคลื่อนมาตรฐานของค่าประมาณของสมการถดถอยพหุนาม คือ

$$S_{y/x} = \sqrt{\frac{S_r}{n - (m+1)}}$$

เมื่อ m คือ อันดับของสมการพหุนาม

สมการถดถอยเชิงเส้นหลายตัวแปร

ตัวอย่างที่ 1.7

จากข้อมูลที่กำหนดให้ จงหาสมการถดถอยเชิงเส้นหลายตัวแปร (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

	y	x_1	x_2	x_1^2	x_2^2	$x_1 x_2$	$x_1 y$	$x_2 y$
	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
$\sum$	54	16.5	14	76.25	54	48	243.5	100

ตาราง 5: แสดงค่าจากการคำนวณของตัวอย่างที่ 1.7

สมการถดถอยเชิงเส้นหลายตัวแปร

จากสมการ 1.46 และตารางที่ 5 จะได้

$$\begin{bmatrix} 6 & 16.5 & 14 \\ 16.5 & 76.25 & 48 \\ 14 & 48 & 54 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 54 \\ 243.5 \\ 100 \end{bmatrix}$$

โดยใช้วิธีกำจัดแบบเกาส์ จะได้ $a_0 = 5, a_1 = 4, a_2 = -3$
ดังนั้น สมการถดถอยเชิงเส้นอย่างง่าย คือ $\hat{y} = 5 + 4x_1 - 3x_2$

สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์ (Multiple Linear Regression in Matrix Form)

สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์

จากสมการทั่วไปของสมการถดถอยเชิงเส้นหลายตัวแปรที่มีตัวแปรอิสระ m ตัว นั่นคือ

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \cdots + a_m x_{im} + e_i \tag{1.47}$$

สำหรับ $i = 1, 2, ..., n$ จากสมการ (1.47) จะได้

$$\begin{aligned} y_1 &= a_0 + a_1 x_{11} + a_2 x_{12} + \cdots + a_m x_{1m} + e_1 \\ y_2 &= a_0 + a_1 x_{21} + a_2 x_{22} + \cdots + a_m x_{2m} + e_2 \\ &\vdots \\ y_n &= a_0 + a_1 x_{n1} + a_2 x_{n2} + \cdots + a_m x_{nm} + e_n \end{aligned} \tag{1.48}$$

สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์

จากสมการ (1.48) เขียนให้อยู่ในรูปแบบเมทริกซ์ จะได้

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_0 + a_1 x_{11} + a_2 x_{12} + \cdots + a_m x_{1m} \\ a_0 + a_1 x_{21} + a_2 x_{22} + \cdots + a_m x_{2m} \\ \vdots \\ a_0 + a_1 x_{n1} + a_2 x_{n2} + \cdots + a_m x_{nm} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

จะได้

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \qquad (1.49)$$

สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์

ดังนั้น สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์ (Multiple Linear Regression in Matrix Form) คือ

$$\mathbf{Y} = \mathbf{X}a + e \tag{1.50}$$

โดยที่

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X}_{n\times(m+1)} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}, \tag{1.51}$$

$$a_{(m+1)\times 1} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}, e_{n\times 1} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{1.52}$$

สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์

โดยใช้กระบวนการเดียวกันกับหัวข้อ 1 จะได้ **สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์ (Multiple Linear Regression in Matrix Form)** คือ

$$a = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{1.53}$$

สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์

ตัวอย่างที่ 1.8

จากข้อมูลที่กำหนดให้ จงหาสมการถดถอยเชิงเส้นหลายตัวแปร (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

สมการถดถอยเชิงเส้นหลายตัวแปรในรูปแบบเมทริกซ์

วิธีทำ

แบบฝึกหัด

1. จากข้อมูลที่กำหนดให้ จงหาสมการถดถอยพหุนามอันดับสอง

x	0	1	2	3	4
y	1	0	3	10	21

2. จากข้อมูลที่กำหนดให้ จงหาสมการถดถอยพหุนามอันดับสอง

x	0	1	2	3	4	5	6
y	71	89	67	43	31	18	9

3. จากข้อมูลที่กำหนดให้ จงหาสมการถดถอยเชิงเส้นหลายตัวแปร

x_1	0	1	1	2	2	3	3	4	4
x_1	0	1	2	1	2	1	2	1	2
y	15.1	17.9	12.7	25.6	20.5	35.1	29.7	45.4	40.2

แบบฝึกหัด

1. จากแบบฝึกหัดข้อที่ 1 จงหาค่าสัมประสิทธิ์สหสัมพันธ์
2. จากแบบฝึกหัดข้อที่ 2 จงหาค่าสัมประสิทธิ์สหสัมพันธ์
3. จากแบบฝึกหัดข้อที่ 3 จงหาค่าสัมประสิทธิ์สหสัมพันธ์
4. จากแบบฝึกหัดข้อที่ 1 จงเขียนภาษาโปรแกรมไพธอนเพื่อหาสมการถดถอยพหุนามอันดับสอง และค่าสัมประสิทธิ์สหสัมพันธ์
5. จากแบบฝึกหัดข้อที่ 2 จงเขียนภาษาโปรแกรมไพธอนเพื่อหาสมการถดถอยพหุนามอันดับสอง และค่าสัมประสิทธิ์สหสัมพันธ์

เฉลยแบบฝึกหัด

1. $\hat{y} = 1 - 3x + 2x^2$
2. $\hat{y} = 81.93 - 8.28x - 0.78x^2$
3. $\hat{y} = 14.46086957 + 9.02521739x_1 - 5.70434783x_2$
4. $r = 1$
5. $r = 0.9503421$
6. $r = 0.9977592$